A Self-Calibrating Approach for the Segmentation of Retinal Vessels by Template Matching and Contour Reconstruction

György Kovács[a,*], András Hajdu[a]

[a]*Faculty of Informatics, University of Debrecen, Debrecen, Hungary*

Abstract

The automated processing of retinal images is a widely researched area in medical image analysis. Screening systems based on the automated and accurate recognition of retinopathies enable the earlier diagnosis of diseases like diabetic retinopathy, hypertension and their complications. The segmentation of the vascular system is a crucial task in the field: on the one hand, the accurate extraction of the vessel pixels aids the detection of other anatomical parts (like the optic disc (Hoover and Goldbaum, 2003)) and lesions (like microaneurysms (Sopharak et al., 2013)); on the other hand, the geometrical features of the vascular system and their temporal changes are shown to be related to diseases, like the vessel turtousity to Fabry disease (Sodi et al., 2013) and the arteriolar-to-venus (A/V) ratio to hypertension (Pakter et al., 2005).

In this study, a novel technique based on template matching and contour reconstruction is proposed for the segmentation of the vasculature. In the template matching step generalized Gabor function based templates are used to extract the center lines of vessels. Then, the intensity characteristics of vessel contours measured in training databases are reconstructed. The method was trained and tested on two publicly available databases, DRIVE and STARE; and reached an average accuracy of 0.9494 and 0.9610, respectively. We have also carried out cross-database tests and found that the accuracy scores are higher than that of any previous technique trained and tested on the same database.

Keywords: template matching, retinal image analysis, vessel segmentation, Gabor filter, contour reconstruction, vascular system, self-calibration

1. Introduction

Retinopathies are indicators of various diseases of the eye and the body. Since the examination of the retina enables in vivo screening, the automated processing of retinal images is a widely researched area in the field of medical image analysis. Due to the wide variability of fundus images and variety of retinopathies, the accurate evaluation of the images requires both expertise and experience. Consequently, the development of automated screening systems is a complex task. The key points are the accurate localization and identification of landmarks; the precise segmentation of the anatomical parts and lesions; and the quantitative characterization of retinopathies. For several reasons, the segmentation of the vascular system is a crucial task in automated retinopathy screening applications:

1. Retinopathies related to diabetes are considered to be the major pathological cause of vision loss and blindness (Taylor and Keeffe, 2001), since the progression of diabetic retinopathies leads to the appearance of large and serious lesions on the retina. Considering that 4.4% of the world's population is forecast to have some level of diabetes in 2030 (Wild et al., 2004), the early diagnosis and prevention is a challenging, but substantial task of healthcare. The first sign of diabetic retinopathy is the presence of tiny capillary dilations called microaneurysms (Marín et al., 2011), appearing usually near thin vessels. Since microaneurysms have similar visual features as high-curvature and junction points of thin vessels, the vast majority of microaneurysm detection algorithms ignore the pixels of the vasculature (Fleming et al., 2006; Sopharak et al., 2013). Accordingly, the accurate segmentation of the vascular system aids the detection of microaneurysms indirectly. We note that the vascular system is also masked out during the segmentation of other retinopathies, like drusen (Bhulyan et al., 2013).

2. The characteristic shape of the vascular system can aid the localization of other, occasionally degenerated anatomical parts like the optic disc (Hoover and Goldbaum, 2003), or macula (Yu et al., 2014).

3. Novel studies have pointed out that the geometrical features of the vascular system are in correlation with various diseases. The authors of (Cheung et al., 2012) analyse the temporal changes of retinal vessel width and tortuosity related to retinopathies of prematurity. In (Sodi et al., 2013) the relation of tortuosity and the Fabry disease is discussed, while in (Li et al., 2010) an automated system for tortuosity diagnosis is described. In (Pakter et al., 2005) the A/V ratio of retinal vessels in

*Corresponding author
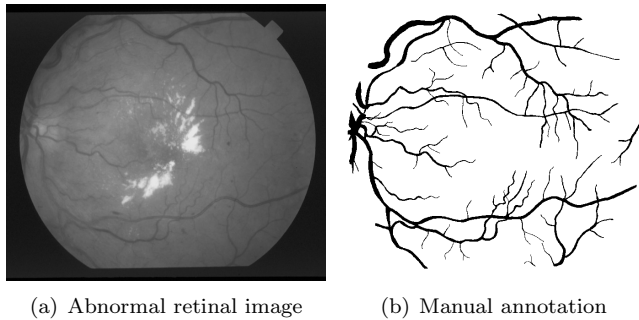
(a) Abnormal retinal image   (b) Manual annotation

Figure 1: The problem of vessel segmentation

hypertension patients is investigated, and accordingly, methods for the classification of artery and vein vessels and the automated calculation of the A/V ratio is presented in (Vázquez et al., 2013) and (Niemeijer et al., 2010a), respectively. In (Chapman et al., 2002), the associations between peripheral vascular disease and the abnormal arteriolar diameters at bifurcations of the vasculature are investigated, and automated methods for the measurement and characterization of bifurcations is described in (Al-Diri and Hunter, 2009). The novel study (Habib et al., 2014) discusses the associations between the changes of retinal vascular geometry and the progression of retinopathies. Age related rarefraction in the fractal dimensions of the vascular system and the correspondence between the fractal dimension and the risk of stroke is discussed in (Azemin et al., 2012; Kawasaki et al., 2011).

As a summary, and also according to the thorough and excellent overview of image processing issues related to retinal images in (Trucco et al., 2013), *there is substantial interest in automatically segmenting the vasculature and measuring its properties.*

The problem of segmenting the vascular system is illustrated in Figure 1 with a diseased fundus image and the corresponding manual (ground truth) annotation.

The paper is organized as follows. In the next section we give a short overview on previously published vessel segmentation techniques, in Section 3, a detailed description of the proposed method is given. A quantitative and comparative evaluation is presented in Section 4 and finally, a short summary is given in 5 and some conclusions are drawn in Section 6.

2. Related work

In the last decades, many algorithms have been published for vessel segmentation in fundus images. In the rest of the section we give a short overview with some remarks on the essence of these techniques.

Probably the first result on the segmentation of the vascular system comes from 1989, where the use of matched line segment filters with Gaussian profiles was proposed

(Chaudhuri et al., 1989). In the first half of the 90's, several papers were published on the segmentation of the vasculature in angiograms, in (Liu and Sun, 1993) a recursive tracking algorithm is described, and the vessel profiles are modelled by Gaussian functions and an adaptive tracking technique is used in (Zhou et al., 1994). Although these methods can be adapted for the segmentation of the vascular system in regular retinal images, many further methods have been proposed in the upcoming two decades. A neural network based method is described in (Gardner et al., 1996) for the classification of rectangular regions of the fundus image, the pixelwise neural network based classification is preceded by principal component analysis in (Sinthanayothin et al., 1999) and fuzzy c-means clustering is applied along vessel profiles for segmentation in (Tolias and Panas, 1998). A method based on the tracking of Gaussian and Kalman filter responses is proposed in (Chutatape et al., 1998) and dual edge tracking in the Canny edge map is applied in (Can et al., 1999).

The work (Hoover et al., 2000) has boosted the development of vessel segmentation techniques by publishing a database (STARE) containing 20 fundus images and two manual annotations for each image. Later on, another database called DRIVE has been made available (Niemeijer et al., 2004), and in a corresponding paper the authors gave a k-nearest neighbor classification based solution in (Staal et al., 2004). Although these two databases can be considered as the *de facto* standards of evaluation, it is worth to notice that they contain images of much lower resolution than the ones used in nowadays practice.

In (Jiang and Mojon, 2003), a threshold probing approach is described, while the authors of (Martinez-Perez et al., 2007) and (Vlachos and Dermatas, 2010) extended the multiscale analysis by some tracking features. In (Soares et al., 2004), 2D Gabor filter responses are used as features for supervised multivariate Gaussian mixture model based classification. Various matched filtering based techniques are proposed in (Gang et al., 2002; Al-Rawi and Karajeh, 2007; Al-Rawi et al., 2007; Cinsdikici and Aydin, 2009) and the vasculature is modelled by Hermite polynomials in (Wang et al., 2007; Lupascu et al., 2013). Unsupervised texture classification is proposed in Bhulyan et al. (2007) and support vector machine (SVM) based classification using line operator features is described in Ricci and Perfetti (2006). Specialized snake algorithms and active contour models are proposed in (McInerney and Terzopoulos, 2000), (Espona et al., 2007) and (Al-Diri et al., 2009). The divergence of gradient vector fields is utilized for segmentation in (Lam and Yan, 2008). Some authors (Walter and Klein, 2001; Zana and Klein, 2006; Mendonça and Campilho, 2006) apply morphological operators utilizing that vessels are usually darker than their background.

Recently, the authors of (Marín et al., 2011) used gray-level features and grayscale invariant-moments with SVM classification. In (Xiao et al., 2013) and (Kovács and Hajdu, 2011) probabilistic Bayesian methods are extended by spatial constraints and an advanced statistical tracing

Manually annotated training database
↓
Preprocessing

Training

| Training of template parameters |
| Training of edge reconstruction parameters |
| Training of thin vessel addition parameters |

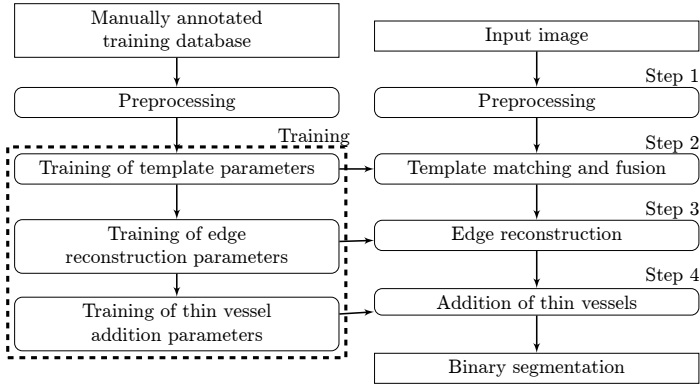| Input image |
|---|
| ↓ |
| Preprocessing | Step 1 |
| ↓ |
| Template matching and fusion | Step 2 |
| ↓ |
| Edge reconstruction | Step 3 |
| ↓ |
| Addition of thin vessels | Step 4 |
| ↓ |
| Binary segmentation |

Figure 2: Flowchart of the proposed method

technique is proposed in (De et al., 2014). The authors of (Yureidini et al., 2012) use RANSAC-fitting of ideal vessel templates, while (Wihandika and Suciati, 2013) integrates the extraction of the optic disc with the segmentation of the vasculature to increase efficiency. In (Lupascu et al., 2010), AdaBoost is used with several features, including Frangi-vesselness (Frangi et al., 1998), while the authors of (Fraz et al., 2012a) apply ensemble classification techniques for segmentation. A novel multi-scale approaches are described in (Moghimirad et al., 2012; Nguyen et al., 2013). In (Orlando and Blaschko, 2014), an SVM based method is described with connectivity constraints on the segmentation. The authors of (Lam et al., 2010) recently introduced a novel publicly available database containing high resolution fundus images (HRF), and proposed a multiconcavity based technique for the segmentation of vessels Surveys and overviews of retinal vessel segmentation algorithms can be found in (Niemeijer et al., 2004) and (Fraz et al., 2012b). A thorough overview of the current issues of retinal image processing (including the segmentation of vessels) can be found in the study (Trucco et al., 2013).

3. The proposed method

The proposed method consists of four consecutive steps. First, some preprocessing is performed including the extraction of the region of interest (ROI) covering the FOV and the extension of the image content over the boundaries of the ROI to prevent the segmentation method from extracting the ROI border as vessel. In the second step, template matching is applied to yield a highly precise binary segmentation. The third step reconstructs the contours of the binary segmentation to have similar characteristics as the contours of manual annotations in the training database. Finally, the fourth step extracts some further thin and low contrast vessels using the already computed results of previous steps. For a better insight, the flowchart of the proposed method is given in Figure 2, as well.

Before starting to discuss the steps of the algorithm in details, we introduce some basic notations. The red and green channels of an RGB retinal image are denoted by I_R and I_G with $I_R, I_G \in \{0, \ldots, 255\}^{R \times C}$, where R and C denotes the number of rows and columns of the images, respectively. Vectors are denoted by boldface typesetting, the coordinates of a pixel are represented by a two dimensional vector $\mathbf{p} = (i, j) \in \{0, \ldots, R-1\} \times \{0, \ldots, C-1\}$ and the intensity of image $I_G \in \{0, \ldots, 255\}^{R \times C}$ at coordinates $\mathbf{p} = (i, j)$ is denoted by $I_G[\mathbf{p}]$ or $I_G[i, j]$, assuming $0 \leq i < R$ and $0 \leq j < C$. Sets are denoted by calligraphic letters, like $\mathcal{R} \subset \{0, 1\}^{R \times C}$. Since the proposed method deals with many local and global quantities and parameters, for the sake of readability we use textual superscripts like t^{length} referring to a threshold on the length of vessels.

3.1. Step 1 – Preprocessing

Preprocessing has two main goals: on the one hand, the region of interest (ROI) covering the field of view is extracted to decrease computational demands and get rid of the unexpected intensity variations of the seemingly zero background; on the other hand, the image content covered by the ROI is extended over the ROI boundary to improve the local statistics of pixels close to the border of the ROI.

3.1.1. ROI extraction

Computational demands can be highly decreased if the operations of the proposed method are carried out only in those regions of the image where vessels can be expected. Therefore, in real applications, the first step of the processing of fundus images is usually the extraction of the ROI. Besides decreasing the computational demands, the proper segmentation of the ROI also affects the evaluation of vessel segmentation methods when no ground truth ROIs are provided with a database. Thus, the detailed description of the extraction of the ROI is also important to make the test results comparable with other techniques.

Due to its high energy content, the ROI extraction is carried out using the red channel of the images. We highlight that a simple global thresholding cannot give satisfactory results, since the illumination properties of images taken under various circumstances and the brightness within a single image can vary widely. We have developed an iterative method based on local thresholding and giving satisfactory results even when the images are not properly illuminated. First, a disk of radius r_0 pixels is fitted to the middle of I_R and an initial threshold t_R is determined as the mean of intensities covered by the disk. A global thresholding of I_R is carried out: the pixels having intensities larger than t_R are added to the ROI region $\mathcal{R}$. Then, the following step is repeated to all pixels of the image until $\mathcal{R}$ is unchanged. To a pixel $\mathbf{p} \notin \mathcal{R}$ a disk of radius r_1 pixels $\mathcal{D}_{r_1}(\mathbf{p})$ is fitted, and the set of pixels $\mathcal{N}(\mathbf{p}) = \mathcal{D}_{r_1}(\mathbf{p}) \cap \mathcal{R}$ covered by the disk and also belonging the ROI $\mathcal{R}$ is determined. The mean intensity of the red channel at the pixels of $\mathcal{N}(\mathbf{p})$ is calculated ($mean_{\mathcal{N}(\mathbf{p})}$) and pixel $\mathbf{p}$ is added to the ROI $\mathcal{R}$ if the intensity of $\mathbf{p}$ is larger than the local threshold $t(\mathbf{p}) = t_0 \cdot mean_{\mathcal{N}(\mathbf{p})}$. After

the iteration, morphological hole filling is applied to the $\mathcal{R}$ to yield the final ROI covering the FOV.

The radii and the threshold used in the ROI extraction method can be varied widely without significantly changing the results. We have used the empirically selected settings $r_0 = 250$ pixels, $r_1 = 13$ pixels and $t_0 = 0.75$ for images with various illumination properties and resolutions (from 0.5 megapixels to 7 megapixels) successfully. A subimage from the red channel of a fundus image and the contour of the extracted ROI are visualized in Figure 3(a). The extracted ROI follows the boundary of the FOV even though the image has several unexpected deficiencies: the middle of the image is extremely bright; the FOV is embedded into a rectangle of varying non-zero intensities; and a dark strip can be observed at the bottom of the image.

3.1.2. Extension of the image content over the ROI border

Since the green channel is the most detailed one regarding the vascular system and vessel related structures (Marín et al., 2011), the rest of the proposed method operates on that channel only. Either filtering or template matching is applied, the environments of pixels near the ROI boundary have improper characteristics: whenever a filter or template exceeds the ROI boundary, the response may be misleading due to the acquisition errors near the boundary (see Figure 3(b)) and the dark background of the FOV. In order to reduce the effects of the ROI border, the acquisition artefacts should be removed, and the image content covered by the ROI should be extended over the boundary of the ROI, preferably preserving the structure of vessels. The method we have developed for the extension of the image content is described below.

Examining a large set of fundus images, we have found that only the region of width 5 pixels along the ROI border is affected by acquisition artefacts, independently from the source or resolution of the images. Thus, the ROI $\mathcal{R}$ extracted by the method described in the previous section is eroded by a disk shaped structuring element of radius 5 pixels and the resulting set of pixels $\mathcal{R}^{erod}$ is considered to contain pixels without artefacts. Initially, let $I_G^{ext} = I_G$ and we introduce the temporary set of pixels $\mathcal{R}^{tmp} = \mathcal{R}^{erod}$. Suppose that the image content is to be extended outside the ROI by L pixels in each direction. For $l = 0, \ldots, L + 5$ the following two steps are repeated:

1. For each outer contour pixel of $\mathcal{R}^{tmp}$ the following operations are carried out:

   (a) Let $\mathbf{p}$ denote an outer contour pixel of $\mathcal{R}^{tmp}$. First, we would like to decide whether $\mathbf{p}$ is an ordinary pixel containing useful image information or $\mathbf{p}$ is deteriorating the boundary of the ROI. This decision is made by taking into consideration the intensity $I_G^{ext}[\mathbf{p}]$ and the statistics of the closest pixels covered by $\mathcal{R}^{erod}$. Particularly, the set of pixels $\mathcal{N}(\mathbf{p}) = D_{l+5}(\mathbf{p}) \cap R^{erod}$ being covered by $\mathcal{R}^{erod}$ and being closer to $\mathbf{p}$ than $l + 5$ pixels is constructed and the mean $mean_{\mathcal{N}(\mathbf{p})}$ and standard

deviation $std_{\mathcal{N}(\mathbf{p})}$ of the intensities at the pixels of $\mathcal{N}(\mathbf{p})$ are determined.

(b) The intensity of pixel $\mathbf{p}$ is considered to belong to the ordinary image content if

$$mean_{\mathcal{N}(\mathbf{p})} - std_{\mathcal{N}(\mathbf{p})} \leq I_G^{ext}[\mathbf{p}] \leq mean_{\mathcal{N}(\mathbf{p})} + std_{\mathcal{N}(\mathbf{p})} \tag{1}$$

holds, that is, the intensity $I_G^{ext}[\mathbf{p}]$ is not an outlier regarding the statistics of the closest pixels covered by $\mathcal{R}^{erod}$. If (1) holds, the intensity at pixel $\mathbf{p}$ is not updated. Otherwise, the intensity $I_G^{ext}[\mathbf{p}]$ is updated by the following step.

(c) First, the characteristic orientation of vessels close to pixel $\mathbf{p}$ is estimated. The local gradients at the pixels of $N(\mathbf{p})$ are computed by the Sobel-operator; the gradients having the largest 20% of magnitudes are determined and the weighted mean of corresponding directions $\alpha(\mathbf{p})$ is computed, where the weights are the magnitudes of the gradients. The angle $\alpha(\mathbf{p})$ can be considered as the normal of vessels near pixel $\mathbf{p}$. In order to elongate the vessels intersecting the border of the ROI, the intensity at the outer contour pixel $\mathbf{p}$ must be updated by the intensity of the closest ordinary pixel at any of the directions $\alpha(\mathbf{p}) \pm \pi/2$. Formally,

$$I_G^{ext}[\mathbf{p}] = \begin{cases} I_G[\mathbf{p} + \mathbf{z}], & \text{if } \mathbf{p} + \mathbf{z} \in \mathcal{R}^{erod}, \\ I_G[\mathbf{p} - \mathbf{z}], & \text{if } \mathbf{p} - \mathbf{z} \in \mathcal{R}^{erod} \\ mean_{\mathcal{N}(\mathbf{p})}, & \text{otherwise}, \end{cases} \tag{2}$$

where $\mathbf{z}$ denotes an offset vector of length $(l + 1)$ pixels and direction $\alpha(\mathbf{p}) + \pi/2$. The first two cases are separated, since although the normal of the vessel is estimated, it is undetermined if $\alpha(\mathbf{p}) + \pi/2$ or $\alpha(\mathbf{p}) - \pi/2$ (both directions tangential with the neighboring vessel structures) points towards the useful image content. The third case is related to the situations when no vessel structures can be found in the neighborhood $\mathcal{N}(\mathbf{p})$. Without elongated structures the estimated normal may be tangential to $\mathcal{R}^{erod}$, leading a situation, where neither $\mathbf{p} + \mathbf{z}$, nor $\mathbf{p} - \mathbf{z}$ points to the useful image content. In this case the intensity of pixel $\mathbf{p}$ is updated by the mean of the intensities of the closest pixels containing useful image information.

2. The outer contour pixels of $\mathcal{R}^{tmp}$ are added to $\mathcal{R}^{tmp}$. After the procedure has finished, I_G^{ext} contains the green channel of the fundus image extended by L layers of pixels.

To illustrate the ROI extension technique, we have visualized a subimage of the border of a fundus image before and after the extension process in Figures 3(b)-3(d). As one can see in the figures, the extension is smooth, keeps the orientation of vessels and removes the tight stripe of bright pixels (imaging artefacts) from the border.

3.2. Step 2 – Template matching and fusion

The main goal of Step 2 is to yield a highly precise segmentation of vessel pixels. The vessel contours are going
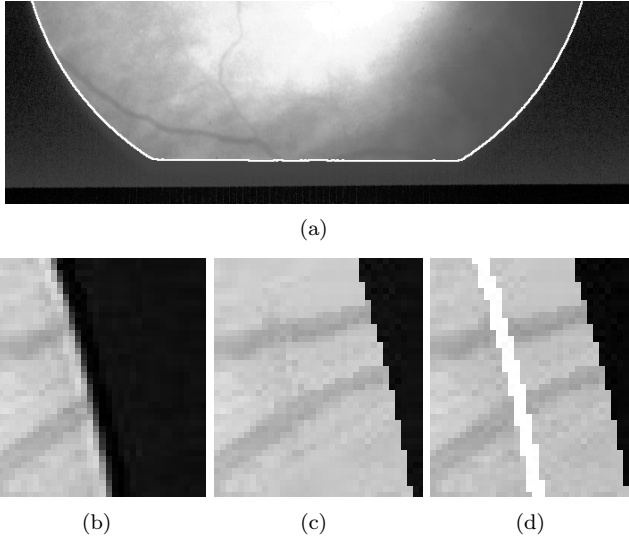
Figure 3: Illustration of ROI detection and extension. A histogram equalized subimage of the red channel of a fundus image and the contour of the extracted ROI (a); acquisition defects near the ROI boundary (b); subimage after the extension of the ROI (c); subimage after the extension of the ROI, the contour of the ROI is visualized by white pixels (d)

to be reconstructed from these pixels in Step 3. During the development of the method we have taken into account the following considerations:

1. Vessels have highly varying profiles: the intensity of vessels changes with the amount of oxygen transported by the blood flowing in them (veins are darker than arteries); images are usually taken under varying illumination conditions and imaging devices have varying optical transfer functions affecting the level of noise and contrast. Even though vessels are relatively simple structures in retinal images (dark stripes on a bright background), these issues lead to many possible shapes of vessel profiles.

2. Vessels vary smoothly: shifting a vessel segment a little along the vasculature it is still highly similar to the underlying image content. Having a template representing one specific type of vessel segments accurately, islands of vessel pixels can be detected by extracting the pixels where the similarity of the template and the image content is sufficiently high.

According to the second observation, a template matching framework may be suitable to identify vessel pixels accurately. However, when the template matching framework is developed, we have to decide how to handle the issues related to the variability of vessel profiles. There are basically three choices:

1. Handling at the level of the image – many authors use some kind of preprocessing step to normalize the intensity variations in retinal images. Global and local intensity normalization techniques (like histogram equalization, background subtraction, etc. (Youssif et al., 2006)) can reduce the intensity variations caused by the varying illumination conditions or large bright lesions but one can

hardly expect that these simple techniques could transform each vessel to have a specific, for example sinusoidal profile (which is implicitly assumed in many techniques using Gabor filters to model vessels).

2. Handling at the level of similarity functions – there are many advanced similarity functions being invariant against even non-linear intensity transforms (like the widely used mutual information (Li, 1990) or the photometric invariant distance function (Hel-Or et al., 2014)). When these similarity functions are used, there is no need for global or local normalization, since one can suspect that the similarity function focuses on the structural similarity of the template and the image content independently from slight variations of intensity profiles. These techniques usually transform the template to have a coarse intensity resolution (e.g. 3-4 different intensity levels), and characterize the variance of the image content covered by those pixels of the modified template that belong to the same intensity level. The drawback of these techniques is that due to the coarse intensity representation the fine details are essentially lost.

3. Handling at the level of templates – the third way is to handle the variability of vessel profiles at the level of templates, namely, one can use a large set of templates covering almost all possible vessel shapes and a relatively simple similarity function to measure the similarity of the image content and the individual templates. If any of the individual similarity scores is high enough, the image content can be considered to contain a vessel. The drawback of this technique is that dealing with a relatively large set of templates increases the computing demands of the template matching.

We follow the third way, particularly, the proposed methods works with a large number of templates representing different kinds of vessel segments, expecting that better segmentations can be achieved by tolerable computational demands than in the cases when the variability of vessel profiles is handled by intensity normalization or advanced similarity functions.

3.2.1. Weak segmentation operators

Along the considerations regarding the template matching based segmentation of vessels, we introduce *weak segmentation operators*, each of them corresponding to a particular kind of vessel template. The term "weak" refers to the expectation that one operator does not extract the entire vasculature, rather, one operator extracts only patches of vessels being similar to the template corresponding to the operator. A large set of these weak segmentation operators (each of them corresponding to a particular vessel structure) is applied to the retinal image, and the resulting patches are fused to get a highly precise segmentation of vessels.

In order to realize the weak segmentation operators, we need to find an appropriate similarity function; some preferably analytic, parameterizable vessel segment templates; and a way to binarize the outcome of template

matching which is expected to be an image containing the similarity scores of a particular template and the image content.

The similarity function. Since the possible vessel structures are assumed to be represented by a large set of templates, the similarity function does not have to handle the possible non-linear deviations of one template and the real vessels. Accordingly, there are only a few requirements against the desired similarity function. The function should be an absolute measure to make the binarization of the results easier; and the function should be invariant against linear intensity transforms to make the template matching independent from the brightness of the background and the variance of the image content covered by the template. Probably the simplest and fastest similarity function satisfying these conditions is the centralized correlation coefficient, also known as Pearson correlation coefficient (PCC): let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^N$ denote a vessel template and a window of the image, respectively. Then, the PCC coefficient becomes

$$PCC(\mathbf{v}, \mathbf{w}) = \frac{\langle \mathbf{v} - \overline{\mathbf{v}}, \mathbf{w} - \overline{\mathbf{w}} \rangle}{\sqrt{\langle \mathbf{v} - \overline{\mathbf{v}}, \mathbf{v} - \overline{\mathbf{v}} \rangle \langle \mathbf{w} - \overline{\mathbf{w}}, \mathbf{w} - \overline{\mathbf{w}} \rangle}}, \quad (3)$$

where $\overline{\mathbf{v}}$ and $\overline{\mathbf{w}}$ denote the mean of the coordinates of vectors $\mathbf{v}$ and $\mathbf{w}$, respectively.

Generalized Gabor function based templates. The use of convolutional Gabor filters is a commonly accepted way to emphasize linear structures in digital images. In its general formulation a Gabor filter is composed as the product of a complex wave and a 2D Gaussian. For linear structures being darker than their background:

$$G(\sigma, \theta, \lambda, \gamma | r, c) = \exp\left(-\frac{\hat{r}^2 + \gamma^2 \hat{c}^2}{2\sigma^2}\right) \cos\left(2\pi \frac{\hat{r}}{\lambda} + \pi\right), \quad (4)$$

$$\hat{r} = r \cos\theta + c \sin\theta, \quad (5)$$

$$\hat{c} = -r \sin\theta + c \cos\theta, \quad (6)$$

where σ denotes the standard deviation of the Gaussian controlling the spatial decay of the filter, θ and λ denote the orientation and wavelength of the linear structure emphasized by the filter, and γ is the aspect ratio of the filter. In practice, the spatial extents of the filter along the local coordinate axis corresponding to coordinate r should be limited to only one valley of the cosine wave. Accordingly, the filter is considered to contain a discrete point $(r, \hat{c})$ if $(\hat{r}, \hat{c}) \in [-\sigma_r, \ldots, \sigma_r] \times [-\sigma_c, \ldots, \sigma_c]$ holds, where $\sigma_r = \sigma$, $\sigma_c = \sigma/\gamma$, and we suppose that for the given wavelength parameter λ the spatial decay σ is selected to let approximately one valley appear in the filter. This choice of spatial extents results filters having approximately sinusoidal profiles with some exponential decay near the boundaries of the filter making them smoothly fade into zero.

Since linear structures can have arbitrary orientations, Gabor filters are usually used in matched filtering frameworks. Particularly, N distinct filters are generated by

fixing the parameters σ, λ, γ, and varying θ in the range $[0, \pi]$. Each of the convolutional filters is applied to a given pixel (i, j) of the input image and the response of matched filtering is defined as the maximum of the responses of the individual filters at pixel (i, j). Clearly, the discretized Gabor filter can be considered as a template, and the matched filtering approach can be turned into an orientation independent template matching technique if the convolution operation is replaced by a similarity function. However, a given Gabor function based template can represent only one specific type of vessel segments accurately: having approximately sinusoidal profiles and fading into the background in both directions $\pm\theta$. In order to enable the representation of a wider range of vessel structures, some further shape parameters and the generalized Gabor function H is introduced. Particularly, in the rest of the paper, we use the templates based on the function

$$H(\sigma, \theta, \lambda, \gamma, q, \beta | r, c) = \exp\left(-\frac{\hat{r}^2 + \beta\gamma^2 \hat{c}^2}{2\sigma^2}\right) A(r), \quad (7)$$

$$A(r) = sgn\left(\cos\left(2\pi \frac{\hat{r}}{\lambda} + \pi\right)\right) \left|\cos\left(2\pi \frac{\hat{r}}{\lambda} + \psi\right)\right|^q, \quad (8)$$

where $\beta \in \{0, 1\}$ and $q \in \mathbb{R}^+$ holds. First, the role of the binary parameter β used to disable the term $\gamma^2 \hat{c}^2$ in the exponent is discussed. Simple Gabor function based templates have an exponential decay in every spatial direction, including $\pm\theta$ and $\theta \pm \pi/2$, so they can represent only those parts of the vascular system properly, where the intensities are increasing along the vessels in both tangential directions $\pm\theta$. By setting $\beta = 0$ and disabling the term $\gamma^2 \hat{c}^2$, the exponential decay in the directions $\pm\theta$ is removed, thus, the template can represent those parts of the vascular system, where the profiles of vessels do not change along the vasculature significantly. We highlight that when the exponential decay in directions $\pm\theta$ is disabled by setting $\beta = 0$, the parameter γ still affects the template by controlling its spatial extent as described before. The role of the exponential parameter q is to enable non-linear deviations from the sinusoidal profile. Clearly, the templates based on H can represent a wider variety of vessel segments than templates based on G, since G is a special case of H, by choosing $q = 1$ and $\beta = 1$. Some templates corresponding to specific parameter settings are visualized in Figure 4.

In order to qualitatively validate the choice of PCC as a similarity function, we have performed template matching by a specific set of orientation invariant Gabor templates using inner product (IP), simple correlation coefficient (CC) and PCC as similarity function and compared the results in Figure 5. As one can see in the subfigures, both IP and CC are affected by the brightness of the image content, therefore, the border of the optic disc gives similar responses as relatively thin vessels. On the contrary, PCC gives low responses at the border of the optic disc, although the noise of the background is definitely larger. Nevertheless, the PCC responses at vessel pixels are still

higher than the responses on the background, thus, we can expect that even a simple binarization technique can differentiate vessels from the background when PCC is used.

Binarization of the outcome of template matching. For a given combination of parameters σ, λ, γ, q, β one can generate an orientation invariant set of generalized Gabor function based templates, carry out template matching with PCC as a similarity function and get an intensity image, where high scores are related to vessel structures. In order to extract patches of vessels being similar to the vessel templates represented by the parameters, one has to binarize the outcome of template matching. A similar problem appears in the well known Canny edge detector, where the image containing edge magnitudes has to be binarized by following the ridges in the image. Thus, the region growing method (also known as hysteresis thresholding) used in the Canny edge detector seems to be an appropriate method for binarization. First, the image containing similarity scores is thresholded at level t_h and the foreground pixels are considered to be the seed points of an iterative region growing method. Then, pixels having intensities larger than t_l and being connected to seed points are added to the set of seed points recursively. When no further pixels can be added to the set of seed points, the operation stops and the final set of seed points is considered as the foreground of binarization. With properly selected t_h and t_l thresholds one can extract vessel pixels having local environments similar to the template described by the parameters σ, λ, γ, q, β and function H. In Figure 6, we have visualized the seed points and the result of hysteresis thresholding applied to the outcome of template matching. As one can observe, some settings of t_h and t_l lead to highly precise segmentations in the sense that only real vessel pixels are segmented as foreground (as desired), while other combinations lead to noisy, practically useless results.

The introduction of weak segmentation operators. As a composition of template matching by generalized Gabor function based templates and hysteresis thresholding, we introduce weak segmentation operators $W(\sigma, \lambda, \gamma, q, \beta, t_h, t_l)$ by the following operation: W applies orientation independent template matching using the templates specified by $H(\sigma, i\pi/N, \lambda, \gamma, q, \beta)$, $i = 0, \ldots, N$, PCC as similarity function, and performs hysteresis thresholding on the outcome of template matching with thresholds t_h and t_l. As a technical detail we mention that the choice $N = 30$ was found to give satisfactory results, thus, the angular resolution 6° is used throughout the paper.

3.2.2. Training of weak segmentation operators

The parameters of the weak segmentation operators can be divided into two groups. Parameters λ, γ, q and β are responsible for the intensity distribution of the template, thus, these parameters determine the shape of the vessel structure recognized by the operator. Later on, we
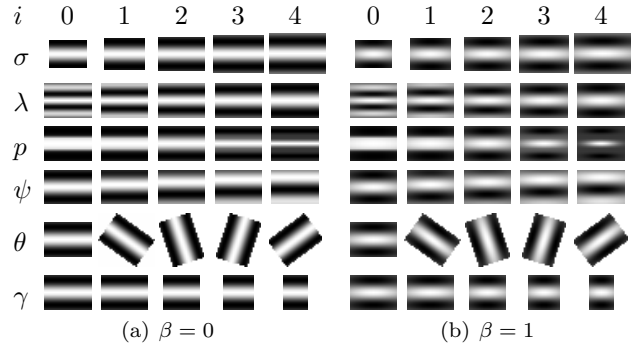

Figure 4: Templates based on various parametrizations of the function $H(\sigma, \theta, \lambda, \gamma, q, \beta)$. In both cases, the settings $\sigma = 10$, $\lambda = 16$, $\theta = 0$, $\gamma = 0.7$, $q = 1$ are varied by changing one of the parameters: $\sigma_i = 8 + i$, $\lambda_i = 8 + 2i$, $p_i = 1^{-i-2}$, $\psi_i = i\frac{\pi}{5}$, $\theta_i = i\frac{\pi}{5}$, $\gamma = 0.7 + 0.2i$

intend to use a large set of weak segmentation operators by specifying the shape descriptor parameters explicitly to make the templates corresponding to the operators cover the possible vessel structures (by making the parameters cover the reasonable range of the parameter space uniformly). Contrarily, the parameters σ, t_h, t_l are related to the nature of the problem: the level of noise in the images, the contrast of the vessel borders, etc. For a given settings of shape descriptors λ, γ, q, β the proper values of σ, t_h, t_l have to be trained.

Training of one weak segmentation operator. In order to simplify the notations, we introduce two multidimensional parameters for the weak segmentation operators, namely $\mathbf{s} = (\lambda, \gamma, q, \beta)$ is related to the shape of the template, $\mathbf{n} = (\sigma, t_0, t_1)$ is related to the nature of the problem and the notation of a weak segmentation operator is simplified to $W(\mathbf{s}, \mathbf{n})$.

Given an explicit setting of the shape descriptor parameter $\mathbf{s}$, the goal of training is to find the parameter $\hat{\mathbf{n}}$ leading to a highly precise segmentation of vessel pixels. The training goes on by following a simulated annealing (Kirkpatrick et al., 1983) based stochastic optimization process respecting the maximization of a target function measuring how precise the segmentation is. In the rest of the paragraph we give the outline of training by simulated annealing. Let $\mathbf{n}^{(k)}$ denote the parameter combination at the kth iteration of the optimization process and $T(\mathbf{n}^{(k)})$ the corresponding value of the target function. For a random combination $\mathbf{n}^{(k+1)}$ the operator $W(\mathbf{s}, \mathbf{n}^{(k+1)})$ is applied to the images of the training database, the segmentation results are evaluated and the target function $T(\mathbf{n}^{(k+1)})$ is determined. The combination $\mathbf{n}^{(k+1)}$ is accepted if the target function has increased ($T(\mathbf{n}^{(k+1)}) > T(\mathbf{n}^{(k)})$), or $\exp\left(T(\mathbf{n}^{(k+1)}) - T(\mathbf{n}^{(k)})/Z\right) > r$ holds, where Z denotes the temperature of the system being decreased as the iterations go on and r denotes a uniformly distributed random number in the range $[0, 1]$. Otherwise, the update of the parameter is reverted: $\mathbf{n}^{(k+1)} = \mathbf{n}^{(k)}$. It is worth noting
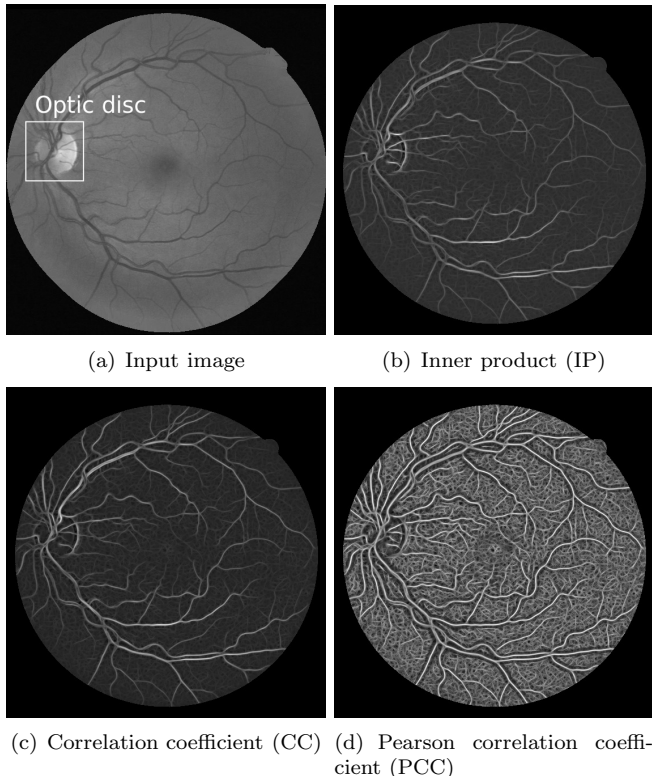
(a) Input image   (b) Inner product (IP)

(c) Correlation coefficient (CC)   (d) Pearson correlation coefficient (PCC)

Figure 5: The extended green channel image I_G and the result of template matching with various similarity functions for the parameter setting $\sigma = 3$, $\lambda = 8$, $\psi = \pi$, $\gamma = 0.5$, $\theta \in \{0, \pi/30, \dots, \pi - \pi/30\}$
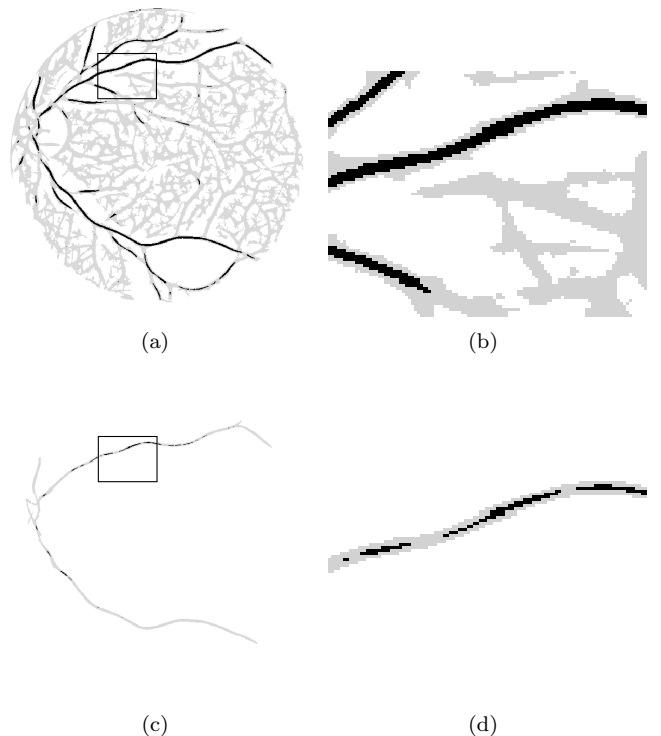

(a)   (b)

(c)   (d)

Figure 6: Seed points (black) and foreground pixels (gray) after segmentation by weak segmentation operators $W(4, 12, 0.5, 1, 0, 1, 0.7, 0.6)$ (a)-(b) and $W(4, 12, 0.5, 1, 0, 1, 0.5, 0.4)$ (c)-(d)

that any other randomized optimization method could be used, as well.

We emphasize again that we do not expect one weak segmentation operator to segment all vessel pixels, rather, we expect one operator to find vessel pixels surely belonging to the foreground. Accordingly, the target function we have chosen to navigate towards the optimal parameter $\hat{n}$ for a given s is based on the *precision* (referring to the performance measure in binary classification), also known as positive predictive value (PPV) of the segmentation: let TP denote the number of correctly classified vessel pixels in the database and AP the overall number of pixels classified as vessel; then, the positive predictive value is defined as $PPV = TP/AP$. Higher PPV means more real vessel pixels and less background pixels classified as foreground. Since segmentation always has errors, the desired highest PPV values are usually reached when TP is small and $TP \simeq AP$. In order to extract more pixels, the conditions are slightly weakened: small errors are allowed by introducing ϵ-*insensitive* PPV as the target function of optimization. Particularly,

$$T(\mathbf{n}) = \begin{cases} PPV, & \text{if } PPV < 1 - \epsilon, \\ TP, & \text{otherwise.} \end{cases} \quad (9)$$

Maximizing the target function means to find the parameter setting $\hat{\mathbf{n}}$ that provides $PPV > 1 - \epsilon$ and TP is as large as possible. Obviously, the target function leads to misclassified foreground pixels, however, we expect that these errors do not interfere when several weak segmentation operators with various parameters are used, and can be removed by simple image processing techniques in a later step. If the condition $PPV > 1 - \epsilon$ cannot be satisfied for a given shape descriptor parameter $\mathbf{s}$, the operator with parameters $\mathbf{s}$ is discarded. This can be interpreted as the setting $\mathbf{s} = (\lambda, \gamma, q, \beta)$ describes templates that cannot be distinguished from the background and noise precisely enough.

The method described in this paragraph can be used to train the parameter $\mathbf{n}$ of one specific weak segmentation operator $W(\mathbf{s}, \mathbf{n})$ for a given shape descriptor $\mathbf{s}$. In order to gain a highly precise segmentation of the entire vasculature, we have to train many weak segmentation operators with various settings of $\mathbf{s}$ related to the various types of vessel structures and unify the results of the individual operators.

We introduce the notation $\mathcal{S} = \Lambda \times \Gamma \times \mathcal{Q} \times \mathcal{B}$ for the set of reasonable parameter combinations of shape descriptors and $\mathcal{W}(\mathcal{S})$ to denote the set of weak segmentation operators $W(\mathbf{s}, \mathbf{n})$, where $\mathbf{s} \in \mathcal{S}$ holds and $\mathbf{n}$ is trained by the method described before. For example, using the parameters $\Lambda = \{6, 7, \dots, 24\}$, $\Gamma = \{0.4, 0.5, \dots, 1.2\}$, $\mathcal{Q} = \{0.166, 0.2, 0.25, 0.33, 0.5, 1, 2, 3, 4, 5, 6\}$, $\mathcal{B} = \{0, 1\}$, and

8

$\epsilon = 0.05$, the cardinality of the set $\mathcal{S}$ becomes 3762. After the training of weak segmentation operators on the 20 training images of DRIVE, the cardinality of $\mathcal{W}(\mathcal{S})$ becomes 3340, showing that there were approximately 400 combinations of shape descriptors that cannot represent vessel structures with sufficiently high PPV scores. The smallest wavelength $\lambda = 6$ pixels appearing in Λ is selected by the following consideration: if the wavelength of a sinusoidal wave is 6 pixels, the width of one valley becomes 3 pixels, that is, the cross section of the template contains one dark pixel in the middle having one bright pixel on both of its sides. Clearly, a template having a profile like this represents vessels of width 1 pixel. As a general rule, we can state that independently from the resolution of the images, the use of weak segmentation operators having wavelengths smaller than 6 pixels is worthless. Similar considerations were used to select the maximum wavelength $\lambda = 24$ pixels, taking into account the width of the thickest vessels appearing in the DRIVE database. The smallest parameter in the set Γ corresponds to vessel segment templates having more than two times larger length than width. Considering the nature of vessels, it is unlikely that much longer straight vessel segments would appear. The largest parameter $\gamma = 1.2$ in the set Γ defines vessel segment templates having slightly larger width than length. If the γ is further increased, the templates do not represent line segments any more. Finally, due to the non-linear effect of the exponential parameter q, and the fast convergence of x^q, $x \in [0, 1]$ as $q \to 0$ or $q \to \infty$, the settings $q \in [0, 0.166[$ or $q > 6$ lead to highly similar templates as the ones $q = 0.166$ or $q = 6$, respectively.

Finding the optimal set of weak segmentation operators. Depending on the cardinality of set $\mathcal{S}$, $W(\mathcal{S})$ can contain thousands of weak segmentation operators, making the proposed method computationally intractable in practice. In order to reduce the computational demands and achieve better segmentation results, the raw set $\mathcal{W}(\mathcal{S})$ of weak segmentation operators is reduced. Computational demands can be decreased by removing the subset $\mathcal{O} \subset \mathcal{W}(\mathcal{S})$ if the pixels classified by the operators in $\mathcal{O}$ as vessel are classified by the operators in $\mathcal{W}(\mathcal{S})/\mathcal{O}$ in the same way. Better segmentation can be achieved by removing the subset $\mathcal{O} \subset \mathcal{W}(\mathcal{S})$ that produces interfering true positive but less interfering false positive matches with the operators in $\mathcal{W}(\mathcal{S})/\mathcal{O}$.

Examining all subsets of $\mathcal{W}(S)$ is computationally infeasible. Suboptimal but acceptable results can be achieved by applying the commonly used backward feature subset selection (BFSS) method to the set $\mathcal{W}(\mathcal{S})$ with a properly chosen target function. Namely, let $\mathcal{W}' = \mathcal{W}(S)$ initially. Then, we repeat the following step while the target function increases: for each $W \in \mathcal{W}'$ the images of the training database are segmented by the weak segmentation operators in the set $\mathcal{W}'/\{W\}$, the binary outcomes are unified by applying pixel-wise "or" operation and the target function is computed; if the highest value of the target function
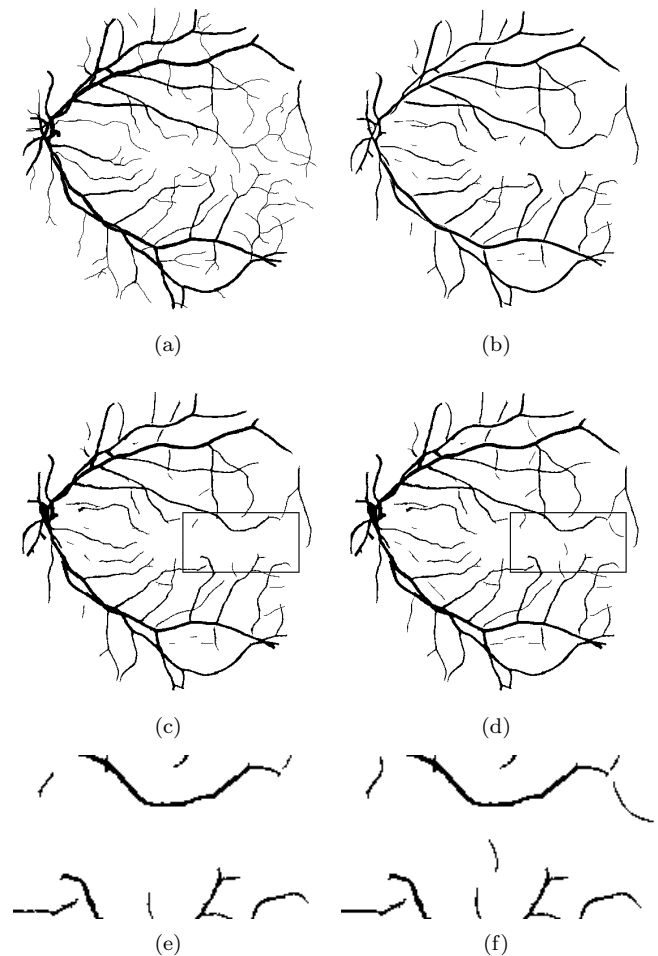

(a) (b)

(c) (d)

(e) (f)

Figure 7: Segmentation results corresponding to specific steps of the proposed method. Manual segmentation (ground truth) (a); Step 2: fused segmentation of the operators $\mathcal{W}^{reduced}(\mathcal{S})$ (b); Step 3: edge reconstruction (c); Step 4: addition of thin vessels (d); the rectangular region enlarged before and after the addition of thin vessels (e)-(f)

is reached at $\mathcal{W}'/\{W^*\}$, let $\mathcal{W}' = \mathcal{W}'/\{W^*\}$. The target function we have chosen for the BFSS method is the accuracy of the segmentation results: the ratio of correctly classified pixels and all pixels of the training database. After the iteration has finished, the set $\mathcal{W}'$ contains the reduced set of weak segmentation operators, denoted by $\mathcal{W}^{red}(\mathcal{S})$ for better readability.

For illustration, we have applied the proposed BFSS technique to the set $\mathcal{W}(\mathcal{S})$ of 3340 operators introduced in the previous paragraph. In the resulting set $\mathcal{W}^{red}(\mathcal{S})$ the number of operators decreased to 384 having tractable computation demands.

3.2.3. Fusion of the binary outcomes of weak segmentation operators

So far, we have trained and selected the computationally tractable $\mathcal{W}^{red}(\mathcal{S})$ set of weak segmentation operators. After the preprocessing of an unseen input image, the op-

erators in $\mathcal{W}^{red}(\mathcal{S})$ are applied to the input image, yielding $|\mathcal{W}^{red}(S)|$ binary images, each of them corresponding to a specific kind of vessel structure. In order to yield one binary image as the outcome of Step 2, the results of the individual operators are fused by applying a pixel-wise "or" operation. However, since the outcomes of the operators may contain some foreground pixels related to noise, the fusion is preceded by some filtering based on the shape of the connected foreground components. Particularly, the connected foreground components in the outcomes of the individual operators are extracted (the objects in Figure 6(c) show examples for these components), and a component is added to the fused result by pixel-wise "or" operation if it is sufficiently elongated. The elongatedness can be measured by the circularity of the component defined by $4\pi A/P^2$, where A and P denote the area and the perimeter of the object. Obviously, the more elongated a component is, the smaller its circularity becomes, thus, an object is added to the outcome of Step 2 if its circularity is smaller than a specific threshold t^{circ}. We have found empirically that the choice $t^{circ} = 0.3$ gives satisfactory results, independently from the source of images being segmented. We also note that small variations of this threshold do not change the results at all.

To illustrate the fusion of the individual binary segmentations and enable the comparison with the ground truth annotation, we have used the set $\mathcal{W}^{red}(\mathcal{S})$ containing 384 operators (being introduced in the previous paragraph) to apply Step 2 of the proposed method to a test image and visualized the binary outcome and the ground truth annotation in figures 7(b) and 7(a), respectively.

3.3. Step 3 – Reconstruction of vessel contours

So far we have introduced some simple weak segmentation operators and a fusion technique to yield highly precise segmentations in the sense that the majority of foreground pixels in the fused outcomes of Step 2 are real vessel pixels. Since the individual weak segmentation operators extract mainly the center lines of vessels, we cannot accept the fused result as final segmentation: many false negative classifications appear along the borders of vessels. In the current step we focus on the reconstruction of vessel contours to gain a highly accurate segmentation of the entire vasculature.

Considering the intensity profile of a vessel, relatively small intensities appear in the middle and increase towards the vessel walls. Ideally, the contour pixels of a segmentation should have the same relative intensity respecting the minimum intensity of the vessel profile and the mean intensity of the background. Since the sensitivity of the visual perception of human observers varies with the wavelength of the stimulus (Deriington and Lennie, 1984) and the images have finite resolution, we assume that this is not the case in practice. We expect that the position of true vessel contours varies with the width and contrast of vessels.

In order to validate these expectations, we introduce three measures to characterize the contour pixels of seg-
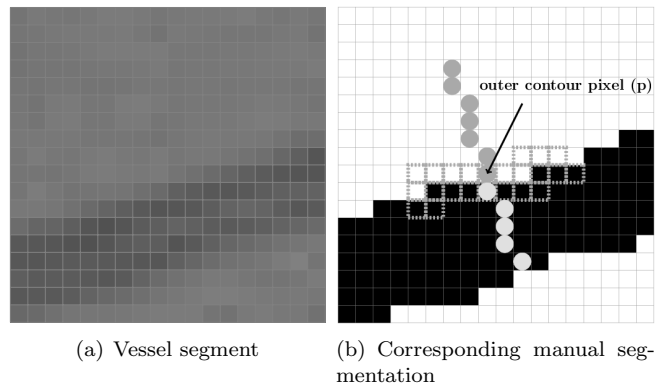

(a) Vessel segment

(b) Corresponding manual segmentation

Figure 8: Measuring the width of a vessel

mented vessels: the width (w) of the segmented vessel passing by, the dynamic range (d) of the vessels and the relative intensity (r) at the contour respecting the minimum intensity of the vessel profile and the mean intensity of the background.

3.3.1. Measuring the width, dynamic range and relative intensity at contour points

Worth to notice that there are some recent studies focusing on measuring the width of vessels in retinal images (Lupascu et al., 2013). These methods usually fit a parametric surface model to the vessel and estimate its width by the analytic parameters of the fitted model. Contrarily, what we need is to estimate the width of *already segmented* vessels at their inner and outer contour points. The method described below gives an extremely fast solution with satisfactory accuracy.

Let $\mathbf{p}$ denote an inner or outer contour pixel of a segmented vessel (for illustration we have visualized a real vessel segment and the corresponding manual annotation in figures 8(a) and 8(b), respectively). In order to find the width of the vessel being tangential to this pixel, we have to get a robust estimation of its normal vector. Sobel operators are applied to the contour pixels of the binary image in a close neighborhood and the mean of the gradient directions is computed. We have found empirically that a disk shaped neighborhood with the radius of 5 pixels gives satisfactory results. In Figure 8(b) pixels with dotted border indicate the region, where the gradient directions are averaged. Let $\alpha(\mathbf{p})$ denote the mean of the gradient directions. A line segment is fitted to $\mathbf{p}$ at angle $\alpha(\mathbf{p})$ (gray dots in Figure 8(b)), and the pixels where the line segment intersects the inner and outer contours of the other side of the vessel are located. Then, if $\mathbf{p}$ is an inner contour pixel, the width of the vessel tangential to $\mathbf{p}$ is estimated as the Euclidean distance of $\mathbf{p}$ and the outer contour pixel intersected by the line segment at the other side of the vessel. Similarly, if $\mathbf{p}$ is an outer contour pixel, the width of the vessel is estimated as the Euclidean distance of pixel $\mathbf{p}$ and the inner contour pixel intersected

by the line segment at the other side of the vessel. In order to make the estimations smooth along the vasculature, the width $w(\mathbf{p})$ of the vessel tangential to $\mathbf{p}$ is defined as the mean of the estimated widths in the close neighborhood.

Measuring the dynamic range of a vessel tangential to $\mathbf{p}$ is carried out similarly to the measurement of width: the line segment normal to the vessel is divided in two regions: the intersection of the line segment and the vessel ($\mathcal{L}_V(\mathbf{p})$ - denoted by light gray dots in Figure 8(b)), and the intersection of the line segment and the background ($\mathcal{L}_B(\mathbf{p})$ - denoted by dark gray dots in Figure 8(b)). The dynamic range is defined as the difference of the minimum intensity $min_{\mathcal{L}_V(\mathbf{p})}$ covered by region $\mathcal{L}_V(\mathbf{p})$ and the mean intensity $mean_{\mathcal{L}_B(\mathbf{p})}$ covered by region $\mathcal{L}_B(\mathbf{p})$:

$$d(\mathbf{p}) = \max\left\{mean_{\mathcal{L}_B(\mathbf{p})} - min_{\mathcal{L}_V(\mathbf{p})}, 0\right\} \qquad (10)$$

Finally, the relative intensity of the contour pixel respecting the mean intensity of the background and the minimum intensity of the vessel profile is defined as:

$$r(\mathbf{p}) = \max\left\{\min\left\{\frac{I_G[\mathbf{p}] - min_{\mathcal{L}_V(\mathbf{p})}}{mean_{\mathcal{L}_B(\mathbf{p})} - min_{\mathcal{L}_V(\mathbf{p})}}, 1\right\}, 0\right\}.$$
$$(11)$$

The value $r(\mathbf{p}) \in [0,1]$ represents the normalized relative intensity of the contour pixel $\mathbf{p}$ within the range $[min_{\mathcal{L}_V(\mathbf{p})}, mean_{\mathcal{L}_B}(\mathbf{p})]$, thus, $r(\mathbf{p})$ characterizes the position of the contour on the normal of the vessel taking into account only the intensity relations.

3.3.2. Characterization of relative intensities at the contour points

As mentioned before, we assume that the relative intensities depend on the dynamic range and width of vessels. Accordingly, we have computed the means of relative intensities at the inner and outer contour points of vessels having specific width (w) and dynamic range (d) properties in the manually annotated training databases DRIVE, STARE and HRF; arranged the values to form two surfaces $r_I(w,d)$ and $r_O(w,d)$ corresponding to the relative intensities at inner and outer contours, respectively; applied some smoothing with a small averaging kernel; and plotted some geodesics in Figure 9.

It can be observed, that the shape and distance of the curves is highly similar, independently from the scale of the images (the resolutions of the images in STARE and HRF are approximately 1.3 and 4.5 times larger than those of DRIVE, respectively) and the dynamic range: for thin vessels the relative intensities are small and increase until the width of approximately 8 pixels is reached. Then, a slower but coherent convergence of the curves can be observed as the width of the vessels increases towards the maximum. Another interesting property is that independently from the dynamic range or database, the distance of the curves decreases as the width of the vessels increases. The most characteristic difference of the curves extracted from various databases is their vertical position. The height of the

curves extracted from STARE and HRF is similar, but the curves of DRIVE are remarkably higher. Based on these observations we can assert that

1. the position of vessel contours (represented by their relative intensities) varies with the width of the vessels and slightly varies with the dynamic range of vessel profile;

2. in order to reach highly accurate segmentation results by the reconstruction of vessel contours, one must take into account the information about the position of the contours represented by the relative intensity surfaces.

3. although the surfaces are basically similar, the most conspicuous difference between them is their vertical position when the surfaces of various databases are compared. We assume that the heights of the surfaces are related to the quality of vessel edges (like sharpness and the level of noise, essentially affected by the characteristics of the imaging device). Accordingly, when the surfaces extracted from a specific set of images are used to reconstruct the vessel contours of images taken under different circumstances, the transformation of the relative intensity surfaces to the unknown ones could possibly increase the accuracy of segmentation. Later on, in Section 3.5, we present a technique for the blind estimation of a multiplicative factor that can be used to transform the relative intensity surfaces extracted from a training database to the unknown surfaces of images taken under different circumstances.

In order to reconstruct the characteristics of contours measured in the training database, the contours of the binary outcome of Step 2 have to be adjusted. Moreover, it can happen that several layers of outer contour pixels have to be added to the foreground to reach the desired similarity of relative intensities extracted from the training database and measured at the contour pixels of the binary segmentation. For the sake of readability the description of the proposed contour reconstruction technique is separated in two subsections: first, the iterative adjustment of one layer of contour pixels is described, then, the scheduling of the iterative adjustment involving one layer of contour pixels is discussed.

3.3.3. The iterative solution for one layer of contour pixels

Probably the most obvious and simple way to recover the desired relative intensities at the contour pixels of a binary segmentation is defining some target function measuring how the surfaces $r_I(w,d)$ and $r_O(w,d)$ fit the current segmentation, and adjusting the binary labels of contour pixels randomly towards the extremum of the target function.

Let $\mathcal{C}$ denote the set of contour pixels being involved in the iterative optimization procedure. The core of the iteration is a simple, pixel-wise operation described below. For a pixel $\mathbf{p} \in \mathcal{C}$ the score $s_0(\mathbf{p}) = s(\mathbf{p})$ is computed; the binary label at pixel $\mathbf{p}$ is negated if the modification does not make a connected vessel component disconnected; and the score of the modified binary image is
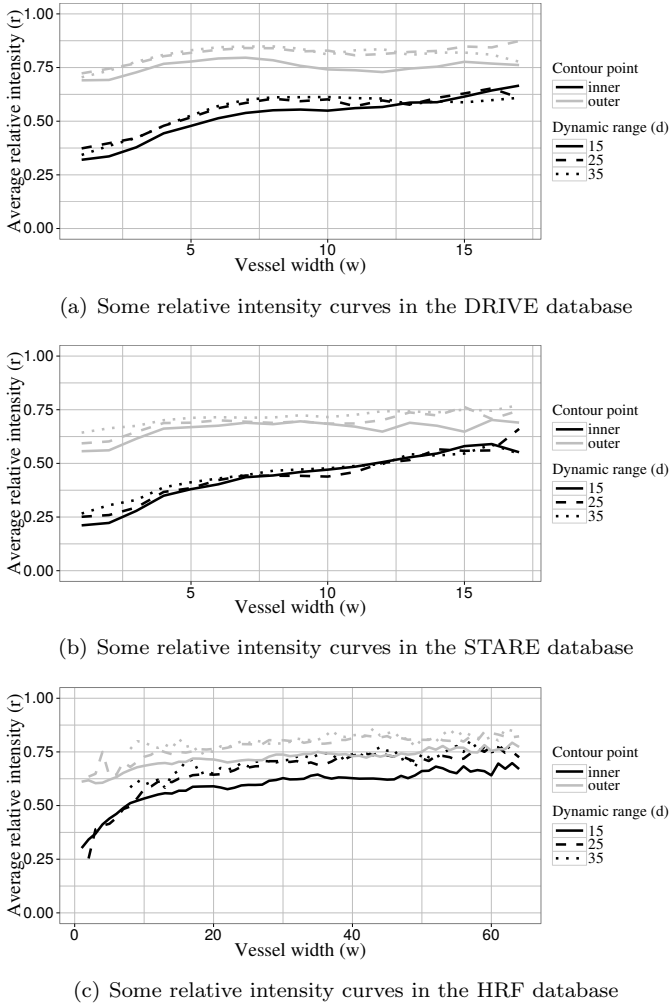
11

(a) Some relative intensity curves in the DRIVE database

(b) Some relative intensity curves in the STARE database

(c) Some relative intensity curves in the HRF database

Figure 9: Relative intensity curves measured on various databases

computed as $s_1(\mathbf{p}) = s(\mathbf{p})$. The new label at pixel $\mathbf{p}$ is kept, if $s_0(\mathbf{p}) > s_1(\mathbf{p})$, otherwise the modification is reverted. This simple operation is repeated for the contour pixels in $\mathcal{C}$ until convergence.

The score function is defined as follows. Let $r(\mathbf{p})$, $w(\mathbf{p})$ and $d(\mathbf{p})$ denote the measured relative intensity, width and dynamic range of the vessel passing by pixel $\mathbf{p}$, respectively, and let $\mathbf{p}$ be an inner contour point. We can use the measured $r_I(w, d)$ surface to estimate the ideal relative intensity at the contour of a vessel having width $w(\mathbf{p})$ and dynamic range $d(\mathbf{p})$. However, since the measured surface is known only at some discrete coordinates, while $w(\mathbf{p})$ and $d(\mathbf{p})$ can take any real value, bilinear interpolation is applied to evaluate the function $r_I(w, d)$ and determine the expected relative intensity $\hat{r}(\mathbf{p}) = r_I(w(\mathbf{p}), d(\mathbf{p}))$ at pixel $\mathbf{p}$. If $\mathbf{p}$ is an outer contour pixel, the desired ideal intensity is derived analogously from the surface $r_O(w, d)$.

Let the term

$$e(\mathbf{p}) = \begin{cases} \max\{0, r(\mathbf{p}) - \hat{r}(\mathbf{p})\}, \text{if } \mathbf{p} \text{ is inner contour point,} \\ \max\{0, \hat{r}(\mathbf{p}) - r(\mathbf{p})\}, \text{if } \mathbf{p} \text{ is outer contour point,} \end{cases} \tag{12}$$

measure the error of the estimated and desired relative intensities. This error term is part of the score function evaluated before and after the binary label of $\mathbf{p}$ is changed. If $\mathbf{p}$ is an inner contour pixel and its relative intensity is smaller than the desired relative intensity, it means that the pixel is closer to the center of the vessel than the inner contour pixels of vessels having similar width and dynamic range in the training database. In this case the error becomes zero, since the pixel should not change its label during the optimization process. However, changing the label of $\mathbf{p}$ to the opposite, it becomes an outer contour pixel, with a large error, thus, the error term suppresses to keep the original foreground label of $\mathbf{p}$. The error term behaves similarly if $\mathbf{p}$ is an outer contour pixel and its relative intensity is larger than the desired one. If $\mathbf{p}$ is an inner contour pixel and $r(\mathbf{p})$ is larger than the ideal, the error term measures the distance of $r(\mathbf{p})$ from the ideal $\hat{r}(\mathbf{p})$ intensity. Changing its label to the opposite, the error term measures the distance of $r(\mathbf{p})$ from the ideal relative intensity of outer contour pixels of vessels with similar width and dynamic range. The smaller value of the error term indicates the true label of $\mathbf{p}$. The error term has similar behavior if $\mathbf{p}$ is an outer contour pixel and $r(\mathbf{p})$ is smaller than the ideal. Consequently, the error term $e(\mathbf{p})$ suppresses a label configuration providing relative intensities being similar to the ideal one. In order to avoid cracked edges, and ensure local smoothness, another term is added to $e(\mathbf{p})$ measuring the similarity of labels in a local neighborhood. Particularly, the smoothness term becomes $w^{nb}(1 - n(\mathbf{p})/8)$, where $n(\mathbf{p})$ measures the number of 8-neighbors of $\mathbf{p}$ having the same binary label as $\mathbf{p}$, and $w^{nb} \in \mathbb{R}$ is a control parameter responsible for the strength of smoothing. Then, the score function $s(\mathbf{p})$ becomes:

$$s(\mathbf{p}) = \frac{e(\mathbf{p})}{r_O(w(\mathbf{p}), d(\mathbf{p})) - r_I(w(\mathbf{p}), d(\mathbf{p}))} + w^{nb}\left(1 - \frac{n(\mathbf{p})}{8}\right). \tag{13}$$

The error term is normalized by the distance of the relative intensity surfaces of inner and outer contours to prevent the score function from preferring thick vessels, where the distance of the curves and consequently the errors of the reconstruction are smaller.

The iterative optimization process repeats the previously described pixel-wise operation based on this score function for all contour pixels in $\mathcal{C}$ until convergence. However, filling $\mathcal{C}$ with inner and outer contour pixels leads to ambiguous results, since the edges of the vasculature may become leaky if an inner contour point is removed, and a neighboring outer contour point is added to the foreground in the same time. Thus, the proposed iterative optimization technique is applied sequentially to layers of inner and outer contour pixels of the foreground as described in the

next subsection.

3.3.4. The contour reconstruction process

The strategy of the proposed contour reconstruction technique is summarized in the following points:

1. We can expect that the adjustment of vessels having extremely small dynamic ranges is ambiguous, since they can hardly be differentiated from the background. Accordingly, a contour pixel is not adjusted if the dynamic range of the vessel passing by is less than t^{dr}.

2. In the first phase, $\mathcal{C}$ contains the inner contour pixels of the vasculature and the iterative optimization described in the previous subsection is applied until convergence. Since the weak segmentation operators introduced before tend to extract the center lines of the vessels, we expect this step to carry out only small adjustments on the contour of the segmented vasculature, and do not expect that more adjustment of inner contour pixels is required. Thus, the next steps involve outer contour pixels only. Two technical details are also noted here:

(a) The neighborhood term in the score function $s(\mathbf{p})$ is disabled if $\mathbf{p}$ is an inner contour pixel and the width of the vessel passing by is smaller than 2 pixels, since the local 8-neighborood of the inner contour pixel $\mathbf{p}$ may contain pixels of the outer contour of the opposite side of the vessel and lead to unexpected behavior.

(b) When a vessel unexpectedly end at the border of a bright lesion or fade into the background, the estimated normal of the vessel at the end point becomes parallel with the vessel and the brightness of the background can suppress the removal of inner contour pixels, shortening the extracted vessels. This undesired behavior is prevented by extracting the end points of the segmented vasculature and forbidding the removal of these foreground pixels during the adjustment of inner contours.

3. In the second phase, $\mathcal{C}$ is filled with the outer contour pixels of the binary segmentation having already adjusted inner contours, and the iterative optimization technique described in the previous section is applied until convergence. We expect this step to increase the width of vessels by adding outer contour pixels to the foreground of the segmentation, if required. In the case of tick vessels the addition of one layer of outer contour pixels can rarely reconstruct the real width of vessels. Thus, the adjustment of outer contour pixels is repeated layer-by-layer until the set of outer contour pixels becomes unchanged.

The proposed approach for the reconstruction of vessel contours has basically two free parameters: the control parameter of the local smoothness term w^{nb} and the threshold of the dynamic range t^{dr}. The optimal value of both parameters is determined by training on the images of a manually annotated training database. For the illustration of the proposed technique, Figure 7(c) presents the test image with reconstructed contours. The free parameters we used are the ones trained on DRIVE: $w^{nb} = 3.145$ and $t^{dr} = 10$. As one can observe, the widths of thick
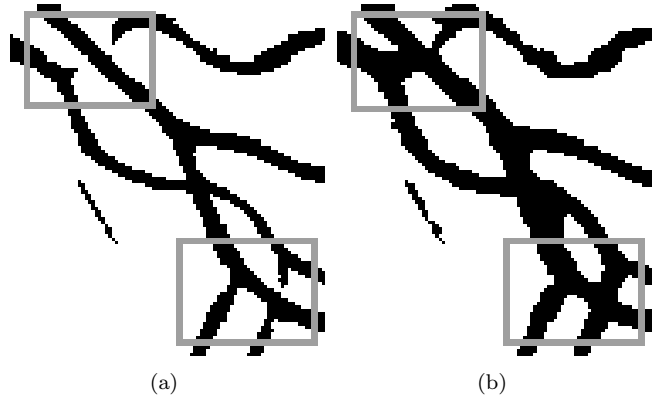

Figure 10: Illustration of the iterative contour reconstruction process: the binary segmentation after Step 2 (a), the binary segmentation after contour reconstruction (b). Some examples for the reconstruction of junction points are emphasized by grey rectangles

vessels became similar to that of the manual annotation, while the widths of relatively thin vessels have not changed significantly, as expected.

A common weakness of template based vessel segmentation techniques is that the similarities near junction points of the vasculature (especially near the junction points of arteries having a bright stripe in the middle) are usually low, resulting fragmented segmentations (see the junction points in Figure 10(a)). As a limitation of the proposed method, these structures cannot be modelled accurately enough with the Gabor function based templates we use. Nevertheless, the contour reconstruction technique implicitly attempts to reconstruct junctions: the relative intensities at the outer contour pixels of the erroneous end points of vessels are close to zero, since the estimated normals are parallel with the vessels. Consequently, outer contour pixels are added near the end points and many of the junctions become reconstructed by the proposed method in this way (see Figure 10(b)). In order to make the entire vasculature connective, the reconstruction of all junctions would be required. This could be accomplished in a Hidden Markov Model based framework by optimizing an energy function consisting of a connectivity term and a probability term describing how well the added vessel pixels fit the distribution of nearby vessels.

3.4. Step 4 – Addition of thin vessels

By thin vessels we mean the vessel segments after the third-fourth bifurcations of the vascular tree, independently from the resolution of the images. These vessels have poor contrast, in many cases fade into the background, further bifurcations can rarely be observed. As one can see in Figure 7(d), many thin vessels are still missing after Step 3. We have identified three issues deteriorating the segmentation of thin vessels by weak segmentation operators:

1. The annotation of thin vessels is ambiguous – we have found many thin vessels unannotated but being highly
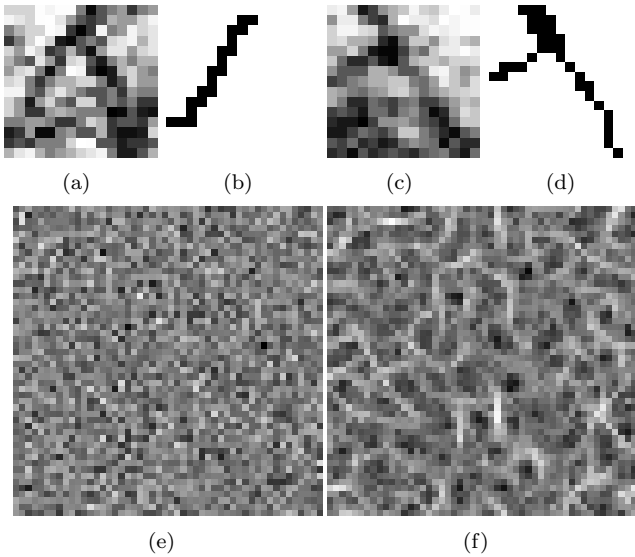
Figure 11: Illustration of the problems related to the segmentation of thin vessels. The ambiguity of annotations (a)-(d); an image containing artificial Gaussian noise (e); and the outcome of template matching on artificial noise (f)

similar to annotated ones. For illustration, we have visualized two thin vessels (after histogram equalization for visibility) and the corresponding manual annotations from database DRIVE in Figures 11(a)-11(d). As one can see comparing the figures, the right hand vessel in Figure 11(a) is not annotated in Figure 11(b), while vessels having similar contrast and scale are clearly annotated in Figures 11(c) and 11(d).

2. Due to the finite resolution of retinal images and the poor contrast of thin vessels, the annotations at the pixel level cannot follow the vessel precisely. If the width of thin vessels is 1-2 pixels, one pixel shift in the annotation results many false positive and false negative matches even if the segmentation seems to be correct visually. This ambiguity of the annotation of thin vessels can also be observed in Figures 11(a)-11(d).

3. If the width of thin vessels is approximately 1-2 pixels, the linear structures formed by the noise of the background can become highly similar to them. For illustration, we have visualized an image containing Gaussian noise of standard deviation $\sigma_G = 3$ pixels in Figure 11(e) and the outcome of template matching by Gabor function based templates with parameters $\sigma = 2.5$, $\lambda = 7$, $\phi = 0.7$, $\gamma = 0.7$, $\beta = 1$ and $q = 1$ in Figure 11(f). As one can see, the structures appearing in the outcome of template matching are highly similar to fragments of thin vessels (the brightest pixels in the outcome of template matching represent the PCC value 0.69), even though the input image in Figure 11(e) contains nothing more but artificial noise.

Although the segmentation of thin vessels is hardened by many circumstances and the direct clinical utilization of segmented thin vessels is limited, we consider this issue being part of the segmentation problem, and believe that the segmentation of thin vessels can indirectly improve various areas of retinal image processing (e.g. the removal of the vasculature for microaneurysm detection or the measurement of the fractal dimension of the vasculature). Therefore, we have appended a simple step to the proposed segmentation method for the extraction of thin vessels being not recognized by Step 2. We note that this step has extremely low computational demands, since it utilizes the results already computed in Step 2.

Due to the low signal-to-noise ratio of thin vessels, the ambiguity of annotations, and the nature of Step 2, only small fractions of thin vessels are extracted by weak segmentation operators, and these small regions are usually filtered out, since the circularities of small fragments are smaller than the predefined threshold. Although one weak segmentation operator cannot extract large fragments of thin vessels, one can expect some interference in the outcomes of template matching when templates having various shapes but wavelengths related to thin vessels are used, and this is what we utilize in this step.

The wavelength related to the thinnest vessels can be estimated in the following way. First, the widths of vessels passing by the inner contour points of the manual annotations are determined and the widths are propagated from the contours to the rest of the foreground pixels. Then, the foreground pixels having the smallest 1% of estimated widths are selected. Clearly, these pixels must belong to thin vessels. At each of the selected pixels template matching is applied by the templates corresponding to the operators in the set $\mathcal{W}^{red}(\mathcal{S})$, and for each pixel the wavelength of the template providing the highest PCC score is determined. The mean of these wavelengths λ^{thin} can be considered as the wavelength related to thin vessels.

Using the estimated wavelength λ^{thin} the addition of thin vessels is carried out in the following way. Template matching is applied by the templates related to the operators in the set $\mathcal{W}^{red}(\mathcal{S})$ and having wavelength parameters $0.8\lambda^{thin} \leq \lambda \leq 1.2\lambda^{thin}$. For each pixel the mean of the non-negative PCC scores is determined. Then, the image containing the pixel-wise means of PCC values is thresholded at various levels in the range $[0.4, 1]$. At each threshold level, connected components are extracted from the binary image and a component is accepted as a thin vessel if the distance of its farthest pixels is larger than t^{length} pixels and the mean of PCC values covered by the component is larger than t^{thin}. The threshold t^{length} corresponds to the observation that thin vessels have usually larger linear extents than the thin, but usually curved structures appearing in the noise of the background. The threshold t^{thin} gives a lower bound on the mean PCC score a component must have to be considered as a real vessel. Both thresholds are selected by training on the images of a manually annotated database. The lower bound of the range of thresholds was selected empirically: we have not found any real vessel component having mean PCC score lower than 0.4.

We emphasize, that due to the ambiguity of annotations we do not expect this step to increase the quantitative accuracy of a segmentation significantly. Nevertheless, it definitely increases the accuracy of segmentation qualitatively. For illustration, we have trained the parameters on the training images of the DRIVE database; found that the settings $t^{thin} = 0.529$ and $t^{length} = 19$ provided the highest accuracy scores; and visualized the outcome of Step 4 in Figures 7(d)-7(f).

3.5. Blind calibration of the trained model

So far we have described the proposed segmentation technique. The training of the method includes the training of weak segmentation operators, the extraction of the relative intensity surfaces and the training of four further parameters: w^{nb} – the control parameter of the smoothness of vessel contours; t^{dr} – the threshold parameter of the dynamic range of vessels when the contours are reconstructed; t^{thin} – the lower bound on the expected PCC values of real thin vessels; and t^{length} – the lower bound on the expected spatial extents of real thin vessels. However, retinal images taken by varying imaging systems have different characteristics. The most significant difference is the scale of the image content, but the differences arising from the varying optical transfer functions also affect the way vessels appear in the images. Thus, when a segmentation model trained on a particular database is used to segment images taken under different circumstances, the accuracy of segmentation deteriorates. In this section, a blind calibration process is described that can adopt a trained model to unseen input images taken under different circumstances than the ones used to train the model.

The idea of blind calibration is based on the observation that one cannot expect accurate results if the scale of the input image and that of the ones used to train the model differ. Thus, the scale of the input image and that of the trained model must be synchronized by the estimation of a scaling factor to improve the accuracy of segmentation. Analogously to the effect the scaling factor, one can estimate further control parameters that can be used to improve the accuracy of segmentation by the transformation of a trained model. The proposed blind calibration process estimates three multiplicative factors that can be used to transform a trained model: $\hat{f}^{sc}$ is related to the scaling of the image content; $\hat{f}^{th}$ is related to the adjustment of threshold parameters in weak segmentation operators and $\hat{f}^{rel}$ is related to the transformation of the relative intensity surfaces. The methods proposed for the estimation of the optimal factors are described in the following subsections. In each case, several factors are tested and the one providing the extrema of some target function is selected as the optimal factor. For the ease of discussion, the estimated optimal factors are denoted by hat symbols (like $\hat{f}^{sc}$), and the ones being tested are denoted without hat symbols (like f^{sc}).

3.5.1. Estimation of the scaling factor

A rough estimation of the scaling factor can be carried out as $f^{sc}_{ROI} = R^{in}_{ROI}/R^{tr}_{ROI}$, where R^{in}_{ROI} denotes the radius of the ROI of the unseen input image and R^{tr}_{ROI} stands for the mean radius of ROIs in the training database. However, this estimation does not take into account that the FOV can vary, that is, the scale of the anatomical parts can differ even if $f^{sc}_{ROI} = 1$. A better, but computationally expensive estimation can be achieved by utilizing the results published in (Dorner et al., 2002): the diameter of the central artery of the retina is approximately $163 \pm 17 \mu m$. Thus, the width of the thickest vessels is a characteristic property of retinal images. Since thick vessels are extracted by the operators having large wavelength parameters, we can expect, that the scale at which these operators perform well is the optimal scale for the segmentation of the vasculature.

Let $\mathcal{W}^{thick}(\mathcal{S})$ denote the set of operators extracting thick vessels, that is, the set of operators having wavelength parameters greater than $0.8\lambda^{max}$, where λ^{max} denotes the largest wavelength appearing in $\mathcal{W}^{red}(\mathcal{S})$. The threshold $0.8\lambda^{max}$ was selected empirically, slight variations of the threshold do not change the results significantly. The criteria used to select the optimal scaling factor are based on two observations: as the scaling factor deviates from the ideal one, less operators from $\mathcal{W}^{thick}(\mathcal{S})$ extract foreground pixels and the width of the extracted vessels also decreases. These observations are implemented in the following estimation technique. Various factors of the range $f^{sc} \in [0.5f^{sc}_{ROI}, 2f^{sc}_{ROI}]$ are used to rescale the templates; at each scale Step 2 is applied to the input image involving the operators $\mathcal{W}^{thick}(\mathcal{S})$; and the scaling factor leading to the largest number of operators extracting foreground pixels is selected as the optimal one. If several factors lead to the same number of operators extracting foreground pixels, the largest median width of extracted vessels is used as the indicator of the optimal one. The method utilizes the rough estimation f^{sc}_{ROI} to decrease the computing demands by reducing the parameter space of the scaling factor.

Although the proposed estimation technique seems to implement the criteria we have formulated before, there is an issue the method cannot handle properly: as the scaling factor of the templates decreases, the operators corresponding to thick vessels begin extracting vessels of medium and small widths. Thus, a rough estimation of the regions where thick vessels reside is required, and the estimation technique described before has to be applied in these regions only. Utilizing that thick vessels are the largest elongated structures appearing in retinal images, some regions where thick vessels reside can be extracted in the following way. Various factors in the range $f^{sc} \in [0.5f^{sc}_{ROI}, 2f^{sc}_{ROI}]$ are applied to rescale the templates; at each scale Step 2 is carried out involving the operators $\mathcal{W}^{thick}(\mathcal{S})$; and the outcome at the largest scale having at least 0.01% of the pixels recognized as foreground is sup-

Table 1: Estimation of the scaling factor ($\hat{f}^{sc}$)

Trained on \ Estimated on	DRIVE	STARE	HRF
DRIVE	-	1.33	5.26
STARE	0.83	-	3.84

posed to contain some center lines of thick vessels. This outcome is dilated by a disk shaped structuring element of radius $0.05R_{ROI}^{in}$ pixels, and we expect that the foreground regions of the dilated image cover regions where thick vessels reside. The previously described estimation of the scaling factor is performed in these regions only.

Once the optimal scaling factor is determined, one can rescale the templates by the factor $\hat{f}^{sc}$ and carry out Step 2 involving the operators $\mathcal{W}^{red}(\mathcal{S})$ to process the input image. A technical detail related to the scaling of templates is remarked here. If $\hat{f}^{sc} \geq 1$, scaled templates can be assumed to represent vessel segments, as expected. On the contrary, if $\hat{f}^{sc} < 1$, some scaled templates may become far too small to recognize anything but the random variations of the background. In order to avoid these situations, after the scaling of templates each weak segmentation operator is checked in the following way: The operator is applied to an image containing artificial Gaussian noise and if the maximum PCC score in the outcome of template matching is larger than $(t_l + t_h)/2$, the operator is not used for segmentation. It is worth noting that Gaussian noise of any standard deviation can be used, since the PCC score is invariant against multiplications by a constant.

To illustrate the operation of the proposed scale estimation technique, we have used the models trained on DRIVE and STARE to estimate the scaling factor of the images of two other databases. The geometric means of the individual factors are summarized in the Table 1. We can draw the conclusions that the estimation method is coherent in the sense that the cross-estimations of scaling factors are approximately inversely proportional (e.g. $1/0.83 = 1.20 \simeq 1.33$), while an approximate transitivity can also be observed, since transforming an image of DRIVE to the scale of HRF means the multiplication of the linear extents by 5.26, while transforming an image of DRIVE to the scale of STARE, then scaling the image to the scale of HRF means the multiplication of linear extents by $1.33 \cdot 3.84 = 5.10 \simeq 5.26$. We emphasize that exact inverse relationships among the estimated scaling factors cannot be expected, since the varying quality of images and the geometrical distortions arising from the projection of curved retinal surfaces may influence the results. However, it can be expected that the scaling of the templates or the inverse scaling of the unseen input image leads to higher segmentation accuracies. In order to validate the scale estimation technique qualitatively, we have resized some images of STARE and HRF to the scale of DRIVE using the estimated scaling factors and enclosed them in Figure 12. As one can see in the figure, although
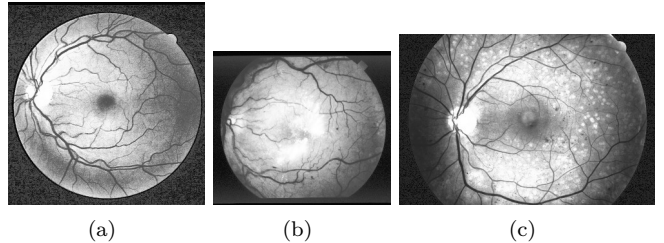

(a)   (b)   (c)

Figure 12: Qualitative illustration of the scale estimation method: image from DRIVE, resolution: 565×584 pixels (a); image from STARE, original resolution: 700×605 pixels (b); image from HRF, original resolution: 3504×2336 pixels (c). In order to increase the visibility of details, histogram equalization was applied to the images

the resolution and the field of view of the images vary, the sizes of main structures (like the diameter of the optic disc and the width of the thickest vessels) are similar.

3.5.2. Estimation of the threshold factor

In fact, fundus cameras have various optical transfer functions. Accordingly, we can suppose that the appearance of vessels also varies when images taken by different instruments are compared. Since the weak segmentation operators are trained to provide as precise segmentations as possible, a slight change in the appearance of the vessels can deteriorate their efficiency. For example, if the level of noise is larger, the PCC responses of template matching become lower and less patches of vessels are extracted. Similarly, if the level of noise is lower, the PCC responses may become higher and beside the patches of vessels, further, unexpected and undesired structures may appear in the outcomes of weak segmentation operators. Thus, adjusting the threshold parameters of weak segmentation operators, may increase the performance of segmentation. By adjustment we mean the estimation of a multiplicative factor f^{th} and the updating the threshold parameters by the rule

$$t'_{\{l,h\}} = \begin{cases} 1 - (1 - t_{\{l,h\}})/f^{th}, & \text{if } f^{th} \geq 1, \\ t_{\{l,h\}} f^{th}, & \text{otherwise.} \end{cases} \quad (14)$$

The updating rule was developed to prevent the thresholds from leaving the interval $[0, 1]$. Accordingly, the adjustment made to a threshold is proportional to its value if $f^{th} < 1$, and the amount of adjustment is proportional to $(1 - t_{\{l,h\}})$ if $f^{th} \geq 1$. The effect of adjustment depends on the value of fth. If $f^{th} > 1$, the thresholds are increased and assumably the number of false positive detections decreases smoothly, even if the threshold parameters of some weak segmentation operators become too high to segment any pixel as foreground. Contrarily, if $f^{th} < 1$, the thresholds are decreased, and the number of extracted foreground pixels increases. However, one has to take into consideration that the excessive decrease of low threshold parameters can deteriorate the overall performance of Step 2 by introducing many false positive detections. In order

to prevent the method from these situations, a lower bound is introduced: the threshold parameters of a weak segmentation operator can be decreased until the lower threshold t_l reaches the lowest unadjusted threshold parameter t_l^* of all weak segmentation operators. Particularly, the lower bound t_l^* is defined as

$$t_l^* = \min\{t_l | W(\mathbf{s}, (\sigma, t_l, t_h)) \in \mathcal{W}^{red}(\mathcal{S})\}. \qquad (15)$$

In order to find the optimal factor $\hat{f}^{th}$, we need to find some measure to characterize how good the outcome of Step 2 is. We have found that the relative intensities measured at the inner and outer contour points of the segmented vasculature can be used for this purpose. Particularly, for various values of $f^{th} \in [0.5, 1.5]$ the threshold parameters of the weak segmentation operators are adjusted and Step 2 is carried out. (The borders of the range are determined empirically: we have not faced any case when a threshold factor outside this range would have been the optimal choice.) For each of the outcomes the mean relative intensities at the inner (m_I) and outer (m_O) contour points are determined. The fth factor providing the largest absolute difference $|m_I - m_O|$ is considered to be the optimal one. The target function can be interpreted in the following way. If the thresholds are too high, only a few pixels are extracted near the center lines of vessels, and $|m_I - m_O|$ becomes small. In this case, we can assume that center lines of many vessels are not extracted at all, thus, the thresholds should be lowered. Similarly, if the thresholds are too low, many vessel pixels are extracted, the inner contours move towards the background of the vessels decreasing the target function $|m_I - m_O|$. In this case, we can assume that many false positive detections are present, thus, the thresholds should be increased. According to this interpretation, the expectation that the optimal $\hat{f}^{th}$ factor is indicated by the largest absolute difference $|m_I - m_O|$ is reasonable.

We note that the outcome of template matching applied by a weak segmentation operator does not depend on the threshold parameters. Consequently, the estimation of the optimal threshold factor can be implemented efficiently, since the only operations being needed to be applied for the various factors f^{th} are hysteresis thresholding and connected component labelling, both of them having linear time complexity.

To illustrate the proposed technique, we have used the models trained on DRIVE and STARE to estimate the optimal threshold factors for the images of two other databases, and the geometric means of the individual factors were determined. The results are summarized in Table 2. We can conclude that the results have some transitivity properties: when the model trained on STARE is used to segment images from DRIVE and HRF, the thresholds are decreased and increased, respectively; when the model trained on DRIVE is used to segment images from STARE or HRF, the thresholds are increased with a larger increase in the latter case. The estimated factors clearly

Table 2: Estimation of the threshold factor ($\hat{f}^{th}$)

Estimated on / Trained on	DRIVE	STARE	HRF
DRIVE	-	1.16	1.22
STARE	0.96	-	1.07

show that the thresholds trained on DRIVE are too low to be applied to image from the other sources, while the thresholds trained on STARE need only a small adjustment to become applicable to the images of DRIVE and HRF. We emphasize that the models trained on DRIVE and STARE have different weak segmentation operators, thus, one cannot expect exact inverse relations among the estimated factors. However, we can expect that the proposed automated adjustment of thresholds increases the accuracy of segmentation. Further quantitative validation of this calibration step is provided in Section 4, where the segmentation accuracies with and without this calibration step are compared.

3.5.3. Estimation of the relative intensity transformation factor

As we have pointed out before, the height of the relative intensity surfaces used to reconstruct the vessel contours varies with the databases. We assume that similarly to the thresholds of weak segmentation operators, the heights of the surfaces are affected by the circumstances under which the images were taken. The aim of this calibration step is to estimate a multiplicative factor f^{rel} that can be used to transform the surfaces $r_{\{I,O\}}(w, d)$ to fit characteristics of the unseen input image better. The transformation rule is similar to the one used to transform the thresholds of the weak segmentation operators:

$$r_{\{I,O\}}(w,d)' = \begin{cases} 1 - (1 - r_{\{I,O\}})(w,d)/f^{rel}, & \text{if } f^{rel} \geq 1, \\ r_{\{I,O\}} f^{rel}, & \text{otherwise,} \end{cases} \qquad (16)$$

that is, the rule prevents the surfaces from exceeding the range $[0, 1]$: as $f^{rel} \to \infty$, the surfaces converge to 1, as $f^{rel} \to 0$, the surfaces converge to 0. In order to estimate the optimal factor $\hat{f}^{rel}$, Step 3 is performed using the relative intensity surfaces transformed by various values of $f^{rel} \in [0.5, 1.5]$, and the outcomes are characterized by a score function. The factor providing the lowest value of the score function is considered to be the optimal choice for the final reconstruction of contours. The score function characterizing the reconstruction of contours is based on the local score function $s(\mathbf{p})$ introduced in equation (13). Since $s(\mathbf{p})$ measures how the relative intensity at pixel $\mathbf{p}$ fits the expectations, all we have to do is to compute the mean of the $s(\mathbf{p})$ scores in the inner (s_I) and outer (s_O) contour points, and use the sum ($s_I + s_O$) to characterize the binary outcome of Step 3 when a given factor f^{rel} is used for the transformation of surfaces. The lowest value of the sum ($s_I + s_O$) indicates the optimal factor $\hat{f}^{rel}$.

Table 3: Estimation of the transformation factor of relative intensities ($\hat{f}^{rel}$)

Trained on \ Estimated on	DRIVE	STARE	HRF
DRIVE	-	0.84	0.77
STARE	1.36	-	0.9

To illustrate the method proposed for the estimation of $\hat{f}^{rel}$, we have used the models trained on DRIVE and STARE to estimate the optimal factors for the images of two other databases, and the geometric means of the individual factors were computed. The results are summarized in Table 3. Again, the results show transitivity properties: when the model trained on DRIVE is used to segment the images from STARE and HRF, the relative intensities are decreased, and the surface transformed to HRF is lower than that of STARE. In accordance, when the model trained on STARE is used to segment images from DRIVE and HRF, the relative intensities are increased and decreased, respectively. Some qualitative validation of the method can be carried out by examining the height of the corresponding curves in Figure 9. Further quantitative validation is provided in Section 4, where segmentation accuracies with and without this calibration step are compared.

3.5.4. Segmentation by the calibrated model

Since the estimation of the scaling factor is computationally expensive, it is not efficient to carry out the estimation for images from the same source individually. Instead, the scaling factor should be estimated for a small set of images, and the geometric mean of the individual scaling factors can be used for the rest of the images from that source. After $\hat{f}^{sc}$ is estimated, all parameters of the proposed method related to the linear extents of vessel structures are scaled by $\hat{f}^{sc}$ (the templates, and the threshold on the length of thin vessels t^{length} in Step 4). With the properly scaled templates the estimation of the threshold factor is carried out and Step 2 is applied using $\hat{f}^{th}$ to adjust the threshold parameters of weak segmentation operators. Then, the optimal transformation factor of the relative intensity surfaces $\hat{f}^{rel}$ is estimated and the vessel profiles are reconstructed. Finally, in Step 4, the threshold parameter t^{thin} (related to the PCC scores like the thresholds of the weak segmentation operators) is updated by the same rule as the thresholds of the weak segmentation operators (14), and the addition of thin vessels is carried out. When a segmentation model trained on low resolution images is applied to images of high resolution, that is, the scale factor $\hat{f}^{sc}$ is much larger than 1, some vessels being thinner than the ones related to $\hat{f}^{sc}\lambda^{thin}$ may appear on the images. In order to give the proposed method a chance for the segmentation of these vessels, Step 4 is repeated by scaling the templates with some scaling factors in the range $[1, \hat{f}^{sc}]$.

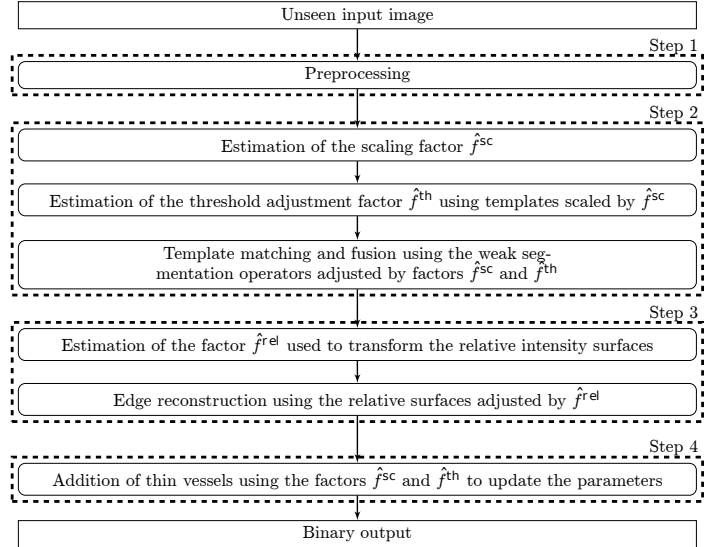For a better insight, the steps of the blind calibration

Figure 13: The blind calibration and segmentation model

and segmentation technique are visualized in a flowchart in Figure 13.

4. Tests and Results

The detailed description of the testing methodology is extremely important to make the results comparable with the state-of-the-art approaches. Although the databases and techniques used for evaluation vary, we try to follow a protocol which makes the proposed method comparable with the majority of previous techniques within the available space.

4.1. Databases

The majority of vessel segmentation methods are evaluated on two publicly available databases. The DRIVE database (Niemeijer et al., 2004) consists of 40 RGB fundus images of size 565×584, divided to training and test sets of equal size, both sets containing healthy and abnormal images, as well. The database contains the manual annotations of two human observers, and the annotations of observer 1 are considered to be the ground truth in the literature. The STARE database (Hoover et al., 2000) contains 10 healthy and 10 abnormal RGB fundus images of size 700×605 pixels. Similarly to DRIVE, database STARE contains the manual annotations of two experts, and the annotations of observer 1 are considered to be ground truth segmentations. In many recent papers the vessel segmentation techniques are evaluated using only DRIVE and STARE (De et al., 2014; Xiao et al., 2013; Mendonça and Campilho, 2006; Moghimirad et al., 2012; Lam et al., 2010; Marín et al., 2011), furthermore, there are some papers where only database DRIVE is used for evaluation (Lupascu et al., 2010). Thus, we can consider these databases as the conventionally accepted way for the

evaluation and comparison of vessel segmentation techniques. Beside testing on DRIVE and STARE, we have evaluated the proposed method on a novel database containing high resolution fundus (HRF) images (Odstrcilik et al., 2013). The HRF database consists of 45 RGB images of size 3504×2336 pixels and the images are divided to three sets of equal size, namely, *healthy*, *glaucomatous* and *diabetic retinopathy* images. This database contains the manual annotations of one human observer.

4.2. Measuring segmentation performance

The segmentation performance is characterized by the commonly accepted measures of binary classification. The test images are segmented by the proposed method, and the numbers of correctly classified vessel pixels (TP), incorrectly classified vessel pixels (FN), correctly classified background pixels (TN) and incorrectly classified background pixels (FP) are determined in the region covered by the ROI. Sensitivity $SE = TP/(TP + FN)$ and specificity $SP = TN/(TN + FP)$ measure the proportion of correctly classified foreground and background pixels to all foreground and background pixels, respectively. For both measures, higher values indicate better results. Positive predictive value $PPV = TP/(TP + FP)$ measures the proportion of correctly classified vessel pixels to all pixels classified as vessel. Negative predictive value $NPV = TN/(TN + FN)$ measures the same for background pixels. Again, higher values indicate better performance. Accuracy $ACC = (TP + TN)/(TP + FP + TN + FN)$ measures the proportion of correctly classified pixels to all pixels covered by the ROI. Most of the authors use ACC as the primary measure to characterize the performance of vessel segmentation. However, the accuracy scores have to be interpreted carefully. Due to the large number of background pixels, the accuracy of a segmentation containing only background labels is 0.83 in DRIVE, which is considered to be a fairly good score in many applications of binary classification. Even in DRIVE – containing the smallest images used for evaluation – a 0.001 increase of accuracy represents the correct classification of approximately 200 additional pixels per image.

The authors of machine learning based approaches usually report another measure: the receiver operating characteristic (ROC) curve plots sensitivity SE as a function of specificity SP, and the corresponding measure is the area under the curve AUC. In fact, the AUC score does not measure the performance of an individual system, although retinal screening systems are based on automated components. Rather, the AUC score is an aggregated measure of classification performance as the parameters of the system are tuned: AUC measures the trade-off of performance when the sensitivity of the method is increased to the detriment of specificity. Although the AUC score is a commonly reported measure, it has to be treated with reservations when vessel segmentation methods are ranked by it. On the one hand, measuring the AUC score supposes that there is a kind of global threshold that can be used to vary SE and SP smoothly. There are algorithms like the proposed one, where neither global probability maps, nor continuous tuning parameters are available to measure various pairs of SE and SP values, plot the ROC curve, and compute the AUC score. On the other hand, there are several techniques used to handle the concavities of ROC curves and measure the area under the curve: one can use simple numerical integration, polynomial fitting, or measure the area under the convex hull of the curve, and each of these techniques leads to different AUC scores. There are also recent criticisms of the AUC score regarding its misleading properties when classifiers are compared by it (Lobo et al., 2008; Hanczar et al., 2010). Although the proposed method is not suitable for the computation of the AUC score, we have generated some SE and SP pairs by manually varying the factors $\hat{f}^{th}$ and $\hat{f}^{rel}$ and calculated the AUC score in the same way as described in (Fraz et al., 2012a). Nevertheless, we treat the AUC score as a secondary measure of vessel segmentation performance and consider the accuracy score (ACC) as the primary indicator.

4.3. Comparison with existing techniques

The performance of the proposed method is compared to the ones reported and used for comparison in recent papers. The majority of authors calculated the measures in the region covered by the ROI and we consider this method to be the preferred way to measure the performance of vessel segmentation techniques. Nevertheless, some authors used all the pixels of the images to calculate the performance measures (Ricci and Perfetti, 2006; Lupascu et al., 2010). This can be easily verified by solving the equations defining ACC, SE and SP for the unknowns TP, TN, FP and FN using the reported values of the measures and the number of foreground pixels in the test images FG. For the sake of completeness, the values TP, TN, FP and FN are determined by the expressions:

$$TP = SE \cdot FG, \quad FN = FG - TP, \qquad (17)$$

$$FP = \frac{TP(ACC - 1) + ACC \cdot FN}{\frac{SP}{1-SP}(1 - ACC) - ACC}, \qquad (18)$$

$$TN = \frac{SP}{1 - SP} FP, \qquad (19)$$

In both of the mentioned cases, the sum $TP + FN + FP + TN$ equals the number of pixels of the images instead of the number of pixels covered by the ROI. In order to make the proposed method comparable with these techniques, we have also computed the performance measures involving all pixels of the images and these results are referred as *Evaluation without ROI*.

4.4. Model selection

The parameters of the proposed method are trained as we have described in Section 3. Especially, for training of template matching operators, we have used the sets Λ, Γ,

$\mathcal{Q}$ and $\mathcal{B}$ as given in Section 3.2.2, assuming that the parameter combinations of these sets densely cover the meaningful and reasonable part of the parameter space. An increased resolution of the parameter space could slightly improve the accuracy of segmentation, since more than 90% of the trained weak segmentation operators were discarded when the optimal set of operators was selected by the BFSS method. According to the vast majority of previous techniques, the primary indicator of performance is the accuracy of segmentation. Thus, the free parameters w^{nb}, t^{dr} (Step 3) and t^{thin}, t^{length} (Step 4) are selected to maximize the accuracy of segmentation using the training images of databases.

4.5. DRIVE database

We have evaluated the proposed method on database DRIVE. At the end of the training process the set $\mathcal{W}^{red}(\mathcal{S})$ contained 384 weak segmentation operators. The surfaces of relative intensities were extracted (some geodesics can be seen in Figure (9) and the free parameters providing the highest accuracy scores on the training images have been determined: $w^{nb} = 3.14$, $t^{dr} = 10$, $t^{thin} = 0.53$ and $t^{length} = 19$.

The proposed method was applied to the images of the test set and the averages of the previously introduced performance measures were computed. The mean accuracy, sensitivity and specificity scores are compared to those of previous techniques in the second column of Table 4. One can observe, that after the reconstruction of vessel contours the accuracy scores are higher than that of any method published before. As expected, the additional segmentation of thin vessels slightly increases the accuracy of the method. However, the addition of thin vessels increases the sensitivity of the method by more than 1%. Considering the gaps between the accuracies of techniques published in the last 10 years, we can assess that the advances of the proposed method are significant. Since the PPV and NPV scores are rarely reported, we cannot compare them to those of previous techniques. Anyway, for the sake of completeness, the average PPV and NPV scores are 0.8397 and 0.9634, respectively. Regarding the AUC score, there is only one technique performing better than the proposed one. For illustration, Figure 14 shows the segmentation results of healthy and abnormal images from DRIVE.

4.6. STARE database

The proposed method has been evaluated also on database STARE. However, due to some reasons, comparing the test results achieved on STARE to that of previous techniques is less precise than that on DRIVE: STARE contains only 20 images; the training set is not defined, different authors used different – usually unreported – subsets for training and testing that makes the interpretation and comparison of the results hard; STARE does not contain ground truth ROIs, larger or smaller ROIs lead to different quantitative results.
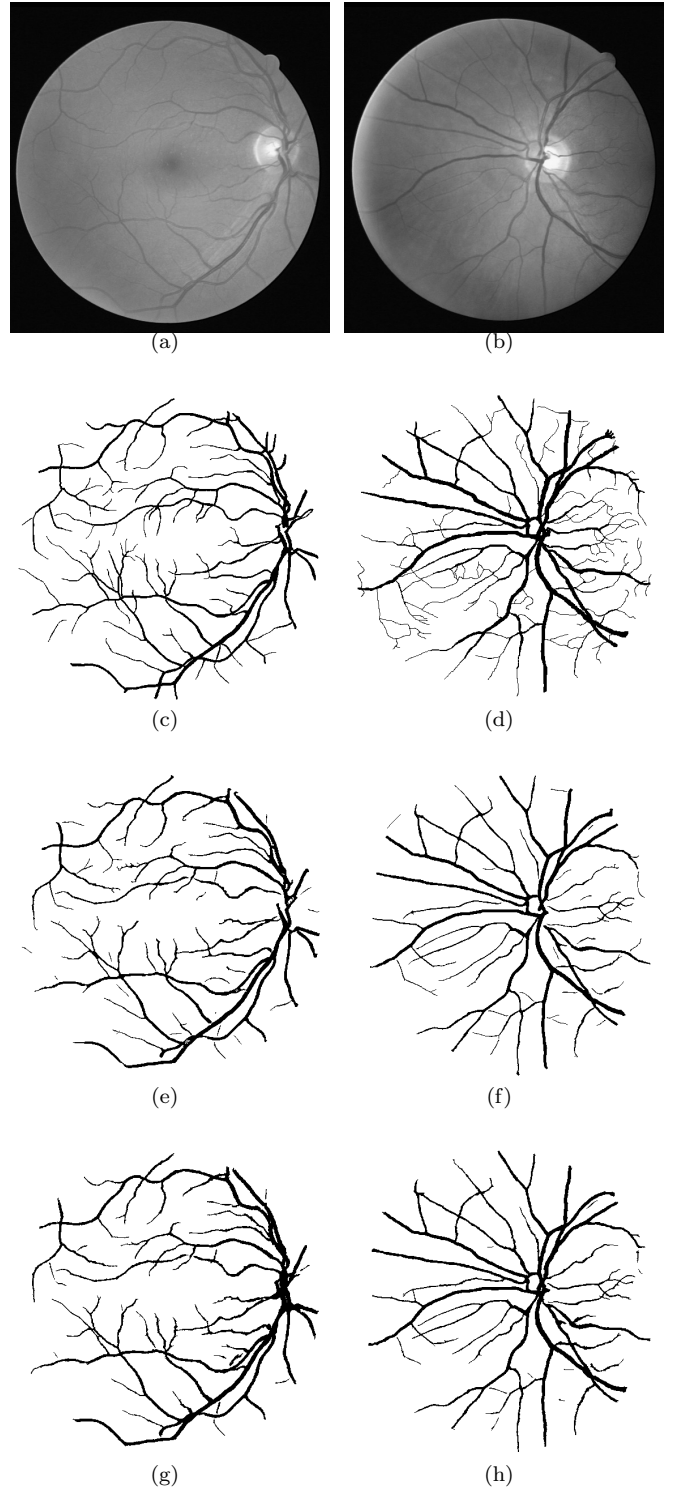

Figure 14: Illustration of the proposed method on the database DRIVE. Healthy (a), abnormal (b) images; ground truth annotation (c), (d); segmentations by the proposed method trained on database DRIVE (e), (f); segmentations by the proposed method trained on database STARE (g), (h)

We have separated the images of the database into two sets: 10 training and 10 test images. The training set con-

Table 4: Comparison of the performance of the proposed method with previous techniques on the databases DRIVE and STARE. In the case of STARE the performance measures were computed on the test set of images and on the entire database (marked by star (∗) symbols), as well. The highest scores in the various categories are indicated by boldface typesetting

Measure	RESULTS ON DRIVE				RESULTS ON STARE			
	ACC	SE	SP	AUC	ACC	SE	SP	AUC
	Trained on DRIVE				Trained on STARE			
Method	Evaluated with ROI							
Second observer	0.9473	-	-	-	0.9550	-	-	-
Jiang and Mojon (2003)	0.9212	-	-	0.9114	0.9009	-	-	0.9298
Vlachos and Dermatas (2010)	0.9291	0.7472	0.9550	-	-	-	-	-
Martinez-Perez et al. (2007)	0.9181	-	-	-	-	-	-	-
Espona et al. (2007)	0.9352	0.7472	0.9615	-	-	-	-	-
Nguyen et al. (2013)	0.9407	-	-	-	-	-	-	-
Niemeijer et al. (2004)	0.9417	0.6898	0.9696	-	-	-	-	-
Staal et al. (2004)	0.9442	0.7194	0.9773	-	0.9516	-	-	0.9614
Marín et al. (2011)	0.9452	0.7067	0.9801	-	0.9526	0.6944	0.9819	0.9769
Mendonça and Campilho (2006)	0.9452	0.7344	0.9764	-	0.9440	0.6996	0.9730	-
Soares et al. (2004)	0.9466	0.7332	0.9751	0.9614	0.9479	0.7207	0.9747	0.9671
Lam et al. (2010)	0.9472	-	-	0.9614	0.9567	-	-	0.9739
Fraz et al. (2012a)	0.9480	0.7406	0.9807	0.9747	0.9534	0.7548	0.9763	0.9768
The proposed method after step 3	0.9491	0.7344	0.9804	-	**0.9610** 0.9609*	0.8012 0.7965*	0.9788 0.9800*	- -
The proposed method after step 4	**0.9494**	0.7450	0.9793	0.9722	**0.9610** 0.9609*	0.8034 0.7991*	0.9786 0.9797*	0.9836 0.9893*
Std. dev. of the proposed method after step 4	0.0053	0.05531	0.0059	-	0.0090	0.0767	0.0104	-
	Evaluated without ROI							
Ricci and Perfetti (2006)	0.9595	-	-	0.9633	0.9646	-	-	0.9680
Lupascu et al. (2010)	0.9597	0.6728	0.9874	0.9561	-	-	-	-
The proposed method after step 4	**0.9652**	0.7446	0.9864	-	**0.9714** 0.9714*	0.8030 0.7987*	0.9847 0.9856*	- -
	Trained on STARE				Trained on DRIVE			
	Evaluated with ROI							
Soares et al. (2004)	0.9397	-	-	-	0.9327	-	-	-
Marín et al. (2011)	0.9448	-	-	-	0.9528	-	-	-
Fraz et al. (2012a)	0.9456	-	-	-	0.9493	-	-	-
The proposed method with scale estimation only	0.9449	0.6515	0.9877	-	0.9485	0.8807	0.9564	-
The proposed method with the estimation of the scale and threshold factors	0.9465	0.6584	0.9885	-	0.9564	0.8086	0.9735	-
The proposed method with the complete calibration	**0.9489**	0.7274	0.9812	-	**0.9605**	0.7700	0.9830	-
	Evaluated without ROI							
Ricci and Perfetti (2006)	0.9266	-	-	-	0.9464	-	-	-
The proposed method with scale estimation only	0.9621	0.6511	0.9919	-	0.9623	0.8802	0.9690	-
The proposed method with the estimation of the scale and threshold factors	0.9632	0.6581	0.9925	-	0.9681	0.8081	0.9812	-
The proposed method with the complete calibration	**0.9648**	0.7270	0.9877	-	**0.9711**	0.7665	0.9879	-

tains the first five healthy and first five abnormal images (with identifiers: 0001, 0002, 0003, 0004, 0005, 0162, 0163, 0235, 0236, 0239), and the rest of the images constitute the test set.

At the end of the training process, the set $\mathcal{W}^{red}(\mathcal{S})$ contained 288 operators. The relative intensity surfaces were extracted, and the free parameters providing the highest accuracy scores on the training images have been selected: $w^{nb} = 3.12$, $t^{dr} = 10$, $t^{thin} = 0.592$ and $t^{length} = 42$. Comparing the optimal values of free parameters to those trained on DRIVE, one can observe, that w^{nb} and t^{dr} are essentially the same, suggesting that these values are related to the nature of the problem of vessel segmentation. Contrarily, there is a significant difference regarding the

parameters t^{thin} and t^{length}. This can be explained easily, considering that t^{thin} is related to the appearance of vessels (affected by the optical transfer function of the imaging device), and t^{length} is closely related to the scale of the image content. Both of these values depend on the circumstances under which the images are taken.

The trained segmentation model was applied to the complete database and the averages of the performance measures were computed for the test images and for all images of the database, as well. The results are compared to those of previously published techniques in Table 4. As one can see in the results, the scores achieved on the test images are essentially the same as the scores computed on the entire database: no overfitting to the training images

can be observed. Furthermore, one can observe, that the accuracy scores of the proposed technique are much higher than that of any other technique published before. Similarly to the case of DRIVE, the addition of thin vessels does not result a significant increase in the accuracy scores, although the sensitivity increases and qualitative improvements can be observed. The average PPV and NPV scores are 0.8200 and 0.9768, respectively. The AUC score of the proposed method is more than 0.5% higher than that of the best performing previous technique trained and tested on STARE. Figure 15 illustrates the segmentation results of healthy and abnormal images of STARE.

4.7. Cross-validation between databases

In the practice of retinal image processing, it is not feasible to train distinct segmentation models for the images coming from various sources. Thus, the cross-validation of a method between databases is a highly important indicator of the practical usefulness of the method, although only a few cross-validation results are reported in previous papers. Accordingly, the model trained on DRIVE was tested on the images of STARE, and vice-versa, the model trained on STARE was applied to the test images of DRIVE. Within the blind calibration process we have used 5-5 images to determine the scaling factors and the factors f^{th} and f^{rel} corresponding to the test images were estimated individually. The means of the estimated factors are summarized in the previously discussed Tables 1, 2, 3. In order to give some further quantitative validation for the steps of the blind calibration process, we have also carried out the calibration and segmentation by disabling the estimation of the threshold factor and the estimation of the transformation factor of relative intensities. Clearly, disabling the estimation of the scaling factor makes no sense, since one cannot expect accurate segmentations without transforming the scale of the templates to that of the vessels. The quantitative results are given in the second part of Table 4. As one can see, the accuracy scores of cross-validation are higher than those of previous techniques in both cases. Moreover, the proposed method trained on 10 images of STARE provides higher accuracy scores on DRIVE than any previous technique trained and tested on DRIVE. Similarly, the accuracy score of the method trained on DRIVE and tested on STARE is much better than that of any previous technique trained and tested on STARE. The accuracy scores of cross-validation are similar to the scores achieved by training and testing on the same databases, indicating that the features we use for the segmentation of the vasculature are robust and the estimations of the blind calibration process are reasonable. One can also observe that both the transformation of the thresholds and that of relative intensities increase the accuracies significantly, without these transformations the results of cross-validation are similar to that of previous techniques. Based on the quantitative results we can conclude that the proposed blind calibration technique is a
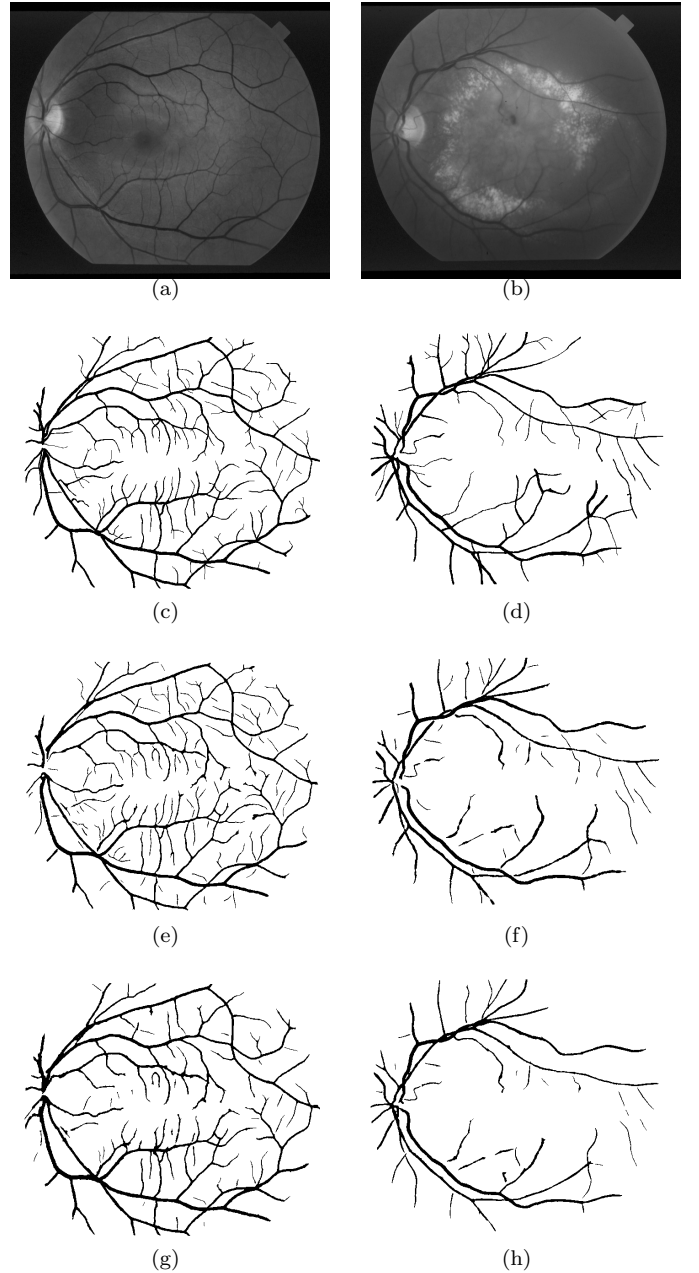

Figure 15: Illustration of the proposed method on the STARE database. Healthy (a), abnormal (b) images; ground truth annotation (c), (d); segmentations by the proposed method trained on database STARE (e), (f); segmentations by the proposed method trained on database DRIVE (g), (h)

reasonable way to increase the accuracy of segmentation when images from various sources are processed.

In Figures 14(g), 14(h), 15(g) and 15(h) one can find some segmentation outcomes of cross-validation and compare them to the results of segmentation when the same database is used for training and testing. Observing the images qualitatively one can assess, that no significant differences can be found.
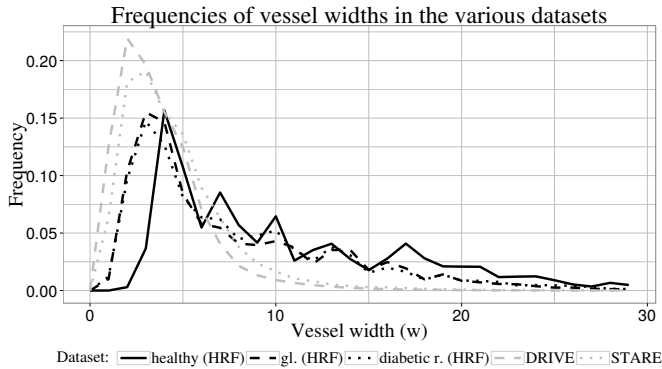
22

Figure 16: Comparison of vessel width distributions measured in DRIVE, STARE and the various classes of HRF

4.8. HRF database

The proposed method was also tested on the recently released database HRF, containing high resolution images. The evaluation was carried out in the spirit of cross-validation, that is, the models trained on DRIVE and STARE were applied to the images of HRF, involving the automated blind calibration process to tune the trained segmentation models. Since HRF has just been released, quantitative comparison is possible only with the method (Odstrcilik et al., 2013, 2009) in Table 5. The first observation is that regarding the mean accuracy and specificity, the proposed method results significantly higher scores than the cited technique in each class of images, independently from the database the segmentation model is trained on. One can also observe that after all steps of calibration are carried out, the accuracy, sensitivity and specificity scores are highly similar within the various groups of images. Considering that the estimation of factors $\hat{f}^{th}$ and $\hat{f}^{rel}$ is carried out for the images individually, these results confirm the robustness of the proposed blind calibration technique.

In the case of DR and glaucomatous images the blind calibration of the model trained on STARE does not increase the accuracy scores remarkably. Contrarily, when the model trained on DRIVE is used for segmentation, the calibration has a significant effect on the accuracy scores of DR and glaucomatous images. This tendency can be interpreted easily by observing the estimated transformation factors in tables 2 and 3: the model trained on STARE is only slightly changed when images of HRF database are segmented, but the model trained on DRIVE is transformed heavily by the estimated transformation factors. Nevertheless, according to the expectations, the calibration of both models increases the accuracy in the case of DR and glaucomatous images.

Interestingly, the performance measures computed in the class of healthy images seem to be contradictory. Although the transformation factors estimated for healthy images are similar to those of glaucomatous and DR images, the blind calibration decreases the accuracy of seg-

mentation. The reason of this unexpected behavior was found in the manual annotations of healthy images. Particularly, we have found that the manual annotations of healthy images are generally thicker than that of DR and glaucomatous images, and this observation is also confirmed by the extremely high specificity and low sensitivity scores achieved in the healthy group of images. In order to show that the annotations of healthy images are highly different from that of glaucomatous and DR images, and also differ from the annotations of the commonly accepted databases DRIVE and STARE, two observations are discussed in details:

1. Contrarily to the smooth distribution of vessel widths measured in the manually annotated images of DRIVE, STARE and that of the glaucomatous and DR images in HRF, in the distribution of vessel widths measured in the healthy images of HRF unexpectedly sharp peaks can be observed at the integer widths 4, 7 and 10 pixels (see Figure 16).

2. Comparing the distribution of vessel widths measured in the manual annotations of healthy, glaucomatous and DR images of HRF, the distribution corresponding to healthy images is shifted to the right by approximately 2 pixels. Considering that the images of HRF have the same resolution, and supposing that the various diseases appearing the sets of glaucomatous and DR images do not decrease the width of vessels in the same manner, we can assess that the annotations of healthy images significantly differ from that of the other images of HRF.

Based on these observations, we believe, that the proposed calibration technique works for the healthy images of HRF, as well, and the decreasing accuracy observed in the case of healthy images can be explained by the uncommon manual annotations. Nevertheless, the results of segmentation involving the complete calibration process still highly majorate the results of (Odstrcilik et al., 2009) in the case of healthy images, as well.

In order to illustrate the results of the proposed method, some images of HRF and the corresponding manual and automated segmentations are visualized in Figure 17.

4.9. Microaneurysm candidate selection

In this subsection we give an insight into how the proper segmentation of the vasculature can improve the detection of microaneurysms. Most of the techniques proposed for the detection of microaneurysms operate in two main steps: first, microaneurysm candidates are extracted, then, high dimensional feature vectors are used to describe them and supervised machine learning based approaches are applied to classify the candidates. In this experiment we investigate the effect of the proposed method on candidate extraction. We have implemented the widely used candidate extraction technique of (Spencer et al., 1996), and applied it to images of the Retinopathy Online Challenge (ROC) database (Niemeijer et al., 2010b). The sensitivity of candidate extraction became 69%. We have applied the

Table 5: Comparison of results on the database HRF. The highest accuracy scores in the various blocks are highlighted by boldface typesetting

	Method	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP	ACC	SE	SP
		Healthy images			Diabetic retinopathy			Glaucomatous images			All images		
	Odstrcilik et al. (2009) (HRF)	0.9539	0.7861	0.9750	0.9445	0.7436	0.9619	0.9497	0.7900	0.9638	0.9494	0.7741	0.9669
Tr. on DRIVE	The proposed method with scale estimation only	**0.9692**	0.8632	0.9811	0.9545	0.8577	0.9620	0.9592	0.8706	0.9661	0.9609	0.8638	0.9696
Tr. on DRIVE	The proposed method with the estimation of the scale and threshold factors	0.9684	0.7973	0.9876	0.9584	0.7951	0.9711	0.9643	0.8013	0.9770	0.9637	0.7978	0.9785
Tr. on DRIVE	The proposed method with complete calibration	0.9690	0.7387	0.9947	**0.9645**	0.7641	0.9800	**0.9700**	0.7602	0.9865	**0.9678**	0.7525	0.9890
Tr. on STARE	The proposed method with scale estimation only	**0.9715**	0.7910	0.9917	0.9630	0.8034	0.9754	0.9677	0.8158	0.9795	0.9674	0.8020	0.9821
Tr. on STARE	The proposed method with the estimation of the scale and threshold factors	0.9698	0.7604	0.9932	0.9631	0.7881	0.9766	0.9684	0.7697	0.9840	0.9671	0.7713	0.9845
Tr. on STARE	The proposed method with complete calibration	0.9685	0.7343	0.9947	**0.9638**	0.7787	0.9781	**0.9700**	0.7438	0.9877	**0.9674**	0.7502	0.9868

Table 6: Execution times when the proposed method is applied to the test images of database DRIVE

Steps of the proposed method	Execution time (avg. ± std. dev.)
Step 1: Preprocessing	2 ± 0.4 s
Step 2: Template matching	215 ± 2 s
Step 3: Contour reconstruction	4 ± 1 s
Step 4: Addition of thin vessels	4 ± 1 s
Overall	225 ± 5 s

proposed technique for the segmentation of the vasculature, and used the binary mask of vessels to decrease the amount of false positive candidate detections. Without Step 4 (the addition of thin vessels), the number of false positive candidates decreased by 38%, and with Step 4 the number of false positive candidates decreased by 39%. The sensitivity remained 69% in both cases. The number of false positive candidates removed by the segmentation of thin vessels is 356. Although the improvement is relatively small, the results show that the additional segmentation of thin vessels can improve the detection of microaneurysms.

4.10. Implementation issues and execution times

The proposed method was implemented in C++. The average execution times reported in Table 6 are measured using one core of an i7-4770K CPU at frequency 3.5 GHz and processing the test images of database DRIVE to make the results comparable with the ones reported in (Lam et al., 2010). As expected, template matching gives the highest portion of computational demands, however, the overall execution time is still reasonable for practical applications.

There are several ways to decrease the execution time of the proposed method, we give a short overview of possibilities below. On the one hand, each step of the proposed method fits the requirements of data parallelism, thus, the execution times can be decreased linearly with the number of processors involved in the computations. On the other

hand, the ROIs can be reduced by excluding the pixels being unlikely to belong to the vasculature (like pixels having local environments with small standard deviation). We also expect that further operators could be removed from the reduced set of trained weak segmentation operators to the slight detriment of segmentation accuracy. Since the template matching step is not expected to give accurate results regarding the contours of vessels, template matching can be performed at low scales, keeping the execution times also low. Finally, we note that the computationally expensive template matching step perfectly fits the nature of GPU (already available in mobile devices) and FPGA computing, since an extremely simple operation (the computation of PCC coefficients) is performed many times, thus, an efficient GPU or FPGA implementation of template matching could decrease the execution times significantly.

5. Summary

Our main results and contributions to the field can be summarized as follows:

1. Many of the previously published techniques are based on probabilistic pixel-wise segmentation frameworks, particularly, some local quantitative features are extracted and advanced machine learning methods are used to differentiate the feature vectors of vessels from that of background pixels (e.g. Hu's invariant features are combined with k-nearest-neighbor classifiers in (Marín et al., 2011); multi-scale Gabor-filter responses are used with ensemble classifiers in (Fraz et al., 2012a); line operators are combined with SVM in (Ricci and Perfetti, 2006); vesselness and ridge features are used with AdaBoost classifier in (Lupascu et al., 2010)). However, statistical learning techniques can hardly learn the visual features of structures being under-represented in the training database, like relatively thin and low contrast vessels or the borders of the optic disc and bright lesions. The novel idea
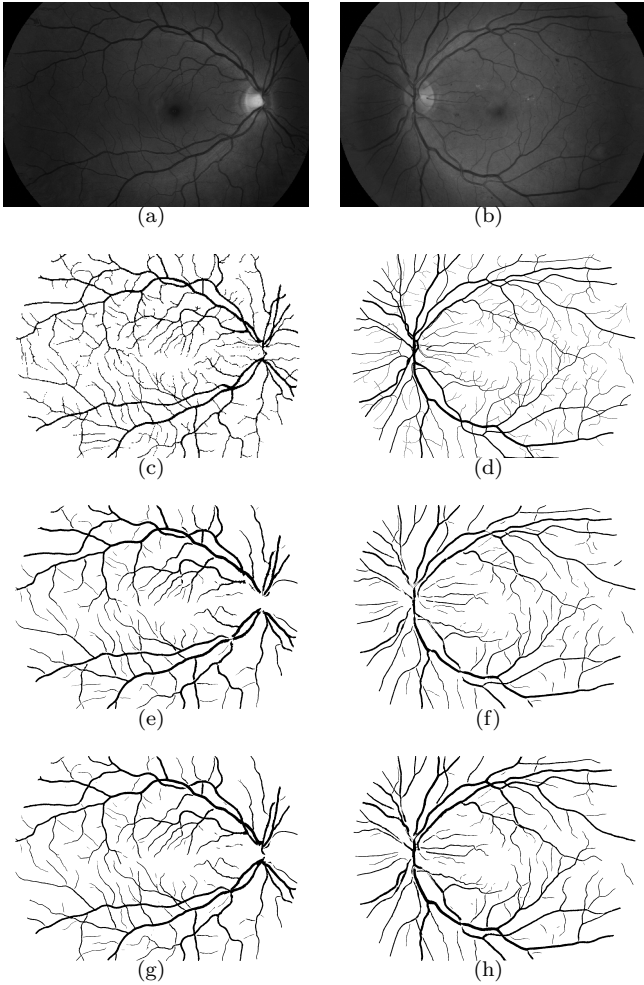
Figure 17: Illustration of the results of proposed method on the database HRF. Images from the classes *healthy* and *diabetic retinopathy* (a), (b); manual ground truth annotations (c), (d); segmentations by the model trained on DRIVE (e), (f); segmentations by the model trained on STARE (g), (h)

of the proposed method is to train dedicated segmentation operators for the various vessel structures appearing in the training database, each one being responsible for the segmentation of one specific kind of vessel only. In this way, one segmentation operator has to recognize only one kind of vessel profile, which is easier than differentiating all possible vessel structures from all undesired structures by one decision boundary as carried out in many previous techniques. Consequently, the deteriorating effects of under-represented structures is also reduced, since under-represented vessel structures also have dedicated segmentation operators. The segmentation operators we introduce are based on template matching using a generalized form of the Gabor function enabling a more accurate modelling of the intensity distribution of vessels than conventional Gabor filters.

2. We have found that the relation of the manually annotated vessel contours and the image content varies with the width and contrast of the vessels. Accordingly, we have developed a method to reconstruct the characteristics of vessel contours measured in the training databases to improve segmentation accuracy. We note that similar observation was made in (Lupascu et al., 2013), but no vessel segmentation algorithm has utilized this feature explicitly so far.

3. The proposed method has a dedicated step for the segmentation of relatively thin vessels having poor contrast and fading into the background.

4. As the test results will show, the accuracy scores of the proposed method on databases DRIVE and STARE are higher than that of any previous technique.

5. We have developed a self-calibration technique which can be used to transform a trained model to retinal images of different resolution, field-of-view (FOV), noise level, etc. Based on this self-calibration technique we have carried out a cross-validation of the trained segmentation models involving the databases DRIVE, STARE and HRF, and found that the accuracy scores still outperform the scores of previous techniques being trained and tested on the same databases.

6. There are many methods published for the segmentation of anatomical parts like the optic disc, or specific lesions, like exudates in retinal images. Most of these methods provide the best accuracy if the content of the image to be segmented has similar scale as the images used for training the segmentation models. In practice, retinal images taken by different devices have different spatial resolution and field-of-view. The most accurate segmentation results can be reached if the unseen images are transformed to the scale of the images used for training. The estimations based on the ratio of the linear extents of images (or that of the FOVs) are biased, since they do not take into consideration the differences in the solid angles of the FOVs: clearly, the content (like the width of the thickest vessels) of images of the same size but differing FOV can highly differ in scale. We have found that the proposed self-calibration technique can be used to estimate the differences in scales accurately, thus, it can be used to improve the accuracy of other methods by automatically scaling the unseen input images to the scale of images used to train the segmentation models.

7. We have given a detailed ranking of many published vessel segmentation techniques, taking into account that the accuracy scores depend on the number of background pixels used for evaluation.

6. Discussion and Conclusions

We have proposed a novel approach for the segmentation of vessels in retinal images: since the center lines of vessels have more characteristic visual features than the vessel walls, we have separated the extraction of pixels near the center lines and the reconstruction of vessel contours; and we have also taken into consideration that some kinds of vessels may be under-represented in the training

25

databases. Considering the special characteristics of low contrast thin vessels, we have also made an attempt to extract some of these vessels in the last step of the method.

The proposed method uses the trained weak segmentation operators and the extracted relative intensity surfaces to represent the visual features of vessels. Although there are some numerical constants wired in the method, most of them do not affect the segmentation directly, instead, they are related to how various quantities (like the width of vessels or relative intensities) are measured. Changing these constants coherently throughout the method does not affect the quality of the outcome remarkably. Aside from these technical constants and supposing that the initial shape descriptor parameters of weak segmentation operators cover the reasonable part of the parameter space, the proposed method has only four free numerical parameters that have to be trained using the images of a training database.

We have developed a method for the blind calibration of trained segmentation models to improve the accuracy when images from various sources are to be processed. The proposed calibration and segmentation technique was evaluated in cross-validation tests among various databases. The corresponding results clearly show the robustness of the proposed method, validate the use of blind calibration to improve the accuracy of segmentation, and confirm that the features used to model the vasculature are characteristic visual properties of retinal vessels, independently from their sources.

The idea of blind calibration – especially the blind estimation of the scaling factor based on the vasculature – can be used in other problems of retinal image processing to improve the accuracy of segmentation or detection by fitting trained models to the characteristics of images from various sources.

It is worth noting that the performance of the proposed method could be further increased by involving weak segmentation operators with templates of other intensity distributions than the ones derived from the generalized Gabor functions. For example, Hermite polynomials can model the intensity profile of arteries (having bright stripes at the center lines due to the high level of oxygen) more accurately than exponential or trigonometric functions (Lupascu et al., 2013). Consequently, the use of weak segmentation operators with templates based on Hermite polynomials could improve the modelling of the vasculature and increase the accuracy of segmentation.

Considering that the accuracy of the proposed method outperformed all the previous techniques in every test case, we can conclude that the method can become a robust and reliable building block of retinal image processing systems.

Acknowledgment

References

Al-Diri, B., Hunter, A., 2009. Automated measurements of retinal bifurcations, in: Proc. of IFMBE World Congress on Medical Physics and Biomedical Engineering, pp. 205–208.

Al-Diri, B., Hunter, A., Steel, D., 2009. An active contour model for segmenting and measuring retinal vessels. IEEE Trans MEd Imaging 28(9), 1488–1497.

Al-Rawi, M., Karajeh, H., 2007. Genetic algorithm matched filter optimization for automated detection of blood vessels from digital retinal images. Comput. Methods Programs Biomed. 87, 248–253.

Al-Rawi, M., Qutaishat, M., Arrar, M., 2007. An improved matched filter for blood vessel detection of digital retinal images. Comput. Biol. Med. 37, 262–267.

Azemin, M.Z., Kumar, D.K., Wong, T.Y., Wang, J.J., Mitchell, P., Kawasaki, R., Wu, H., 2012. Age-related rarefaction in the fractal dimension of retinal vessel. Neurobiology of Aging 33(1), 194.e1–194.e4.

Bhulyan, A., Kawasaki, R., Sasaki, M., Lamoreux, E., Ramamohanarao, K., Guymer, R., Wong, T.Y., Yogesan, K., 2013. Drusen detection and quantification for early identification of age related macular degeneration using color fundus images. J. Clin Exp Ophthalmol 4(5), 1–6.

Bhulyan, A., Nath, B., Chua, J., Kotagirl, R., 2007. Blood vessel segmentation from color retinal images using unsupervised texture classification, in: International Conference on Image Processing.

Can, A., Shen, H., Turner, J.N., Tanenbaum, H.L., Roysam, B., 1999. Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms. IEEE Trans. Inform. Technol. Biomed. 3, 125–138.

Chapman, N., Dell'omo, G., Sartini, M.S., Witt, N., Hughes, A., Thom, S., Pedrinelli, R., 2002. Peripheral vascular disease is associated with abnormal arteriolar diameter reslationships at bifurcations in the human retina. Clin Sci (Lond). 103(2), 111–116.

Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., Goldbaum, M., 1989. Detection of blood vessels in retinal images using two-dimensional matched filters. IEEE Trans. Med. Imag. 8, 263–269.

Cheung, C.S., Butty, Z., Tehrani, N.N., Lam, W.C., 2012. Computer-assisted image analysis of temporal retinal vessel width and tortuosity in retinopathy of prematurity for the assessment of disease severity and treatment outcome. J AAPOS 16(5), 431–436.

Chutatape, O., Zheng, L., Krishman, S., 1998. Retinal blood vessel detection and tracking by matched gaussian and kalman filters, in: IEEE Int. Conf. Eng. Biol. Soc., pp. 3144–3149.

Cinsdikici, M.G., Aydin, D., 2009. Detection of blood vessels in ophthalmoscope images using mf/ant (matched filter/ant colony algorithm. Comput. Methods Programs Biomed. 96, 85–95.

De, J., Li, H., Cheng, L., 2014. Tracing retinal vessel trees by transductive inference. BMC Bioinform. 15.

Deriington, A.M., Lennie, P., 1984. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. J. Physiol. 357.

Dorner, G.T., Polska, E., Garhfer, G., Zawinka, C., Frank, B., Schmetterer, L., 2002. Calculation of the diameter of the central retinal artery from noninvasive measurements in humans. Current Eye Research 25(6), 341–345.

Espona, L., Carreira, M.J., Ortega, M., Penedo, M.G., 2007. A snake for retinal vessel segmentation. Pattern Recognit. and Image Anal. 4478, 178–185.

Fleming, A.D., Philip, S., Goatman, K.A., Olson, J.A., Sharp, P.F., 2006. Autmoated microaneurysm detection using local contrast

normalization and local vessel detection. IEEE Trans Med Imaging 25(9), 1223–1232.

Frangi, A., Niessen, W.J., Vincken, K., Viergever, M., 1998. Multiscale vessel enhancement filtering, in: in Proc. 1st MICCAI, pp. 130–137.

Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., 2012a. An ensemble classification-based approach applied to retinal blood vessel segmentation. IEEE Trans. Biomed. Eng. 59(9), 2538–2548.

Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A., 2012b. Blood vessel segmentation methodologies in retinal images – a survey. Comput. Methods Programs Biomed. 108, 407–433.

Gang, L., Chutatpe, O., Krishnan, S.M., 2002. Detection and measurement of retinal vessels in fundus images using amplitude modified second-order gaussian filter. IEEE Trans. Biomed. Eng. 49, 168–172.

Gardner, G.G., Keating, D., Williamson, T.H., Elliott, A.T., 1996. Automatic detection of diabetic retinopathy using an artificial neural network: A screening tool. Br. J. Ophthalmol. 80, 940–944.

Habib, M.S., Al-Diri, B., Hunter, A., Steel, D.H.W., 2014. The association between retinal vascular geometry changes and diabetic retinopathy and their role in prediction of progression – an exploratory study. BMC Ophthalmology 14(89), 1–11.

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., Dougherty, E.R., 2010. Small-sample precision of roc-related estimates. Bioinformatics 26(6), 822–830.

Hel-Or, Y., Hel-Or, H., David, E., 2014. Matching by tone mapping: Photometric invariant template matching. IEEE Trans. on Pattern Analysis and Machine Intelligence 36(2), 317–330.

Hoover, A., Goldbaum, M., 2003. Locating the optic nerve in a retinal image using the fuzzy convergence of blood vessels. IEEE Trans Med Imaging 22(8), 951–958.

Hoover, A., Kouznetsova, V., Goldbaum, M., 2000. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imag. 19, 203–210.

Jiang, X., Mojon, D., 2003. Adaptive local thresholding by verification based multithreshold probing with application to vessel detection in retinal images. IEEE Trans. Pattern Anal. Mach. Intell. 25, 131–137.

Kawasaki, R., Azemin, M.Z.C., Kumar, D.K., Tan, A.G., Liew, G., Wong, T.Y., Mitchell, P., Wang, J.J., 2011. Fractal dimension of the retinal vasculature and risk of stroke: a nested case-control study. Neurology 76, 1766–1767.

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Sci. 220, 671–680.

Kovács, G., Hajdu, A., 2011. Extraction of vascular system in retina images using averaged one-dependence estimators and orientation estimation in hidden markov random fields, in: IEEE Symposium on Biomedical Imaging: From Nano to Macro, pp. 693–696.

Lam, B.S.Y., Gao, Y., Liew, A.W.C., 2010. General retinal vessel segmentation using regularization-based multi-concavity modeling. IEEE Trans. Med. Imag. 29(7), 1369–1381.

Lam, B.S.Y., Yan, H., 2008. A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields. IEEE Trans. Med. Imag. 27, 237–246.

Li, Q., You, J., Jinghua, W., Wong, A., 2010. A fully automated system for retinal vessel tortuosity diagnosis using scale dependent vessel tracing and grading, in: Proc. of IEEE 23rd International Symposium on Computer-Based Medical Systems (CGMS), pp. 221–225.

Li, W., 1990. Mutual information functions versus correlation functions. J. Stat. Phys 60(5-6), 823–837.

Liu, I., Sun, Y., 1993. Recursive tracking of vascular networks in angiograms based on the detection-deletion scheme. IEEE Trans. Med. Imag. 12, 334–341.

Lobo, J.M., Jlmnez-Valverde, A., Real, R., 2008. Auc: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17, 145–151.

Lupascu, C.A., Tegolo, D., Trucco, E., 2010. Fabc: Retinal vessel segmentation using adaboost. IEEE Trans. Inf. Technol. in Biomed. 14(5), 1267–1274.

Lupascu, C.A., Tegolo, D., Trucco, E., 2013. Accurate estimation of retinal vessel width using bagged decision trees and an extended multiresolution hermite model. Medical Image Analysis 17, 1164–1180.

Marín, D., Aquino, A., Gegndez-Arias, M.E., Bravo, J.M., 2011. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. IEEE Trans. Med. Imag. , 146–158.

Martinez-Perez, M.E., Hughes, A.D., Thom, S.A., Bharath, A.A., Parker, K.H., 2007. Segmentation of blood vessels from red-free and fluorescein retinal images. Med. Imag. Anal. 11, 47–61.

McInerney, T., Terzopoulos, D., 2000. T-snakes: Topology adaptive snakes. Med. Imag. Anal. 4, 73–91.

Mendonça, A.M., Campilho, A., 2006. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. IEEE Trans. Med. Imag. 25, 1200–1213.

Moghimirad, E., Rezatofighi, S.H., Soltanian-Zadeh, H., 2012. Retinal vessel segmentation using a multi-scale medialness function. Comput. Biol. Med. 42(1), 50–60.

Nguyen, U.T.V., Bhuiyan, A., Park, L.A.F., Ramamohanarao, K., 2013. An effective retinal blood vessel segmentation method using multi-scale line detection. Pattern Recognit. 46, 703–715.

Niemeijer, M., van Ginneken, B., Abrmoff, M.D., 2010a. Automatic determination of the artery vein ratio in retinal images, pp. 76240I–76240I–10.

Niemeijer, M., van Ginneken, B., Cree, M.J., Mizutani, A., Quellec, G., Snchez, C.I., Zhang, B., Hornero, R., Lamard, M., Muramatsu, C., Wu, X., Cazuguel, G., You, J., Mayo, A., Li, Q., Hatanaka, Y., Cochener, B., Roux, C., Karray, F., Garca, M., Fujita, H., Abrmoff, M.D., 2010b. Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs. IEEE Trans. Med. Imaging 29, 185–195.

Niemeijer, M., Staal, J., v. Ginneken, B., Loog, M., Abramoff, M.D., Fitzpatrick, J., Sonka, M., 2004. Comparative study of retinal vessel segmentation methods on a new publicly available database. SPIE Med. Imag. 5370, 648–656.

Odstrcilik, J., Jan, J., Kolar, R., Gazarek, J., 2009. Improvement of vessel segmentation by matched filtering in colour retinal images, in: IFMBE Proc. World Cong. Med. Phys. Biomed. Eng., pp. 327–330.

Odstrcilik, J., Kolar, R., Budai, A., Hornegger, J., Jan, J., Gazarek, J., Kubena, T., Cernosek, P., Svoboda, O., Angelopoulou, E., 2013. Retinal vessel segmentation by improved matched filtering: evaluation on a new high-resolution fundus image database. IET Imag. Proc. 7(4), 373–383.

Orlando, J.I., Blaschko, M., 2014. Learning fully-connected crfs for blood vessel segmentation in retinal images, in: Proc. of Medical Image Computing and Computer Assisted Intervention (MICCAI) 2014.

Pakter, H.M., Ferlin, E., Fuchs, S.C., Maestri, M.K., Moraes, R.S., Nunes, G., Moreira, L.B., Gus, M., Fuchs, F.D., 2005. Measuring arteriolar-to-venus ratio in retinal photography of patients with hypertension: development and application of a new semi-automated method. Am J Hypertens 18(3), 417–421.

Ricci, E., Perfetti, R., 2006. Retinal blood vessel segmentation using line operators and support vector classification. IEEE Trans. Med. Imag. 26, 1357–1365.

Sinthanayothin, C., Boyce, J.F., Cook, H.L., Williamson, T.H., 1999. Automated localisation of the optic disc, fovea and retinal blooc vessels from digital colour fundus images. Br. J. Ophthalmol. 83, 902–910.

Soares, J.V.B., Leandro, J.J.G., Jr, R.M.C., Jelinek, H.F., Cree, M.J., 2004. Retinal vessel segmentation using the 2d gabor wavelet and supervised classification. IEEE Trans. Med. Imag. 25, 1214–1222.

Sodi, A., Guarducci, M., Vauthier, L., Ioannidis, A.S., Pitz, S., Abbruzzese, G., Sofi, F., Mecocci, A., Miele, A., Menchini, U., 2013. Computer assisted evaluation of retinal vessels tortuosity in fabry disease. Acta Ophthalmologica 91(2), 113–119.

Sopharak, A., Uyyanonvara, B., Barman, S., 2013. Simple hybrid

method for fine microaneurysm detection from non-dilated diabetic retinopathy retinal images. Computerized Medical Imaging and Graphics 37(5-6), 394–402.

Spencer, T., Olson, J.A., McHardy, K.C., Sharp, P.F., Forrester, J.V., 1996. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. Comput Biomed Res 29(4), 284–302.

Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., v. Ginneken, B., 2004. Ridge based vessel segmentation in color images of the retina. IEEE Trans. Med. Imag. 23, 501–509.

Taylor, H.R., Keeffe, J.E., 2001. World blindness: A 21st century perspective. Br. J. Ophthalmol. 85, 261–266.

Tolias, Y.A., Panas, S.M., 1998. A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering. IEEE Trans. Med. Imag. 17, 263–273.

Trucco, E., Ruggeri, A., Karnowski, T., Giancardo, L., Chaum, E., Hubschman, J.P., al Diri, B., Cheung, C.Y., Wong, D., annd G. Lim, M.A., Kumar, D., Burlina, P., Bressler, N.M., Jelinek, H., Meriaudeau, F., Quellec, G., MacGillivray, T., Dhillon, B., 2013. Validating retinal fundus image analysis algorithms: issues and a proposal. Investigative Ophthalmology and Visual Science 54(5), 3546–3559.

Vázquez, S.G., Cancela, B., Barreira, N., Penedo, M.G., Rodríguez-Blanco, M., Seijo, M.P., de Tuero, G.C., Barceló, M.A., Saez, M., 2013. Improving retinal artery and vein classification by means of a minimal path approach. Machine Vision and Applications 24(5), 919–930.

Vlachos, M., Dermatas, E., 2010. Multi-scale retinal vessel segmentation using line tracking. Comp. Med. Imag. Graph. 34, 213–227.

Walter, T., Klein, J.C., 2001. Segmentation of color fundus images of the human retina: Detection of the optic disc and the vascular tree using morphological techniques. Med. Data Anal., ser. Lect. Notes in Comput. Sci. , 282–287.

Wang, L., Bhalerao, A., Wilson, R., 2007. Analysis of retinal vasculature using a multiresolution hermite-gaussian model. IEEE Tran Med Imaging 26(2), 137–152.

Wihandika, R.C., Suciati, N., 2013. Retinal blood vessel segmentation with optic disc pixels exclusion. I. J. Imag., Graph. and Signal Proc. 5, 26–33.

Wild, S., Roglic, G., Green, A., Sicree, R., King, H., 2004. Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030. Diabetes Care 27, 1047–1053.

Xiao, Z., adel, M., Bourennane, S., 2013. Bayesian method with spatial constraint for retinal vessel segmentation. Comput. and Math. Methods in Med. 2013, 9.

Youssif, A.A., Ghalwash, A.Z., Ghoneim, A.S., 2006. Comparative study of contrast enhancement and illumination equalization methods for retinal vasculature segmentation, in: Cairo International Biomedical Engineering Conference (CIBEC).

Yu, C.Y., Liu, C.C., Yu, S.S., 2014. A fovea localization scheme using vessel origin-based parabolic model. Algorithms 7(3), 456–470.

Yureidini, A., Kerrien, E., Cotin, S., 2012. Robust ransac-based blood vessel segmentation, in: Hanor, D.R., Ourselin, S. (Eds.), SPIE Med. Imag., SPIE Press. p. 8314M.

Zana, F., Klein, J.C., 2006. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. IEEE Trans. Med. Imag. 25, 1200–1213.

Zhou, L., Rzeszotarski, M.S., Singerman, L.J., Chokreff, J.M., 1994. The detection and quantification of retinopathy using digital angiograms. IEEE Trans. Med. Imag. 13, 619–626.