Initial Results of an Automatic Blood-Vessel Segmentation Procedure in Digital Fundus Photographs via Multiscale Line Operators and Global Threshold Selection

D.J.J. Farnell[1]

Academic Department of Radiation Oncology, Division of Cancer Studies, Faculty of Medical and Human Science, University of Manchester, c/o The Christie NHS Foundation Trust, Manchester M20 4BX, United Kingdom

Abstract. A multiscale line operator (MSLO) approach was employed in order to segment blood vessels in digital fundus images. This approach was used for images in the STARE retinal image archive (www.ces.clemson.edu/~ahoover/stare/). The STARE dataset contained both healthy (10 images) and diseased eyes (10 images) at a variety of field widths. Otsu threshold selection was employed in order to provide a global threshold for the MSLO-filtered images. A Boxcar average filter was also used to segment the images. We determined the sensitivity (true positive rate) and specificity (1 – false positive rate) with respect to a "gold standard" provided by manual segmentations. The best results for the entire dataset were found to be (sensitivity=0.80; specificity=0.97). Our results for the normal set of images were better than those the abnormal set. A "topographic" map of the sensitivity and specificity as a function of position on the retina showed that the sensitivity was highest at the fovea, i.e., the middle of the retina (sensitivity=0.89). This area might provide an ideal place for extraction of clinically relevant biometric parameters. The accuracy of vessel extraction techniques is more complex than that suggested by simple averages of sensitivity and specificity across the entire field-of-view.

1 Introduction

Methods of blood-vessel segmentation in medical images [1] range from pattern recognition (multiscale approaches, skeletons, region growing, ridge-based approaches, blood-vessel tracking, matched filters, differential geometry, and mathematical morphology), model-based approaches (e.g. active snakes), tracking systems, to artificial intelligence and neural networks. Many of these approaches utilise image information at a variety of length scales. The multiscale line operator [2-5] (MSLO) has been used with great success in detecting linear structures in mammograms [2,3]. In the context of blood-vessel segmentation in retinal images [4-8], one might expect it to work well because of the approximately linear nature of blood vessels locally. This method has the additional advantage of being conceptually straightforward. The MSLO algorithm is fast and is also relatively straightforward to implement. The subject of automatic global threshold selection is a common one in image processing. The interested reader is referred, e.g., to Refs. [9,10] for a review of some of the standard threshold selection procedures.

2 Method

2.1 The STARE Retinal Image Archive

The STARE dataset is an archive of fundus images that is publicly available (see www.ces.clemson.edu/~ahoover/stare/). We use the same dataset as in previous treatments [4-7] that contains 20 images with healthy, diabetic, and AMD image features. According to information provided by the above website, 10 of these images were classified as normal and 10 as abnormal. The images were in colour and were 700×605 pixels in size and were at a variety of field widths. Two independent observers created manual tracings of the blood vessels using a graphical user interface, and these tracings were also downloaded from the STARE website. As in previous studies (see Ref. [6]), we use the manual tracings of the first such observer as our "gold standard" for the STARE dataset. The green channel of the fundus images is used in order to find the blood vessels.

2.2 Sensitivity, Specificity, and Receiver-Operator Characteristic (ROC) Analysis

In order to quantify the accuracy of our results we determined the sensitivity and specificity. We were able to determine the number of pixels classified as "positive" by our automated routines that were, in fact, either "true" or "false" compared to a "gold standard" provided by manual segmentations of the blood vessels. The entire vessel width was used here. We were able to find the true positive (sensitivity) and false positive (1-specificity) fractions, and one of the aims of this study was to illustrate how our results for the sensitivity and specificity vary across the entire image when using the entire vessel width. Furthermore, in previous studies [4,5] we were able to perform an ROC analysis by plotting the true positive fraction on the ordinate axis against the false positive fraction on the

[1]Email: damian.farnell@manchester.ac.uk

abscissa for thresholds varying independently for each image. These results are also provided as a comparison for our present results. We note that the "ideal limit" is thus given by values for both sensitivity and specificity of 1. However, a "practical limit" is provided by comparing the results of the second observer against those of the first observer that we take as our gold standard and finding values for the sensitivity and specificity. Another limit is that provided by the point of closest approach of the MSLO ROC results of Ref. [4,5] to the ideal limit of both sensitivity and specificity equal to 1.

2.3 Multiscale Line Operator (MSLO)

We implemented the MSLO [2-5] in order to enhance the retinal blood vessels in the image. For each pixel in the original image at point (i,j), the mean average value, P, of the greyscale values of a line of pixels centred on (i,j) is obtained. However, we also determine the average greyscale value, Q, of those pixels in a rectangular area surrounding this line. The response, S, of the line operator at point (i,j) at a given orientation of the line (and scale) is given by, $S=P-Q$. In order to keep our treatment fully consistent with those calculations outlined in Refs. [2-5], we used a line operator of length 5 pixels obtained at 12 orientations. The largest value of S for all 12 orientations is taken to be the value of the line operator at that pixel. The subtraction of the mean greyscale value of local environment, Q, from the mean value for the line, P, ensures that changes in background illumination should also be removed in the filtered image. The line operator is applied at varying levels of scale by constructing a "pyramid of images" at consecutively coarser scales via Gaussian sampling. The number of levels of the Gaussian pyramid was taken to be an explicit variable here. The final result for the MSLO was the sum of all line operator filtered images in the Gaussian pyramid. Equal weighting was given to each length scale in the final MSLO image. The line operator enhances linear features in the image. Moreover, blood vessels of varying width are, in principle, treated on an equal footing. This algorithm was speeded by saving those configurations or "masks" for the line of central pixels and its neighbours in the "rectangular environment" (used in finding P and Q, respectively) at the 12 different orientations at the start of the entire run in a look-up table. The number of operations involved in the application of the line operators was thus kept to a minimum. The application of the MSLO filter for an image pyramid with 3 levels took approximately one minute for a 3 GHz CPU. The interested reader is referred to Refs. [2-5] for more detailed descriptions of the MSLO algorithm.

2.5 Selection of the Global Threshold

A single global threshold (see, e.g., Refs. [9,10]) was applied in order to segment the retinal blood vessels. In order to select the threshold automatically, three standard procedures were tested. These were namely maximum entropy, Otsu, and, Kittler-Illingworth thresholding. For maximum entropy binary thresholding, the image is split into (two) pieces based on grey scale levels with respect to a given threshold. Shannon's information entropy is separately for those levels below the threshold and those equal to or above the threshold. The total entropy is the sum of the two parts and the total entropy may be maximised with respect to the threshold. Otsu thresholding allows us to carry out binary thresholding. It is equivalent to the maximization of the likelihood of a conditional distribution, again with respect to the grey scale threshold. Otsu thresholding assumes that the distributions of two types of pixels are normal distributions with common variance. By contrast, it may be shown that Kittler and Illingworth's criterion is equivalent to the maximization of the likelihood of the joint distribution under the assumption of normal distributions but with different variances. Multi-level thresholding (i.e., using more than one threshold in order to segment the image into more than one type of object) was not considered here. We note that the Kittler-Illingworth and maximum entropy approaches were found to provide relatively poor results for the images considered here, and so, for the sake of brevity, these results are not discussed further here. As blood vessels constituted typically only a small percentage of the pixels in the images, we found that an iterative approach in order to segment the blood vessels using the Otsu threshold selection method worked well for the MSLO-filtered images. This approach was found to converge within ten to twenty seconds on a 3GHz processor. This approach was found to provide a consistently a good estimate of the blood vessels. The results of this approach shall be referred to as the MSLO/Otsu segmentations. This initial estimate was further refined by automatically segmenting of a Boxcar-filtered image using the maximisation of sensitivity and specificity with respect to the initial MSLO/Otsu segmentation. This approach was also found to work reasonably accurately, and its results shall be referred to as Boxcar segmentations. (Otsu threshold selection applied directly to Boxcar filtered image was found to be less reliable.) Finally, a "geographical map" averaged over the normal and (separately) abnormal images of the sensitivity (i.e., the true positive rate) and specificity (i.e., one minus the false positive rate) was determined for the Boxcar segmentations with respect to the gold standard provided by manual tracings of the blood vessels.

3 Results

A colour fundus photograph of a normal subject is shown in Figs. 1a and 1b. A blood-vessel segmentation (shown in white) found after applying the MSLO with three levels and Otsu threshold selection (to the green channel of the

image) is superimposed on the original colour image in Fig. 1a. The corresponding Boxcar segmentation is shown in Fig. 1b. The Boxcar algorithm captures more of the finer vessel detail than the MSLO/Otsu segmentation. The edge of the optic disk led to false classifications. Furthermore, the centre of the macula in some healthy subjects was particularly dark in "normal" subjects, and this also caused false classifications. A colour fundus image of an "abnormal" subject with a similar blood-vessel segmentation (MSLO + and Otsu threshold selection) superimposed on it is shown in Fig. 1c, and the Boxcar segmentation is shown in Fig. 1d. Again, we see in Figs. 1c and 1d that more of the vessel detail is captured by the Boxcar segmentation, although more false classifications are also made. Haemorrhages were a general and strong source of artefacts in the vessel segmentations. Lesions were also found to be the main cause of false classifications in the abnormal images, as may be observed in Figs. 1c and 1d; the edges of the bright disease feature caused false classifications to be made. Large choroidal neovascular (CNV) lesions are a common feature of "wet" age-related macular degeneration (AMD). Microaneurisms were also seen in some images. They were small red features and they were also found to lead to falsely classified pixels.

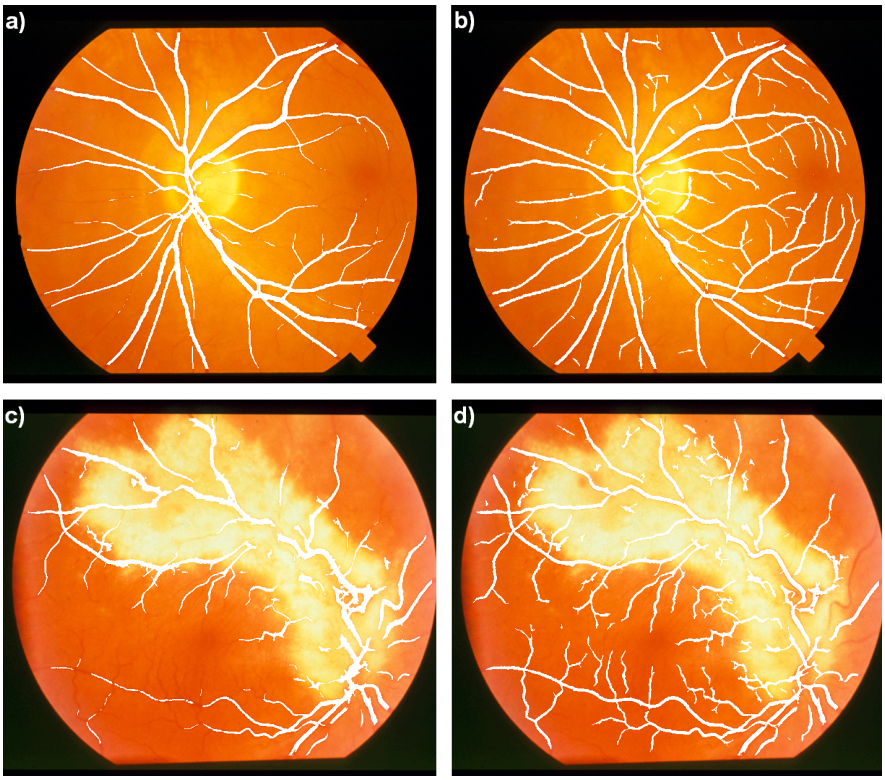

Figure 1. Extracted blood vessels (shown in white) superimposed on the original colour fundus image. a) MSLO/ Otsu segmentation for a normal image. b) Boxcar segmentation for the same normal image. c) MSLO/Otsu segmentation for an abnormal image. d) Boxcar segmentation for the same abnormal image.

As reported previously [4,5], ROC analysis (see Fig. 2) was carried out for our blood-vessel segmentations when compared to a "gold standard" of blood vessels traced out by a retinal image interpretation expert by varying the threshold separately for each image. Results for the sensitivities (true positive rate) and specificities (1-false positive rate) are shown in Fig. 2 for the MSLO/Otsu and Boxcar segmentations. We find that the results produced by automatic threshold selection are either on or near to the lines for those results produced by ROC analysis for the entire data set (Fig. 2a) and for those results for the normal and abnormal images (Fig. 2b). By visual inspection of Figure 2a, we see that our ROC results compared well to those the results of another segmentation algorithm of Ref. [7] for the STARE dataset. As expected, we see also from Fig. 2b that we obtain better results for the normal set of images than the abnormal set. The results for sensitivity and specificity for the entire STARE dataset for the MSLO/Otsu and Boxcar segmentations are also shown Table 1. Our results lie reasonably close to those results provided by the "point of closest approach" of the ROC obtained by varying the threshold separately for each image to ideal solution (i.e., a true positive rate equal to one and a false positive rate equal to zero). Values for the sensitivity and specificity for each image were obtained for the manual tracings of the blood vessels of the second observer also compared to the "gold standard" results of the first observer. Averaged over the entire set of images, they gave a single value for the sensitivity and specificity and our results are also reasonably close to this upper limit. We found that the sensitivity and specificity of the Boxcar segmentations were nearer to those results for the point-of-closest approach of the ROC to the ideal point and also to the inter-observer estimate than the results for the MSLO/Otsu segmentations. A reliable method of segmenting the Boxcar-filtered images was thus obtained by using the MSLO/Otsu segmentation in order to inform "optimal" threshold selection of the Boxcar-filtered images.

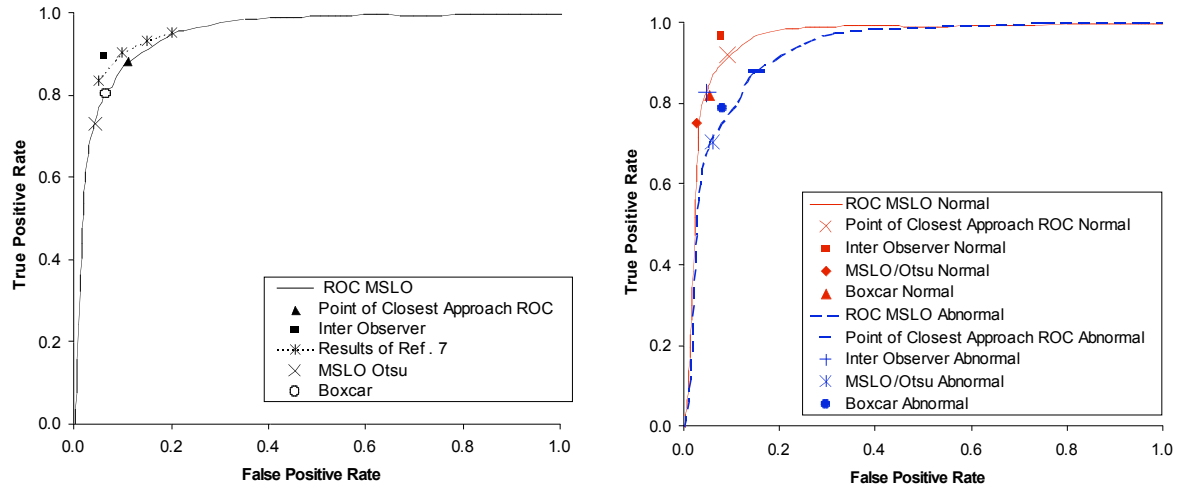

Figure 2. Results for the true and false positives rates. (a; left) For all images in the STARE dataset compared to independent results of Ref. [7]. (b; right) For the normal and abnormal images of the STARE dataset separately.

A map of the sensitivity (true positive rate) and specificity (1-false positive rate) averaged over the normal and (separately) abnormal images was also found for the Boxcar segmentations, see Fig. 3. We found that the sensitivity was highest (sensitivity=0.89) in regions close to the centre of the macula for both the normal and abnormal images. The sensitivity was found to be lower near to the edges of the fundus image for both normal and abnormal images. The specificity was slightly lower in this area centred on the macula (specificity=0.91). However, this effect was predominantly observed for the abnormal images (specificity=0.89) and not the normal images (specificity=0.93). The centre of the macula contained lesions in some of the abnormal images and it was often also slightly darker than surrounding tissues in the normal images. Both of these aspects led to false classifications in these two groups at the centre of the macula. Finally, we note that area around the optic disk had lower sensitivity (=0.75).

	FPR MSLO/ Otsu	TPR MSLO/ Otsu	FPR Boxcar	TPR Boxcar	FPR ROC PCA	TPR ROC PCA	FPR Inter-Observer	TPR Inter-Observer
Normal	0.03	0.75	0.05	0.82	0.09	0.92	0.08	0.97
Abnormal	0.06	0.70	0.08	0.79	0.15	0.88	0.05	0.83
All Images	0.04	0.73	0.07	0.80	0.11	0.89	0.06	0.90

Table 1. True positive rate (TPR = sensitivity) and false positive rate (FPR = 1-specificity) for the STARE dataset. An estimate of the TPR and FPR for the results of manual tracings for the second observer compared to the first observer is also given. (MSLO = Multiscale Line Operator; ROC PCA = Point of Closest Approach of ROC curve to the ideal point (i.e., sensitivity=1 and specificity=1).)

4 Conclusions

General Method: The MSLO filter can be used to enhance blood vessel segmentation. Global thresholding (here using Otsu threshold selection) can be used to segment MSLO filtered digital fundus images automatically. A reasonably accurate and robust (albeit for the small number of images used here) method of vessel segmentation was obtained by using the output of the MSLO/Otsu segmentations in order to inform "optimal" threshold selection for Boxcar filtered images. However, results for the sensitivity and specificity indicated that further improvements can still be made. Global thresholding into a binary image worked well for relatively "clean" images such as the normal images or images with small amounts of retinopathy, although less well for those images with larger amounts of retinopathy such as lesions (e.g., for wet AMD).

Medical Context: Central vision is located at the centre of the macula, i.e., at the fovea. This is also a physiologically sensitive area. A "topographic" map of the sensitivity and specificity indicated that the best results were obtained for a region centred on the macula, but not including the optic disk.

Implications: The centre of the macula might be optimal for extraction of biometric parameters, especially for early-stage diabetic retinopathy and "dry" AMD. However, more work on larger data sets needs to be carried out in order to establish this more firmly. This will be carried out for the larger ARIA dataset (www.eyecharity.com/aria_online/) collected by us. Outlying areas of the macula demonstrated poorer results. However, this might have been partially due to the range of field-or-view angles (thus influencing the magnification) being used in the photographs in the STARE dataset. Those photographs with larger field-of-view were more likely to contain these outlying areas and

they also had the lowest magnification; vessels appeared finer and so were harder to segment. In any case, considerable variation in sensitivity and (to a lesser extent) specificity was observed. These "topographic maps" showed that the accuracy of our segmentations was more complex than that suggested by a simple average of the sensitivity and specificity over the entire field-of-view. This aspect should be noted when analysing the results of segmentation procedures, especially if the image dataset contains different field widths.

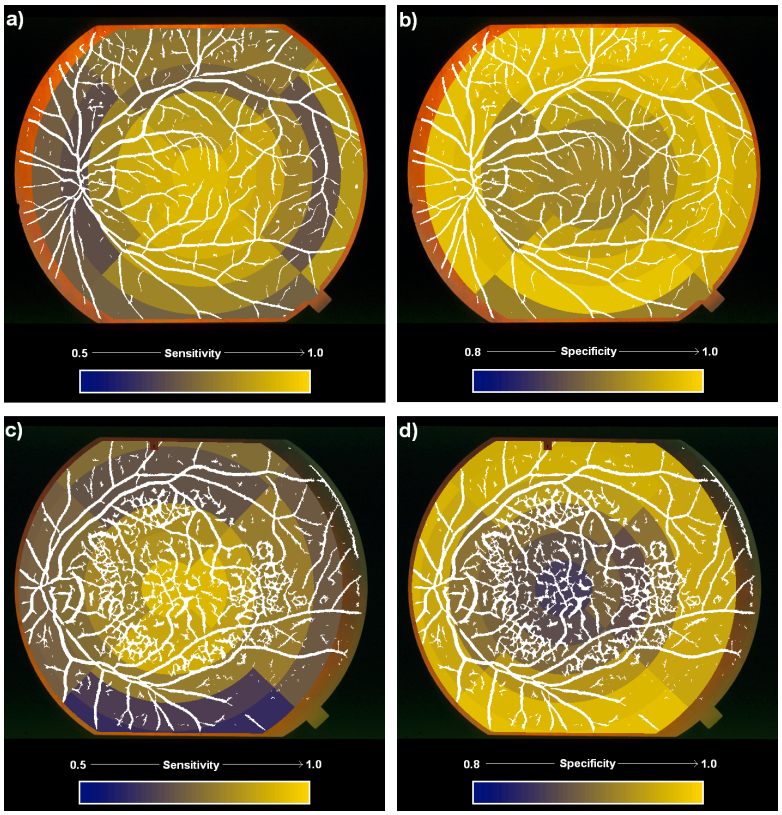

Figure 3. Topographic maps of the sensitivity (true positive rate) and specificity (1-false positive rate) averaged over all the normal and (separately) abnormal images in the STARE data set. This is superimposed on a test image from the normal/abnormal sets with Boxcar vessel segmentation shown in white. a) Sensitivity for the normal images. b) Specificity for the normal images. c) Sensitivity for the abnormal images. d) Specificity for the abnormal images.

References

1. C. Kirbas and F.H. Quek. "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys* **36**, pp. 81–121, 2004.
2. R. Zwiggelaar, S.M. Astley, C.R. Boggis, C.J. Taylor, "Linear structures in mammographic images: detection and classification*," IEEE Transactions on Medical Imaging* Vol. **23**, pp. 1077–1086, 2004.
3. E.M. Hadley, E.R.E. Denton, R. Zwiggelaar, "Mammographic Risk Assessment Based on Anatomical Linear Structures," in Digital Mammography, *Lecture Notes in Computer Science* Vol. **4046** (Springer Verlag, Heidelberg), pp. 626—633, 2006.
4. D.J.J. Farnell, F.N. Hatfield, P.C. Knox, M. Reakes, S.P. Harding, "Initial Results of an Application of Multiscale Line Operators for Blood-Vessel Segmentation in Digital Fundus Photographs," in the Proceedings of the *Medical Image Understanding and Analysis (MIUA) Conference*, pp. 141–145, 2007.
5. D.J.J. Farnell, F.N. Hatfield, P.C. Knox, M. Reakes, D. Parry, S. Spencer, and S.P. Harding. "Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators," *Journal of the Franklin Institute* 345, pp. 748–765, 2008.
6. A.D. 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* **19**, pp. 203–210, 2000.
7. N.M. Salem and A.K. Nandi. "Segmentation of retinal blood vessels using scale-space features and k-nearest-neighbour classifier," in the Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. *ICASSP 2006* **2**, pp. 1001–1004, 2006.
8. S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum, "Detection of Blood Vessels in Retinal Images Using Two Dimensional Matched Filters," *IEEE Transactions on Medical Imaging* **8**, pp. 263—269, 1989.
9. T. Kurita, N. Otsu and N. Abdelmalik. "Maximum likelihood thresholding based on population mixture models," *Pattern Recognition* **25**, pp. 1231–1240, 1992.
10. P.-Y. Yin. "Maximum entropy-based optimal threshold selection using deterministic reinforcement learning with controlled randomization," *Signal Processing* **82**, pp. 993–1006, 2002.