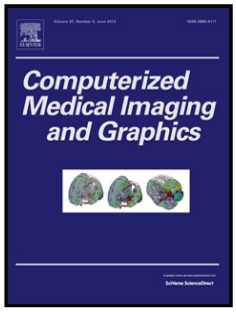
Performance Comparison of Publicly Available Retinal Blood Vessel Segmentation Methods

Vostatek, Pavel ; Claridge, Ela; Uusitalo, Hannu; Hauta-Kasari, Markku; Fält, Pauli; Lensu, Lasse

DOI:
10.1016/j.compmedimag.2016.07.005

Document Version
Peer reviewed version

Citation for published version (Harvard):
Vostatek, P, Claridge, E, Uusitalo, H, Hauta-Kasari, M, Fält, P & Lensu, L 2017, 'Performance Comparison of Publicly Available Retinal Blood Vessel Segmentation Methods', *Computerized Medical Imaging and Graphics*, vol. 55, pp. 2-12. https://doi.org/10.1016/j.compmedimag.2016.07.005



Title: Performance Comparison of Publicly Available Retinal Blood Vessel Segmentation Methods

Author: Pavel Vostatek Ela Claridge Hannu Uusitalo Markku Hauta-Kasari Pauli Fält Lasse Lensu

PII: S0895-6111(16)30070-2
DOI: http://dx.doi.org/doi:10.1016/j.compmedimag.2016.07.005
Reference: CMIG 1456

To appear in: *Computerized Medical Imaging and Graphics*

Received date: 29-2-2016
Revised date: 18-7-2016
Accepted date: 21-7-2016

Please cite this article as: Pavel Vostatek, Ela Claridge, Hannu Uusitalo, Markku Hauta-Kasari, Pauli Fält, Lasse Lensu, Performance Comparison of Publicly Available Retinal Blood Vessel Segmentation Methods, *<![CDATA[Computerized Medical Imaging and Graphics]]>* (2016), http://dx.doi.org/10.1016/j.compmedimag.2016.07.005

Performance Comparison of Publicly Available Retinal Blood Vessel Segmentation Methods

Pavel Vostatek[a], Ela Claridge[b], Hannu Uusitalo[c], Markku Hauta-Kasari[d], Pauli Fält[d], Lasse Lensu[a]

[a] *Machine Vision and Pattern Recognition Laboratory*
Department of Mathematics and Physics
LUT School of Technology
Lappeenranta University of Technology
PO Box 20, FI-53851 Lappeenranta, Finland
[b] *School of Computer Science, University of Birmingham, United Kingdom*
[c] *Department of Ophthalmology/SILK, University of Tampere, Finland*
[d] *School of Computing, University of Eastern Finland*

Abstract

Retinal blood vessel structure is an important indicator of many retinal and systemic diseases, which has motivated the development of various image segmentation methods for the blood vessels. In this study, two supervised and three unsupervised segmentation methods with a publicly available implementation are reviewed and quantitatively compared with each other on five public databases with ground truth segmentation of the vessels.

Each method is tested under consistent conditions with two types of preprocessing, and the parameters of the methods are optimized for each database. Additionally, possibility to predict the parameters of the methods for each database by the linear regression model is tested. Resolution of the input images and amount of the vessel pixels in the ground truth are used as predictors.

The results show the positive influence of preprocessing on the performance of the unsupervised methods. The methods show similar performance for segmentation accuracy, with the best performance achieved by the method by Azzopardi et al. (Acc 94.0) on ARIADB, the method by Soares et al. (Acc 94.6, 94.7) on CHASEDB1 and DRIVE, and the method by Nguyen et al. (Acc 95.8, 95.5) on HRF and STARE. The method by Soares et al. performed better with regard to the area under the ROC curve. Qualitative differences between the methods are discussed. Finally, it was possible to predict the parameter settings that give performance close to the optimized performance of each method.

Keywords: Fundus, Retinal imaging, Vessel segmentation

Email address: lasse.lensu@lut.fi (Lasse Lensu)

1. Introduction

Visual examination of the retina provides a non-invasive view into the eye and at the same time into the central nervous system. One of the distinct features of the retina are the blood vessels, whose structure is an important indicator of disorders such as diabetes, hypertension and cardiovascular disease [1]. Retinal imaging has been used to characterize the vessel structure, and diagnose, monitor and document abnormal conditions [2]. With current technology, it is already possible to produce quantitative information of the signs of eye diseases like diabetic retinopathy and glaucoma, as well in many cardiovascular and neurovascular diseases. A review of retinal imaging and its medical implications has been provided in [3].

To diagnose incipient abnormalities and diseases in their early stages, screening programs with systematic protocols are being implemented for groups at risk. As the screening programs become more extensive, the amount of data increases and in many cases manual diagnosis becomes a bottleneck. To remedy the problem with the increasing workload, computer-aided diagnosis tools can be used to provide access to retinal images and enable high-throughput workflows for the screening programs. To enable automatic or semi-automatic image analysis and the structural characterization of the blood vessels, various approaches have been proposed for segmenting the vessels from retinal images, see [4] and [5]. A review of general vessel extraction techniques has been published in [6].

The aim of this study is to review and compare the performance of pixel-wise blood vessel segmentation methods designed for retinal images and for which an implementation is publicly available. Five methods, two supervised and three unsupervised, were studied (see Section 2.1), and their performance was assessed using five publicly available databases with ground truth for vessel segmentation (see Section 2.2). To assess the performance potential of each method, an experiment was set up so that the parameter space of each method was sampled for each database and searched for optimal performance using grid search (see Section 3.1). In order to keep the conditions of the experiment consistent, the image preprocessing parts of the methods were separated and applied to all the methods.

Analysis of the algorithms' performance over the parameter ranges allows to get several valuable results: (1) comparison and assessment of the generalization capability of the methods over several different databases while exploiting the segmentation potential of each method, (2) comparison of the performance and settings of the methods with the original publications and with the state-of-the-art, and (3) baseline settings suitable for application of the methods to other data.

This study is an extended and enhanced version of the research published in [7]. The method presented by Azzopardi et al. was added to the review, as well as a step for image preprocessing. The Matthew correlation coefficient and area under the receiver-operating characteristic (ROC) curve were added to the set of performance measures. Improvements were made in the parameter optimization, i.e. the parameter spaces were sampled more densely and more

widely. Finally, corrections were made to the implementation of the method by Nguyen et al. and to the contents of the table reviewing the image databases.

2. Databases and methods

2.1. Retinal blood vessel segmentation methods

The method proposed by Soares et al.[1] in [8] (*Soares method*) is a supervised classification algorithm that uses the Morlet (Gabor) wavelet filter response as the classification feature. Three types of classifiers – Gaussian mixture model (GMM), K-nearest neighbor (KNN) and least mean square error (LMSE) – are available to use. Furthemore, the green channel of an input image is by default added to the feature set. All features are normalized individually to zero mean and unit standard deviation. The parameters of the method are: the set of Morlet scales (Λ_{mor}) to define the classification features; the number of training samples (n_s), number of Gaussians for modeling the vessels and non-vessels (n_{g1}, n_{g2}), and number of iterations of the expectation maximization (EM) algorithm (n_i) to define the GMM. The authors emphasize efficiency of the Gabor transform to enhance the vessel contrast. At the same time they conclude long training yet short classification time of the GMM classifier. Simplicity of the implemented algorithms is emphasized. The reported disadvantages are false detections around high-contrast structures like pathologies or the optic disc, and in rare occasions, bad tolerance to uneven illumination.

The method proposed by Sofka et al.[2] in [9] (*Sofka method*) is a supervised classification algorithm based on multiscale matched filtering, and confidence and edge measures. The method extracts the vessel centerlines, and originally, its pixel-wise segmentation performance was not evaluated. However the method offers pixel-wise vessel likelihood (LRV) as an output. The LRV measure with subsequent binarization by thresholding was used in our experiments. The method is available as a pretrained executable with no parameters to set. The authors claim statistically significant improvement of the vessel segmentation performance over Frangi's vesselness measure and matched filter. Particular focus is on the detection of low-contrast and narrow vessels and improvement of the classifier performance on pathologies. The performance of the method is, however, assessed on thinned vessels due to the wider response of the filter around the vessel edges.

The method proposed by Azzopardi et al.[3] in [10] (*Azzopardi method*) is an unsupervised algorithm that employs a bar-selective version of a 'combination of a shifted filter responses' (COSFIRE) filter – B-COSFIRE – which first filters the image with a difference of Gaussians (DoG) mask and then through the COSFIRE mechanism emphasizes the line patterns (creating response R). The final segmentation is obtained by thresholding. For proper detection of vessel

[1] `https://sourceforge.net/projects/retinal/`
[2] `https://www.cs.rpi.edu/~sofka/vessels_exec.html`
[3] `http://www.mathworks.com/matlabcentral/fileexchange/37395`

endings, R of two types, symmetric and asymmetric, are combined by averaging. Each R is defined by 4 parameters: Standard deviation of the DoG filter (σ), length of the line pattern (ρ) and parameters allowing for spatial tolerance in computation of R (σ_0, α). The authors emphasize versatility of the COSFIRE filter as it can be easily rearranged to detect shapes other than lines which were used in the case of the vessel segmentation. The efficiency of the implementation and robustness to noise are also emphasized.

The method proposed by Bankhead et al.[4] in [11] (*Bankhead method*) is an unsupervised algorithm based on isotropic undecimated wavelet transform (IUWT) [12] and binarization by percentile-computed threshold. After the binarization, post-processing by removing isolated objects and filling holes is done. The parameters are a set of wavelet levels (Λ_{ban}) for IUWT, percentile (p_t) used to compute the threshold value, and sizes of the isolated objects and holes (ξ_s, ξ_h) for the post-process. The authors emphasize high processing speed of the method and simplicity of the setup where the most important parameter – Λ_{ban} – has a small range of values even for images of very different resolutions. The disadvantage of the method is a slightly worse segmentation performance compared to the state-of-the-art methods.

The method proposed by Nguyen et al.[5] in [13] (*Nguyen method*) is an unsupervised algorithm based on line operators [14]. Vessel pixels are amplified by filtering the image with a mask of defined size (W) that enhances pixels along lines with different orientations. Multiple filters with varying length of the line ($l_{1..n}$) together with the green channel of the input color image are averaged to produce a single response with enhanced vessel contrast. The response is normalized to zero mean and unit standard deviation. The number of filters is defined by step ω. The output of the algorithm is a gray-scale map. Thresholding (threshold τ) is used to produce the binary map. The authors emphasize the classification speed as an advantage of the method. Also local accuracy (segmentation near the vessel pixels) is claimed to be high. The method is supposed to handle well such areas that are often merged by other segmentation methods. The method is claimed to perform 'extremely well on non-pathological images'.

2.2. Databases with blood vessel ground truth

Medical image databases with an appropriate ground truth about the image contents enable the development and proper evaluation of automatic image analysis methods. Information about the databases selected for testing the retinal blood vessel segmentation methods is summarized in Table 1. The number of images, image dimensions, field of view (FOV) angle and diameter, subsets and percentage of vessel pixels in the ground truth are presented.

DRIVE and STARE are currently the datasets most commonly used to evaluate retinal vessel segmentation methods. All methods reviewed in this paper were originally evaluated using DRIVE and STARE, with an exception of the

[4] http://petebankhead.github.io/ARIA/

[5] http://people.eng.unimelb.edu.au/thivun/projects/retinal_segmentation/

Table 1: Summary of database information. N_i is the number of images and N_{GT} is the number of experts and percentage of annotated vessel pixels (per expert) in the ground truth segmentation. Abbreviations of the image subsets are age-related macular degeneration (AMD) and diabetic retinopathy (DR).

Name, ref.	N_i	FOV [°]	Dimensions FOV ⌀	Subsets	N_{GT}
ARIADB [15]	143	50°	768x576 739 px	AMD (23) Healthy (61) DR (59)	2 (9.6%, 8.5%)
CHASEDB1 [16]	28	30°	999x960 916 px	Left eye (14) Right eye (14)	2 (10.1%, 9.7%)
DRIVE [17]	40	45°	565x584 540 px	Training (20) Test (20)	2 (12.7%, 12.3%)
HRF [18]	45	60°	3504x2336 3262 px	Healthy (15) DR (15) Glaucoma (15)	1 (9.13%)
STARE [19]	20	35°	700x605 649 px	–	2 (10.3%, 14.8%)

Bankhead method, which was evaluated only using DRIVE. Our experiments were done on three other databases with a range of different properties that lead to more robust comparison of the methods and allow to model and predict settings of the methods' parameters.

3. Experiment setup

3.1. Setup of the grid search

A grid search was used to find the best segmentation performance in a subspace created by a cartesian product over the sampled sets of the parameter values in Table 2. The parameters of the methods were optimized for highest accuracy (Acc)[6]. Below we discuss the details for each method that differ from standard grid search procedure.

With the *Soares method*, parameter settings of the classifiers were determined by preliminary experiments and the reported values were set as indicated in Table 2. To search for the best Λ_{mor}, a greedy optimization approach was used in order to avoid evaluating poorly performing Λ_{mor}.

[6]Optimization for highest area under the ROC curve (AUC) was also done, although this paper deals only with the results of the Acc optimization except for Figure 2. Complete results from the optimization procedure are available at `http://www.it.lut.fi/mvpr/medimg` along with the testing framework.

Table 2: Sampled values of the parameters. Superscript over the set of wavelet levels (e.g., $\{1,2\}^{\leq 2}$) represents subsets of the indicated size (e.g. $\{1,2\}^{\leq 2} = \{\{1\},\{2\},\{1,2\}\}$). *Azzopardi method*: Assessment of the algorithm performance was done separately for individual parameters beginning at the 'Starting point' and adding values indicated in the column 'Test ranges'. *Bankhead method*: ξ_s was sampled logarithmically in the interval $\langle 0, \text{max value (all vessels removed)}\rangle$ on each database. AR, CH, DR, HR, ST are abbreviated names of the databases.

Soares method		*Bankhead method*		*Nguyen method*	
Λ_{mor}	$\{1,2\cdots 17\}^{\leq 3}$	Λ_{ban}	$\{1,2,\cdots,4\}^{\leq 4}$ (AR,CH,DR,ST)	W	$\{12,13,\cdots,16\}$ (AR, DR, ST)
n_s	$\{2\cdot 10^5\}$ (AR, DR, ST)		$\{1,2,\cdots,5\}^{\leq 3}$ (HR)		$\{45,46,\cdots,55\}$ (HR)
	$\{3\cdot 10^5\}$ (CH, HR)	p_t	$\{.05,.1,\cdots,.3\}$		$\{25,26,\cdots,35\}$ (CH)
n_{g1}	$\{30\}$	ξ_s	40 values in log. scale	ω	$\{2,4,\cdots,W-1\}$
n_{g2}	$\{40\}$	ξ_h	$\{0\}$	τ	$\{0.5,0.55,\cdots,1.5\}$

Sofka method		*Azzopardi method*						
τ, pad	$\{0.50,0.52,\cdots,2.0\}$ (AR)		Starting point					
preprocessing	$\{0.50,0.52,\cdots,1.5\}$ (CH,DR)	Parameter	AR	CH	DR	HR	ST	Test ranges
	$\{0.50,0.52,\cdots,1.5\}$ (HR,ST)	σ_1	2.5	4.8	2.4	7.2	2.7	$\{0,\pm 0.2,\pm 0.4\}$
τ, CLAHE	$\{20.0,20.02,\cdots,21.0\}$ (AR)	σ_2	2	4.3	1.8	6.8	2.1	$\{0,\pm 0.2,\pm 0.4\}$
preprocessing	$\{18.0,18.02,\cdots,20.5\}$ (CH)	r_1	10	18	8	26	12	$\{0,\pm 3,\pm 6\}$
	$\{8.0,8.02,\cdots,14.0\}$ (DR)	r_2	24	34	22	50	24	$\{0,\pm 3,\pm 6\}$
	$\{31.50,31.52,\cdots,33.5\}$ (HR)	σ_{01}	2	3	3	2	1	$\{0,\pm 1,\pm 0.5\}$
	$\{20.0,20.02,\cdots,22.0\}$ (ST)	σ_{02}	1	1	2	1	1	$\{0,\pm 1,\pm 0.5\}$
		a_1	0.4	0.2	0.7	0.4	0.6	$\{0,\pm 0.4,\pm 0.2\}$
		a_2	0.1	0.1	0.1	0.1	0.1	$\{0,\pm 0.4,\pm 0.2\}$
		τ	$\{0.1,0.105,\cdots,0.25\}$					

With CHASEDB1 and HRF, it was infeasible to execute the *Sofka method* on the images in the original resolution. Therefore, the images were downscaled so that the longer side of an image was 600 pixels.

The *Azzopardi method* is set up using 8 parameters, which is a relatively large number. As a consequence, an exhaustive search in the parameter space requires evaluating the performance at too many points. Therefore, it was decided to explore the performance for the parameters individually beginning at a point defined in the paper [10] or selected randomly. Then, the ranges defined in Table 2 were followed from the starting point only in the direction of the parameter axes. The whole procedure was repeated from the best observed point as long as there was an improvement in the performance.

Performance of the *Bankhead method* was observed to be influenced very little by parameter ξ_{h}, and thus, ξ_{h} was not used in the experiment. It should be noted that the AUC measure was assessed on the vessel-enhanced image before the binarization and cleaning as the post-processing step. As a consequence, it yielded low values. It is possible to observe the influence of the cleaning on the ROC characteristics in Figure 2.

3.2. Training data

Subsets of the databases were used as training data to optimize the parameters and to train the classifier of the *Soares method*. The number of training images differed for each database: With DRIVE, the dedicated training set was used. With HRF, 15 random images were selected for training. With ARIADB, 30 random images were selected. With STARE and CHASEDB1, each database was randomly divided into two subsets of the same length and each subset was used to train a classifier. Each of the classifiers was then used to classify images from the training set of the other classifier.

3.3. Image preprocessing

Two different preprocessing approaches of the input images were identified among the methods. The first one is the 'pad only' preprocessing method which pads the edges of the FOV [8] and forms a part of the *Soares method*. The second one is the 'CLAHE' preprocessing method, where the image is padded as above and then contrast limited adaptive histogram equalization (CLAHE) is applied. This approach comes as a part of the implementation of the Azzopardi method. With the pad only preprocessing, padding by 50 pixels was used. With the CLAHE preprocessing, the image was padded to the image edges before applying the CLAHE algorithm with 6x6 tiles per image.

Each method except Sofka was applied to the green channel of the input image. The *Sofka method* can be applied to the green channel only or to the full color image. It produced better results when applied to the color image.

3.4. Performance measures

The segmentation performance of each algorithm was assessed using the manual segmentation of a database's first observer (sorted in alphabetic order)

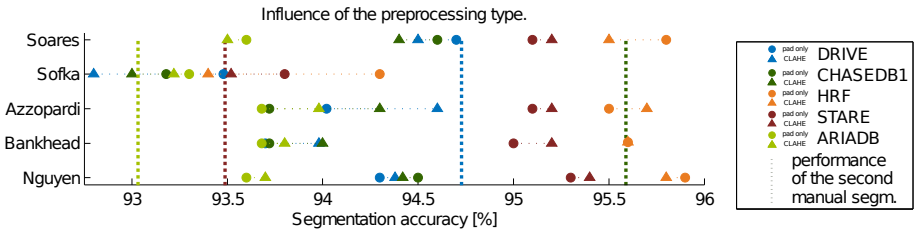


Figure 1: Comparison of the segmentation performance: 'pad only' (circles) and 'CLAHE' (triangles). The databases are marked with different colors. Solid dots and triangles mark the algorithm accuracy. Vertical lines mark the performance of the second observer when available.

as the ground truth and Matthew's correlation coefficient (MCC) [20], Acc, sensitivity (Sn), specificity (Sp) and AUC as the performance measures. Acc, Sn, Sp and AUC are established measures for assessing the vessel segmentation performance. MCC appeared recently in the vessel segmentation literature (for example, [10]) and can give more insight into the evaluation when the sample sizes of the classes are skewed, which is the case in vessel segmentation. The performance measures are expressed as percentages. If two sets of manual segmentations were provided in a database, the performance of the second observer (second manual segmentation) was assessed and compared to the performance of the automatic methods. Measurement of the performance was always done only on pixels inside the FOV. In the case of databases without a FOV mask (ARIADB, CHASEDB1, STARE), the mask was generated using edge detection and ellipse fitting.

4. Results

This section is organized as follows; Section 4.1 reports the best performance on each database for each algorithm and compares the algorithms, Section 4.2 reports parameters corresponding to the results reported in Section 4.1 and deals with prediction of the parameters, and Section 4.3 provides a comparison of the results to current state-of-the-art methods.

4.1. Performance of the algorithms

The performance of the algorithms is presented in Table 3. It is possible to refer to Figure 1 for visual comparison of the resulting accuracy and its change with change of the preprocessing. ROC characteristics of the methods are provided in Figure 2 for a more general comparison of the methods. In the ROC comparison, each method is presented only with the preprocessing type that resulted in better performance, applied to all databases, to keep the amount of data reasonable.

First we compare the influence of preprocessing on each method and focus only on the preprocessing approach that leads to better performance. Influence of the preprocessing selection is illustrated in Figure 1. Applying the CLAHE

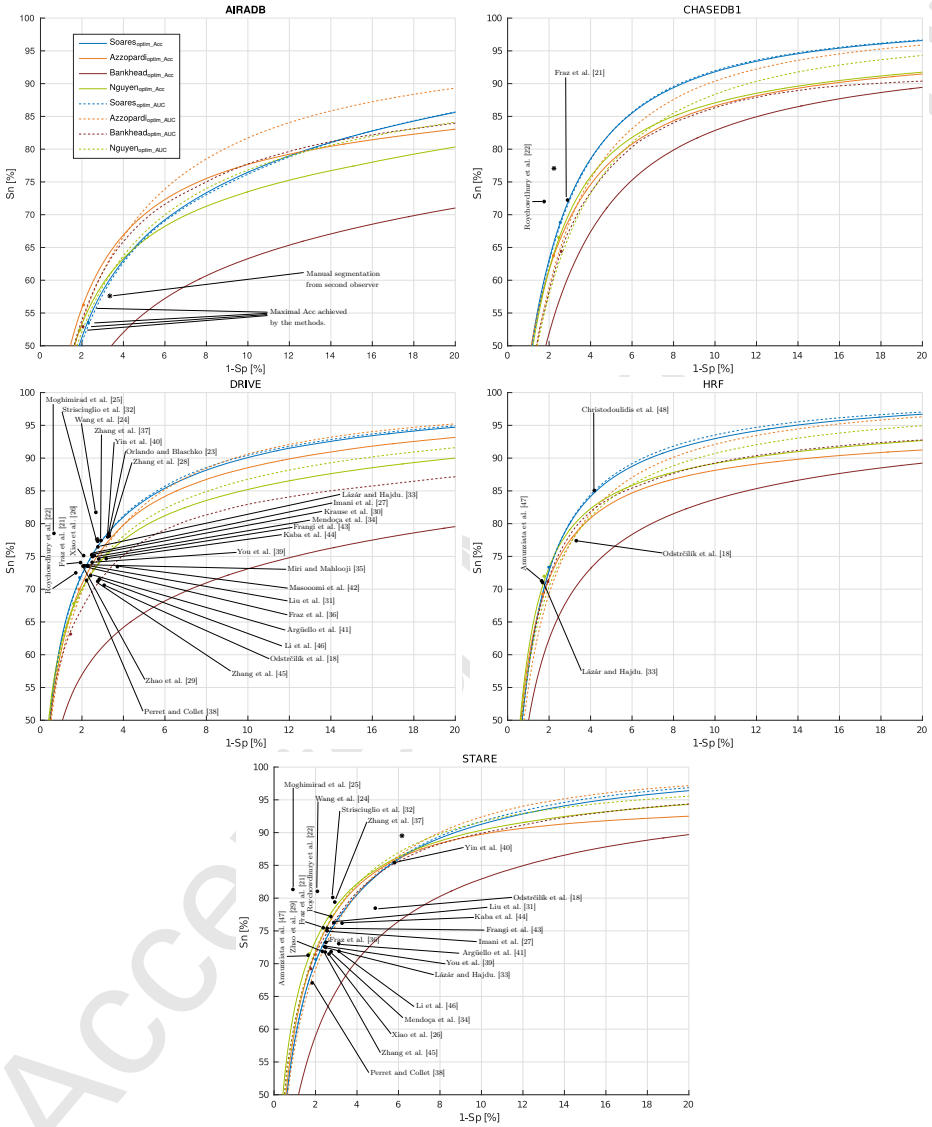


Figure 2: ROC characteristics of the studied methods. Manual segmentation by second observer is marked with an asterisk. The ROC curves correspond to parameters optimized by Acc (solid line) and AUC (dotted line) measures. The *Bankhead method* is different due to its postprocessing: one is the ROC curve of the IUWT response (solid line) and the other is the convex hull of all possible performances from the parameter search (dotted line). Legend relevant for all sub-figures is placed in the ARIADB sub-figure.

Table 3: Performance of the methods, each assessed with selected type of preprocessing – *Soares and Sofka methods* with pad only and *Azzopardi, Bankhead and Nguyen methods* with CLAHE. Corresponding parameter settings are given in Section 4.2. The results published in the original papers are shown in italics under the current results. Best scoring result in sense of Acc and AUC for each database.

Pad only – Soares, Sofka. CLAHE – Azzopardi, Bankhead, Nguyen.

	Soares method					*Sofka method*					*Azzopardi method*				
—	Acc	Sn	Sp	MCC	AUC	Acc	Sn	Sp	MCC	AUC	Acc	Sn	Sp	MCC	AUC
AR	93.6	53.6	97.7	57.6	**90.7**	93.3	44.5	98.3	53.3	86.3	**94.0**	56.2	97.9	60.8	89.2
CH	**94.6**	69.0	97.4	68.9	**96.4**	93.2	50.9	98.0	58.0	90.8	94.3	63.7	97.8	66.7	93.2
	—	—	—	—	—	—	—	—	—	—	*93.9*	*75.9*	*95.9*	*68.0*	*94.9*
DR	**94.7**	71.7	98.1	75.0	**96.1**	93.5	60.9	98.2	67.9	91.5	94.5	70.0	98.1	74.0	95.6
	94.7	—	—	—	—	—	—	—	—	—	*94.4*	*76.6*	*97.0*	*74.8*	*96.1*
HR	95.8	73.4	98.0	73.3	**97.0**	94.3	58.3	97.8	61.4	93.7	95.7	69.3	98.3	72.0	95.6
ST	95.1	70.3	98.0	72.6	**96.7**	93.8	56.5	98.1	62.7	92.4	95.3	71.4	98.0	73.3	95.2
	94.8	—	—	—	—	—	—	—	—	—	*95.0*	*77.2*	*97.0*	*73.4*	*95.6*

	Bankhead method					*Nguyen method*					Second manual segm.			
—	Acc	Sn	Sp	MCC	AUC	Acc	Sn	Sp	MCC	AUC	Acc	Sn	Sp	MCC
AR	93.8	56.9	97.6	50.0	85.7	93.8	52.4	98.1	58.7	87.5	93.0	57.6	96.7	57.0
CH	94.0	64.4	97.4	65.4	91.7	94.4	66.5	97.5	67.4	93.5	95.6	77.0	97.8	75.7
DR	94.0	63.1	98.6	70.7	90.5	94.5	67.8	98.4	73.3	93.4	94.7	77.6	97.3	76.0
	93.7	*70.2*	*97.2*	—	—	*94.1*	—	—	—	—	—	—	—	—
HR	95.6	71.2	98.1	71.8	92.1	**95.8**	72.0	98.2	73.3	94.7	—	—	—	—
ST	95.2	69.2	98.2	72.3	93.7	**95.5**	71.5	98.3	74.3	95.8	93.5	89.5	93.8	72.2
	—	—	—	—	—	*93.2*	—	—	—	—	—	—	—	—

preprocessing had a positive effect on the *Azzopardi and Bankhead methods*. With the *Nguyen method*, the effect was positive on average, but the results were inconsistent. In general, the choice of the preprocessing approach had the smallest effect on the *Nguyen method*. The effect on the *Soares method* was negative, with the exception of the STARE database. The effect on the *Sofka method* was negative for all databases. The absolute difference between the accuracy measured with CLAHE and pad only was up to 0.5 in percentage units. As a result, the comparison that follows will consider the results from CLAHE preprocessed images for the *Nguyen, Azzopardi and Bankhead methods*, and the results based on pad only for the *Soares and Sofka methods*.

Comparison of the methods reveals relatively similar performance. For the four methods except *Sofka method*, the absolute difference between the best and the worst accuracy on individual databases was up to 0.5 in percentage units. No method seemed to be clearly superior. The best and second best performance was achieved by the *Azzopardi and Bankhead methods* on ARIADB, the *Soares and Nguyen methods* on CHASEDB1, the *Soares and Azzopardi methods* on DRIVE, the *Nguyen and Soares methods* on HRF and the *Nguyen and Azzopardi methods* on STARE.

The search for optimal parameters brought small improvement in the performance of the algorithms compared to the performance published in the original papers. Nguyen et al. [13] obtained lower τ value than was obtained in the experiments presented in this paper which led to significantly worse performance on STARE. Compared to the original papers, the performance of the *Bankhead*

Table 4: Parameters corresponding to the results reported in Table 3.

	ARIADB	CHASEDB1	DRIVE	HRF	STARE
Soares	pad only	pad only	pad only	pad only	pad only
Λ_{mor}	{2,3,5}	{3,7,9}	{2,4}	{5,13,15}	{2,4,5}
n_s	2e5	3e5	2e5	3e5	2e5
Sofka	pad only	pad only	pad only	pad only	pad only
τ	1.86	1.3	0.82	3.74	1.75
Azzopardi	CLAHE	CLAHE	CLAHE	CLAHE	CLAHE
σ_1, σ_2	3.3, 1.6	5.3, 3.9	2, 1.6	7.2, 6.4	2.8, 1.6
r_1, r_2	19, 27	21, 22	8, 25	23, 44	12, 28
σ_{01}, σ_{02}	1, 0.5	2.5, 0	3, 1.5	0.5, 0	0.5, 0.5
a_1, a_2	0.4, 0.1	0.2, 0	0.5, 0.1	0.4, 0	0.5, 0.1
τ	0.14	0.16	0.16	0.16	0.15
Bankhead	CLAHE	CLAHE	CLAHE	CLAHE	CLAHE
Λ_{ban}	{2,3}	{3,4}	{2,3}	{3,4}	{2,3}
p_t	0.12	0.12	0.12	0.12	0.12
ξ_s	522	780	150	1030	270
Nguyen	CLAHE	CLAHE	CLAHE	CLAHE	CLAHE
W	19	33	17	45	17
ω	18	16	4	22	16
τ	1.05	0.95	0.9	1.05	1.05

and Nguyen methods were slightly improved by CLAHE preprocessing.

4.2. Prediction of the parameters

Parameters obtained from the state-space search are reported in Table 4. These parameters correspond to the results in Table 3. Taking advantage of having multiple test databases with different image resolutions, we aimed to estimate how well is it possible to predict the optimized parameters using linear models. Two different predictors were chosen: angular resolution of the segmented database $d_r = \frac{\text{FOV}\varnothing}{\text{FOV}[^\circ]}$ and the percentage of the pixels in the ground truth $d_n = N_1$ using the values from Table 1.

The parameters d_r and d_n were used to create two linear models using least squares fitting: model M_r with d_r as a predictor, and model M_{rn} with two predictors d_r and d_n. As illustrated in Figure 3, many of the parameters are well correlated with d_r. The ability of the models to fit the parameter values was measured by the segmentation accuracy when the predicted parameters are used. The results are reported in Table 5. Selection of M_r or M_{rn} in Table 5 depends on the segmentation performance achieved by the predicted settings. Model M_{rn} was selected when it gave better results than M_r; otherwise M_r was preferred.

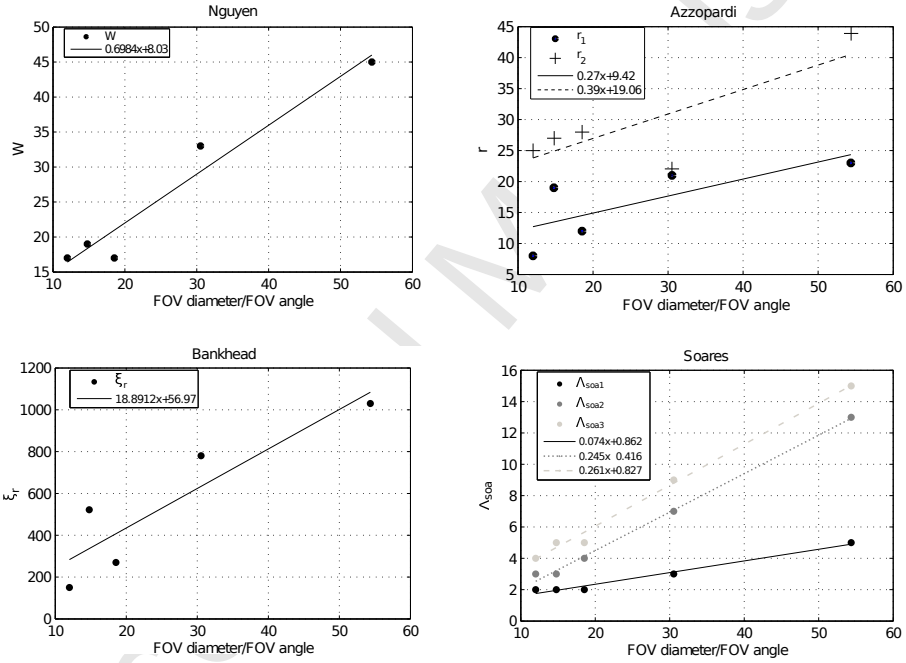


Figure 3: Illustration of the linear models. Selected parameters are correlated with angular resolution of the databases.

Table 5: Prediction of the method parameters by angular resolution of the databases (M_r) or by the angular resolution and the percentage of vessel pixels in the ground truth (M_{rn}). Least squares method was used to fit the models. *Assessment* reports the segmentation accuracy when the predicted parameters were used.

database	ARIA	CHASEDB1	DRIVE	HRF	STARE
Soares, Λ_{mor}	{2,3,5}	{3,7,9}	{2,3,4}	{5,13,15}	{2,4,5}
M_r	{2,3,5}	{3,7,9}	{2,3,4}	{5,13,15}	{2,4,6}
assessment	93.6	94.6	94.7	95.8	95.1
Azz., s_1, s_2	3.3,1.6	5.3,3.9	2.0,1.6	7.2,6.4	2.8,1.6
M_{rn}	3.2,1.4	4.7,3.6	2.0,1.6	7.4,6.5	3.4,2.0
Azz., r_1, r_2	19,27	21,22	8,25	23,44	12,28
M_{rn}	17.7,24.9	18.0,31.0	7.8,23.4	23.8,40.5	16.0,26.3
Azz., $s0_1, s0_2$	1,0.5	2.5,0	3,1.5	0.5,0	0.5,0.5
M_{rn}	0.8,0.3	1.3,0.4	3.0,1.4	0.87,0	1.4,0.5
Azz., a_1, a_2	0.4,0.1	0.2,0	0.5,0.1	0.4,0	0.5,0.1
M_{rn}	0.4,0.1	0.4,0	0.5,0.1	0.3,0	0.4,0.1
Azz., τ	0.14	0.16	0.16	0.16	0.16
M_{rn}	0.14	0.15	0.16	0.16	0.15
assessment	94.0	94.0	94.5	95.5	95.3
Bankhead, ξ_r	522	780	150	1029	270
M_r	336	634	284	1084	407
assessment	93.8	94.1	93.9	95.6	95.2
Nguyen, τ	1.05	0.95	0.90	1.05	1.05
M_{rn}	1.05	1.01	0.90	1.03	1.01
Nguyen, W	19	33	17	45	17
M_{rn}	18	29	17	46	21
Nguyen, ω	18	16	4	22	16
M_{rn}	16	16	4	22	16
assessment	93.8	94.1	94.5	95.8	95.4

All the methods scale well among images of different resolution. Of the multiscale parameters that were modeled—Λ_{mor}, r_1, r_2, s_1, s_2 and W—the parameters r_1, r_2 of the *Azzopardi method* were the most difficult to predict. With the *Soares method*, three models were trained—one for each wavelet level in Λ_{mor}—and they provided very good prediction of the parameters. Other modeled parameters included τ of the *Azzopardi and Nguyen methods*, which were well predicted by $\mathrm{M_{rn}}$. Parameter ξ_r of the *Bankhead method* was predicted sufficiently by $\mathrm{M_r}$. The parameters p_t, Λ_{ban} of the *Bankhead method* spans small range of values and fixed value 0.12 was used as the modeled one.

4.3. Comparison with state-of-the-art

Here we provide a brief comparison of the tested methods with state-of-the-art methods which are not available with implementation. Papers for the comparison that were published before year 2011 were gathered from the review by Fraz et al. [4]. The more recent papers were gathered from the list of papers that cite [15], [16], [17], [18], [19], the publications introducing the databases reviewed in Section 2.2. We observed that many methods report high performance but without providing clear methodology of the performance assessment. To ensure that the comparison is fair, only methods explicitly stating that performance was measured on pixels inside the FOV were included.

Typically the performance of retinal vessel segmentation algorithms is reported on DRIVE and STARE and, thus, many results are available using those databases. Few methods were identified that reported performance also on CHASEDB1 and HRF. Comparison of those state-of-the-art methods for which accuracy was reported is presented in Tables Table 6, 7 and 8. When sensitivity and specificity were also provided, the performance was plotted in Figure 2. The latter way of comparing the methods enables clearer and more fair way of comparison.

5. Discussion

As a consequence of lacking standard evaluation methodology for automatic retinal image processing methods [53], factors affecting the outcome of the evaluation (for example, test set size, image preprocessing, evaluation metrics) vary among studies. Here, segmentation accuracy was used as the primary measure for comparing the segmentation performance. It is the most widely used measure for assessing the segmentation performance and it makes interpreting the results intuitive. Using accuracy alone can lead to simplifying conclusions when, in general, accuracy of the segmentations with different sensitivity are compared. Typically, human raters tend to produce results with higher sensitivity than automatic algorithms, but their accuracy can be lower. Performance on STARE database is a typical example [8]. Automatic methods, on the other hand, might not be able to compete with manual segmentation at the same sensitivity level. Therefore, we found it necessary to provide also the ROC characteristics (Figure 2) of the algorithms so that the relative performance of the methods at different sensitivity is demonstrated.

Table 6: Overview of state-of-the-art methods evaluated on DRIVE and STARE. The methods are sorted by mean performance on both databases.

	DRIVE				STARE			
Algorithm	Sn	Sp	Acc	AUC	Sn	Sp	Acc	AUC
Wang et al. [24]	81.7	97.3	97.7	94.8	81.0	97.9	98.1	97.5
Moghimirad et al. [25]	78.5	99.4	96.6	95.8	81.3	99.1	97.6	96.8
Imani et al. [27]	75.2	97.5	95.2	—	75.0	97.5	95.9	—
Al-Rawi et al. [49]	—	—	95.4	94.4	—	—	—	—
Lam et al. [50]	—	—	94.7	96.1	—	—	95.7	97.4
Liu et al. [31]	73.5	97.7	94.7	—	76.3	97.1	95.7	—
Annunziata et al. [47]	—	—	—	—	71.3	98.4	95.6	96.6
Roychowdhury et al. [22]	72.5	98.3	95.2	96.2	77.2	97.3	95.2	96.9
Fraz et al. [21]	74.1	98.1	94.8	97.5	75.5	97.6	95.3	97.7
Xiao et al. [26]	75.1	97.9	95.3	—	71.5	97.4	94.8	—
Zhang et al. [28]	78.1	96.7	95.0	—	—	—	—	—
Strisciuglio et al. [32]	77.3	97.2	94.7	95.9	80.1	97.2	95.4	96.3
Zhao et al. [29]	73.5	97.9	94.8	—	71.9	97.7	95.1	—
Krause et al. [30]	75.2	97.4	94.7	—	—	—	—	—
Soares method	71.7	98.1	94.7	96.1	70.3	98.0	95.1	96.7
Zhang et al. [37]	77.4	97.1	94.5	—	79.4	97.1	95.1	—
Nguyen method	67.8	98.4	94.5	93.4	71.5	98.3	95.5	95.8
Azzopardi method	70.0	98.1	94.5	95.6	71.4	98.0	95.3	95.2
Orlando and Blaschko [23]	78.5	96.7	—	—	—	—	—	—
Staal et al. [17]	—	—	94.4	95.2	—	—	95.2	96.1
Miri and Mahlooji [35]	73.5	98.0	94.6		—	—	—	—
Fraz et al. [36]	73.5	97.7	94.5	96.7	73.3	97.5	95.0	96.7
Perret and Collet [38]	71.4	97.8	94.4	—	67.1	98.2	95.1	—
Lázár and Hajdu. [33]	76.5	97.2	94.6	—	72.5	97.5	94.9	—
You et al. [39]	74.1	97.5	94.3	—	72.6	97.6	95.0	—
Tagore et al. [51]	—	94.2	95.3	—	—	—	95.0	96.1
Masooomi et al. [42]	73.5	96.3	94.3	—	—	—	—	—
Frangi et al. [43]	74.6	97.2	94.2	—	75.4	97.4	95.0	—
Mendoça et al. [34]	75.0	97.5	94.6	—	71.8	97.3	94.6	—
Bankhead method	63.1	98.6	94.0	90.5	69.2	98.2	95.2	93.7
Argüello et al. [41]	72.1	97.6	94.3	—	73.1	96.9	94.5	—
Zhang et al. [45]	71.2	97.2	93.8	—	71.8	97.5	94.8	—
Kaba et al. [44]	74.7	96.8	94.1	—	76.2	96.7	94.6	—
Yin et al. [40]	78.0	96.8	94.3	—	85.4	94.2	93.3	—
Sofka method	60.9	98.2	93.5	91.5	56.5	98.1	92.4	93.8
Li et al. [46]	71.5	97.2	93.4	—	71.9	96.9	94.1	—
Odstrčilík et al. [18]	70.6	96.9	93.4	95.2	78.5	95.1	93.4	95.7

Table 7: Overview of state-of-the-art methods evaluated on CHASEDB1.

Algorithm	Sn	Sp	Acc	AUC
Roychowdhury et al. [22]	72.0	98.2	95.3	95.3
Fraz et al. [21]	72.2	97.1	94.7	97.1
Soares method	69.0	97.7	94.6	96.4
Nguyen method	66.5	97.5	94.4	93.5
Azzopardi method	63.7	97.8	94.3	93.2
Bankhead method	64.4	97.4	94.0	91.7
Sofka method	45.6	98.3	93.0	89.1

Table 8: Overview of state-of-the-art methods evaluated on HRF.

Algorithm	Sn	Sp	Acc	AUC
Cheng et al. [52]	70.4	98.6	96.1	—
Soares method	73.4	98.0	95.8	97.0
Christodoulidis et al. [48]	85.1	95.8	94.8	—
Nguyen method	72.0	98.2	95.8	94.7
Annunziata et al. [47]	71.3	98.4	95.8	—
Azzopardi method	69.3	98.3	95.7	95.6
Bankhead method	71.2	98.1	95.6	91.3
Lázár and Hajdu. [33]	71.0	98.3	95.3	—
Odstrčilík et al. [18]	77.4	96.7	94.9	96.7
Sofka method	58.3	97.8	94.3	93.7

The performance of all the methods, except the *Sofka method*, was very much similar on all the databases studied, and the order of best-performance varied based on the different databases analyzed. *Soares method* provides a slightly better AUC—it offers better accuracy when higher sensitivity is considered. The comparison with the state-of-the-art methods (Figure 2) revealed only few significantly better methods and showed the methods perform close to each other. There might thus be a need to improve the whole assessment methodology. One option could be to use skeletonized vessels [9], which results in higher demands on the detection of narrower vessels.

As stated above, ROC curves offer a meaningful way to compare the performance of the automatic methods and manual segmentations. It was shown in Figure 2 that the manual segmentation of CHASEDB1 and STARE outperform the automatic segmentation markedly, and they can be regarded as challenging datasets. For DRIVE, the Soares and Azzopardi methods offer performance very close to the manual segmentation, and for ARIA, the manual segmentation is outperformed by all the methods.

Considering the advantages and disadvantages of the methods reviewed from the original publications in Section 2.1, most authors claimed fast segmentation. The exact segmentation times were not measured in our study, but it is possible to conclude that each method classified an image of STARE or DRIVE within the order of seconds, the fastest method being the *Bankhead method* and followed by the *Nguyen and Azzopardi methods*. It is worth noting that

training phase of the *Soares method* was speeded up approximately 20 times by optimizing the source code.

The *Sofka method* was the only method designed to prevent false positive classifications around high-contrast areas like pathologies or the optic disc. Performance of the method was, however, well below the others. Qualitative inspection of the results showed that the reason for the poor performance can be mainly because of the method is not well suited for pixel-wise classification [9]. The resulting vesselness gains high values near the vessels thus producing classification wider than the one used as the ground truth. Therefore, in [9] the method was used to detect the vessel centerlines. However, the method was not inspected closer as it is provided as an executable binary, and our focus was more on the methods provided with the source code.

Qualitative inspection[7] of the other methods showed that with the *Soares method* false positive detections at the edge of the optic disc are rare but appear more often with the unsupervised methods. This was, however, largely corrected by applying the CLAHE preprocessing. The preprocessing, on the other hand, increased the noise in the images as well as the subsequent classifications. The *Azzopardi method* was the most resistant to the noise and the only one benefiting significantly from the preprocessing. Uneven illumination does not seem to influence the classification of the unsupervised methods. A more important factor appears to be that the vessels in the over or under illuminated areas have reduced contrast. On the other hand, the *Soares method* is affected more by the uneven illumination, possibly due to fact that there are not enough examples of such images in the training data.

Considering the individual methods, the *Soares method* provides very good performance if part of the classified data is available for training. Training on a different database leads to a performance drop [8] and when databases with different image resolution are considered, it is infeasible to properly train the classifier unless the images are resampled. The *Soares method*, when properly trained, provides high AUC which is valuable when segmentation with higher sensitivity is desired. Of the unsupervised methods, the *Azzopardi method* yielded the highest AUC, that in some databases was even comparable with the *Soares method*. The *Azzopardi method* seems to produce robust segmentation of the wider vessels and also has the best robustness against pathologies and other high-contrast areas. Narrow vessels cause lower response of the filter and are usually missed. The *Nguyen method* provides performance similar to the *Azzopardi method*, but looses performance around the pathologies with high contrast. On the other hand, it gives higher response for narrow vessels thus providing a more balanced segmentation. AUC of the *Nguyen method* is lower than the *Soares and Azzopardi methods*. Lastly, the *Bankhead method* provides the worst performance although not by a high margin. At the same time, it is the fastest method. Also the *Bankhead method* exhibits performance improve-

[7]The vessel segmentation in all tested images is available at `http://www.it.lut.fi/mvpr/medimg`.

ment from simple postprocessing such as the removal of isolated objects up to a certain size, as can be seen in the performance gain of the method in Figure 2.

Most of the parameters of the methods proved to be possible to model and predict by using the database resolution and the expected percentage of vessel pixels in the ground truth. Also, strong correlation of most of the parameters with the predictors (Figure 3) supports this hypothesis. As a result, the proposed models are expected to provide close-to-optimal parameters for the methods. However, the models were not on databases not used to establish the models. This was due to the limited number of the databases. Therefore, the models could fail on other unseen databases. In that case, the *Bankhead method* seems to be the option with the smallest number of choices required about the parameter settings and the best robustness against parameter variation as it needs a small range of wavelet coefficients, uses the percentile-based thresholding and the size of the removed objects can be defined proportional to the ROI size.

With regard to the sets of the tested parameter values, the range limits were carefully selected. In the case of the *Bankhead and Nguyen methods*, the parameters are expected to be sampled properly to get close-to-optimal settings on each database. The Azzopardi method is harder to optimize because it has eight parameters. The whole parameter subspace defined by the ranges was thus difficult to evaluate, so there might be room for improvement. Azzopardi et al. [10] have, however, optimized the parameters already for DRIVE, STARE and CHASEDB1. With the *Soares method*, no parameter search was done for the parameters of the Morlet wavelet other than Λ_{mor}, which was limited to integer values. This provides room for further improvement of the Morlet wavelet based vessel segmentation methods and the current tested method in particular; setting the ϵ and k_0 parameters can produce a filter with similar properties as the line operator that is implemented in the *Nguyen method*. Advantages of this type of response could be integrated into the framework of the *Soares method*. One observation with respect to the wavelet filter mask is that between the Λ_{mor} values 1 and 1.5, the wavelet mask has a different mean value for different orientations and using a different normalization of the mask might enable use of the response with wavelet level size <2.

6. Conclusion

This study reviewed retinal vessel segmentation methods with publicly available implementation and publicly available databases of color fundus photographs containing ground truth for vessel segmentation. Two supervised and three unsupervised methods were studied and quantitatively compared using five publicly available databases. Two types of image preprocessing approaches were tested and the method parameters were optimized for the best performance on each database. In addition, the studied methods were compared to recent state-of-the-art approaches.

The results show that the parameter optimization does not significantly improve the segmentation performance of the methods when the original data is

used. However, the performance of the methods in new image data differs significantly. The performance of the tested methods with respect to accuracy was very close; highest performance was achieved on ARIADB by the Azzopardi method (Acc 94.0), on CHASE and DRIVE by the *Soares method* (Acc 94.6, 94.7) and on HRF and STARE by the *Nguyen method* (Acc 95.8, 95.5). The *Soares and Azzopardi methods* usually provides higher area under the ROC curve than the other methods. Preprocessing of the images with CLAHE improved the overall performance of the unsupervised methods. Parameters yielding the reported performance are also provided to give reasonable parameter ranges and starting points to support optimization on new data. Finally, it was possible to predict parameters that give best segmentation performance for each method.

7. Conflict of interest

None.

8. Acknowledgements

The authors would like to thank the Academy of Finland for the financial support of the ReVision project (No. 259560).

[1] J. J. Kanski, B. Bowling, Synopsis of Clinical Ophthalmology, Elsevier Health Sciences, 3rd edition, 2013.

[2] G. Liew, J. J. Wang, P. Mitchell, T. Y. Wong, Retinal vascular imaging a new tool in microvascular disease research, Circulation: Cardiovascular Imaging 1 (2008) 156–161.

[3] M. Abràmoff, M. Garvin, M. Sonka, Retinal imaging and image analysis, IEEE transactions on medical imaging 3 (2010) 169–208.

[4] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, S. A. Barman, Blood vessel segmentation methodologies in retinal images - a survey, Comput. Methods Prog. Biomed. 108 (2012) 407–433.

[5] R. Bernardes, P. Serranho, C. Lobo, Digital ocular fundus imaging: A review, Ophthalmologica 4 (2011) 161–81.

[6] C. Kirbas, F. Quek, A review of vessel extraction techniques and algorithms, ACM Comput. Surv. 36 (2004) 81–121.

[7] P. Vostatek, E. Claridge, P. Fält, M. Hauta-Kasari, H. Uusitalo, L. Lensu, Evaluation of publicly available blood vessel segmentation methods for retinal images, in: C. X, G. MK, L. J, T. E, X. Y (Eds.), Proceedings of the Ophthalmic Medical Image Analysis Second International Workshop, OMIA 2015, Held in Conjunction with MICCAI 2015, Iowa Research Online, 2015, pp. 137–144.

[8] J. Soares, J. Leandro, R. Cesar, H. Jelink, M. Cree, Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification, IEEE Transactions on Medical Imaging 25 (2006) 1214–1222.

[9] M. Sofka, C. V. Stewart, Retinal vessel extraction using multiscale matched filters, confidence and edge measures, IEEE Transactions on Medical Imaging 25 (2006) 1531–1546.

[10] G. Azzopardi, N. Strisciuglio, M. Vento, N. Petkov, Trainable cosfire filters for vessel delineation with application to retinal images, Medical image analysis 19 (2015) 46–57.

[11] P. Bankhead, C. N. Scholfield, J. G. McGeown, T. M. Curtis, Fast retinal vessel detection and measurement using wavelets and edge location refinement, PLoS ONE 7 (2012) e32435.

[12] J. Starck, J. Fadili, F. Murtagh, The undecimated wavelet decomposition and its reconstruction, IEEE Transactions on Image Processing 16 (2007) 297–309.

[13] U. T. Nguyen, A. Bhuiyan, L. A. Park, K. Ramamohanarao, An effective retinal blood vessel segmentation method using multi-scale line detection, Pattern recognition 46 (2013) 703–715.

[14] E. Ricci, R. Perfetti, Retinal blood vessel segmentation using line operators and support vector classification, Medical Imaging, IEEE Transactions on 26 (2007) 1357–1365.

[15] D. J. Farnell, F. Hatfield, P. Knox, M. Reakes, S. Spencer, D. Parry, S. Harding, Enhancement of blood vessels in digital fundus photographs via the application of multiscale line operators, Journal of the Franklin institute 345 (2008) 748–765.

[16] C. G. Owen, A. R. Rudnicka, R. Mullen, S. A. Barman, D. Monekosso, P. H. Whincup, J. Ng, C. Paterson, Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program, Investigative ophthalmology & visual science 50 (2009) 2004–2010.

[17] J. Staal, M. Abrámoff, M. Niemeijer, M. Viergever, B. van Ginneken, Ridge-based vessel segmentation in color images of the retina, IEEE Transactions on Medical Imaging 23 (2004) 501–509.

[18] J. Odstrcilik, R. Kolar, A. Budai, J. Hornegger, J. Jan, J. Gazarek, T. Kubena, P. Cernosek, O. Svoboda, E. Angelopoulou, Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database, IET Image Processing 7 (2013) 373–383.

[19] A. Hoover, V. Kouznetsova, M. Goldbaum, Locating blood vessels in retinal images by piece-wise threhsold probing of a matched filter response, IEEE Transactions on Medical Imaging 19 (2000) 203–210.

[20] P. Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, H. Nielsen, Assessing the accuracy of prediction algorithms for classification: an overview, Bioinformatics 16 (2000) 412–424.

[21] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, S. Barman, et al., An ensemble classification-based approach applied to retinal blood vessel segmentation, Biomedical Engineering, IEEE Transactions on 59 (2012) 2538–2548.

[22] S. Roychowdhury, D. D. Koozekanani, K. K. Parhi, Blood vessel segmentation of fundus images by major vessel extraction and subimage classification, IEEE journal of biomedical and health informatics 19 (2015) 1118–1128.

[23] J. I. Orlando, M. Blaschko, Learning fully-connected crfs for blood vessel segmentation in retinal images, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014, Springer, 2014, pp. 634–641.

[24] S. Wang, Y. Yin, G. Cao, B. Wei, Y. Zheng, G. Yang, Hierarchical retinal blood vessel segmentation based on feature and ensemble learning, Neurocomputing 149 (2015) 708–717.

[25] E. Moghimirad, S. H. Rezatofighi, H. Soltanian-Zadeh, Retinal vessel segmentation using a multi-scale medialness function, Computers in biology and medicine 42 (2012) 50–60.

[26] Z. Xiao, M. Adel, S. Bourennane, Bayesian method with spatial constraint for retinal vessel segmentation, Computational and mathematical methods in medicine 2013 (2013).

[27] E. Imani, M. Javidi, H.-R. Pourreza, Improvement of retinal blood vessel detection using morphological component analysis, Computer methods and programs in biomedicine 118 (2015) 263–279.

[28] L. Zhang, M. Fisher, W. Wang, Retinal vessel segmentation using multi-scale textons derived from keypoints, Computerized Medical Imaging and Graphics 45 (2015) 47–56.

[29] Y. Q. Zhao, X. H. Wang, X. F. Wang, F. Y. Shih, Retinal vessels segmentation based on level set and region growing, Pattern Recognition 47 (2014) 2437–2446.

[30] M. Krause, R. M. Alles, B. Burgeth, J. Weickert, Fast retinal vessel analysis, Journal of Real-Time Image Processing (2013) 1–10.

[31] X. Liu, Z. Zeng, X. Wang, Vessel segmentation in retinal images with a multiple kernel learning based method, in: Neural Networks (IJCNN), 2014 International Joint Conference on, IEEE, pp. 507–511.

[32] N. Strisciuglio, G. Azzopardi, M. Vento, N. Petkov, Multiscale blood vessel delineation using b-cosfire filters, in: International Conference on Computer Analysis of Images and Patterns, Springer, pp. 300–312.

[33] I. Lázár, A. Hajdu, Segmentation of retinal vessels by means of directional response vector similarity and region growing, Computers in biology and medicine 66 (2015) 209–221.

[34] A. Mendonça, B. Dashtbozorg, A. Campilho, Segmentation of the vascular network of the retina, Image Analysis and Modeling in Opthalmology (2014) 85–109.

[35] M. S. Miri, A. Mahloojifar, Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction, Biomedical Engineering, IEEE Transactions on 58 (2011) 1183–1192.

[36] M. Fraz, P. Remagnino, A. Hoppe, S. Barman, Retinal image analysis aimed at extraction of vascular structure using linear discriminant classifier, in: Computer Medical Applications (ICCMA), 2013 International Conference on, IEEE, pp. 1–6.

[37] J. Zhang, E. Bekkers, S. Abbasi, B. Dashtbozorg, B. ter Haar Romeny, Robust and fast vessel segmentation via gaussian derivatives in orientation scores, in: International Conference on Image Analysis and Processing, Springer, pp. 537–547.

[38] B. Perret, C. Collet, Connected image processing with multivariate attributes: an unsupervised markovian classification approach, Computer Vision and Image Understanding 133 (2015) 1–14.

[39] X. You, Q. Peng, Y. Yuan, Y.-m. Cheung, J. Lei, Segmentation of retinal blood vessels using the radial projection and semi-supervised approach, Pattern Recognition 44 (2011) 2314–2324.

[40] B. Yin, H. Li, B. Sheng, X. Hou, Y. Chen, W. Wu, P. Li, R. Shen, Y. Bao, W. Jia, Vessel extraction from non-fluorescein fundus images using orientation-aware detector, Medical image analysis 26 (2015) 232–242.

[41] F. Argüello, D. L. Vilariño, D. B. Heras, A. Nieto, Gpu-based segmentation of retinal blood vessels, Journal of Real-Time Image Processing (2014) 1–10.

[42] R. Masooomi, A. Ahmadifard, A. Mohtadizadeh, Retinal vessel segmentation using non-subsampled contourlet transform and multi-scale line detection, in: Intelligent Systems (ICIS), 2014 Iranian Conference on, IEEE, pp. 1–5.

[43] A. F. Frangi, W. J. Niessen, K. L. Vincken, M. A. Viergever, Multiscale vessel enhancement filtering, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, pp. 130–137.

[44] D. Kaba, C. Wang, Y. Li, A. Salazar-Gonzalez, X. Liu, A. Serag, Retinal blood vessels extraction using probabilistic modelling, Health Information Science and Systems 2 (2014) 2.

[45] B. Zhang, L. Zhang, L. Zhang, F. Karray, Retinal vessel extraction by matched filter with first-order derivative of gaussian, Comput. Biol. Med. 40 (2010) 438–445.

[46] Q. Li, J. You, D. Zhang, Vessel segmentation and width estimation in retinal images using multiscale production of matched filter responses, Expert Systems with Applications 39 (2012) 7600–7610.

[47] R. Annunziata, A. Garzelli, L. Ballerini, A. Mecocci, E. Trucco, Leveraging multiscale hessian-based enhancement with a novel exudate inpainting technique for retinal vessel segmentation (2015).

[48] A. Christodoulidis, T. Hurtut, H. B. Tahar, F. Cheriet, A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images, Computerized Medical Imaging and Graphics (2016).

[49] M. Al-Rawi, M. Qutaishat, M. Arrar, An improved matched filter for blood vessel detection of digital retinal images, Computers in Biology and Medicine 37 (2007) 262–267.

[50] B. S. Lam, Y. Gao, A. W.-C. Liew, General retinal vessel segmentation using regularization-based multiconcavity modeling, Medical Imaging, IEEE Transactions on 29 (2010) 1369–1381.

[51] M. Tagore, G. B. Kande, E. K. Rao, B. P. Rao, Segmentation of retinal vasculature using phase congruency and hierarchical clustering, in: Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on, IEEE, pp. 361–366.

[52] E. Cheng, L. Du, Y. Wu, Y. J. Zhu, V. Megalooikonomou, H. Ling, Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features, Machine Vision and Applications 25 (2014) 1779–1792.

[53] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancarlo, E. Chaum, J. P. Hubschman, B. Al-Diri, C. Y. Cheung, D. Wong, M. Abramoff, G. Lim, D. Kumar, P. Burlina, N. M. Bressler, H. Jelinek, F. Maiaudeau, G. Quellec, T. MacGillivray, B. Dhillon, Validating retinal fundus image analysis algorithms: Issues and a proposal, Investigative Ophthalmology & Visual Science (2013) 3546–3559.