A fully automated pipeline of extracting biomarkers to quantify vascular changes in retina-related diseases

Citation for published version (APA):

DOI:
10.1080/21681163.2018.1519851

Document status and date:
Published: 02/11/2019

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
A fully automated pipeline of extracting biomarkers to quantify vascular changes in retina-related diseases

Jiong Zhang, Behdad Dashtbozorg, Fan Huang, Tao Tan & B. M. ter Haar Romeny

To cite this article: Jiong Zhang, Behdad Dashtbozorg, Fan Huang, Tao Tan & B. M. ter Haar Romeny (2019) A fully automated pipeline of extracting biomarkers to quantify vascular changes in retina-related diseases, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 7:5-6, 616-631, DOI: 10.1080/21681163.2018.1519851

To link to this article: https://doi.org/10.1080/21681163.2018.1519851

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 15 Oct 2018.

Submit your article to this journal

Article views: 467

View related articles

View Crossmark data
A fully automated pipeline of extracting biomarkers to quantify vascular changes in retina-related diseases

Jiong Zhang, Behdad Dashtbozorg, Fan Huang, Tao Tan and B. M. ter Haar Romeny

Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands

ABSTRACT

This paper presents an automated system for extracting retinal vascular biomarkers for early detection of diabetes. The proposed retinal vessel enhancement, segmentation, optic disc (OD) and fovea detection algorithms provide fundamental tools for extracting the vascular network within the predefined region of interest. Based on that, the artery/vein classification, vessel width, tortuosity and fractal dimension measurement tools are used to assess a large number of quantitative vascular biomarkers. We evaluate our pipeline module by module against human annotations. The results indicate that our automated system is robust to the localisation of OD and fovea, segmentation of vessels and classification of arteries/veins. The proposed pipeline helps to increase the effectiveness of the biomarkers extraction and analysis for the early diabetes, and therefore, has the large potential of being further incorporated into a computer-aided diagnosis system.

1. Introduction

Diabetes mellitus (DM) is a chronic disease in which the high blood sugar concentration is presented. The world diabetic population is reaching epidemic proportions worldwide, especially in Asia due to fast lifestyle changes and genetic factors. People with all types of diabetes (type 1, type 2 and gestational) are at risk for diabetic retinopathy (DR). The longer a person has diabetes, the higher the risk is. In the late stage of diabetes, medical treatment is no longer effective to cure the disease, while patients can still be controlled from the early diabetes in time (Tuso 2014). Early diabetes is the status of diabetes but before its complications (hypoglycaemia, heart disease, nerve damage and amputation, and vision problems) appear (Yang and Chan 2016). Many (Tuomilehto et al. 2001; World Health Organization 2006) show that prevention of type 2 diabetes mellitus (T2DM) would eliminate a large proportion of the risk for visual loss from DR. Examination of retinas on patients diagnosed with T2DM is therefore performed in order to check if the disease has not yet lead to DR (damage to the retina due to T2DM). Thus, early detection is the key to prevention and successful treatment of these forms of blindness. However, the fact is that many cases still go unnoticed and are not treated in time, especially in rural areas where few medical experts are present.

Many ocular and systemic diseases including T2DM and DR can cause geometrical or pathological changes in the retina. A special property of the retina is that it is one of the only places in the human body where the vascular system can be directly observed. Clinical examination of the retina can be achieved via different techniques, in which the retinal imaging through fundus photography provides a non-invasive way to ophthalmologists for investigating different eye-related and systemic diseases (Abrahamoff et al. 2010) including DR, age-related macular degeneration and glaucoma (Wong et al. 2008; Lim et al. 2012; Amerasinghe et al. 2008). The advantage of retinal imaging is that it provides direct access to the vascular abnormalities and enables further quantitative analysis of the retinal vasculature.

As a major component of the retina, retinal blood vessels include rich geometrical information and can provide important clinical biomarkers to assist ophthalmologists. Vessel calibre/width (Ikram et al. 2006), tortuosity (Han 2012; Bekkers et al. 2015) and fractal dimension (FD; Aliahmad et al. 2014) changes of the retinal arteries and veins have been shown to be important indicators for the assessment of diabetic disease. However, in the literature there is no general consensus about retinal vasculature changes. For example, in Cheung et al. (2012) a positive, and in Sasongko et al. (2011) a negative association of vessel tortuosity with progression towards diabetes is found. Sasongko et al. suggested that this inconsistency could be due to the duration of diabetes of the patients involved: it might be that an increase in tortuosity only occurs after long exposure (over 10 years) to diabetes.

Computer-aided diagnosis (CAD) systems (Doi 2007; Van Ginneken et al. 2001; Giger 2000; Tan et al. 2013; Tan et al. 2012; Kozegar et al. 2017; Tan et al. 2015; Somasundaram and Alli 2017; Niemeijer et al. 2010; Welikala et al. 2014; Sintnanayothin et al. 2002; Tavakoli et al. 2013; Antal and Hajdu 2014; Pourreza-Shahri et al. 2014) have been developed for different diagnosis purposes in medical imaging field. Due to the importance of the retinal vasculature, different computer-assisted platforms like the Integrative Vessel Analysis (IVAN) (Niemeijer et al. 2011), the Singapore I Vessel Assessment (Ng et al. 2014) and the RetinaCAD (Dashtbozorg et al. 2014a) were developed for the assessment of vascular changes. These softwares apply different retinal vessel analysis methods to identify the status of diabetes. Currently, the demands of early CAD systems for automatic screening programmes in different eye-
related disease conditions are still growing. Thus, advanced retinal vessel analysis techniques need to be developed to set up more complete and reliable CAD infrastructures.

In this paper, our main focus is to set up a fully automatic system of processing retinal images and extracting retinal vascular biomarkers for further analysis on retina-related disease like type-2 diabetes and DR. The whole pipeline includes locating optic disc (OD) and fovea, vessel segmentation, artery/vein classification and biomarker extraction related to vessel width, tortuosity and FD measurements. This is the first time that all our proposed high-performance algorithms have been combined and adapted to a complete pipeline for retinal vessel analysis. In addition, this is also the first time that all the vessel biomarkers are calculated for arteries and veins separately using the proposed infrastructure. The proposed pipeline tools provide a comprehensive description about the automatic extraction of important geometric and topological biomarkers using our state-of-the-art vessel processing and analysis techniques. Hence, they can produce a large amount of retinal vascular biomarkers which reflect many of the possible vascular changes. The whole system implies high potential and reliability of being used to process and analyse large-scale clinical data sets. In Section 2, we introduce the infrastructure and the methodologies for retinal vessel analysis and biomarkers extraction. In Section 3.1, different data sets for evaluating our system will be described firstly. Afterwards, quantitative analysis will be separately performed on the automated results of each module, and we show that our automated pipeline tools are more effective compared to other state-of-the-art methods. We will discuss the results and finally conclude this paper in Section 4.

2. Methodology
In the framework of the RetintaCheck project (Dashtbozorg et al. 2016a), we have developed several retinal image analysis tools including automatic retinal vessel enhancement (Zhang et al. 2016b; AbbasiSureshjani et al. 2015), segmentation (Zhang et al. 2015, 2016a), OD/fovea detection (Dashtbozorg et al. 2016b), artery/vein classification (Huang et al. 2017a, 2018), calibre calculation (Huang et al. 2017b), vessel curvature measurement (Bekkers et al. 2015) and fractal analysis (Huang et al. 2015, 2016) to obtain important vessel biomarkers like central retinal arterial equivalent (CRAE), central retinal venous equivalent (CRVE), the arterial-venous diameter ratio (AVR), vessel tortuosity and FD.

In Figure 1, we illustrate how these modules interact with each other. All of these modules and the computed biomarkers are explained in the following subsections. With the availability of a large amount of biomarkers, we identify critical biomarkers with statistical analysis.

2.1. Vessel enhancement and segmentation
Retinal vascular analysis requires a well-extracted vessel tree from the original image \( f \). In the context of large-scale screening programmes, we need an efficient and accurate vessel segmentation algorithm to assist ophthalmologists. Here, we employ two different filters, i.e. the multi-scale left-invariant derivative (LID) filters and the locally adaptive derivative (LAD) filters in orientation scores for the enhancement and segmentation of blood vessels (Zhang et al. 2015, 2016a). The second-order multi-scale LID filters on the orientation score is denoted by

\[
\Phi_{\eta, s} (U_f) := -\mu^{-2} \frac{\partial^2}{\partial \eta^2} G_{\alpha_s} U_f, \tag{1}
\]

where the LID frame of reference is defined by

\[
\{ \partial_x, \partial_\eta, \partial_y \} := \{ \cos \theta \partial_x + \sin \theta \partial_y, \cos \theta \partial_x - \sin \theta \partial_y \}, \tag{2}
\]

and the conversion factor \( \mu \) with physical unit 1/length is also used here to balance the orientation and spatial direction that have different units, and where \( \alpha_s > 0 \) and \( \alpha_\eta > 0 \) provide the 2D spatial scale \( \frac{1}{2} \sigma_s^2 \) and 1D angular scale \( \frac{1}{2} \sigma_\eta^2 \) of the Gaussian kernel. The rotating derivatives are taken in the \( \eta \) directions that are perpendicular to the vessel structures at their corresponding
orientation planes $\theta$. Disentangled vessel segments in the 3D orientation scores $U_f$ are enhanced by the rotating filters with proper scale samples $S$, and afterwards the 2D enhanced vessel map $Y(f)(x)$ is obtained by taking the maximum filter response over all orientations per position $x = (x, y)$. The final image reconstruction from the multi-scale filtered orientation scores is written as

$$Y(f)(x) := \max_{\theta \in R^{\{1, \ldots, No\}}} \left\{ \sum_{\sigma \in S} \Phi_{b, \text{norm}}(U_f)(x, \theta) \right\}, \quad (3)$$

where $No$ represents the number of orientations with $\theta = i\frac{\pi}{No}$.

The LID frame is defined globally for each orientation plane rather than locally for each pixel in the score domain. Therefore, the LID filters are not always aligned perfectly with all local orientations. This inflexibility makes it difficult to achieve an optimal vessel enhancement with a limited number of orientation samplings $No$. Following the theory of exponential curve fit as described in Zhang et al. (2016a), we set up the LAD frame $\{\partial_x, \partial_y, \partial_\theta\}$ which is adaptive to each position and orientation of the data. In our final vessel enhancement pipeline, the vessels are enhanced via second-order derivatives $\partial_\theta$ (in the direction perpendicular to the vessels). An important efficiency consideration here is that, once the left-invariant Hessian $\mathcal{H}U_f$ and the $\{\partial_x, \partial_y, \partial_\theta\}$ frame are computed, the $\partial_\theta^2$ derivatives can be efficiently computed by projecting the Hessian matrix onto the direction vector $e_\theta$ via

$$\partial_\theta^2 U_f = e_\theta^T (\mathcal{H}U_f) e_\theta, \quad (4)$$

where $\mathcal{H}U_f$ represents the Hessian matrix in the orientation score domain which is computed via the LID frame $\{\partial_x, \partial_y, \partial_\theta\}$:

$$\mathcal{H}U_f := \begin{pmatrix} \partial_x^2 U_f & \partial_x \partial_y U_f & \partial_x \partial_\theta U_f \\ \partial_y \partial_x U_f & \partial_y^2 U_f & \partial_y \partial_\theta U_f \\ \partial_\theta \partial_x U_f & \partial_\theta \partial_y U_f & \partial_\theta^2 U_f \end{pmatrix}. \quad (5)$$

Compared with the LID filter, the newly proposed LAD provides perfect alignment to local structures, and therefore it is more robust to local orientation changes, and even allows to use a small number of orientations. The scale-normalised LAD filter can be written as

$$\Phi_{b, \text{norm}}^\sigma(U_f) := -\mu^{-2} \partial_\theta^2 G_{\sigma, \theta_0} * U_f. \quad (6)$$

The tubular structure enhancement via the locally adaptive frame is illustrated in Figure 2. The final image reconstruction from the multi-scale filtered orientation scores is given by

$$Y(f)(x) := \max_{\theta \in R^{\{1, \ldots, No\}}} \left\{ \sum_{\sigma \in S} \Phi_{b, \text{norm}}^\sigma(U_f)(x, \theta) \right\}. \quad (7)$$

By defining a proper threshold value on the enhanced image, the binary vascular map is finally obtained. Both of the LID and LAD filters are validated on the DRIVE and STARE data sets (Staal et al. 2004; Hoover et al. 2000) with different image types, on which competitive segmentation performances are achieved.

### 2.2. OD and fovea detection

In retinal image analysis, the OD and fovea locations are important landmarks to decide the protocolised region of interest (ROI) for the measurement of vascular geometry changes, such as the FD, tortuosity, CRAE, CRVE and central AVR. A predefined special ROI is able to provide consistent and reliable measurements for later biomarker analysis. In this paper, we rely on the automatic OD and fovea detection technique proposed by Dashbozorg et al. (2016b), where a new convergence index operator, called super-elliptical filter (SEF) is presented. Furthermore, a set-up for the simultaneous localisation of the OD and fovea is introduced, in which the detection result of one landmark facilitates the detection of the other one. The response of the PSEF for pixel $(x, y)$ can be formulated as

$$PSEF(x, y) = \begin{cases} \text{SEF}_{\text{OD}}(x, y) \quad \text{if } x \leq \frac{M}{2} \text{ and } y \leq \frac{M}{2}, \\ \text{SEF}_{\text{FC}}(p, q) \text{ if } x \geq \frac{M}{2}, \\ \text{SEF}_{\text{FC}}(p, q) \text{ if } x < \frac{M}{2}, \end{cases} \quad (8)$$
where \( p, q \in \mathbb{N} \), the \( \text{SEF}_{\text{OD}}(x, y) \) and the \( \text{SEF}_{\text{FC}}(p, q) \) are the pair of filters for the detection of OD and fovea centralis and \( M \) represents the width of the input image. If \( x \geq M/2 \), the fovea is on the left side of the OD; otherwise the fovea is on the right side. For the \( \text{SEF}_{\text{OD}}(x, y) \) filter, the inner and outer limits of the support region are set as \( 0.8\text{r}_{\text{OD}} \) and \( 1.2\text{r}_{\text{OD}} \), while these limits for the \( \text{SEF}_{\text{FC}}(p, q) \) filter are set as \( 0.5\text{r}_{\text{OD}} \) and \( \text{r}_{\text{OD}} \). The number of radial lines, \( N \), is set to an optimal value of 32 for both filters and the band widths of \( 0.1\text{r}_{\text{OD}} \) and \( 0.2\text{r}_{\text{OD}} \) are used for the \( \text{SEF}_{\text{OD}}(x, y) \) and the \( \text{SEF}_{\text{FC}}(p, q) \), respectively. Note that the \( \text{SEF}_{\text{FC}} \) filter is applied on the inverted green channel since the fovea usually appears darker than the background in retinal fundus images. In Figure 3, we show a typical example of accurate OD and fovea detection based on the above filtering-based approach. The paired SEF approach has been evaluated on the MESSIDOR data set (MESSIDOR 2014) and achieves success rates of 99.75% and 98.87% for the OD and fovea detection, respectively. The ROI is defined as ring sector centred on OD centre within 2–5 disc radius \( r_{\text{OD}} \) from the OD margin.

2.3. Artery/vein classification

The retinal vasculature can be categorised into arteries and veins while the retinal arteries and veins behave differently under pathological conditions. Their geometrical changes are respectively considered as signs of several diseases. Therefore it is important to define and study biomarkers separately for arteries and veins. In Figure 4, we show the A/V pixel-wise classification results on an example image using different feature subsets based on our previously developed approach (Huang et al. 2017a) in the RetinaCheck group.

In our vessel analysis pipeline, a new supervised approach (Huang et al. 2018) is employed to classify the vessels into arteries and veins. In this method, we first obtain the segmented vessel pixels from the vessel segmentation method in Section 2.1. Then for each pixel, the vessel calibre is roughly measured in order to categorise them into small, medium or large vessel groups. We assign to each pixel a vessel calibre value based on the vessel calibre of the closest centreline point. Calibre of centreline pixels are found via thinning and a distance transform on the binary segmentation. The detection and parameterisation of the OD in Section 2.2 provides a binary mask which can be used to remove the vessels within the OD region. In addition, the centralis position is used in calculating some of the spatial features for the artery/vein classification. For each vessel centreline pixel, we extract in total 455 features including the local intensity of RGB and HSB colour channels; the mean, standard deviation, median, minimum and maximum of the intensities inside small, medium and large circular regions; the intensity values along each vessel centreline; and the intensity inside each vessel segment (as explained in Table 1).

The extracted feature vector with 455 features will result in an extremely high-dimensional feature space. Therefore, a genetic search-based feature selection approach is used to select a subset of features giving the highest classification accuracy. In the training phase, a linear discriminate analysis (LDA) classifier using a 10-fold cross-validation is trained with the optimal subset of features and validated on the testing sets. More details about the experimental settings and the validation results are given in the Section 3.4.

2.4. Vessel width measurement

Vessel width is very important in the clinical study. The changes in vessel calibre directly reflect the change of blood flow viscosity and blood pressure in the vessels. In Figure 5, we provide examples of a healthy subject and a diabetes type 2 subject to show the vessel width-based AVR difference in a specific ROI.

In this paper, we measure the CRAE, CRVE and AVR of the retinal vasculature in a specific ROI. An automatic multi-scale active contour technique (Huang et al. 2017b) is used to compute the retinal vessel calibre in fundus images. The full vessel calibre map is calculated based on the vessel segmentation
results obtained from Section 2.1. Afterwards, an approach similar to the one described in Dashtbozorg et al. (2014b) is applied for the estimation of the three biomarkers within predefined ROIs.

The proposed method for automatic vessel calibre measurement uses the centreline of a segmented blood vessel to initialise a deformable enclosed contour. Then the contour is evolved iteratively to find the best fit to the vessel boundaries. Finally, the vessel calibre is measured by computing the distance from one detected vessel edge to the other one. All the stages are summarised in Figure 6.

In the vessel calibre measurement step, the geodesic active contour model proposed in Caselles et al. (1997) for solving a global optimisation problem is exploited to locate the left and right edges of each vessel segment. An enclosed and deformable contour \( x(t) = (x(t), y(t))(t \in (0, 1)) \) is initialised using the extracted centreline pixels. Afterwards, the surface is iteratively deformed to minimise the energy function:

\[
\text{energy}(x) = \int_{0}^{1} \left( \frac{\partial x}{\partial t} \right)^2 dt + \lambda \int_{0}^{1} |\nabla x|^2 dt
\]

where \( \lambda \) is a user-defined parameter to control the smoothness of the contour.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vessel width</td>
<td>The vessel width estimated at each pixel.</td>
<td>F12</td>
</tr>
<tr>
<td>Spatial coordinate</td>
<td>Euclidean distance to OD and image centre, angle with respect to OD.</td>
<td></td>
</tr>
<tr>
<td>Circular zone A</td>
<td>The mean, std, min, min, med and max of the 11 local intensities within a circular region with radius 0.5*vessel width.</td>
<td>F13 F15</td>
</tr>
<tr>
<td>Circular zone B</td>
<td>The mean, std, min, med and max of the 11 local intensities within a circular region with radius 1.0*vessel width.</td>
<td>F16 F70</td>
</tr>
<tr>
<td>Circular zone C</td>
<td>The mean, std, min, med, max of the 11 local intensities within a circular region with radius 2.0*vessel width.</td>
<td>F126 F180</td>
</tr>
<tr>
<td>Centrelines</td>
<td>The mean, std, min, med and max of the 11 local intensities along every centreline.</td>
<td>F181 F235</td>
</tr>
<tr>
<td>Vessel segments</td>
<td>The mean, std, min, med and max of the 11 local intensities within every vessel segment.</td>
<td>F236 F290</td>
</tr>
<tr>
<td>All centrelines</td>
<td>The mean, std, min, med and max of the 11 local intensities of all centreline pixels.</td>
<td>F291 F345</td>
</tr>
<tr>
<td>All vessel segments</td>
<td>The mean, std, min, med and max of the 11 local intensities of all vessel segment pixels.</td>
<td>F346 F400</td>
</tr>
<tr>
<td>Whole field of view</td>
<td>The mean, std, min, med and max of the 11 local intensities of all pixels inside the field of view.</td>
<td>F401 F455</td>
</tr>
</tbody>
</table>

Table 1. The complete set of features extracted for each centreline pixel.
The vessels within the ROIs are used to compute the CRAE and CRVE values: (a) the widths of selected vessels are measured by the proposed method; (b) the six largest arteries and veins are then selected for the calculation of CRAE, CRVE and AVR values.

\[
E(x) = \int_0^1 \left( E_{\text{int}}(x(t)) + E_{\text{ext}}(x(t)) + E_{\text{con}}(x(t)) \right) dt
\]  

(9)

where \( E_{\text{int}} \) (internal energy) represents the force of the interaction between adjacent control points, which preserves the smoothness of the surface. While \( E_{\text{ext}} \) (external energy) indicates the image gradient which pulls the contour towards vessel boundary, and \( E_{\text{con}} \) (constraint energy) is used as a constraint for the external force. Therefore, at each iteration, the control points follow the contour evolution equation which is written as

\[
\frac{\partial x(t)}{\partial t} = a g(l)(c + \kappa) n + \beta (\nabla g(l) \cdot n)n + y \nabla x(t) \cdot \nabla g(l).
\]  

(10)

where \( I(x,y) \) is the image and \( \nabla I \) gives the first-order Gaussian derivative of \( I \). \( \kappa \) and \( n \) are the Euclidean curvature and the unit normal vector of \( x(t) \), respectively. \( g(l) \) is the speed function given \( \nabla I \) and \( c, a, \beta \) and \( y \) are weighting parameters. Smooth vessel boundaries are obtained after a certain number of iterations for contour evolution.

After removing the control points at the endings of vessel segment after contour evolution, vessel calibre is estimated using the left and right vessel boundaries. For each control point of one boundary, a corresponding nearest point is found on a B-spline interpolated curve of two nearest points on the other boundary. The Euclidean distance between each two points is computed and converted to micrometre (µm) using the physical pixel size of each image. The vessel calibre is measured by averaging the distances after removing outliers with extreme values.

The vessel width-based biomarkers including the CRAE and the CRVE (see Figure 7) are calculated based on the vessel calibres within a specific ROI. The OD centre and diameter are obtained using the technique in Section 2.2. The vessels within the ROIs of 0.5–1.0 disc diameter around the OD centre are selected and the width-based biomarkers are calculated using the Knudtson’s revised formulas (Knudtson et al. 2003). Finally, the AVR value is defined as the ratio between the CRAE and CRVE.

2.5. Vessel tortuosity measurement

The vessel tortuosity is measured based on the local tortuosity (curvature) of blood vessels, which indicates the rate of change in orientation, and which might be increased because of diseases. The common practice for measuring the vessel curvature is based on a lifted vessel segmentation. By finding the vascular skeleton, splitting the skeleton at junction positions, fitting a curve to each segment and parameterisation of these curves, the local curvatures of these segments are obtained. Then we obtain the curvature at each vessel centreline pixel. The vessel tortuosity degree is defined by the ratio between the actual length of a vessel travelling from position \( P1 \) to \( P2 \), and the straight/shortest distance between \( P1 \) and \( P2 \) (Patton et al. 2006).

However, the above technique is limited for accurate tortuosity measurement due to accumulated errors associated with retinal vessel segmentation and imperfect skeletonisation. To avoid propagation of errors in each step of such pipelines, a new alternative local curvature extraction method was proposed (Bekkers et al. 2015). The tortuosity of vascular structure is computed to study its association with type 2 diabetes. For each retinal image, a corresponding curvature map, where the pixel intensity indicates the local curvature value at each pixel, is obtained by using the exponential curvature estimation technique which relies on the theory of exponential curve fits in orientation scores (Bekkers et al. 2015; Zhang et al. 2016a). In orientation scores, exponential curves are circular spirals, which are essentially straight lines with respect to the curved geometry of the \( SE(2) \) rotation and translation space. The best fit is obtained by eigensystem analysis of the Gaussian Hessian in the lifted space and their spatial projections directly define the curvature at each pixel. The curvature values are computed directly from tangent vectors of exponential curves that locally best fit the data.

Figure 7. The vessels within the ROIs are used to compute the CRAE and CRVE values: (a) the widths of selected vessel are measured by the proposed method; (b) the six largest arteries and veins are then selected for the calculation of CRAE, CRVE and AVR values.
Figure 8. Validation of the SE(2) exponential curvature estimation on a typical synthetic image. From left to right: input image (SNR = 1), ground truth colour-coded curvature map, measured curvature map with resp.

\[ \kappa = \frac{c^0 \text{sign}(c^1)}{\sqrt{(c^0)^2 + (c^1)^2}}, \]  

where the coefficients of the tangent vector \( c^1, c^0 \) and \( c^0 \) are expressed in the moving frame of reference \( \{ \partial_1, \partial_2, \partial_3 \} \) (see Section 2.1). Figure 8 shows an example of using our method to calculate the curvature map of an image with curvilinear structures. The visual comparison shows a remarkable agreement of curvature estimation between our SE(2) curvature estimation technique and the ground truth.

After obtaining the curvature map, the curvature-related biomarkers are defined as the average (\( \mu_n \)) and the standard deviation (\( \sigma_n \)) of all the centre pixels of large size vessels (> 6.5 pixels), medium size vessels (4–6.5 pixels), small size vessels (< 4 pixels) and all vessels on the conventional curvature map (Conv) and the exponential curvature map (SE(2)). In addition, the biomarkers are extracted from the arterial tree, venous tree and both type of vessels.

2.6. FD measurement

The concept of FD was initially defined and developed in mathematics. It measures the complexity of self-similar objects that have the same patterns across different scales, e.g., trees, snowflakes and vessel networks. The self-similar property can be described as

\[ N(r) = r^{-F}, \]  

where \( N(r) \) represents some measurements applied on the complicated pattern of the object at a scale \( r \), and \( F \) is the FD that implies how many new similar patterns are observed as the scale changes. For the calculation of the FD, we use the method proposed by Huang et al. (2016). The FD is computed on the binary segmented images using three classic FD measurements that are widely used in the literature. These three measurements are the box dimension (BD), information dimension (ID) and correlation dimension (CD), which measure different properties of the self-similar pattern of the retinal vessel structures, respectively.

In this paper, we extensively estimate the FD of a vascular network by using the BD, ID, CD (Huang et al. 2016), Renyi FD spectrum, Mandelbrot singularity spectrum and the lacunarity (Kinsner 2008). BD is the direct implementation of the Hausdorff dimension in mathematics. The BD is defined as the real number \( D_B \) such that the number \( N(r) \) of balls with radius \( r \) that is needed to cover an object grows with \( (1/r)^{D_B} \) as \( r \to 0 \):

\[ BD = \lim_{r \to 0} \frac{\log N(r)}{\log 1/r}. \]  

So in the image domain, the measurement \( N(r) \) is the number of boxes with side length \( r \) that overlap with the vessel segmentation. ID is defined as:

\[ ID = \lim_{\delta \to 0} \frac{\sum_{i=1}^{N} p_i \log p_i}{\log 1/\delta}, \]  

where \( N \) is the number of boxes with size \( \delta \) overlapped with the object, the numerator \( \sum_{i=1}^{N} p_i \log p_i \) is the first-order Shannon entropy, \( p_i=n_i/M \) is the probability for finding a part of the object in the \( i \)th box, \( M \) is the total mass of it and \( n_i \) is the part of the object in the box. The limit of Equation (15) is estimated as the slope of the regression line of the logarithmic points.

The CD estimates the FD via the relationship between two pixels inside a region. A correlation integral can be approximately expressed in terms of the probability density

\[ C_k = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} Q(r_k - ||x_i - x_j||) \approx \sum_{j=1}^{N} p_k^2, \]  

where \( Q(x) \) is the Heaviside step function, \( x_i \) is the \( i \)th pixel belonging to an object, \( p_k = n_k/M \) is the probability density of the object with mass \( M \) in the \( j \)th box with size \( r_k \). The CD \( D_C \) is defined via the relationship between \( C_k \) and \( r_k \) as \( D_C = \lim_{r_k \to 0} \frac{\log C_k}{\log r_k} \).

The related biomarkers are computed using FDs on the full vascular network and also on the arterial and venous network separately, using the result of artery/vein separation. We also consider the different ROIs as defined in Subsections 2.2 and 2.4 for the statistical analysis. In Figure 9, two examples of a healthy subject and a diabetes type 2 subject are shown and the BD is calculated using box-counting technique on the arterial tree, venous tree and both.
3. Validation and experimental results

3.1. Material

In this work, in total seven (including four established publicly available and three private) colour retinal fundus data sets, MESSIDOR (MESSIDOR 2014), DRIVE (Staal et al. 2004), STARE (Hoover et al. 2000), INSPIRE-AVR (Niemeijer et al. 2011), NIDEK, CANON and TOPCON, are used for evaluating different modules of our proposed pipeline. The best performance of each measurement is labeled as boldface in all tables.

3.1.1. MESSIDOR data set

MESSIDOR data set (MESSIDOR 2014) is the only data set with publicly available ground truth for both the OD and the fovea centres. This data set includes 1200 retinal fundus colour images acquired from a Topcon non-mydriatic retinograph with 45° field of view (FOV). The images have resolutions of 1440 × 960, 2240 × 1488 or 2304 × 1536 pixels. The manually delimited OD boundaries for all 1200 images and the annotations of fovea centres for 1136 images are provided by the University of Huelva which are publicly available (MESSIDOR 2014).

3.1.2. DRIVE data set

DRIVE data set is a public data set provided by Staal et al. (2004). The images are fovea-centred and were acquired by a Canon CR5 non-mydriatic 3CCD camera with a 45° FOV at resolution of 768 × 584 pixels. The data set is originally split into a training set and a testing set, each of which contains 20 images. The manually delimited OD boundaries for all 1200 images and the annotations of fovea centres for 1136 images are provided by the University of Huelva which are publicly available (MESSIDOR 2014).

3.1.3. STARE data set

STARE data set (Hoover et al. 2000) includes 20 colour fundus images with a resolution of 700 × 605 pixels, in which 10 images contain pathologies. Two groups of manually segmented binary maps are provided by two observers. The FOV masks of the STARE data sets are created manually in our work since the FOV boundary in the retinal image is obvious.

3.1.4. INSPIRE-AVR data set

INSPIRE-AVR data set (referred as INSPIRE) is a public data set provided by Niemeijer et al. (2011). It contains 40 OD-centred images at resolution 2392 × 2048, where the vessel centres and the vessel types are labelled by Dashtbozorg et al. (2014a).

3.1.5. NIDEK data set

NIDEK data set consists of 200 retinal images, where 100 are fovea-centred and 100 are OD-centred images, with size of 3744 × 3744 acquired in the Ophthalmology Department of the Academic Hospital Maastricht (AZM) the Maastricht Study (Schram et al. 2014) in the Netherlands. These images were captured using a NIDEK AFC-230 non-mydriatic fundus camera. The blood vessels are segmented by the automatic vessel segmentation described in previous section and the vessel types were labelled by experts using the manual annotation tool in ‘RHINO’ software developed by Dashtbozorg et al. (2017).

3.1.6. CANON data set and TOPCON data set

CANON and TOPCON data sets both contain 60 fovea-centred and 60 OD-centred retinal images. The images were captured by a Topcon NW300 on 12 healthy subjects, and each subject received five acquisitions. The images of Topcon camera have size of 2048 × 1536 pixels. The A/V labels are obtained using the manual annotation tool in ‘RHINO’ software (Dashtbozorg et al. 2017).

3.2. Vessel segmentation

We exploit the following performance measurements: the vessel pixels in the ground truth that are correctly classified as vessels in the segmentation results are denoted as true positives (TP). If they are labelled as non-vessels, they are considered as false negatives (FN). The non-vessel pixels in...
the ground truth that are correctly classified as non-vessels in the segmentation results are denoted as true negatives (TN). If they are predicted as vessels, they are counted as false positives (FP). To make a global performance comparison of the proposed algorithm with state-of-the-art vessel segmentation methods, we evaluate the performance measures: Sensitivity (Se), Specificity (Sp), Accuracy (Acc) given by

\[
Se = \frac{TP}{TP + FN}, \quad Sp = \frac{TN}{TN + FP}, \quad Acc = \frac{TP + TN}{N},
\]

where \(N = TN + TP + FN + FP\). To evaluate the quality of our vessel enhancement, the receiving operator characteristics (ROC) curve is computed with the TP ratio (Se) versus the FP ratio \((1 - Sp)\) with respect to the varying threshold value \(T_r\). The area under the ROC curve (AUC) is calculated to quantify the segmentation performance, where the AUC value 1 represents a perfect segmentation.

Figure 10 shows examples of vessel segmentation results of the DRIVE and STARE data sets. We also show the ground truth images that were annotated by the first human observer as references. We can see that our LAD filter-based segmentation approach is able to preserve most of the small vessel structures, which has great clinical significance for the assessment of biomarkers like FD (complexity of the vasculature) (Patton et al. 2006) and detection of neovascularisation (formation of new microvasculature) (Lee et al. 2013). To compare with the state-of-the-art vessel segmentation algorithms, we evaluate the proposed methods with the performance metrics Se, Sp, Acc and AUC for the DRIVE and STARE data sets as shown in Table 2. The performance measurements, particularly, Acc and AUC values demonstrate that the proposed LAD-OS method outperforms most of the unsupervised and supervised methods on both of the colour fundus image data sets. More detailed comparisons and discussions about the segmentation performance are given in Section 4.

3.3. OD and fovea detection

For the evaluation of the OD localisation technique, a detected position is considered correct if it is inside the manually annotated OD boundary. Table 1 summarises the results for the OD detection, where the PSEF approach provides similar or better performance than most of the state-of-the-art methods. Our results are comparable to those obtained by the OD detection methods presented by Marin et al. (2015), Mendonca et al. (2013) and Lu (2011). These methods as well as our approach fail to detect the OD in only 3 cases out of 1200 images, while the method proposed by Dashtbozorg et al. (2015) fails in only 2 images.

For the evaluation of the fovea detection, an obtained fovea location is considered correct if the Euclidean distance to the location of the manually annotated fovea is less than the OD radius \(r_{OD}\). For further evaluation, three additional distances are also included as criteria. These distances are \(0.25r_{OD}\), \(0.5r_{OD}\) and \(2r_{OD}\) which are also used by Gegundez-Arias et al. (2013) and Aquino (2014). We analyse the performance of our fovea detection results on 1136 images to compare with the results reported in Aquino (2014), Giachetti et al. (2013) and Gegundez-Arias et al. (2013), and we also repeat the experiment for the whole set to compare with other methods where all the 1200 images have been used (Kao et al. 2014; Yu et al. 2011; Niemeijer et al. 2009). As shown in Table 2, the PSEF approach achieves a success rate of 98.87% for the distance less than \(r_{OD}\) which is higher than ones reported by Kao et al. (2014), Niemeijer et al. (2009) and Gegundez-Arias et al. (2013) for all 1200 images. For the set of 1136 images, we obtain a success rate of 99.65% which is relevantly higher than those achieved by Aquino (2014), Giachetti et al. (2013) and Gegundez-Arias et al. (2013). The results presented in Table 2 demonstrate that better performance is achieved by the proposed approach for fovea detection compared to other methods.

Figure 11 demonstrates example results of OD and fovea detection on images which are acquired using different type of fundus cameras. In order to standardise the biomarkers, all values are calculated from the vessels inside specified ROI around the fovea or OD centres (Figure 11). For the fovea-centred images, the regions are defined by discs centred at the fovea location with the radii of 3, 5 and 6 times the OD radius \(r_{OD}\), while for the OD-centred images, the regions are specified as the standard ring areas within 2 to 3, 5 or 6 \(r_{OD}\) from the OD centre. The \(r_{OD}\) is set in accordance to the average of OD radius in the human eye.
Table 2. Segmentation results on the DRIVE and STARE data sets.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Year</th>
<th>DRIVE</th>
<th>STARE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second human observer</td>
<td>2010</td>
<td>0.7760</td>
<td>0.7120</td>
</tr>
<tr>
<td>Zhang et al. (2010)</td>
<td>2010</td>
<td>0.7120</td>
<td>0.9724</td>
</tr>
<tr>
<td>You et al. (2011)</td>
<td>2011</td>
<td>0.7410</td>
<td>0.9751</td>
</tr>
<tr>
<td>Fraz et al. (2012b)</td>
<td>2012</td>
<td>0.7152</td>
<td>0.9759</td>
</tr>
<tr>
<td>Roychowdhury et al. (2015)</td>
<td>2015</td>
<td>0.7395</td>
<td>0.9782</td>
</tr>
<tr>
<td>Azzopardi et al. (2015)</td>
<td>2015</td>
<td>0.7655</td>
<td>0.9704</td>
</tr>
<tr>
<td>Yin et al. (2015)</td>
<td>2015</td>
<td>0.7246</td>
<td>0.9709*</td>
</tr>
<tr>
<td>Proposed LID</td>
<td>2016</td>
<td>0.7473</td>
<td>0.9764</td>
</tr>
<tr>
<td>Proposed LAD</td>
<td>2016</td>
<td>0.7743*</td>
<td>0.9725</td>
</tr>
<tr>
<td>Supervised</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lupa,scu et al. (2010)</td>
<td>2010</td>
<td>0.7200</td>
<td>0.9597*</td>
</tr>
<tr>
<td>Mar’in et al. (2011)</td>
<td>2011</td>
<td>0.7067</td>
<td>0.9452</td>
</tr>
<tr>
<td>Fraz et al. (2012a)</td>
<td>2012</td>
<td>0.7406</td>
<td>0.9480</td>
</tr>
<tr>
<td>Orlando et al. (2016)</td>
<td>2016</td>
<td>0.7897*</td>
<td>0.9684</td>
</tr>
<tr>
<td>Li et al. (2016)</td>
<td>2016</td>
<td>0.7569</td>
<td>0.9527*</td>
</tr>
</tbody>
</table>

*Best values in comparison with the unsupervised methods; **best values in comparison with the supervised methods.

3.4. Artery/vein classification

Retinal images from the DRIVE, INSPIRE, NIDEK, Canon and Topcon data sets are used to evaluate the artery/vein classification method. Each data set is trained and tested individually. Half of the images are used for feature selection and classifier training, and the rest are used for testing. Vessel centrelines including large, medium and tiny size of vessels are extracted, and the features are obtained as described in Section 2.3. Afterwards, the optimal feature subset is found via the genetic-search feature selection technique and a final LDA classifier is trained using all training data with the optimal features. For the test phase, the trained classifier assigns a probability value (between 0 (vein) and 1 (artery)) to each centreline pixel, and a threshold value of 0.5 is used for the A/V label decision. The performance of classification is evaluated by computing the sensitivity (arteries classified correctly), the specificity (veins classified correctly) and the accuracy (the average of sensitivity and specificity).

In Table 3, we show the quantitative evaluations of the classification performance on different data sets. We categorise the data sets into fovea-centred images and OD-centred images, where the DRIVE data set contains only fovea-centred images, the INSPIRE data set contains only OD centred and the rest contain both types. The results are shown in terms of accuracy (Acc), sensitivity (Se), specificity (Sp) and AUC. Figure 12 shows the example of A/V classification results for the DRIVE and INSPIRE data sets separately using the proposed framework. At each row, we show the original retinal image, the A/V ground truth for the vessel centreline, pixel-wise classification and segment-wise classification results. The vessels with red and blue colour represent the correctly classified artery and vein, while the yellow colour represents the wrongly classified vessels.

3.5. Vessel width measurement

For the vessel width-based biomarkers measurement, we take the system error as a metric to show the robustness of each tool by using the same vasculature acquired in multiple acquisitions. Here, we employ five acquisitions on 15 subjects to compute the errors by taking the average of the relative errors (REs) (mean/standard deviation) of them. The results are shown in Table 4. We can see that the Vampire annotation tool produces the largest variation as expected among the three tools, which is two times
larger than the other two tools. The calibres obtained by manual vessel annotation are clearly prone to human error.

In Figure 13, the Bland–Altman plots are provided to show the comparison between our method and the Vampire annotation tool by taking the IVAN software as a reference. The CRAE values measured using our method show better agreement with the values obtained from the IVAN software, where it presents a lower bias than the Vampire. Both of the tools achieve similar performance with approximately zero bias for the CRVE measurement, though the error of the Vampire is lower than our method.

For measuring AVR, which is an important clinical relevant biomarker in large-scale setting, our fully automatic method produces much accurate results than the human annotation tool, with lower bias and variation. For more details about the validation, we refer to Huang et al. (2017b).

### 3.6. Curvature measurement

In Figure 14, we provide two typical cases to show the curvature-based vessel tortuosity difference between a healthy subject and a diabetes type 2 subject.

Tortuosity measures $\mu_{\kappa}$ and $\sigma_{\kappa}$ were computed on images of the MESSIDOR data set (MESSIDOR 2014), consisting of 1200 images of diabetes patients which are graded for DR: R0 (no retinopathy), R1, R2 and R3 (severe retinopathy). All images are made with 45° FOV cameras, however, with varying image resolutions. In order to have approximately the same physical pixel size in all images, they are cropped and resized such that the FOV area spans a width of 1024 pixels. Table 5 shows the distribution of feature values for different subgroups of the MESSIDOR data set. We observe an increase in both $\mu_{\kappa}$ and $\sigma_{\kappa}$ with increasing severity grading of DR. Based on a Mann–Whitney U test (p-values reported in Table 6), we conclude that all subgroups show a significant increase in $\mu_{\kappa}$ and $\sigma_{\kappa}$ in comparison to the corresponding base groups (R0). For more details and validations, we refer Bekkers et al. (2015).

### 3.7. FD measurement

To validate our measurements, we perform stability analysis of our fractal methods in terms of the choice of manual annotations, various ROI and different imaging modalities. The RE

![Figure 12. A/V classification results from the DRIVE and INSPIRE data sets. From left to right: the original images, the A/V label of the vessel centrelines, the pixel-wise classification and the segment-wise classification. Correctly classified arteries are in red, correctly classified veins are in blue and the wrongly classified vessels are in yellow.](image)

Table 4. Comparisons of the success rate of the proposed framework for OD detection on all 1200 MESSIDOR images.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Success rates</th>
<th>Number of fails</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSEF Dashbozorg et al. (2016b)</td>
<td>99.75%</td>
<td>3</td>
</tr>
<tr>
<td>Marin et al. (2015)</td>
<td>99.75%</td>
<td>3</td>
</tr>
<tr>
<td>Dashbozorg et al. (2015)</td>
<td>99.83%</td>
<td>2</td>
</tr>
<tr>
<td>Bekkers et al. (2014a)</td>
<td>99.50%</td>
<td>6</td>
</tr>
<tr>
<td>Giachetti et al. (2013)</td>
<td>99.67%</td>
<td>4</td>
</tr>
<tr>
<td>Mendonça et al. (2013)</td>
<td>99.75%</td>
<td>3</td>
</tr>
<tr>
<td>Aquino et al. (2012)</td>
<td>98.83%</td>
<td>14</td>
</tr>
<tr>
<td>Yu et al. (2012)</td>
<td>99.08%</td>
<td>11</td>
</tr>
<tr>
<td>Shijian (2011)</td>
<td>99.75%</td>
<td>3</td>
</tr>
</tbody>
</table>

![Figure 13. The Bland–Altman plots for comparing the (a) CRAE, (b) CRVE and (c) AVR values obtained by our method and the Vampire tool with the IVAN.](image)
with respect to the binary images annotated by Observer 1 as the reference is used to investigate the variation of FDs. The obtained FD values are compared with the coefficient of variation (relative standard deviations (RSD)) of all subjects in the DRIVE data set, which are 2.3%, 2.1% and 2.0% for BD, ID and CD, respectively. We computed variations of the inter-group and intra-group FD (BD) for the different groups of DR in the MESSIDOR data set. For images with different DR grades, the BD is calculated on the full image and also inside the ROI around the fovea. The averages and RSD of FD values for each separate DR group are shown in Table 7. We can see that the differences between the mean of FD values for different DR groups are small compared to the RSD of each DR group. The average of RSD in the different groups is higher than 2.5%.

The results of multiple one-way ANOVA tests are shown in Table 9. With this test, we study whether a pair of sub-groups have different distributions. In the case of using the full FOV as ROI, there are no significant mean differences between any two groups, except in group pairs R0-R2 and R2-R3. For the circle ROI around the fovea, the mean difference is significant between the R0 and R2 and between the R1 and R2 groups. To validate our FD computation, we compared the FD values that were calculated from reference standards given by two experts within the circular ROI with 5/C2OD. We used the FDs of Observer 1 as reference (Staal et al. 2004). The result is shown in Table 10.

To investigate the associations of different ROIs with the fractal calculation, we also calculate the FD in various circular regions around the fovea centre of the DRIVE ground truth images annotated by Observer 1. The ROI radii are considered as 4 × OD, (ROI1), 5 × OD, (ROI2) and 6 × OD, (ROI3), and ROI3 is used as reference for the RE calculation. The REs of changing the ROI are shown in Table 11. When FDs are calculated in

![Figure 14](image_url) Two typical examples of showing the tortuosity (mean curvature) difference between one healthy subject and one diabetes type 2 subject. Blue indicates veins and red indicates arteries.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Mean ± (STD)</th>
<th>Mean ± (STD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>μ_jj (10^-2)</td>
<td>σ_jj (10^-2)</td>
</tr>
<tr>
<td>R0</td>
<td>1.624 ± (0.120)</td>
<td>2.333 ± (0.134)</td>
</tr>
<tr>
<td>R1</td>
<td>1.657 ± (0.124)</td>
<td>2.365 ± (0.131)</td>
</tr>
<tr>
<td>R2</td>
<td>1.698 ± (0.122)</td>
<td>2.436 ± (0.144)</td>
</tr>
<tr>
<td>R3</td>
<td>1.795 ± (0.160)</td>
<td>2.674 ± (0.235)</td>
</tr>
<tr>
<td>p-Value^</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

^Compared to R0.

Table 7. The mean and standard deviation of FD values (D^2) for different DR grades.

<table>
<thead>
<tr>
<th>DR grade</th>
<th>Number</th>
<th>Mean</th>
<th>SD</th>
<th>RSD</th>
<th>Mean</th>
<th>SD</th>
<th>RSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI: full FOV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R0</td>
<td>546</td>
<td>1.3864</td>
<td>0.0224</td>
<td>2.34%</td>
<td>1.3285</td>
<td>0.0211</td>
<td>2.38%</td>
</tr>
<tr>
<td>R1</td>
<td>153</td>
<td>1.3852</td>
<td>0.0245</td>
<td>2.49%</td>
<td>1.3317</td>
<td>0.0204</td>
<td>2.29%</td>
</tr>
<tr>
<td>R2</td>
<td>247</td>
<td>1.3781</td>
<td>0.0364</td>
<td>2.64%</td>
<td>1.3215</td>
<td>0.0364</td>
<td>2.91%</td>
</tr>
<tr>
<td>R3</td>
<td>254</td>
<td>1.3869</td>
<td>0.0384</td>
<td>2.77%</td>
<td>1.3276</td>
<td>0.0375</td>
<td>2.82%</td>
</tr>
<tr>
<td>Total</td>
<td>1200</td>
<td>1.3846</td>
<td>0.0350</td>
<td>2.52%</td>
<td>1.3273</td>
<td>0.0343</td>
<td>2.59%</td>
</tr>
</tbody>
</table>

| ROI: 5 × OD |
| R0       | 546 | 1.3864 | 0.0224 | 2.34% | 1.3285 | 0.0211 | 2.38% |
| R1       | 153 | 1.3852 | 0.0245 | 2.49% | 1.3317 | 0.0204 | 2.29% |
| R2       | 247 | 1.3781 | 0.0364 | 2.64% | 1.3215 | 0.0364 | 2.91% |
| R3       | 254 | 1.3869 | 0.0384 | 2.77% | 1.3276 | 0.0375 | 2.82% |
| Total    | 1200 | 1.3846 | 0.0350 | 2.52% | 1.3273 | 0.0343 | 2.59% |

SD: standard deviation; RSD: relative standard deviation.
Table 8. The relative error of the CRAE, CRVE and AVR values obtained by the proposed method, IVAN and Vampire tools.

<table>
<thead>
<tr>
<th>Software</th>
<th>CRAE</th>
<th>CRVE</th>
<th>AVR</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Our method</td>
<td>2.84%</td>
<td>2.40%</td>
<td>3.20%</td>
<td>2.81%</td>
</tr>
<tr>
<td>IVAN</td>
<td>2.32%</td>
<td>1.91%</td>
<td>2.65%</td>
<td>2.29%</td>
</tr>
<tr>
<td>Vampire</td>
<td>4.09%</td>
<td>3.63%</td>
<td>5.73%</td>
<td>4.48%</td>
</tr>
</tbody>
</table>

Table 9. Comparison between FD values in different DR groups.

<table>
<thead>
<tr>
<th>DR grade</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>p-Value</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI: full FOV</td>
<td>R0 R1</td>
<td>0.00123</td>
<td>0.00193</td>
<td>0.981</td>
<td>0.0070</td>
</tr>
<tr>
<td>R2</td>
<td>0.00834</td>
<td>0.00267</td>
<td>0.010</td>
<td>0.981</td>
<td>0.0015</td>
</tr>
<tr>
<td>R3</td>
<td>0.00046</td>
<td>0.00625</td>
<td>0.998</td>
<td>0.0073</td>
<td>0.00063</td>
</tr>
<tr>
<td>R1</td>
<td>0.00711</td>
<td>0.00358</td>
<td>0.195</td>
<td>0.9021</td>
<td>0.00163</td>
</tr>
<tr>
<td>R2</td>
<td>0.00169</td>
<td>0.00356</td>
<td>0.965</td>
<td>0.0109</td>
<td>0.0075</td>
</tr>
<tr>
<td>R3</td>
<td>0.00087</td>
<td>0.00311</td>
<td>0.925</td>
<td>0.0168</td>
<td>0.0098</td>
</tr>
<tr>
<td>ROI: 5 × OD</td>
<td>R0</td>
<td>0.00324</td>
<td>0.00313</td>
<td>0.720</td>
<td>0.0113</td>
</tr>
<tr>
<td>R2</td>
<td>0.00026</td>
<td>0.00263</td>
<td>0.040</td>
<td>0.992</td>
<td>0.0002</td>
</tr>
<tr>
<td>R3</td>
<td>0.00086</td>
<td>0.00260</td>
<td>0.987</td>
<td>0.0058</td>
<td>0.0076</td>
</tr>
<tr>
<td>R1</td>
<td>0.01020</td>
<td>0.00352</td>
<td>0.020</td>
<td>0.0011</td>
<td>0.0193</td>
</tr>
<tr>
<td>R2</td>
<td>0.00410</td>
<td>0.00350</td>
<td>0.946</td>
<td>0.0049</td>
<td>0.0131</td>
</tr>
<tr>
<td>R3</td>
<td>0.00610</td>
<td>0.00306</td>
<td>0.191</td>
<td>0.0140</td>
<td>0.0018</td>
</tr>
</tbody>
</table>

*The mean difference is significant at the 0.05 level.
*One-way ANOVA test with null hypothesis that the means of distributions are equal.

Table 10. The comparison of FD between two human observers and the automated vessel segmentation method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Max</th>
<th>MRE</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 1</td>
<td>7.1%</td>
<td>2.0%</td>
<td>0.0585</td>
</tr>
<tr>
<td>Zhang et al. (2015)</td>
<td>7.4%</td>
<td>3.9%</td>
<td>0.4950</td>
</tr>
<tr>
<td>Frangi et al. (1998)</td>
<td>9.3%</td>
<td>4.3%</td>
<td>0.8035</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Max</th>
<th>MRE</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 2</td>
<td>6.7%</td>
<td>1.9%</td>
<td>0.0851</td>
</tr>
<tr>
<td>Zhang et al. (2015)</td>
<td>7.4%</td>
<td>3.8%</td>
<td>0.8506</td>
</tr>
<tr>
<td>Frangi et al. (1998)</td>
<td>9.4%</td>
<td>4.3%</td>
<td>0.8802</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Max</th>
<th>MRE</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer 1</td>
<td>6.2%</td>
<td>1.8%</td>
<td>0.0974</td>
</tr>
<tr>
<td>Zhang et al. (2015)</td>
<td>7.3%</td>
<td>3.8%</td>
<td>0.6919</td>
</tr>
<tr>
<td>Frangi et al. (1998)</td>
<td>9.4%</td>
<td>4.3%</td>
<td>0.6990</td>
</tr>
</tbody>
</table>

*Max: Maximum relative error with respect to Observer 1.
*MRE: Mean of relative error with respect to Observer 1.
*Pearson correlation test with null hypothesis that the correlation coefficient is zero.

level, respectively. The LAD-based vessel segmentation method is an unsupervised technique, which does not rely on any labour-intensive training phase, but is still able to achieve similar performance compared to many of the supervised algorithms as shown in Table 0. In the vessel analysis pipeline, we consider to use the LAD filter-based segmentation method to achieve more accurate segmentation performance.

The OD and fovea locations are required to determine protocolised ROI for the assessment of signs related to vascular changes. In the proposed pipeline, the PSEF set-up is used for the simultaneous detection of OD and fovea. Compared with other techniques, this method does not require retinal blood vessel extraction and it is robust to imaging artefacts and different types of retinal lesions.

Artery/vein classification plays a critical role on follow-up steps of biomarker extraction. The proposed framework achieved an average pixel-wise accuracy of 83% on high-resolution retinal images (more than 3 megapixels) including the INSPIRE, NIDEK, Topcon and Canon data sets. Our module achieved an average accuracy of 92.0%, sensitivity of 89.6% and specificity of 91.3%, which is slightly better than the state-of-art methods. More importantly, our method performs both on both large and small vessels near the OD. These vessels are important in biomarker measurements such as CRVE, CRAE and AVR.

Based on vessel segmentation, an automatic technique for the vessel calibre measurement is designed for further CAD. In Section 3.5, we perform quantitative evaluations for our method and compare it with the semi-automatic tool IVAN and the manual vessel annotation tool Vampire. The experimental result shows the superiority of the proposed automatic vessel calibre measurement. The proposed method is able to provide automatic calibre measurements with a comparable system error and similar CRAE, CRVE measurements to IVAN. Additionally, evaluation of the computational efficiency is also provided to compare with the other two softwares which require time-consuming human intervention particularly for large-scale screening settings. The automatic vessel segmentation, artery/vein classification and vessel calibre measurement steps respectively take 2, 4 and 2 min on a single-core CPU. The calculation time is reduced to less than 1 min per image when a parallel setting with 12 cores and 2.30 GHz CPU is used.

Table 11. The comparison of Dk values using different region of interest.

<table>
<thead>
<tr>
<th>Method</th>
<th>Radius</th>
<th>Max</th>
<th>MRE</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROI1</td>
<td>4 × OD</td>
<td>3.8%</td>
<td>2.4%</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>ROI2</td>
<td>5 × OD</td>
<td>1.0%</td>
<td>0.4%</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>ROI3</td>
<td>6 × OD</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Pearson correlation test with null hypothesis that the population correlation coefficient is zero.
In our proposed pipeline, the vessel tortuosity-based biomarkers are obtained by exploiting the curvature estimations based on the best exponential curve fits in orientation scores, as explained in Section 2.5. Both qualitative examples as shown in Figures 8 and 13 and quantitative evaluations as shown in Table 5 show that our method is a reliable tool for vessel curvature extraction. Evaluation on the clinical MESSIDOR data set shows strong positive correlations of the tortuosity measurements with different stages of DR, which implies high potential of using our method as an important component of the whole retinal vessel analysis pipeline for diabetes and DR diagnosis in large-scale screening settings.

The FD biomarkers in this paper are designed for large-scale retinal screening. It is essential to validate the robustness of the technique before using it in clinical study. Our results show that the classic FDs shall be calculated under very strict conditions, and tiny changes on the images and vessel segmentation can cause significant variations. We also compared the FD values based on the vessel annotations from DRIVE data set by two experts within the circular ROI with $5 \times OD$. We used the FDs of Observer 1 as reference. As shown in Table 6, the main difference between the two manual annotations is the presence of the tiny vessels. We found that missing the tiny vessels does affect the FD. The mean REs of 1.97%, 1.88% and 1.77% and maximal differences of 7.11%, 6.70% and 6.23% are obtained for BD, ID and CD, respectively, which are noticeable compared to the calculated RSDs. Moreover, the results in Table 7 show that FDs calculated in different ROIs are associated with each other, with p-values less than 0.01. This observation means the FDs calculated in different ROIs are significantly associated. The fact that a smaller ROI produces a lower FD in general implies that a fixed ROI is necessary in order to obtain comparable FD values.

In this paper, a fully automatic pipeline including vessel segmentation, the OD and fovea detection, artery and vein classification is applied to extract important vessel biomarkers. These biomarkers fall into three groups: vessel width, curvature (tortuosity) and FD. We have performed quantitative evaluation and analysis on the results from different modules in our pipeline. The validation is based on in total seven data sets. Currently, the needs of CAD systems for automatic screening programmes are still growing. An accurate and reliable feature computation pipeline plays a key role in a CAD framework. The validation results in this work show that the biomarkers/features computed from our pipeline tools are accurate compared to human measurements. Hence, it has the potential of being used in a CAD system to make large-scale screening programmes more effective.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work is part of the NWO-Hé Programme of Innovation Cooperation [No. 629.001.003] and the European Foundation for the Study of Diabetes/Chinese Diabetes Society/Lilly project.

Notes on contributors
Jiong Zhang received the master’s degree in computer science from the Northwest A&F University, Yangling, China, and the Ph.D. degree from the Eindhoven University of Technology, The Netherlands. He then joined as a Post-Doctoral Researcher with the Medical Image Analysis Group, Eindhoven University of Technology, The Netherlands. His research interests include ophthalmologic image analysis, medical image analysis, computer-aided diagnosis, and machine learning.

Behdad Dashtbozorg received the master’s degree from Yazd University, Iran, and the Ph.D. degree in electrical engineering from the University of Porto, Portugal. He is currently a Post-Doctoral Researcher with the Medical Image Analysis Group, Eindhoven University of Technology, The Netherlands. His research interests include medical image analysis, image processing, machine learning, and computer vision.

Fan Huang received the M.Sc. degree in biomedical engineering from the Eindhoven University of Technology, The Netherlands, where he is currently pursuing the Ph.D. degree in biomedical engineering with the Medical Image Analysis Group. His Ph.D. focus is on multiple biomarkers analysis for diabetes retinopathy using retinal images. His research interest is in digital image analysis, machine learning, and pattern recognition.

Tao Tan is currently an assistant professor with Biomedical Image Analysis (BIA) at the Department of Biomedical Engineering at the Eindhoven University of Technology, and a consultant of Qiview Medical Inc. Dr. Tan obtained his PhD from Diagnostic Image Analysis Group, Radiology, Radboud University Medical Center. His main research focuses on ultrasound image analysis including denoising, landmark segmentation, lesion segmentation, registration, computer-aided diagnosis and detection. He has led the development of one FDA approved and one CE approved product in medical imaging. He has served as a committee member, editorial member and reviewer for numerous international conferences and journals in the area of medical imaging.

Bart M. ter Haar Romeny received the M.Sc. degree in applied physics from the Delft University of Technology in 1979 and the Ph.D. degree from Utrecht University in 1983. He is currently a Professor of medical image analysis with Northeastern University, Shenyang, China, and with the Eindhoven University of Technology, The Netherlands. He is also a Project Leader of the RetinaCheck project. His research interests include brain-inspired medical image analysis and computer-aided diagnosis.

ORCID
Behdad Dashtbozorg http://orcid.org/0000-0003-4443-1614

References


