

2008 International
Symposium on

UBIQUITOUS MULTIMEDIA COMPUTING

13 - 15 October, 2008
Hobart, Australia



IEEE



UNIVERSITY
OF TASMANIA



KOREA MULTIMEDIA SOCIETY
Ubiquitous
Multimedia **RESEARCH GROUP**
Computing

Feature Extraction for Classification from Images: A Look at the Retina

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

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

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

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

Abstract

A method is proposed for the representation of localised features using disjoint sub-images taken from several datasets of retinal images for use within 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 done using the composite of these components and high quality fundal images. Vessel segmentation performance was evaluated on the DRIVE and STARE datasets achieving average specificity of 0.9379 and sensitivity of 0.5924.

1. Introduction

Feature extraction and image classification techniques are hot topics in current image research, particularly with the wealth of higher quality digital images that are now becoming more accessible with the increased availability of digital camera technologies. However, Artificial Intelligence (A.I.) based image analysis techniques are in their infancy, and as with most A.I. methods it is necessary to take small steps towards the ultimate goal of automatic feature extraction and classification from images.

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. In this feature, cardiovascular disease, retinopathy of prematurity (ROP) and diabetes are some diseases that may manifest in abnormal retinal topology and microvasculature.

Until recently, medical imaging has been focused on methods for 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.

This study is to investigate whether it is 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.

2. Previous Literature

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 background regions. The process of image segmentation provides this ability and allows for feature identification.

The vascular network is perhaps one of the most significant features in the retina. 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.

One main issue that must be overcome is the problem of varying intensity from images taken with the fundus camera. This property is due to the curvature of the retina, the medium through which light needs to travel through and optical properties of the cornea. An approach to counter the intensity variation problem is to use the second derivative of the intensity. This can give topological information with a two stage region growing process [1].

Most techniques apply rules found *a priori* to produce their annotations. Another approach would be with supervised methods. 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.

After a review of current techniques in retinal imaging, it was apparent that the classification and segmentation from these techniques have been done as a single process leading to a diagnosis which one could consider as expert knowledge. Focusing on retinal vasculature, some produced binary classifiers for vessel segmentation [2, 3] while others produced quantitative information in the form of bifurcation angles [4].

If it were possible to approach the feature extraction process incrementally on a case-by-case basis, a more flexible mechanism would be available; one where the context in which an intervention was required could have corrections applied in an incremental fashion. It has been shown that organization of computer vision processes can be done with knowledge-based systems [5, 6]. 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 exists. Thus it was proposed that this problem domain would be a suitable candidate for the application of the Multiple Classification Ripple Down Rules (MCRDR) method for the maintenance and acquisition of knowledge [7].

3. Methodology

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 in the system.

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 below) for determining segmentation performance.

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 [8] 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 equation:

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

Where $W(i, j)$ is a sub-image of $N \times N$ dimensions centred on the pixel (i, j) . $I(i, j)$ is the original intensity of a pixel located at (i, j) , $\mu_{desired}$ the desired mean intensity and $W_{\mu}(i, j)$ 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 an illumination-equalised image.

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 with a constant value $\mu_{desired}$. This produces 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 removes the variation for any sub-image with a size greater than these dimensions.

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 48×48 pixels each. Each histogram for unsupervised threshold selection was formed using a neighbourhood of 3×3 tiles overlapping by $8/9^{th}$.

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 $i = [0, 1, 2, \dots, L-1]$ 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 $N = n_0 + n_1 + \dots + n_{L-1}$. Each normalised histogram can be represented as a discrete probability density function with the probability distribution p_i written as:

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

3.2. Threshold Selection Algorithm

The first approach for obtaining a threshold that was tested was Otsu's method [9]. However, it was found this method was outclassed by the k-means method described below.

The second method tested was described by Gonzalez, Woods and Eddins [10]. 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. This threshold was used to divide the histogram into two classifications, $C_0 = [0, \dots, T-1]$ and $C_1 = [T, \dots, L-1]$ to produce two distributions G_0 and G_1 each with N_0 and N_1 pixels respectively. Within these distributions the weighted means μ_0 and μ_1 were calculated by:

$$\mu_0 = \frac{\sum_{i=0}^{T-1} n_i i}{N_0} \quad \mu_1 = \frac{\sum_{i=T}^{L-1} n_i i}{N_1}$$

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

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

4. If the new threshold T was not the same as the previously calculated threshold, steps 2 to 4 were repeated until there was no change in T between steps.

Results of this segmentation process using k-means clustering are shown in Figure 1.

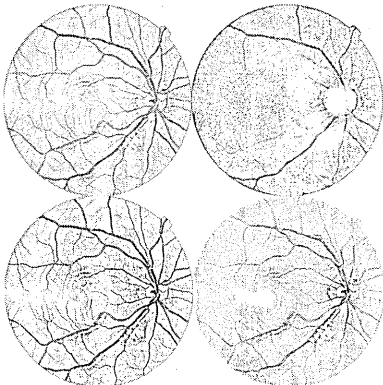


Figure 1 shows a tile-level adaptive threshold

using 1D k-means clustering with illumination equalisation.

From the two segmentation methods with illumination equalisation, one-dimensional k -means clustering produced fewer tile-based artefacts and was superior to other methods tested. This segmentation method was chosen as the basis for building case information generated through connected components analysis of these binary images.

3.3. Connected Components Analysis

After segmentation of the original image, 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 representation containing information both would recognise and provided a link from which both could work.

The resultant binary images from each colour component were partitioned into 48×48 pixel tiles. Segment analysis to label individual regions was performed on a tile-by-tile basis to create localised feature space information. These labelled 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.

4. 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.

4.1. 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 to provide ground truth and an assessment of the observers' annotation accuracy.

Visualization of the comparative pixel classifier performances was achieved by plotting receiver operator characteristic (ROC) curves for each segmentation method [11]. In this scenario, each pixel is an instance with a predicted class and an 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 classification has one of four possible outcomes; true positive, false positive, true negative or false negative.

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

4.2. 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 [12, 13].

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.

4.3. The STARE Dataset

The STARE (Structured Analysis of the Retina) database consists of 20 images containing ten pathological and ten normal retinas [2, 14].

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

The STARE dataset also provided the pixel classifications using matched filter response (MFR) for each image to allow for comparative assessment of segmentation performance for MFR hard classifications.

5. Results

Experiments of the proposed method for vessel segmentation were conducted on the MacCana

dataset using a tile size of $N=48$ while the DRIVE and STARE datasets had a tile size of $N=16$.

The initial performance of the segmentation algorithm provided a baseline. This assessment provided an objective method for 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.

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 |

5.1. DRIVE dataset

Figure 2 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* [12] that there is some disagreement between the two sets of manual segmentations. This provided a reference point from which a statistical comparison could be made.

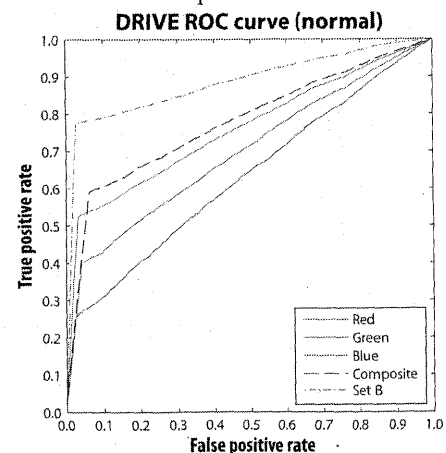


Figure 2 ROC curve formed by merging the twenty samples in the DRIVE dataset.

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 pairwise manner. The results from a Wilcoxon rank sum test comparing mean accuracy rates of each component and the composite showed that there was a statistical difference with 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 2 shows the performance of the composite segmentation method when compared with previous 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 [3] | 0.9452 (0.0062) | 0.7344 | 0.0236 |
| Mendonça – Luminosity [3] | 0.9463 (0.0065) | 0.7315 | 0.0219 |
| Staal <i>et al.</i> – PBM [12] | 0.9441 | 0.6780 | 0.0170 |
| Martínez-Pérez <i>et al.</i> – MS [15] | 0.9344 | 0.7246 | 0.0345 |

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 2. These locations are visible on the merged ROC curve plotted from experimental results shown in Figure 2 for their respective segmentation methods. The performance from the segmentation of the blue component was relatively poor given that a pure random classifier would produce an A_z 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 can be seen in Table 2.

5.2. STARE dataset

Tests with the STARE dataset provided segmentation performance on images with pathology

and a comparison to images without. These results are not shown due to space restrictions, but can be found in full in an earlier publication [16].

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 A_z 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 3.

Table 3 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 | 0.00058 | 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 artefacts introduced into the STARE images as a by-product of digitisation.

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 they 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.

6. 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.

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.

7. Further Work

With the focus of this paper on 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.

The logical next step of this work would be to apply an incremental learning algorithm, such as MCRDR allowing the expert to select the automatically extracted features as conditions for rules to produce a knowledge base capable of diagnosing from the image.

8. References

- [1] M. E. Martínez-Pérez, A. D. Hughes, A. V. Stanton, S. A. Thom, A. A. Bharath, and K. H. Parker, "Segmentation of retinal blood vessels based on the second directional derivative and region growing," presented at IEEE International Conference on Image Processing, 1999.
- [2] 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.
- [3] 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.
- [4] M. E. Martínez-Pérez, A. D. Hughes, A. V. Stanton, S. A. Thom, N. Chapman, A. A. Bharath, and K. H. Parker, "Retinal vascular tree morphology: A semi-automatic quantification," *IEEE Transactions on Biomedical Engineering*, vol. 49, pp. 912-917, 2002.
- [5] J. Kerr and P. Compton, "Toward Generic Model-Based Object Recognition by Knowledge Acquisition and Machine Learning," presented at Workshop on Mixed-Initiative Intelligent Systems, International Joint Conference on AI, 2003.
- [6] P. K. Singh, "Knowledge-based Annotation of Medical Images," in *Department of Artificial Intelligence*, vol. PhD. Sydney: The University of New South Wales, 2006, pp. 224.
- [7] 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.
- [8] 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.
- [9] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-9, pp. 62-66, 1979.
- [10] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image processing using MATLAB*. Upper Saddle River, N. J.: Pearson Prentice Hall, 2004.
- [11] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, pp. 861-874, 2006.
- [12] J. Staal, M. D. Abramoff, 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.
- [13] M. Niemeijer and B. Van Ginneken, "DRIVE: Digital Retinal Images for Vessel Extraction," 2002.
- [14] A. Hoover, "STARE Project website," vol. 2007, 2002.
- [15] M. E. Martinez-Perez, A. D. Hughes, S. A. Thom, A. A. Bharath, and K. H. Parker, "Segmentation of blood vessels from red-free and fluorescein retinal images," *Medical Image Analysis*, vol. 11, pp. 47-61, 2007.
- [16] F. C. F. Chui, "Intelligent Feature Extraction from Retinal Images," in *Computing*. Hobart: Tasmania, 2007, pp. 83.