Volumetric Image Classification using Homogeneous Decomposition and Dictionary Learning: A Study Using Retinal Optical Coherence Tomography for Detecting Age-related Macular Degeneration

Abdulrahman Albarrak[a,*], Frans Coenen[b], Yalin Zheng[c]

[a] *Department of Computer Science, College of Computer and Information Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia*
[b] *Department of Computer Science, University of Liverpool, Ashton Building, Ashton, Street, Liverpool L69 3BX, United Kingdom*
[c] *Department of Eye and Vision Science, University of Liverpool, Apex Building, 6 West Derby Street, Liverpool L7 8TX, United Kingdom*

Abstract

Three-Dimensional (3D) (volumetric) diagnostic imaging techniques are indispensable with respect to the diagnosis and management of many medical conditions. However there is a lack of automated diagnosis techniques to facilitate such 3D image analysis (although some support tools do exist). This paper proposes a novel framework for volumetric medical image classification founded on homogeneous decomposition and dictionary learning. In the proposed framework each image (volume) is recursively decomposed until homogeneous regions are arrived at. Each region is represented using a Histogram of Oriented Gradients (HOG) which is transformed into a set of feature vectors. The Gaussian Mixture Model (GMM) is then used to generate a "dictionary" and the Improved Fisher Kernel (IFK) approach is used to encode feature vectors so as to generate a single feature vector for each volume, which can then be fed into a classifier generator. The principal advantage offered by the framework is that it does not require the detection (segmentation) of specific objects within the input data. The nature of the framework is fully described. A wide range of experiments were conducted with which to analyse the operation of the proposed framework and these are also re-

[*] Corresponding author

Email address: `aabkb@ccis.imamu.edu.sa; aabkb@yahoo.com` (Abdulrahman Albarrak)

ported fully in the paper. Although the proposed approach is generally applicable to 3D volumetric images, the focus for the work is 3D retinal Optical Coherence Tomography (OCT) images in the context of the diagnosis of Age-related Macular Degeneration (AMD). The results indicate that excellent diagnostic predictions can be produced using the proposed framework.

Keywords: Feature Selection, Image Classification, Dictionary Learning,
Homogeneous Decomposition, Optical Coherence Tomography (OCT),
Age-related Macular Degeneration (AMD)

1. Introduction

Medical imaging is one of the most important breakthroughs in the management of various conditions. Such images are extensively used in day-to-day clinical practice to support the diagnosis, management and prognosis of medical conditions. The images of interest have tended to be two-dimensional (2D); but 3D or volumetric images, such as those produced using Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and Optical Coherence tomography (OCT) are becoming increasingly prevalent. However, there has been no corresponding increase in the availability of software tools to support the analysis of large amounts of medical volumetric data. Further, what tools are available frequently incorporate some form of object segmentation, a resource intensive process that is often error prone. To address this issue this paper presents a software framework designed to provide end-to-end support for 3D (volumetric) medical image diagnosis, founded on the idea of image classification, that obviates the need for segmentation. Image classification is concerned with the generation and application of predictors/classifiers to allocate class labels to images. The process of building classifiers is generally well understood. The challenge is in converting the input data into a form appropriate for classifier generation.

The proposed framework is founded on the idea of homogeneous image decomposition and dictionary learning influenced partially by the work described in (Albarrak et al., 2014) although this only considered dictionary learning with respect to image decomposition; the work presented in this paper is directed at the advantages that can be gained by coupling homogeneous decomposition with dictionary learning. A central issue with respect to homogeneous hierarchical spatial decomposition is the termination condition. The most straightforward approach is to specify a maximum level of decomposition. An alternative approach is to terminate the decomposition whenever homogeneous regions are arrived at. The benefit of the second approach is that it avoids needless decomposition. The second approach is therefore adopted with respect to the proposed framework. The challenge is how to determine whether a region is homogeneous or not. This is done using what is known as a *critical function*. The work described in this paper includes consideration of a number of proposed critical functions for use with respect to the presented framework. The operation of these functions is compared with respect to a number of existing functions in terms of classification effectiveness.

Once the hierarchically decomposed is complete, using the proposed framework, each region is represented using a feature vector. Feature vectors can be generated in various ways; in this paper the use of HOG (Lazebnik et al., 2006; Yang et al., 2009) is advocated (with some refinement to allow applicability to 3D data volumes). Thus each image is represented in terms of a set of feature vectors and the entire input set as a *Bag-of-Features* (BoF) (Csurka et al., 2004). For the image classifier generation/application process one feature vector per image is required. To achieve this a dictionary learning based approach was incorporated into to the proposed framework.

Using dictionary learning a discriminative subset of features, referred to as the "dictionary", is identified from a collection of features; and then used as a guide for generating a feature vector for each image. Dictionary learning has been successfully applied elsewhere (Yang et al., 2009; Perronnin et al., 2010; Wang et al., 2010; Zhou et al., 2010) and can be conducted in a number of manners, but with respect to the work presented in this paper the IFK (Perronnin et al., 2010) method is suggested. IFK relies on a GMM to generate the desired "dictionary". The dictionary is then used to encode the feature vectors for each image so as to form a single feature vector for each image to be fed into a traditional classifier generator. The application of the dictionary learning concept, to the volumetric image classification problem, is a central theme of the investigation presented in this paper.

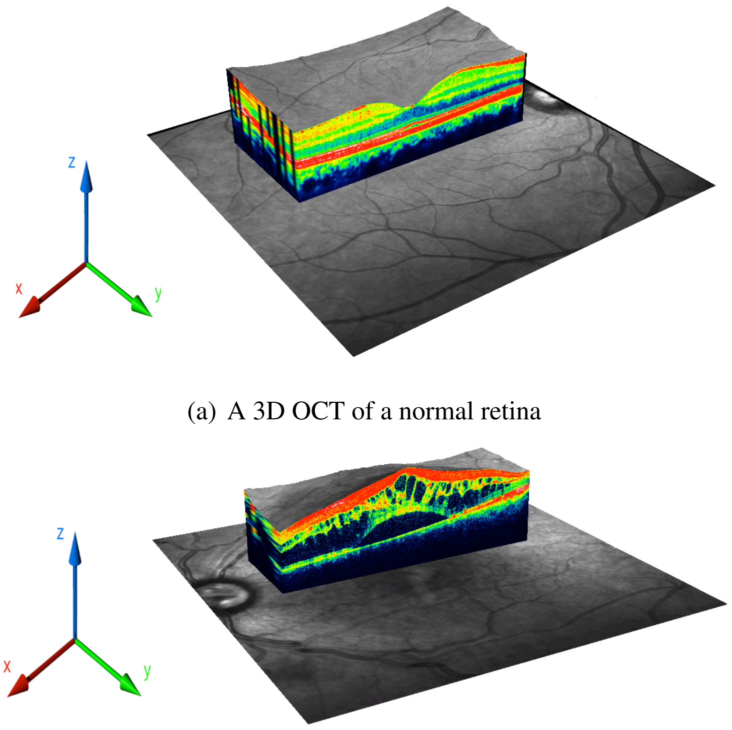
Although the proposed framework has general applicability in the context of 3D data, it was specifically designed for application to collections of 3D retinal OCT images with the objective of detecting the presence (or absence) of Age-related Macular Degeneration (AMD). AMD is the most common global cause of blindness with respect to people aged 50 years and over. It affects the centre of the retina, the "macula", resulting in central vision loss. Traditionally, 2D colour fundus images have been used to detect AMD, and a number of mechanisms for automating this process have been proposed (Hijazi et al., 2011; Zheng et al., 2012). More recently OCT has become widely used for the management of

AMD. Analogous to ultrasound, OCT is a non-invasive imaging technique based on the principle of interference of low coherence light (Huang et al., 1991). OCT can produce cross-sectional images of biological tissues at a high level of resolution and speed. Given the transparent nature of the eye, OCT has become an indispensable tool with which to manage retinal diseases. Figure 1 shows two 3D OCT volumes. Figure 1(a) shows a 3D OCT image of a normal retina where the retina has smooth contours and a regular arrangement of individual retinal layers. Figure 1(b) shows a 3D OCT retinal image with AMD, showing the abnormal change in the retina associated with AMD where fluid and disruption of the retina tissue is evident. Current technical advances enable the generation of a volumetric scan of the retina in seconds, this means that it is relatively straightforward to produce retinal OCT volumes. However, the large quantity of volumetric retina data produced makes it difficult for clinicians to process the data in a timely manner. As noted above, the availability of automated analysis tools has not kept pace with corresponding hardware development.

This paper makes a number of contributions. Firstly the proposed framework obviates the need for resource intensive segmentation. Secondly, the work demonstrates that by adopting the proposed framework it is possible to improve the performance of 3D image classification in the context of AMD screening and diagnosis. Thirdly the paper demonstrates that classification advantages can be gained by adopting hierarchical spatial decomposition based representation methods. More specifically it is argued that by decomposing images into homogeneous regions it is possible to produce effective solutions to the 3D retinal image classification problem. Various ways of extracting homogeneous regions are considered in the paper; the proposed 3D HOG feature vector generation mechanism, for representing individual regions, is of note. Fourthly the paper establishes that IFK based dictionary learning can be effectively employed, in the context of AMD diagnosis, to limit the feature space.

Given the above the main technical and practical contributions, and the novelty of the work, presented in this paper can be summarised as follows:

1. The coupling of homogeneous decomposition with Dictionary Learning for forming feature vectors.
2. A novel and robust approach to 3D retinal image classification using a sophisticated framework for generating classifiers applicable to 3D volumetric data.
3. A mixed oct- and quad-tree decomposition mechanism specifically designed for retinal OCT data volumes.



(a) A 3D OCT of a normal retina

(b) A 3D OCT of an AMD retina

Figure 1: Examples of two 3D OCT images showing the difference between: (a) a “normal” and (b) an AMD retina.

4. A number of histogram based critical functions to support decomposition.
5. A 3D variation of the HOG representation (commonly applied in the context of 2D data) for use with respect to the 3D regions identified during a decomposition.
6. A dictionary learning mechanism applicable to the 3D regions resulting from a homogeneous decomposition.
7. A novel approach to the screening of eye conditions, such as AMD, that can provide real time diagnosis at point of care (point of data acquisition).

In the context of the above the motivations for adapting a hierarchical decomposition based approach to 3D OCT image representation are as follows:

- The decomposing of a space into subspaces helps to identify the most significant localised patterns in the context of the subspace; whereas otherwise, by considering the space in its entirety, localised patterns may be missed. Note that in the context of OCT images the patterns we are interested in generally tend to be in those parts of the image (sub-images) that feature disease.

- Hierarchical decomposition allows for a more "complete" analysis in that the analysis can be directed at different levels of decomposition.

- There is evidence to suggest that image representation methods that rely on localised features tend to produce better performance than those that use global features.

- Spatial relationships between regions can be maintained. Regions that share the same parent remain identifiable (and may be indicative of disease).

The rest of this paper is structured as follows. Section 2 provides a review of related research. The proposed framework is then detailed in Section 3. Section 4 reports on experiments conducted to evaluate the proposed framework. The paper is concluded in Section 5 with a summary of the main findings. Note that in the remainder of this paper when we refer to image data we mean volumetric (3D) image data, and when we refer to regions we mean sub-volumes of a 3D image (volume).

2. Related Work

In this section, we review some of the related work concerning: (i) retinal OCT image classification, (ii) methods for decomposing image data, (iii) image representation and (iv) dictionary learning methods.

2.1. Retinal OCT Classification

Most of the current retinal disease diagnosis tools are designed for application to 2D images. Two examples directed at OCT images, but considering only a single "slice", can be found in Gossage et al. (2003) and Liu et al. (2011). In Gossage et al. (2003) a texture based method was proposed using a Spatial Grey-Level Dependence Matrix (SGLDM) and the Discrete Fourier Transform (DFT). A set of statistical functions was applied to the SGLDM such as energy, entropy, correlation, local homogeneity and inertia to form a feature vector. A Bayesian classifier was used to classify images. In Liu et al. (2011) a method for detecting retinal diseases, including AMD, was presented using a Multi-Scale Spatial Pyramid (MSSP) representation. Histograms of Local Binary Patterns (LBPs) were generated from each sub-block of the MSSP. Dimensionality reduction was then applied to the collection of LBPs using Principal Component Analysis (PCA). All the selected LBPs were then concatenated together to build a single feature vector (one per image). A Support Vector Machine (SVM) classifier was then used for categorising these feature vectors. For both the above two approaches only a single OCT slice across the centre of the retina was chosen for the analysis. This reduces the overall computational cost, but can lead to inaccurate results because of information outside of the selected slice being missed (Albarrak et al., 2012, 2013). It is therefore suggested in this paper that volumetric image mining, using the proposed framework, is more desirable.

The study reported in Quellec et al. (2010) proposed a method for 3D OCT retinal image classification. Here the retinal layers were segmented using a multi-scale 3D graph algorithm. A set of sub-volumes were generated, one sub-volume per layer. A set of statistical features was then generated for each sub-volume. These features included: intensity distribution, run length and features generated using co-occurrence matrices and wavelets. The generated features were used as input to a K-Nearest Neighbour (KNN) classifier. The principal disadvantage of the method is that it relies on the quality of the layer segmentation.

2.2. Image Decomposition Methods

Image decomposition can be conducted in a variety of ways, the simplest is to use a fixed grid (Marszałek et al., 2007; Lazebnik et al., 2006; Yang et al.,

2009) to demarcate a set of equal sized square (2D), or cube (3D), regions. From the literature we can identify two popular forms of grid based decomposition: (i) fixed-sized “window” and (ii) hierarchical. In the first, a predefined rectangular shaped window is used to extract regions from an image (Yang et al., 2009). A problem in some cases is that the edges of the given image may not be covered if the window size is not directly compatible with the image size (although typically we are not interested in things that occur at the edges of images). In the context of hierarchical tree-based decomposition one popular method is the Spatial Pyramid (SP) approach (Lazebnik et al., 2006, 2009). Here, at each level of the decomposition, each region is iteratively divided into four sub-regions (2D) to form a quad-tree. In the 3D case an oct-tree will be formed (Sundar et al., 2007). It is argued that hierarchical spatial decomposition provides for a robust representation because the selected representation method is applied to regions rather than the entire image. In addition issues associated with occlusion are likely to be reduced because of the way images are divided into regions (Tuytelaars and Mikolajczyk, 2008). Image decomposition has been used in various application domains: such as volume rendering (Frisken et al., 2000; Schneider and Westermann, 2003), animation (Chen et al., 2003), segmentation (Lin and Davis, 2010; Voisin et al., 2013) and geographic information systems (Bruzzone and Carlin, 2006). Image decomposition is central to the proposed framework.

2.3. Image Representation

Little work on image representation has been directed at 3D images. One example can be found in Zhao and Pietikainen (2007) who proposed the use of Three Orthogonal Plane LBPs (LBP-TOP) to represent 3D images. The LBP-TOP representation uses LBPs only with respect to neighbouring voxels located in the XY, XZ and YZ planes. An advantage of the LBP-TOP representation is that it avoids the rotational invariant problem (Zhao and Pietikainen, 2007). The Local Phase Quantisation (LPQ) 3D image representation was proposed in Paivarinta et al. (2011). LPQ uses low frequency local Fourier transforms whereby a histogram of the quantised Fourier transform can be generated (Ojansivu and Heikkilä, 2008). In this paper a variation of the HOG representation, adjusted so as to be compatible in the context of 3D data, is proposed.

2.4. Region-based Representations

The approach advocated in this paper for representing regions, identified as the result of an adopted decomposition process, is to represent each region in terms of a feature vector so that the entire image is represented by a set of feature vectors.

Once the sets of feature vectors have been identified, describing the entire input data set, a global set of features can be collected together to form a BoF vector (Csurka et al., 2004). However, for the purpose of classification a single feature vector per image is desirable. An issue is that different images may end up with different length feature vectors if the images have different numbers of regions as a result of a homogeneous decomposition. The proposed dictionary learning strategy avoids this issue. Example methods used to generated dictionaries include: (i) Vector Quantization (VQ) (Lazebnik et al., 2006), (ii) Sparse Coding (SC) (Yang et al., 2009), (iii) Locality-constrained Linear Coding (LLC) (Wang et al., 2010), (iv) IFK encoding (Perronnin et al., 2010) and (v) Super Vector encoding (SV) (Zhou et al., 2010). In Huang et al. (2014) an evaluation is presented on the relative performance of these different coding methods; it was found that the IFK method outperformed the others. Hence this method was adopted with respect to the framework presented in this paper.

3. The Volumetric Data Classification Framework

From the foregoing, in this paper we present a framework for conducting volumetric data classification using homogeneous decomposition and dictionary learning directed at 3D retinal image diagnosis. Figure 2 presents a block diagram describing the framework. The input is set of image volumes, more specifically retinal OCT image volumes. A four stage process is then followed prior to classifier generation/application. The four stages are: (i) image decomposition, (ii) decomposition representation, (iii) dictionary learning and (iv) single feature vector generation. For each stage there are a number of techniques that can be adopted. For the image decomposition the central issues are the nature of the decomposition and the nature of the critical function to be adopted, this is discussed further in Subsection 3.1. Once we have our decomposition each decomposition can be represented according to the individual identified regions using a set of feature vectors (one per region). The proposed feature vector extraction methods are discussed in further detail in Subsection 3.2. To reduce the number of identified features dictionary learning is used, this is detailed in Subsection 3.3. The final stage is single feature vector generation, one per image. This is also described in Subsection 3.3. For completeness the classifier generation process is briefly presented in Subsection 3.4.

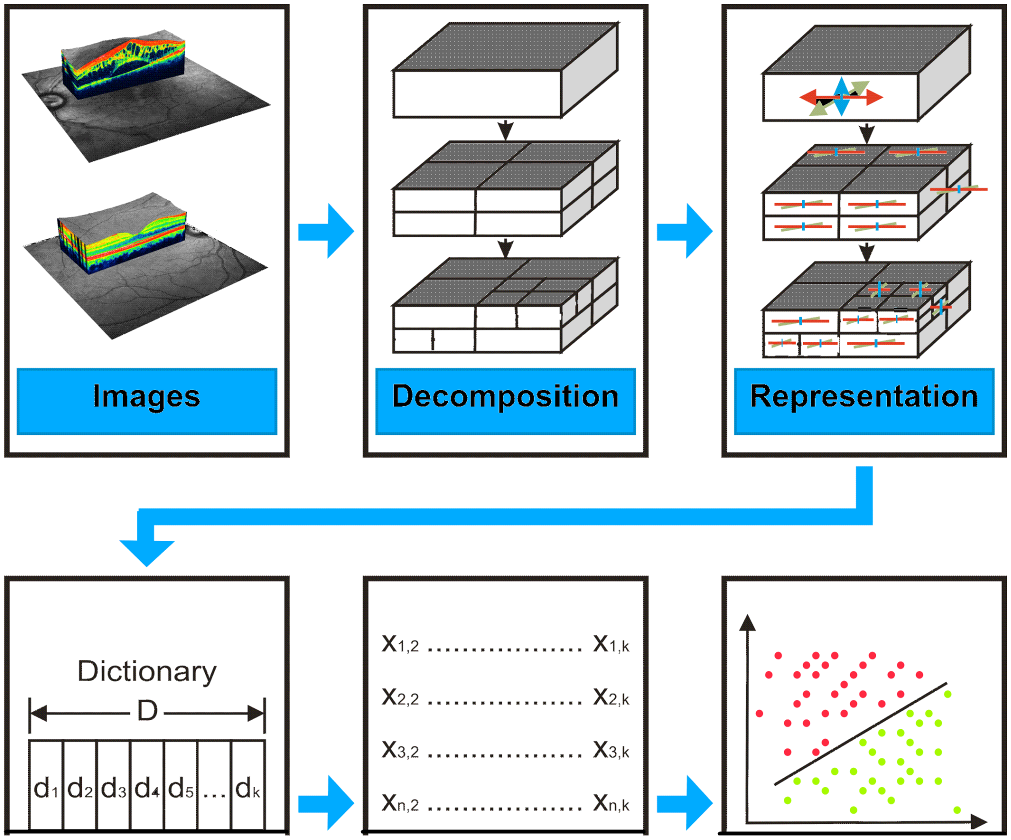

Images
Decomposition
Representation
Dictionary
D
d1 d2 d3 d4 d5 ... dk
x1,2 .... x1,k
x2,2 .... x2,k
x3,2 .... x3,k
xn,2 .... xn,k

Figure 2: Block diagram outlining the stages in the proposed volumetric data classification framework.

3.1. Image Decomposition

Stage one of the proposed framework is image decomposition. A mixed oct- and quad-tree decomposition is proposed, specially designed for use with OCT volumes because of their “oblong” nature whereby the number of pixels (or the number of A scans in a B scan) in one of the dimensions is significantly longer than the other two (see Figure 1). From example, the OCT volumes used in this work typically measure $1024 \times 496 \times 20$ pixels. Thus after the initial oct-decomposition each volume will measure $512 \times 248 \times 10$. If we carried on with the oct-decomposition the volumes would quickly become too thin to be useful, hence the oct-decomposition was stopped after the first iteration and quad-decomposition was adopted instead. However, given some other kind of volume there is no reason why the oct-decomposition cannot be continued, the proposed framework will operate equally well. The decomposition algorithm is presented in Subsection 3.1.1 below. The algorithm will operate with any one of a number of critical functions, the nature of such critical functions is therefore considered in Subsection 3.1.2.

3.1.1. Decomposition

As noted above, in order to be appropriate to the nature of OCT images a mixed oct- and quad-tree decomposition is proposed. On the first iteration of the decomposition the volume is divided into eight sub-volumes; on all of the following iterations each identified subregion is then further divided into only four sub-volumes. As the decomposition progresses the identified regions are stored in a tree data structure T. Algorithm 1 describes the decomposition process. Note that this is a recursive process. The inputs to the algorithm are an input data volume V and a maximum level of decomposition *maxLevel*. The output is the tree data structure T holding the decomposition. We indicate the child i of a node in the tree using the notation $node.i$. The root of the tree is $root$. The algorithm commences by dividing the input volume V int eight sub-volumes $\{s.1, s.2, \ldots, s.8\}$ (line 8). Following this, for each identified sub-volume the $decompose$ procedure is called (line 11). The $decompose$ procedure first tests if the *maxLevel* has been reached and whether the current sub-volume is homogeneous or not (by applying a *critical function*). If so the current volume is not decomposed any further. Otherwise it is divided into four sub-volumes $\{s.1, s.2, \ldots, s.4\}$ (line 18) and the process repeated for each of these sub-volumes with another call to the $decompose$ procedure. Eventually, either the *maxLevel* or homogeneous sub-volumes are arrived at, the recursion then “unwinds”.

Algorithm 1 Pseudocode for the proposed decomposition methods.

```
**Intput:**
$V$ = The input volume
$maxLevel$ = The maximum level,forth decomposition
**Output:**
$T$ = Tree data structure holding the final decomposition

$level \leftarrow 1$
$root$ = Pointer to root of tree $T$
$s.1 \ldots s.8 \leftarrow$ volume $V$ decomposed into eight sub-regions
**for** $i = 1$ to $i = 8$ **do**
    $root.node_i \leftarrow$ detail of $s_i$
    $decompose(level + 1, root.node_i)$
**end for**
exit with $T$

**procedure** $decompose(level, node)$
    **if** $level \geq maxLevel$ **or** $Homogeneous(node)$ **then**
        **return**
    **end if**
    $s.1 \ldots s.4 \leftarrow$ volume at $node$ decomposed into four sub-regions
    **for** $i = 1$ to $i = 4$ **do**
        $node.node_i$ = detail of $s_i$
        $decompose(level + 1, node.node_i)$
    **end for**
**end procedure**
```

3.1.2. Critical Functions for Regional Homogeneity

The decomposition algorithm (Algorithm 1) will operate with respect to a number of critical functions. Recall that a critical function is a mechanism for identifying whether a particular region is homogeneous or not. The point is that they can be used to control the decomposition so that unnecessary decomposition resulting in additional unrequired volumes is not undertaken. As noted above, determining regional homogeneity is an important issue in the context of image decomposition as it defines when the decomposition should be stopped. For the proposed framework five alternate critical functions were considered: (i) Average Intensity Values (AIV) (Hijazi et al., 2011), (ii) Kendall's Coefficient Concordance (KCC) (Zang et al., 2004), (iii) Euclidean Distance (ED), (iv) Kullback-Leibler divergence (KLD) (Johnson and Sinanovic, 2001) and (v) Longest Common Subsequence (LCS) (Vlachos et al., 2003). Note that the last three functions are histogram based.

Using AIV the average intensity values for the region to be decomposed and its potential child regions are first computed. Then a homogeneity value ω is calculated using Equation 1, where s is the total number of child regions, AIV_p indicates the parent AIV and AIV_i the AIV for child region i, $i = [1,s]$. If ω is greater than a specified threshold t then the decomposition is valid and the child regions are added to the collection of regions in the level and the decomposition continues.

$$\omega = \frac{1}{s}\sum_{i=1}^{s}|AIV_p - AIV_i| \tag{1}$$

The KCC function (Zang et al., 2004) operates as follows. For each voxel, a point series is derived comprised of the intensity values of the voxel's nearest neighbours. Then the KCC function is applied with respect to each generated point series and a value, indicative of the homogeneity (similarity) of the point series, calculated. KCC is calculated using Equation 2.

$$KCC = \frac{\sum_{i=1}^{n}(R_i^2 - n\bar{R}^2)}{1/12K^2(n^3-n)}, \tag{2}$$

where: (i) R_i is the sum of the point series for the ith voxel, (ii) $\bar{R}$ is the mean of each point series ($\bar{R} = \frac{(n+1)K}{2}$), (iii) K is the size of the point series (number of selected neighbours for each voxel) and (iv) n is the number of voxels in a given region. The resulting KCC value will range from 0 to 1, where 1 indicates that the sub-volume is a completely homogeneous region and 0 an entirely un-homogeneous region. If *KCC* is less than a specified threshold *t* then the decomposition is valid and the child regions are added to the collection of regions in the level and the decomposition continued.

Histogram based methods are generally considered to be more robust than when single values are used as in the case of AIV and KCC. Using the histogram based methods histograms of intensity values for the region under consideration for decomposition and its potential child regions are first computed. A homogeneity value for each parent-child pair is then computed based on the "distance" (difference) between these histograms. With respect to the work presented in this paper three alternate distance measures are considered: (i) ED, (ii) KLD (Johnson and Sinanovic, 2001) and (iii) LCS (Vlachos et al., 2003). Euclidean distance is the most obvious measure and is calculated as per Equations 3 where *hp* is the histogram of the parent and hc_i is the histogram of the *i*th child (region), both with length *hl*. With respect to the experiments reported later in this paper *hl* was set to 256 bins. If the *ed* distance value, with respect to one child, is greater than a specified threshold *t* then the decomposition is valid and the child regions are added to the collection of regions and the decomposition continues.

$$ed(hc_i,hp) = \sqrt{\sum_{j=1}^{hl}(hc_{i_j} - hp_j)^2} \tag{3}$$

KLD is calculated using Equations 4, 5 and 6, where: (i) *hp* is the parent histogram, (ii) hc_i is the child histogram for child node *i* and (iii) *nc* is the number of child nodes (4 or 8 in our case). If the calculated *KLD* value is greater than a pre-specified threshold *t* then the decomposition is valid and the child regions are added to the collection of regions in the level and the decomposition continued.

$$KLD = (KL1 + KL2)/2 \tag{4}$$

$$KL1 = \sum_{i=1}^{nc} hp*)log(hp) - log(hc_i)) \tag{5}$$

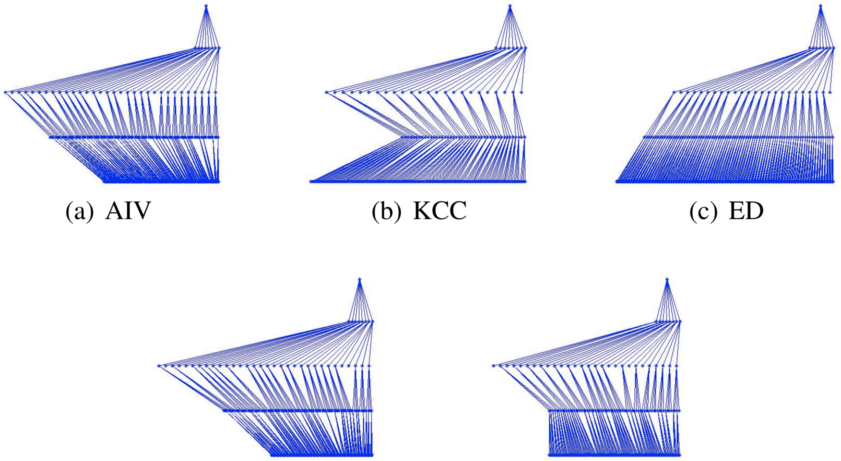
$$KL2 = \sum_{i=1}^{nc} hc_i * (log(hc_i) - log(hp)) \tag{6}$$

LCS is a time series matching method which may be used to compare the similarity between two point series. LCS computes the similarities between two series using the concept of a Minimum Bounding Envelope (MBE) defined by two parameters δ and ε (Vlachos et al., 2003). The LCS value is computed using Equation 7; note that Equation 7 is recursive. With reference to the equation the LCS value is 0 if either hp or hc_i are empty. If the difference between the last value in hp and the last value in hc is less than ε, and the difference in overall size (number of points) of hp and hc is less than or equal to δ, in other words the two point series lie within the boundary of the MBE, we call Equation 7 again but with the last element of hp and hc_i removed (the LCS result is added to 1). Otherwise, if either (or both) of the point series are not entirely contained within the MBE we recalculate two LCS values and choose the maximum; one with hp shortened by one, and one with hc_i shortened by one. The final difference value $D_{\delta,\varepsilon}(hp,hc_i)$ is then calculated using Equation 8 where hs is the size of the two given histograms for the parent node hp and the child node hc_i. If the calculated $D_{\delta,\varepsilon}(hp,hc_i)$ value is less than some threshold t, then the parent region is considered to be homogeneous and so it is not decomposed further. The advantage offered by the LCS mechanism, compared to the other point series comparison mechanisms considered here, is that the complexity of the problem is reduced through the use of the MBE concept.

$$LCS_{\delta,\varepsilon}(hp,hc_i) = \begin{cases} 0 & if hp\ or\ hc_i\ is\ empty \\ 1 + LCS_{\delta,\varepsilon}(hp_{|hp|-1}, hc_{i_{|hc_i|-1}}) & if \mid hp_{|hp|} - hc_{i_{|hc_i|}} \mid < \varepsilon \\ & \mid |hp| - |hc_i| \mid \leq \delta \\ max\{LCS_{\delta,\varepsilon}(hp_{|hp|-1}, hc_i), LCS_{\delta,\varepsilon}(hp, hc_{i_{|hc_i|-1}})\} & otherwise \end{cases} \tag{7}$$

$$D_{\delta,\varepsilon}(hp,hc_i) = 1 - \frac{LCS_{\delta,\varepsilon}(hp,hc_i)}{hs} \tag{8}$$

For the experiments the threshold value t used with the critical functions, as presented above, was set to 0.5. This value was selected because experiments conducted indicated that this produced the best classification outcomes. Figure 3 provides a comparison of the operation of each of the above five critical functions. The figure was generated by applying each critical function to decompose a given data volume down to a maximum level of five (the root is level one). From the figure it can be seen, from the general shape of the trees, that there are clear differences in the operation of these critical functions. For example at the third level it can be seen that the regions are decomposed further for some critical functions and not for others.



(a) AIV

(b) KCC

(c) ED

(d) KLD

(e) LCS

Figure 3: Illustration of the decomposition outcomes using the five different critical functions considered and a given image volume of interest (maximum level of decomposition = 5).

3.2. Region-based Representation

A number of alternative mechanisms for representing regions, generated as described above, can be adopted. In this paper, we propose the use of a HOG representation based on work presented in (Dalal and Triggs, 2005) where the HOG concept was used in the context of 2D data. In order to generate a HOG, the region gradients are first computed followed by the angles between the gradients. These angles are accumulated in *histograms bins* (Dalal and Triggs, 2005). The x-axis of the histogram represents the angles and the y-axis the sum of the gradient magnitudes. For a region, I, the gradient of each voxel, $\nabla I(x,y,z)$ is given as follows:

$$\nabla I(x,y,z) = \frac{\partial I}{\partial x}\mathbf{i} + \frac{\partial I}{\partial y}\mathbf{j} + \frac{\partial I}{\partial z}\mathbf{k} \tag{9}$$

where $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$, and $\frac{\partial I}{\partial z}$ are partial derivatives along the x, y and z directions respectively. These partial derivatives are usually estimated by finite difference schemes such as the *Forward* difference scheme. Each gradient magnitude $|\nabla I(x,y,z)|$ is computed using Equation 10. Following this the orientation "angles" $\theta(x,y,z)$ in each location of the region are extracted using Equation 11 (Scovanner et al., 2007; Lowe, 1999). The values for these angles range from between 0 and 2π. In order to fix the number of bins in each histogram for each region, the range of possible angles was discretized to 8, the angles were thus quantized using $a = angle \times 2 * \pi/8$. For each quantized angle a between 0 and 8, $h = cos(\theta(x,y,z) - a)^{\alpha}$, where a is the selected angle and α is a constant set to 9 (Sivic and Zisserman, 2003), thus forming an orientation histogram for each quantized angle a, $hist(a) = hist(a) + (h. * magnitude)$.

$$|\nabla I(x,y,z)| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2 + \left(\frac{\partial I}{\partial z}\right)^2} \tag{10}$$

$$\theta(x,y,z) = atan2\left(\frac{\partial I}{\partial z}, \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}\right) \tag{11}$$

An entire image volume can thus be represented by a collection of histograms corresponding to the number of sub-volumes (regions) identified during the decomposition which in turn depends on the *maxLevel* value and the homogeneity of the regions (defined in terms of some critical function). Histogram data can be encapsulated in terms of feature vectors. Given a collection of data volumes the entire data space can be represented in terms of a set of feature vectors, one feature vector per each identified region. Collectively the set of features used describes a BoF.

3.3. Dictionary Learning and Feature Vector Generation

From the foregoing, each volume is represented by a collection of feature vectors (drawn from a global BoF). So as to be compatible with classifier generation each volume (image) needs to be represented by a single feature vector. We would also like to reduce the overall size of the BoF so that we are left with a highly discriminative set of features. In the context of the proposed framework the proposed mechanism for achieving both objectives is to use: (i) a dictionary learning mechanism to “learn” a set of highly discriminative set of features and (ii) a coding mechanisms whereby the learnt dictionary is applied to curate single feature vectors (one per image). The dictionary learning was founded on a GMM to model the distribution of the feature vectors describing a collection of images (volumes). For the coding the IFK encoding mechanism was adopted (introduced in Subsetion 2.4 above).

The dictionary learning process is as follows. Let $X = \{\mathbf{x_1}, \dots, \mathbf{x_N}\}$ be the set of feature vectors describing the decomposed regions over a collection of images.

The distribution of the feature vectors within X is estimated using GMM, a mixture of K multivariate Gaussian distributions where K is the number of desired elements in the dictionary. For a given vector $\mathbf{x} \in X$ the conditional probability $p(\mathbf{x}|\Lambda)$ is calculated as follows:

$$p(\mathbf{x}|\Lambda) = \sum_{k=1}^{K} w_k g(\mathbf{x}|\mu_k, \Sigma_k) \tag{12}$$

$$g(\mathbf{x}|\mu_k, \Sigma_k) = \frac{1}{\sqrt{(2\pi)^D \det \Sigma_k}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_k)^T \Sigma_k^{-1} (\mathbf{x}-\mu_k)\right] \tag{13}$$

$\Lambda = \{w_k, \mu_k, \Sigma_k, k = 1, \ldots, K\}$ where w_k, μ_k and Σ_k are the prior probability, mean and covariance matrix of the Gaussian g_k. D is the number of dimensions of the feature vector describing a sub-volume. In general, the parameter Λ is unknown at the beginning of the process and has to be learnt by maximizing the log-likelihood of the set of feature vectors X. This is usually conducted using the Expectation Maximization (EM) algorithm, interested readers are referred to the original paper for further details (Dempster et al., 1977). At the end of the process the dictionary comprises a set of K clusters each defined by w_k, μ_k and Σ_k parameters.

Once the dictionary has been learnt the IFK single feature vector encoding can be commenced. The reasons for using IFK with respect to the work described in this paper are: (i) that (as noted earlier in this paper in Subsetion 2.4) it has been shown to perform well with respect to other image classification applications (Huang et al., 2014), and (ii) the anticipated relatively small size of dictionary required to represent all the possible features. More specifically, IFK uses the gradients of the mean $G_{\mu,i}^{N}$ and covariance $G_{\Sigma,i}^{N}$ of Gaussian i for feature vectors $i = 1 \ldots K$ as shown in Equations 14 and 15.

$$G_{\mu,i}^{N} = \frac{1}{N\sqrt{w_i}} \sum_{n=1}^{N} \gamma_n(i) \left(\frac{x_n - \mu_i}{\Sigma_i}\right) \tag{14}$$

$$G_{\Sigma,i}^{N} = \frac{1}{N\sqrt{2w_i}} \sum_{n=1}^{N} \gamma_n(i) \left[\frac{(x_n - \mu_i)^2}{\Sigma_i{}^2} - 1\right] \tag{15}$$

$$\gamma_n(i) = \frac{w_i \mu_i(x_n)}{\sum_{j=1}^{K} w_j \mu_j(x_n)} \tag{16}$$

Here N is the number of regions of an image. A single feature vector (one per image/volume) is formed by concatenating the two gradients ($G_{\mu,i}^{N}$ and $G_{\Sigma,i}^{N}$). Note that all the generated single feature vectors, for each image, are of size $2 * K * D$.

3.4. Classifier Generation

From the foregoing, a single feature vector $FV(I)$ is formed to describe each image I. For training purposes each feature vector $FV(I)$ was combined with a class label $c_I \in C = \{c_1, c_2, \dots\}$. In the context of AMD screening, the focus for the work presented in this paper, $C = \{+1, -1\}$ was used, where $+1$ indicates a retina with AMD and -1 a normal retina (for Diabetic Retinopathy screening an alternative class set might be used). The class labels were allocated by medical experts. The feature vector representation is compatible with a great many widely available classifier generators. With respect to the evaluation presented in the following section an SVM classifier was used, obtained from the Library for Support Vector Machines (LIBSVM) package (Chang and Lin, 2011)[1]. Once a classifier is trained it can be used to classify new unseen images. SVM classification was adopted because of its widely reported good performance, but other classification models could equally well be adopted.

4. Experiments and Results

In this section we present an evaluation of the proposed volumetric image classification framework. For the purpose of the evaluation a retinal OCT volume data set was used comprised of 140 3D OCT volumes, 68 "normal" (control) volumes and 72 AMD volumes. All the images were acquired using a Spectralis OCT camera[2]. The size of each volume was approximately 1024×496 pixels $\times 19$ slices representing a $6 \times 6 \times 2$ mm retinal volume.

The evaluation was conducted by applying the generated classifiers, built using the proposed framework, to the test data and comparing the predicted labels with the known labels. Ten-fold Cross Validation (TCV) was used throughout, whereby the image dataset is randomly divided into ten sub-sets (each with approximately the same number of images for each class) and the process run ten times, each time with a different $\frac{1}{10}$th of the data as the test set. Note that the use of TCV mitigates against any tendency for overfitting. The Area Under the receiver operator characteristic Curve (AUC) was recorded on each occasion. AUC was used as the evaluation measure because it takes into account the class priors (whereas measures such as accuracy do not). The overall objective was to identify

[1] In the context of LIBSVM the polynomial kernel was used with a coefficient of one and a complexity constant of one.

[2] Manufactured by Heidelberg Engineering, Heidelberg, Germany.

the most appropriate techniques for use in the context of the proposed volumetric classification framework. More specifically the objectives of the classification effectiveness testing were to determine:

- In the context of 3D image decomposition, whether there is any distinction between using a critical function and not using a critical function; and, assuming there is some benefit to using a critical function, the most appropriate critical function to adopt in terms of the five critical functions considered (AIV, KCC, ED, LCS, KLD).
- The most appropriate value for the *maxLevel* parameter to be used with respect to the decomposition. A range of values was considered, $\{3,4,5,6\}$, intuitively selected so that the generated decomposed volumes would be neither too large or too small to be useful.
- The most appropriate dictionary size K to be used (again a range of values was considered $\{32,64,128,256,512\}$).

The obtained results are presented in Subsection 4.1 below. These are then discussed in the context of the above listed objectives in the following three subsections, Subsections 4.2 to 4.4. Computation time is considered in Subsection 4.5. To determine whether, with respect to each objective, the results were indeed statistically significant, the ANalysis Of Variance (ANOVA) test and the Tukey Honestly Significant Difference (HSD) Post-Hoc Test were employed to evaluate the pairwise difference between groups of techniques (Demšar, 2006).

4.1. Evaluation Results

Table 1 presents the average AUC evaluation results obtained with respect to the conducted TCV. It is noteworthy that a best average AUC value of 1.00 is recorded. The column marked NCF (No Critical Function) gives the AUC results when no critical function was used (only a maximum level of decomposition). Recall that the parameter K specifies the dictionary size.

4.2. Use of Critical Function Versus No Critical Function

From the results presented in Table 1 it is clear that better results are obtained when a critical function is used than when a critical function is not used. Average AUC values of 1.00 were recorded with respect to the ED and LCS critical functions. The best overall performing approach was that using the LCS critical function, while ED also produced good results. Thus it can be concluded that the

Table 1: AUC evaluation results for different configurations of the proposed volumetric image classification framework, where $L = \mathit{maxLevel}$ and the ± value is the standard deviation.

K	L	NCF	AIV	KCC	ED	LCS	KLD
32	3	0.93 ± 0.06	0.97 ± 0.00	0.93 ± 0.05	0.96 ± 0.03	0.96 ± 0.02	0.93 ± 0.05
	4	0.97 ± 0.00	0.95 ± 0.00	0.99 ± 0.03	0.99 ± 0.01	0.98 ± 0.02	0.89 ± 0.05
	5	0.94 ± 0.02	0.98 ± 0.00	0.96 ± 0.03	0.99 ± 0.00	0.97 ± 0.02	0.98 ± 0.02
	6	0.91 ± 0.03	0.98 ± 0.00	0.91 ± 0.02	0.99 ± 0.02	0.99 ± 0.00	0.88 ± 0.00
64	3	0.94 ± 0.04	0.94 ± 0.02	0.88 ± 0.00	0.95 ± 0.03	0.98 ± 0.03	0.91 ± 0.06
	4	0.90 ± 0.03	0.94 ± 0.02	0.96 ± 0.00	1.00 ± 0.00	0.98 ± 0.00	0.90 ± 0.03
	5	0.95 ± 0.00	0.99 ± 0.02	0.98 ± 0.02	1.00 ± 0.00	0.99 ± 0.00	0.95 ± 0.00
	6	0.93 ± 0.05	0.98 ± 0.02	0.91 ± 0.00	0.98 ± 0.00	0.98 ± 0.00	0.95 ± 0.02
128	3	0.93 ± 0.05	0.91 ± 0.07	0.94 ± 0.03	0.96 ± 0.03	0.99 ± 0.00	0.93 ± 0.08
	4	0.88 ± 0.02	0.95 ± 0.00	0.99 ± 0.00	0.98 ± 0.00	0.99 ± 0.00	0.90 ± 0.03
	5	0.88 ± 0.02	0.98 ± 0.02	0.97 ± 0.00	0.99 ± 0.00	0.98 ± 0.00	0.94 ± 0.00
	6	0.93 ± 0.02	0.98 ± 0.02	0.88 ± 0.00	0.97 ± 0.00	0.98 ± 0.00	0.97 ± 0.00
256	3	0.91 ± 0.07	0.99 ± 0.02	0.93 ± 0.05	0.96 ± 0.00	0.96 ± 0.00	0.92 ± 0.05
	4	0.91 ± 0.02	0.98 ± 0.00	0.97 ± 0.02	0.98 ± 0.00	0.98 ± 0.00	0.94 ± 0.00
	5	0.92 ± 0.02	0.99 ± 0.00	0.99 ± 0.00	1.00 ± 0.00	0.96 ± 0.00	0.96 ± 0.02
	6	0.86 ± 0.03	0.99 ± 0.00	0.90 ± 0.00	0.99 ± 0.00	0.99 ± 0.00	0.98 ± 0.02
512	3	0.91 ± 0.05	0.93 ± 0.00	0.95 ± 0.04	0.93 ± 0.03	0.95 ± 0.03	0.94 ± 0.02
	4	0.86 ± 0.03	0.96 ± 0.00	0.99 ± 0.00	0.97 ± 0.00	1.00 ± 0.00	0.89 ± 0.02
	5	0.94 ± 0.02	0.99 ± 0.00	0.97 ± 0.03	0.98 ± 0.00	0.99 ± 0.00	0.97 ± 0.00
	6	0.92 ± 0.00	0.99 ± 0.00	0.88 ± 0.02	0.99 ± 0.00	0.97 ± 0.00	0.99 ± 0.02

similarity matching techniques used with respect to the LCS and ED critical functions are more effective than those used with respect to the other critical functions considered.

A box plot representing the results from an associated Tukey test applied using the AUC values presented in Table 1 is given in Figure 4. The figure should be interpreted as follows. Along the x-axis are listed the "groupings" of techniques under consideration, in this case the groups are critical functions. The y-axis represents the recorded AUC classification results. The red line in each box represents the median AUC value while the top and bottom of the box represents the 75% and 25% quartiles with respect to each group. The notch in each box represents the 95% confidence intervals of the measured median AUC value. The whiskers mark the highest/lowest AUC values with respect to each group that are within 1.5 times the interquartile range of the box edges. The red plus signs represent the outliers beyond the data range. In particular, when the notches of two methods do not overlap, the median AUCs can be considered to be significantly different at a 0.05 significance level.

With respect to Figure 4, the ANOVA p-value for comparing the results in the context of the critical functions (and NCF) was $2.5805e-53$, which is much less than 0.01 indicating that there is indeed a substantial statistical difference in the results obtained. From the box plot it can also be seen that the AUC results obtained with NCF were statistically different from the results where critical functions were used. The figure confirms that the best critical function was LCS, which has associated with it the narrowest AUC confidence interval and a median of 0.9744; but it can also be noted that, although the LCS result was slightly better, the results using AIV, ED, KCC and LCS were not statistically different. Whatever the case, it was concluded that LCS was the most appropriate critical function to be used in the context of the proposed framework.

4.3. Best Maximum Level Parameter (L)

From Table 1 it can also be seen that the use of larger values of L produced a better classification performance. In terms of statistical significance the recorded AUC results, with respect to the nature of the value of L, were statistically different with an ANOVA p-value of $2.7988e-10$ (< 0.01). A box plot presenting the results from the associated Tukey test is presented in Figures 5. Along the x-axis, in this case, is listed the "groupings" of techniques according to value for L. The y-axis, as before, records the AUC values. From Figure 5 it can be seen that the results obtained using $L = 3$ were statistically different from the other L values considered. The figure also indicates that $L = 5$ produced the best overall AUC

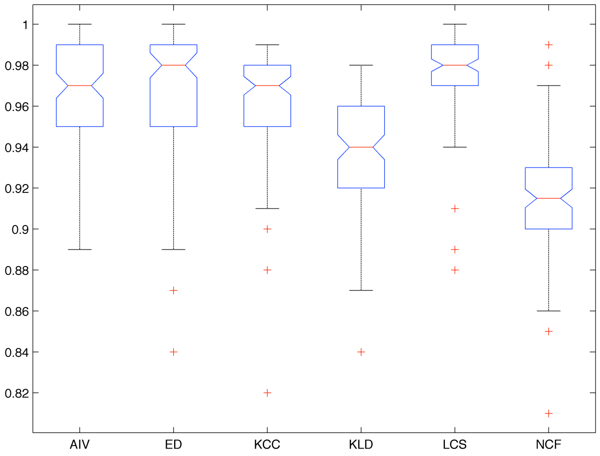


Figure 4: Box plot illustrating the difference in operation between critical functions (and no critical function).

results in that it provided the shortest range with the narrowest confidence interval. It is suggested that $L = 5$ produced good results because it decomposed the image down to a larger number of regions than when lower values of L were used, thus producing a more discriminative set of features. Thus $L = 5$ was adopted as the most appropriate value for L with respect to the proposed framework.

4.4. Best Dictionary Size

With respect to the most appropriate dictionary size K, a range of values for K was experimented with ($\{32, 64, 128, 256\}$). Again, with reference to Table 1, it can be seen that good results were produced regardless of the K value used. However, the calculated ANOVA p-value was 0.0325 (< 0.05) indicating some statistical difference as confirmed by the box plot given in Figure 6. The box plot, presenting the results from the associated Tukey test, has the groupings using different dictionary sizes along the x-axis while the y-axis represents AUC values (as before). From the plot it can be seen that, although the results obtained using different dictionary sizes were similar, a slight improvement featured when $K = 32$ (an AUC median of 0.9537). The similarity between the results using the different dictionary sizes is probably due to the use of IFK which encodes the feature vectors regardless of the dictionary size K used for the GMM.

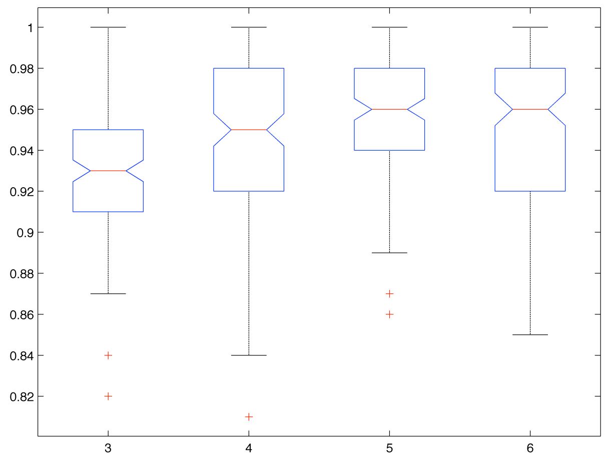


Figure 5: Box plot illustrating the difference in operation using different values for the maximum decomposition level *L*.

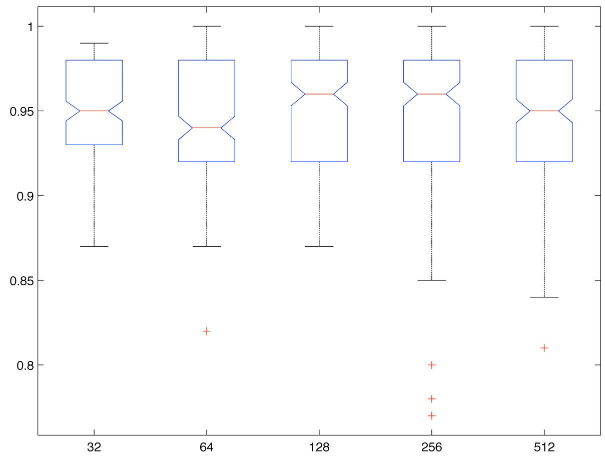


Figure 6: Box plot illustrating the difference in operation when using different dictionary sizes *K*.

4.5. Computation Time

Table 2 shows the average required run time (over ten runs) for the proposed framework to complete (using $K = 32$ because previous experiments, reported above, had indicated that this tended to produce best results). The table shows the Average Decomposition Time (ADT), summed Average Feature Vector Generation Time and Classification Time (AFVGCT), and average Total Execution Time (TET). ADT is the average time required to decompose an image into a set of regions. AFVGCT is the average time required to generate feature vectors for each image plus the time to generate and test the associated classifier. From the table it can be seen that usage of AIV and KCC critical functions tended to be the most efficient (best recorded average time was TET $= 1.75$ seconds using AIV and $L = 3$), while usage of the LCS critical function required more time than the rest of the critical functions considered (TET $= 31.89$ seconds when $L = 6$).

5. Discussion and Conclusions

In this paper we have proposed a framework for classifying volumetric OCT retinal image data using hierarchical decomposition and dictionary leaning. The concept of a critical function is used and a *maxLevel* parameter (L) to control the decomposition. The usage of five different critical functions was considered. Experiments were conducted using a range of values for L. In addition, for the dictionary learning, a value K was used to define the size of the dictionary, thus experiments were also conducted using a number of alternate values for K. The proposed framework was applied to a retinal OCT image collection in the context of the diagnosis (screening) of AMD. The results indicated that the best recorded AUC results (1.00) was obtained using the ED and LCS critical function. Further analysis indicated that the LCS critical function tended to produce the best overall performance. The most appropriate value for L was found to be 5, and the most appropriate value for K was found to be 32.

In the context of AMD detection the experiments demonstrated an excellent performance (best recorded AUC of 1.00). It is thus argued that the proposed framework can be usefully employed for AMD screening purposes (it may also have a role as a training platform). It is also argued that the proposed framework has more general applicability. The same process can be used for the screening of other eye related conditions such as diabetic maculopathy, diabetic retinopathy and glaucoma, also all common causes of vision loss. With respect to diabetic retinopathy many countries have established screening programmes for patients

Table 2: Run time results (seconds) in terms of Average Decomposition Time (ADT), Average Feature Vector Generation Time and ClassificationTime (AFVGCT) and Total Execution Time (TET).

CF	L	ADT	AFVGCT	TET
NCF	3	0.55	12.52	13.07
	4	0.61	15.73	16.34
	5	0.69	26.26	26.95
	6	1.02	27.14	28.16
AIV	3	1.32	0.43	1.75
	4	1.35	1.67	3.02
	5	2.14	6.44	8.58
	6	3.78	16.72	20.5
KCC	3	1.61	0.87	2.48
	4	1.67	2.44	4.11
	5	1.67	4.78	6.45
	6	1.78	13.56	15.34
KLD	3	8.60	0.76	9.36
	4	8.75	2.72	11.47
	5	8.81	4.52	13.33
	6	9.33	15.95	24.92
ED	3	2.56	0.83	3.39
	4	2.57	2.05	4.62
	5	2.69	4.12	6.81
	6	2.80	17.41	20.21
LCS	3	15.63	0.78	16.41
	4	15.82	2.79	18.61
	5	16.05	3.80	19.85
	6	17.29	14.6	31.89

with diabetes. Whatever the case the proposed framework can be usefully employed to provide point of care diagnosis, thus providing patient and resource benefits.

The strengths of the proposed framework may be summarised as follows: (i) it avoids the use of segmentation-based methods, such as that advocated in Quellec et al. (2010), which are time consuming and often unreliable; (ii) the mixed oct- and quad-tree decomposition process which is well suited to OCT image data; (iii) the applicability to 3D OCT images data, in contrast to the work presented by Gossage et al. (2003) and Liu et al. (2011), which was applied to 2D images; and (iv) the excellent resulting classification performance (AUC values of 1.00).

Although the work presented demonstrates clear potential the presented evaluation has some limitations. Firstly only a relatively small dataset was used due to the absence of readily available OCT benchmark data for the evaluation. Secondly, due to hardware limitations, the *maxLevel* parameter was limited to a maximum value of 6, it may thus be the case that the level of decomposition whereby an optimal classification performance is achieved may not always have been reached (although AUC results of 1.00 were recorded). Thirdly, although the results were encouraging, the parameters for the SVM classifier were not optimized, nor were experiments with alternative classifier generators conducted. Future work will thus be directed at more comprehensive large-scale experimentation. However, from our experience to date, our expectation is that the performance will be similar if not better to that reported in this paper. It might also be interesting to investigate the outcomes when much larger values of K are used, for example $K = 1024$, however our experiments to date have indicated that the computational resource and run time required make such experiments impractical. For the multiclass classification problem, there are many methods that could be applied if the binary classification performed well, such as one-vs-rest or one-vs-one. These methods make use of repeated applications of a binary classification model. Alternatively a multi-class classifier can be adopted.

In conclusion a novel framework, founded on homogeneous decomposition and dictionary learning, for 3D image (volumetric) classification has been presented. The framework was evaluated using a 3D retinal OCT image set for the diagnosis of AMD, and some excellent results were produced. Given that the proposed framework is generic in nature it is anticipated that it will be equally applicable to other ophthalmological 3D retinal image classification problems and more general 3D classification problems.

References

Albarrak, A., Coenen, F., Zheng, Y., 2013. Classification of volumetric retinal images using overlapping decomposition and tree analysis, in: IEEE 26th International Symposium on Computer-Based Medical Systems, pp. 11–16.

Albarrak, A., Coenen, F., Zheng, Y., 2014. Dictionary learning-based volumetric image classification for the diagnosis of age-related macular degeneration, in: Machine Learning and Data Mining in Pattern Recognition. Springer, pp. 272–284.

Albarrak, A., Coenen, F., Zheng, Y., Yu, W., 2012. Volumetric image mining based on decomposition and graph analysis: An application to retinal optical coherence tomography, in: IEEE 13th International Symposium on Computational Intelligence and Informatics, pp. 263–268.

Bruzzone, L., Carlin, L., 2006. A multilevel context-based system for classification of very high spatial resolution images. IEEE Transactions on Geoscience and Remote Sensing 44, 2587–2600.

Chang, C.C., Lin, C.J., 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2, 27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/čjlin/libsvm`.

Chen, M., Silver, D., Winter, A.S., Singh, V., Cornea, N., 2003. Spatial transfer functions: A unified approach to specifying deformation in volume modeling and animation, in: Proceedings of the 2003 Eurographics/IEEE TVCG Workshop on Volume Graphics, pp. 35–44.

Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C., 2004. Visual categorization with bags of keypoints, in: Workshop on Statistical Learning in Computer Vision, pp. 1–2.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 886–893.

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of Royal Statistical Society, Series B 39, 1–38.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30.

Frisken, S.F., Perry, R.N., Rockwood, A.P., Jones, T.R., 2000. Adaptively sampled distance fields: A general representation of shape for computer graphics, in: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 249–254.

Gossage, K.W., Tkaczyk, T.S., Rodriguez, J.J., Barton, J.K., 2003. Texture analysis of optical coherence tomography images: feasibility for tissue classification. Journal of Biomedical Optics 8, 570–575.

Hijazi, M.H.A., Jiang, C., Coenen, F., Zheng, Y., 2011. Image classification for age-related macular degeneration screening using hierarchical image decompositions and graph mining, in: Machine Learning and Knowledge Discovery in Databases. Springer, pp. 65–80.

Huang, D., Swanson, E., Lin, C., Schuman, J., Stinson, W., Chang, W., Hee, M., Flotte, T., Gregory, K., Puliafito, C., et, a., 1991. Optical coherence tomography. Science 254, 1178–1181.

Huang, Y., Wu, Z., Wang, L., Tan, T., 2014. Feature coding in image classification: A comprehensive study. IEEE Transactions on Pattern Analysis and Machine Intelligence 36, 493–506.

Johnson, D.H., Sinanovic, S., 2001. Symmetrizing the Kullback-Leibler distance. IEEE Transactions on Information Theory 1, 1–10.

Lazebnik, S., Schmid, C., Ponce, J., 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2169–2178.

Lazebnik, S., Schmid, C., Ponce, J., et al., 2009. Spatial pyramid matching. Object Categorization: Computer and Human Vision Perspectives 3, 4.

Lin, Z., Davis, L., 2010. Shape-based human detection and segmentation via hierarchical part-template matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, 604–618.

Liu, Y.Y., Chen, M., Ishikawa, H., Wollstein, G., Schuman, J., Rehg, J.M., 2011. 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.

Lowe, D.G., 1999. Object recognition from local scale-invariant features, in: The proceedings of the seventh IEEE international conference on Computer vision, pp. 1150–1157.

Marszałek, M., Schmid, C., Harzallah, H., van de Weijer, J., 2007. Learning object representations for visual object class recognition. Visual Recognition Challange workshop, in conjunction with ICCV.

Ojansivu, V., Heikkilä, J., 2008. Blur insensitive texture classification using local phase quantization, in: Proceedings of the 3rd International Conference on Image and Signal Processing, pp. 236–243.

Paivarinta, J., Rahtu, E., Heikkilä, J., 2011. Volume local phase quantization for blur-insensitive dynamic texture classification, in: Proceedings of the 17th Scandinavian Conference on Image Analysis, pp. 360–369.

Perronnin, F., Sánchez, J., Mensink, T., 2010. Improving the Fisher kernel for large-scale image classification, in: Computer Vision, pp. 143–156.

Quellec, G., Lee, K., Dolejsi, M., Garvin, M.K., Abràmoff, M.D., Sonka, M., 2010. Three-dimensional analysis of retinal layer texture: identification of fluid-filled regions in sd-oct of the macula. Medical Imaging, IEEE Transactions on 29, 1321–1330.

Schneider, J., Westermann, R., 2003. Compression domain volume rendering, in: IEEE Visualization, pp. 293–300.

Scovanner, P., Ali, S., Shah, M., 2007. A 3-dimensional SIFT descriptor and its application to action recognition, in: Proceedings of the 15th international conference on Multimedia, pp. 357–360.

Sivic, J., Zisserman, A., 2003. Video google: a text retrieval approach to object matching in videos, in: Proceedings of the Ninth IEEE International Conference on Computer Vision, pp. 1470–1477.

Sundar, H., Sampath, R.S., Adavani, S.S., Davatzikos, C., Biros, G., 2007. Low-constant parallel algorithms for finite element simulations using linear octrees, in: Proceedings of the 2007 ACM/IEEE Conference on Supercomputing,, pp. 1–12.

Tuytelaars, T., Mikolajczyk, K., 2008. Local invariant feature detectors: A survey. Found. Trends. Comput. Graph. Vis. 3, 177–280.

Vlachos, M., Hadjieleftheriou, M., Gunopulos, D., Keogh, E., 2003. 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, pp. 216–225.

Voisin, A., Krylov, V., Moser, G., Serpico, S., Zerubia, J., 2013. Classification of very high resolution sar images of urban areas using copulas and texture in a hierarchical Markov random field model. IEEE Geoscience and Remote Sensing Letters 10, 96–100.

Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y., 2010. Locality-constrained linear coding for image classification, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3360–3367.

Yang, J., Yu, K., Gong, Y., Huang, T., 2009. Linear spatial pyramid matching using sparse coding for image classification, in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1794–1801.

Zang, Y., Jiang, T., Lu, Y., He, Y., Tian, L., 2004. Regional homogeneity approach to FMRI data analysis. NeuroImage 22, 394–400.

Zhao, G., Pietikainen, M., 2007. Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 915–928.

Zheng, Y., Hijazi, M.H.A., Coenen, F., 2012. Automated “ disease/no disease” grading of age-related macular degeneration by an image mining approach. Investigative Ophthalmology & Visual Science 53, 8310–8318.

Zhou, X., Yu, K., Zhang, T., Huang, T.S., 2010. Image classification using super-vector coding of local image descriptors, in: Computer Vision. Springer. Lecture Notes in Computer Science, pp. 141–154.