International Journal of Signal Processing, Image Processing and Pattern
Vol. 2, No. 1, March, 2009

Applying Feature Extraction for Classification Problems

Foon Chi Francis Chui
University of Tasmania
francis.chui@gmail.com

Ray Williams
University of Tasmania
R.Williams@utas.edu.au

Ivan Bindoff
University of Tasmania
ibindoff@utas.edu.au

Byeong Ho Kang
University of Tasmania
bhkang@utas.edu.au

Abstract

With the wealth of image data that is now becoming increasingly accessible through the advent of the world wide web and the proliferation of cheap, high quality digital cameras it is becoming ever more desirable to be able to automatically classify images into appropriate categories such that intelligent agents and other such intelligent software might make better informed decisions regarding them without a need for excessive human intervention. However, as with most Artificial Intelligence (A.I.) methods it is seen as necessary to take small steps towards your goal. With this in mind a method is proposed here to represent localised features using disjoint sub-images taken from several datasets of retinal images for their eventual use in an incremental learning system. A tile-based localised adaptive threshold selection method was taken for vessel segmentation based on separate colour components. Arteriole-venous differentiation was made possible by using the composite of these components and high quality fundal images. Performance was evaluated on the DRIVE and STARE datasets achieving average specificity of 0.9379 and sensitivity of 0.5924.

1. Introduction

Up until recently, medical imaging has been focused on methods in the annotation of images. Their purpose was to encompass both diagnosis and feature extraction into one complete system. A flaw in this method of knowledge representation for medical imaging is that the distinction between a clinical diagnosis and the feature extraction steps are not clear. They often convey the same information but it may not always be the case. Some situations require more knowledge from an expert for a full diagnosis.

This poses two questions, is it possible to delineate the process of feature extraction and clinical diagnosis clearly and use an incremental learning method for maintaining the knowledge required for feature extraction in image processing? Given these questions, the following hypothesis was made: An incremental learning method can be used to develop and maintain the knowledge required for extracting features from retina images so these can be used in the automated interpretation of the images.

One particular field that makes frequent use of and is well suited to imaging is ophthalmology, the sub domain of medicine that studies the eye. A diagnostic test performed routinely by optometrists and ophthalmologists is the inspection of the fundus of the eye using ophthalmoscopes and fundus cameras.

There are three main features of interest in a fundal image: the vascular tree, optic disc and macula. These features are found within the posterior fundus of the eye. A factor that makes these areas of particular interest is the visibility of abnormalities that can occur, such as: tortuous vessels, optic disc irregularities and colored irregularities around the macula.

Within retinal images are many features from which an expert is trained to perform diagnostic tests; a major feature is the vascular system of the retina. From these vessels: cardiovascular disease, retinopathy of prematurity (ROP) and diabetes are some diseases that may manifest in abnormal retinal topology and microvasculature [1].

2. Previous literature

In image processing, the general approach taken is frequently made in a stepwise fashion. Fundamentally, information is extracted by the relationship between a pixel and its. Generation of new pixels from neighbouring pixel properties is called neighbourhood processing or more commonly spatial filtering [2].

The goal for any segmentation algorithm is to find pixel boundaries at locations where an object to be identified and the background can be distinguished. Classifiers of these boundaries often exploit intensity properties of pixels. The properties that these classifiers use can be categorized into two types, the rate of change in photometric intensities, and the absolute difference between two regions of high and low intensity.

2.1. Retinal images

To identify features such as arterioles and veins, the optic disc, pathological features or other retinal landmarks, requires the differentiation of these object regions from the background regions. Using the process of image segmentation provides this ability and allows for feature identification.

Initially, low-level image information must be gathered and then incorporated with higher-level knowledge for the appropriate interpretations and diagnosis. In retinal images, locating the optic disc, fovea and retinal vessels provides the initial low-level image information [3].

The optic disc is the region where the optic nerve, retinal artery and veins emanate providing the source of circulation and innervations. Hoover and Goldbaum [4] found that the only consistent feature for identifying the optic disc was the convergence of retinal blood vessels to this region. Using a previous technique of vessel segmentation they presented a novel algorithm called fuzzy convergence for the localisation of the optic disc.

The vascular network is perhaps one of the most significant features in the retina. Located in the posterior segment of the ocular fundus, it is unique to other vascular tissue in that it can be observed non-invasively in vivo. Vessels emanate from the optic disc and radiate out in a tree structure. There are two types of vessels, veins and arterioles, which carry blood to and from the heart. Oxygenation of blood in these vessels has an effect on their visual appearance with veins having a relatively darker bluish tint relative to the brighter red arterioles [5].

The structure of these vessels also causes morphological difference in visual appearance. An arteriole has thicker walls with more smooth muscle than a vein. Conversely the blood pressure within a vein is significantly lower than in an arteriole. The effects of these properties cause arterioles to be rounder and thus their reflective properties are different from

that of veins, which are flatter. These visual properties are what many feature extraction techniques use for identifying vascular structures.

One main issue that must be overcome is the problem with varying intensity from images taken with the fundus camera. This property is due to the rounded nature of the retina, the medium through which light needs to travel through and optical properties of the lens, this is demonstrated in Figure 1.

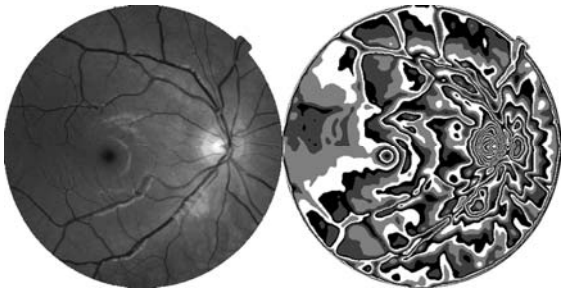

Figure 1. Illustrating the intensity variation within a retinal image. Left - raw intensity levels in the green channel in the original image. Right - iso-contours of the intensity level.

Gaussian or derivative kernels in multiple orientations have been used to match vessel profile in many vessel segmentation techniques. Hoover's vessel segmentation method used a Gaussian function as the model for vessel profiles. Piecewise Gaussian models have also been used for differentiating vessel types. It was found that the central reflex was more apparent in arterioles and that it was possible to uses two Gaussian functions to depict the vessel and the central reflex [5].

Given that images do not tend to be equally sized and their resolutions may differ, a fixed sized filter is often not appropriate for segmentation of features. Multi-scale matched filters attempt to solve this by the use of scalable kernels. This had the added advantage for vessel segmentation as calibers decrease along the tree structure the further away a segment is from the optic disc.

Mendonça & Campilho [6] presented an automated algorithm for the segmentation of the retinal vascular network. They used a mixture of matched filters for finding centerlines and multi-scale filters for segmenting regions that were later used for region growing and masking. The segmentation process followed three phases.

The backgrounds of retinal images usually have a variable intensity from region to region. This irregularity in local intensity can interfere with the vessel segmentation process, which relies on intensity profiles. Thus the first step taken was to normalize for each background region. To achieve this, two methods were used: the original image was converted to monochrome using the green channel of the RGB colour space and the luminance channel in L*a*b* colour space. These conversions extracted the highest contrasting monochrome image directly from the original. After this, the subtraction of an estimate of the background using a large arithmetic mean kernel normalized the image. This provided a uniform luminance for all regions while keeping the contrast within each region.

The approaches taken in the previously discussed techniques apply various combinations of imaging methods, through rules found a priori, to arrive at their annotations. Another approach would be with a supervised method of machine learning. This process required classified images as exemplar cases for the purposes of training a system that has the ability to learn patterns within this data.

This approach was taken by Sinthanayothin et al. [7] where a multi-layer perceptron neural network was used. The image had been pre-processed using a Canny edge operator to increase the separability between vessel and non-vessel pixels. Training was done with a dataset based on 10x10 sized tiles from these pre-processed images. The corresponding tiles of manually labeled vessels were provided to train the system.

Staal et al. [8] believed that a pixel representation was not optimal for vessel structures. They used a k-nearest neighbor classifier for automated vessel labeling they called PBM (primitive-based method).

2.2. Summary

The current solutions in retinal imaging have been to design specific imaging processes to extract features for an expert to review as annotations. These annotations may or may not relate directly to clinical diagnoses, but often highlight features of significance. Features may exist in multiple scales, ranging from the pixel-level up to the image-level. They can also provide metrics for clinical evaluation that would not have been possible before.

After the review of some current techniques in retinal imaging, it was apparent that the classification and segmentation from these technique have be done as a single process with the correlation of a diagnosis which one could consider as expert knowledge. Focusing on retinal vasculature some produced binary classifiers for vessel segmentation while others produced quantitative information in the form of bifurcation angles.

It has been shown that organization of computer vision processes can be done with knowledge-based systems. While these have not been directly related in that induction techniques were used for knowledge acquisition, it has been shown that the maintenance of knowledge bases in medical fields can be done by other means.

In the field of retinal images in particular, diagnoses are often related to the context of where other features are located not whether a feature is existent. Therefore it is a suitable candidate for the application of Multiple Classification Ripple Down Rules (MCRDR) [9] or similar case based knowledge acquisition method for the maintenance and acquisition of knowledge.

3. Methodology

The image analysis techniques discussed in the previous section all work with knowledge organized into rules of some description. These rules were created either a priori or found through training. Both techniques described require complete knowledge for an effective rule set.

In the case of rule-based image segmentation, prior knowledge of image size, scale, along with what features could be extracted and their corresponding techniques were needed before a system to be developed. Any form of maintenance on these rules would require a redesign of the algorithm to ensure correctness of classifications and diagnoses were preserved.

In supervised training, manual annotations were used as exemplar cases. Labeling of these images on a pixel-by-pixel basis is laborious, requiring hours of work. This would be time that could be better spent by medical experts with patients. Thus maintenance of this knowledge could be equally as time consuming and expensive.

For incremental learning, a symbolic representation of image information is required for manipulation by machine and human. This symbolic representation exists as case information relating to the image containing information such as image features. A method for extracting case information was required such that an incremental method could be used as a basis for further improvements of the system.

An inherently symbolic representation is a per-pixel basis, but this is unsuitable for working with humans, so it is necessary to introduce a form of image abstraction to create another level of symbolic representation that is recognizable by human and readily produced by machine. This symbolic representation attempts to interpret how one obtains meaning from an image. That is, not on the pixel level but through the use of higher-level morphological and intensity differences based on multiple pixels. A bootstrapping process was needed to produce this initial information from which both human knowledge can be obtained and identified by the machine in an automated process.

The imaging techniques discussed in the previous section followed a paradigm of automated image annotation with the goal of replicating identifiable features a human would find useful and significant. An extension to this would be to retain the reasoning behind each annotation, much like traditional expert systems where in this case the imaging expert is providing the reasoning for choosing each step or process.

To test the feasibility of using automatic annotation as a form of symbolic information, experiments were conducted using retinal images provided by Finian MacCana, an Optometrist, and further evaluation used images from the DRIVE database and STARE database (described further below) for finding segmentation performance.

The majority of previous vessel extraction methods from retinal images have worked with only the green band of an image. The reasoning for this was that major colour contrast was found in this band. Mendonça and Campilho [6] however used luminosity, which weights each colour bands intensity according to normal human perception of that colour. It was found that it increased the accuracy of vessel segmentation when compared with the green band alone.

As part of the knowledge engineering process, to reduce an image into manageable amounts of data, quantization of the original image was performed. The method chosen was the segmentation of each image using thresholds to produce discrete spatial regions. For this to occur, an illumination equalisation was required to produce a localised intensity mean that was consistent prior to threshold selection. This was followed by the classification of each region into foreground and background. The decision was made to retain all three separate bands in the process of segmentation to retain as much information as possible for later processes.

3.1. Illumination

Due to the nature of retinal fundus, illumination across these images is often irregular. To overcome this, the process used by Hoover & Goldbaum [4] was applied to each pixel. In this

process, the quantity $I_{eq}(i, j)$, which is the equalised intensity of pixel (i, j), is given by the following equation:

$$I_{eq}(i,j) = I(i,j) + \mu_{desired} - W_{\mu}(i,j)$$

Where $W(i, j)$ is a sub-image of NxN dimensions centred on the pixel (i, j). I(i, j) is the original intensity of a pixel located at (i, j), µdesired the desired mean intensity and Wµ(i, j) the actual mean intensity for the window centred on pixel (i, j). In this case N=49 and $\mu_{desired}$=128 where N must be an odd number.

Initially, the black outer edge was masked with the mean of pixel intensities above a threshold of 30 in any band. After this, a mean image was obtained using a two pass filtering approach. Finally, subtracting the mean image from the edge-masked image using an offset of 128 produced the illumination-equalised image that is illustrated in bottom right of Figure 2.

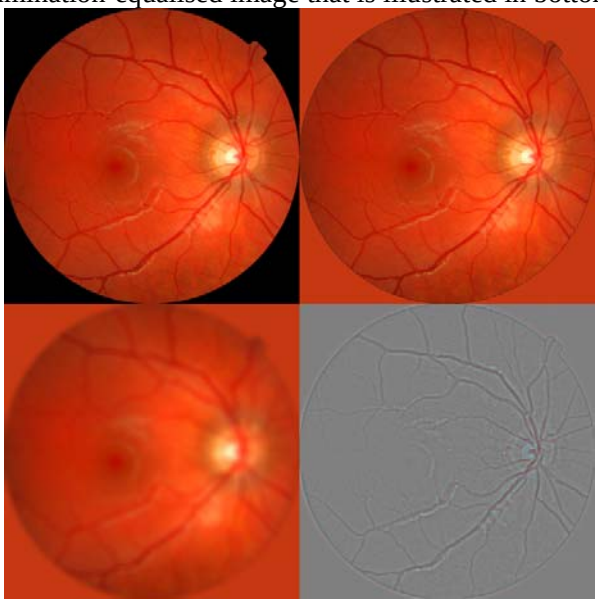

Figure 2. Shows the steps taken to produce the illumination-equalised image. The images correspond to: tl) Original, tr) Edge masked, bl) Mean intensity using two pass filtering with a window size of 49x49 and br) Illumination equalised.

From the illumination-equalised image, it was possible to perform localised operations without the gradual intensity gradation affecting results while retaining local contrast of vessel segments. This was made possible by replacement of the original mean intensities and with a constant value µdesired. Producing an image that retains the localised intensity variation for each pixel (i, j) and the surrounding sub-image $W(i, j)$ with N dimensions but removing the variation for any sub-image greater than these dimensions.

A process similar to boundary detection by Chow & Kaneko [10] was applied to retinal images with the goal of vessel segmentation. This required global equalisation to produce localised sub-images that overlapped at the pixel level. Using the equalised image, a localised intensity histogram obtained from overlapping tiles was used as the approximation of the discrete probability density function of intensity levels. This allowed for the automatic

selection of thresholds. Segmentation using an adaptive threshold selection process was performed on each band using their respective intensity level distributions. The selected threshold obtained from these histograms was applied to the centre tile within the neighbourhood. In the case of tiles at the edge of an image, missing tiles were omitted.

The process for computing a localised histogram was performed on a tile-by-tile basis; partitioning the image into sub-images of NxN dimensions with N=48. Each histogram for unsupervised threshold selection was formed using a neighbourhood of 3x3 tiles overlapping by 8/9th.

Using nine tiles for formation of the histogram was a mechanism for averaging the intensity distributions while retaining more local intensity information when compared to global histogram distributions. By averaging the intensity distributions, it was possible to prevent any over-fitting to an intensity profile of any single tile while preserving the continuity of vessel regions from tile to tile.

Consider each histogram of the intensity levels where $i = [0, 1, 2, \ldots, L-1]$ and L is number of discrete intensity levels. The number of pixels at intensity level i is n_i and the total number of pixels $N = n0 + n1 + \ldots + nL-1$. Each normalised histogram can be represented as a discrete probability density function with the probability distribution pi written as:

$$p_i = {}^{n_i}\!/_{N}, \quad p_i \geq 0, \sum_{i=0}^{L} p_i = 1$$

Histograms for each tile was generated in stripes of current, previous and next row in order to prevent the need for recalculating intensity totals for any single tile. The tile neighbourhood for each individual tile was retained with each step adding and removing the new and old tile histograms respectively.

4. Threshold selection

The second method of automatic threshold selection tested was described by Gonzalez, Woods and Eddins [2]. In this segmentation process only two classifications were required, background and foreground pixels (non-vessel and vessel). Clustering of the histogram into these two classifications gives the boundary for automatic threshold selection. The following steps were used to find the threshold value for each colour component:

1. An initial threshold T was selected at the mean of the distribution weighted by n_i given by:

$$T = \frac{\sum_{i=0}^{L-1} n_i i}{N}$$

2. Where by dividing the histogram into two classifications, $C0 = [0, \ldots, T - 1]$ and $C1 = [T, \ldots, L - 1]$ to produce two distributions $G0$ and $G1$ each with $N0$ and $N1$ pixels respectively. Within these distributions are the weighted means μ_0 and μ_1 calculated by:

3. Using these two means, a new T is calculated as follows:

$$\mu_0 = \frac{\sum_{i=0}^{T-1} n_i i}{N_0}$$

$$\mu_1 = \frac{\sum_{i=T}^{L-1} n_i i}{N_1}$$

$$T = \frac{\mu_0 + \mu_1}{2}$$

4. If the new threshold T is not the same as the previously calculated threshold, repeat the steps 2 to 4 until there is no change in T between these steps.

This algorithm for threshold selection can provide multiple thresholds for a predefined k number of classifications. This differentiation may not produce any useful classifications and in this case was not considered for vessel segmentation. It would also have been possible to include multiple dimensions for clustering, such as (i, j) image coordinate space, in conjunction with histogram.

Results of this segmentation process using k-means clustering on the histogram are shown in Figure 3. By applying the illumination equalisation process prior to threshold segmentation, the tile artifacts from binarizing an image with intensity gradation were reduced significantly.

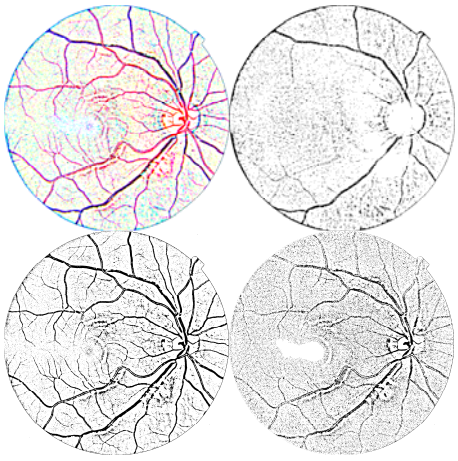

Figure 3. Shows a tile-level adaptive threshold using 1D k-means clustering with illumination equalisation.

4.1. Connected components analysis

After segmentation of the original image, a process producing raw binary images in red, green and blue bands, an analysis of pixel connectivity was performed to identify objects within these segments. This produced regions each with attributes that were human-machine manageable. This was the intermediary containing information both would recognize and provided a link from which both could work.

The resultant binary images from each colour component were partitioned into NxN dimension tiles with N=48. An exception to this was for the last tile in each row and column where a tiles height and width could be less. Segment analysis to label individual regions was performed on a tile-by-tile basis to create localised feature space information. These labeled regions within each tile would allow for the creation of rules based on localised feature space information, which would contain knowledge specific to this locale.

To visualize the area differences from region to region, each region was assigned a hue between red and blue (0° and 240° respectively in HSV model) with saturation and value at 100%). This is illustrated on Figure 8 for the example image from the MacCana dataset.

From initial inspection, it was possible to identify a coverage pattern where background regions with tile coverage close to 50% or less were more likely to contain blood vessels. Tiles with background regions having coverage higher had a lower likelihood of containing vessels. Masking the background regions allowed for the identification of regions that were classified as vessels and their respective sizes (Figure 4 bottom row). This assisted in the identification of false positives that have a tendency to be disconnected from the vascular network.

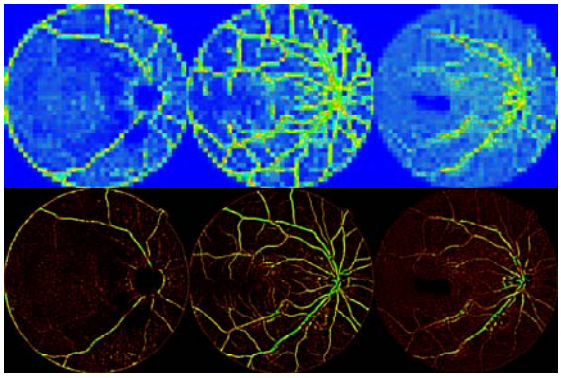

Figure 4. Shows the distinct regions within each tile as hues. Red corresponding to single pixel sized regions, green 50% coverage and blue corresponding to 100% tile coverage.

5. Evaluation

Tests were performed using a range of retinal images from high-resolution images using current retinal imaging technology to standard datasets used for comparative assessment of classifiers.

5.1. Assessment of segmentation

For the assessment of the vessel segmentation algorithm, datasets from two publicly available retinal image databases were used. This provided a set of results that were comparable with the performance of other vessel segmentation algorithms. Both sets included manual annotations by separate observers for both a grounds truth and an assessment of their annotation accuracy.

Analysis of segmentation methods was through the use of receiver operator characteristics for the pixel classifier. This was done by using the manual annotations as the ground truth of each pixel and the segmented image pixels as the hypothesized classifications.

Comparison of the pixel classifier performance and their visualization was done using receiver operator characteristics (ROC) of each segmentation method. In this scenario, each pixel is an instance with a predicted class and the actual class. These pixels can be either positive or negative for being vessel or non-vessel respectively for both actual and predicted.

Given an instance, each pixel has one of four possible outcomes, these were marked as true positive (TP), false positive (FP), true negative (TN) or false negative (FN). The truth in this case was assigned classifications from the manually segmented images in Set A. Hypothesized pixel classification was determined by the segmentation method being tested.

Where TPR is the true positive rate, which provides an indicator of how accurate the classifier is at identifying and segmenting vessel pixels. This is sometimes referred to as the sensitivity of a classifier. FPR is the false positive rate, which indicates how inaccurate it is when classifying non-vessel pixels. Another term commonly used is the specificity of a classifier, which is 1 – false positive rate.

For the comparison of classifiers that produced soft-classifications, classifiers that provide a score or probability instead of a binary value, calculation of the *AZ* (area under the receiver operator characteristics curve) was performed. The *AZ* provides another indication of the classifier performance where a purely random classifier producing an *AZ* of 0.5 when there are two classes.

5.2. The DRIVE dataset

The DRIVE (Digital Retinal Images for Vessel Extraction) database is a collection of 40 retinal images, 7 of which contain pathological indicators, obtained from a screening program at the University Medical Centre Utrecht, Netherlands [11]. This set was originally compressed in JPEG-format captured by a Canon CR5 non-mydriatic 3CCD camera with a 45° field of view, which produced images 540 pixels in diameter.

The test set in this database (20 images) contains vessels manually segmented by three observers: a computer science student, a clinical expert and an image-processing expert. The clinical expert, an ophthalmologist, trained the observers to mark pixel-by-pixel as vessels when they were 70% certain that a vessel was present. The 40 hand-segmented vessel images were divided into two sets that will be referred to as ‘Set A’ and ‘Set B’. Set A contains 14 images annotated by the student and 7 by the ophthalmologist. Set B contains 20 images annotated by the image-processing expert. Mask images were in this set, denoting the regions of outside the FOV.

5.3. The STARE dataset

The STARE (Structured Analysis of the Retina) database consists of 20 images containing ten pathological and ten normal retinas. These consisted of slides that were captured with a TopCon TRV-50 fundus camera having a FOV of 35°. The slides were digitized to produce 700x605 pixel images with an approximate FOV diameter of 650 pixels .

In this set, Adam Hoover (AH) and Valentina Kouznetsova (VK) performed manual vessel labeling for all twenty images. Textual annotations by an expert as to disease state for each image were also provided in this dataset. Using these annotations, it was also possible to discriminate the segmentation performance on images with pathology and those that were normal [5].

The STARE dataset also provided the pixel classifications using MFR for each image to allow for comparative assessment of segmentation performance for MFR hard classifications. Mask images were not included in this set and were created for each image and applied before performance was assessed.

6. Results

Experiments of the proposed method for vessel segmentation were conducted on the DRIVE and STARE datasets with a tile size of N=16.

The initial performance of the segmentation algorithm provided a baseline from which improvements could be made. This assessment provided an objective method in identifying the quality of features extracted as representations of vessel segments when compared with human annotations. The DRIVE and the STARE datasets were used in this process. These datasets contain twenty test images, each with two sets of manual annotations done by separate observers to provide a ground truth and assessment of subjectivity.

6.1. The DRIVE dataset

The following table shows the mean accuracy, specificity and sensitivity for composite and colour component vessel segmentation using images from the DRIVE dataset. Set B was included in the comparison with Set A as a measure of human accuracy in manual annotations. The results of set B confirmed the findings by Staal et al. [8] that there is some disagreement between the two sets of manual segmentation. This provided a reference point from which a statistical comparison could be made as is seen in Table 1. That is, the accuracy of the segmentation method was being tested against a ground truth that had some inherent subjectiveness from the human observer.

On initial review of these results, it was apparent that the green component produced the highest number of true positives followed by red then blue. To find if this difference between the mean accuracies was statistically significant a non-parametric test was done in a pair wise manner. The method chosen was a Wilcoxon rank sum test, producing the p-value matrix shown Table 2. The results of this test shows that the there was a statistical difference between all mean accuracies and rates having a significance level of $\alpha = 0.05$ with two exceptions: the overall accuracy of the blue component when tested against the red, which had a p-value of 0.1895, and the overall accuracy of the red component tested against the composite, which had a p-value of 0.1075.

Table 1. Shows the accuracies between each colour component with the first human observer as a point of reference.

DRIVE database: Segmentation accuracy against Set A			
Segmentation method	*Mean Accuracy (Std. dev.)*	*True positive rate (TPR)*	*False positive rate (FPR)*
Set B	0.9473 (0.0048)	0.7760	0.0275
Composite	0.8936 (0.0149)	0.5924	0.0621
Red	0.8871 (0.0150)	0.4015	0.0416
Green	0.9108 (0.0107)	0.5266	0.0328
Blue	0.8817 (0.0120)	0.2587	0.0273

Table 2. Shows the p-values from a Wilcoxon rank sum test from comparing the means. P-values with the null hypothesis accepted are crossed out ($\alpha = 0.05$).

Overall Accuracy Means			
	Red	***Green***	***Blue***
Composite	~~0.1075~~	1.7936e-004	0.0060
Red		8.5974e-006	~~0.1895~~
Green			5.2269e-007
Blue			
True Positive Rate Means			
	Red	***Green***	***Blue***
Composite	1.0646e-007	0.0028	6.7956e-008
Red		3.4995e-006	2.6898e-006
Green			6.7956e-008
Blue			
False Negative Rate Means			
	Red	***Green***	***Blue***
Composite	1.9971e-004	1.8030e-006	1.4309e-007
Red		0.0123	2.4706e-004
Green			0.0207
Blue			

Since the pixel classifier is discrete in nature, producing only a class label of either vessel or non-vessel, its performance exists in ROC space as a single point at location (FPR, TPR) corresponding to the values in Table 1. These locations are visible on the merged ROC curve plotted from experimental results shown in Figure 5 for their respective segmentation methods.

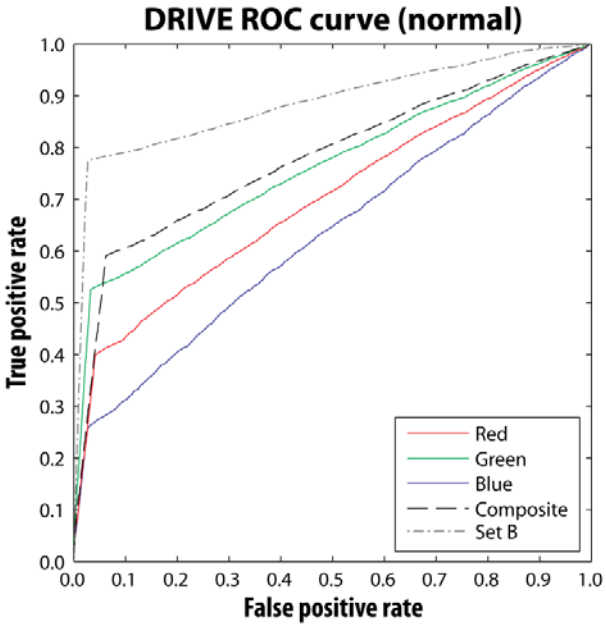

Figure 5. ROC curve formed by merging the twenty samples in the DRIVE dataset.

Plotting the individual ROC curves illustrates the variance between images in the dataset. The performance from the segmentation of the blue component was relatively poor given that a pure random classifier would produce an AZ of 0.5 but it must be noted that it contributes to the true positive rate when considering the composite of all components.

The composite segmentation results were compared with previous vessel segmentation techniques that were tested against the DRIVE dataset. Comparison of the mean true positive rates showed that the initial results were worse than these techniques as seen in Table 3.

Table 3. Shows the performance of the composite segmentation method compared with past vessel segmentation techniques with set B as a point of reference.

DRIVE database: Segmentation accuracy against Set A			
Segmentation Method	*Mean Accuracy (Std. dev.)*	*True positive rate*	*False positive rate*
Set B	0.9473 (0.0048)	0.7760	0.0275
Composite	0.8936 (0.0149)	0.5924	0.0621
Mendonça – Green	0.9452 (0.0062)	0.7344	0.0236
Mendonça – Luminosity	0.9463 (0.0065)	0.7315	0.0219
Staal *et al.* – PBM	0.9441	0.6780	0.0170
Martinez-Perez *et al.* – MS	0.9344	0.7246	0.0345

6.2. The STARE dataset

Tests with the STARE dataset provided segmentation performance on images with pathology and a comparison to images without. These results support the findings from tests on the DRIVE dataset; that of all the colour components, the green component provides the highest sensitivity and specificity.

Comparing the results from images without pathology, the composite image yielded a greater sensitivity but lower specificity in comparison with results from MFR hard classification. This difference was shown to be statistically significant with a p-value less than 0.0002 using a Wilcoxon rank-sum test. It must be noted that use of multiple thresholds in MFR produces a soft classifications so using AZ would produce a fairer comparison.

In the segmentation results from images with pathology, each component fell in accuracy, specificity and sensitivity. To test the statistical significance of this difference, a Wilcoxon rank-sum test was performed comparing the mean accuracy, TPR and FPR from these two subsets. The results of this test are shown in Table 4.

Table 4. Comparison between results from abnormal and normal subset of the STARE images.

p-values from Wilcoxon Rank-Sum test on the means: normal vs. abnormal				
	Composite	*Red*	*Green*	*Blue*
Accuracy	0.1859	0.0017	0.0017	0.0046
True positive rate	0.0140	0.0173	0.0140	0.1405
False positive rate	0.7913	0.0091	5.8284e-004	0.2123

Comparing the variance in accuracies between colour components obtained from the STARE dataset has shown that the false positives increase with each additional colour component. This suggests that images in the STARE dataset had inherently more chromatic noise in comparison with segmentation results from the DRIVE dataset, where luminosity contributed a greater proportion of the noise. This disparity in relative component accuracies between datasets may be due to artifacts introduced into the STARE images as a by-product of digitization.

Given that the DRIVE dataset was without textual annotations, showing which images contained pathology, it was not known if the skewed nature of these results were due to the presence of pathology or if it were due to an insufficient sample size. However, it could be seen from the STARE results that the overall accuracies in composite segmentation were adversely affected by the lowered specificity of red and blue components when compared to the results from the DRIVE dataset. This suggests that the pathology within an image has an adverse affect upon the proposed vessel segmentation method.

7. Conclusions

In this paper, a method for locally representing vessel features was identified for the purpose of incremental learning in retinal image feature extraction. The proposed technique for feature extraction used a localised adaptive threshold selection method, which was then

applied to a range of retinal images. Using this approach for vessel segmentation, it was also possible to perform arteriole-venous differentiation with high-quality retinal images based on the composite output.

Evaluation of the correctness in vessel segmentation, from which features were derived, was performed on two standard retinal image datasets. It was shown to perform below previous segmentation methods on the same datasets with the exception of the STARE dataset with no pathology having a greater sensitivity than that of MFR hard classifications.

It should be noted that individual colour components have not been tested in the literature for the purposes of vessel segmentation and the experiments have shown that they contribute to the accuracy of the pixel classifier.

8. Further work

With the focus in this paper towards segmentation and feature extraction for the purpose of improving feature set accuracy, the inclusion of medical knowledge was a path left unexplored. It would be possible to use the same feature set to derive clinical information, such as the differentiation of vessel types which was done qualitatively but by defining a rule set for their classification, this process could be automated.

With the given vessel segmentation providing some localised features, further work could involve the organization of neighbouring regions within and between tiles into a graph like structure as a form of vessel tracking. This would provide an expanded feature set from which information could be obtained and provide another feature space for knowledge acquisition.

It has been demonstrated in the literature that incremental learning using RDR and MCRDR always increases in classification accuracy; this is the nature of these algorithms [9, 12]. Demonstration of this process in image feature extraction would be the next step, a process in which the learning rate could be assessed based on the provided feature set showing how segmentation accuracy would improve with additional rules and features.

9. References

[1] R. N. Frank, "Diabetic Retinopathy," *New England Journal of Medicine*, vol. 350, pp. 48-58, 2004.

[2] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image processing using MATLAB*. Upper Saddle River, N. J.: Pearson Prentice Hall, 2004.

[3] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan, and I. J. Constable, "Retinal image analysis: Concepts, applications and potential," *Progress in Retinal and Eye Research*, vol. 25, pp. 99-127, 2006.

[4] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 951-958, 2003.

[5] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Transactions on Medical Imaging*, vol. 19, pp. 203-210, 2000.

[6] A. M. Mendonça and A. Campilho, "Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction," *IEEE Transactions on Medical Imaging*, vol. 25, pp. 1200-1213, 2006.

[7] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson, "Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *British Journal of Ophthalmology*, vol. 83, pp. 902-910, 1999.

[8] J. Staal, M. D. Abràmoff, M. Niemeijer, M. A. Viergever, and B. Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, pp. 501-509, 2004.

[9] B. Kang, P. Compton, and P. Preston, "Multiple Classification Ripple Down Rules: Evaluation and Possibilities," in *the Proceedings of the 9th AAAI-sponsored Banff Knowledge Acquisition for Knowledge Based Systems Workshop*. Banff, 1995.

[10] C. K. Chow and T. Kaneko, "Automatic boundary detection of the left ventricle from cineangiograms," *Computers and Biomedical Research*, vol. 5, pp. 388-410, 1972.

[11] M. Niemeijer and B. Van Ginneken, "DRIVE: Digital Retinal Images for Vessel Extraction," 2002.

[12] P. Compton, L. Peters, G. Edwards, and T. G. Lavers, "Experience with Ripple-Down Rules," *Knowledge-Based Systems*, vol. 19, pp. 356-362, 2006.