3

MORPHOMETRIC MEASUREMENTS OF THE RETINAL VASCULATURE IN FUNDUS IMAGES WITH VAMPIRE

Emanuele Trucco[1], Andrea Giachetti[2], Lucia Ballerini[1], Devanjali Relan[3], Alessandro Cavinato[1], and Tom MacGillivray[3]

[1] *VAMPIRE/CVIP, School of Computing, University of Dundee, Dundee, United Kingdom*
[2] *Department of Computer Science, University of Verona, Verona, Italy*
[3] *Clinical Research Imaging Centre, University of Edinburgh, Edinburgh, United Kingdom*

Much research is being directed toward investigating links between quantitative characteristics of the retinal vasculature and a variety of outcomes to identify biomarkers. The interest for retinal biomarkers lies in the fact that the retina is easily observed via fundus photography. Outcomes considered for research of biomarkers in the literature include conditions such as diabetes and lacunar stroke, and also cognitive performance and genetic expression [1–5]. The need for measuring large volumes of images, needed to power biomarker discovery studies, makes semiautomatic software systems desirable. This chapter reports recent algorithms developed by the VAMPIRE group for vasculature detection and quantification, including recent developments on landmark detection. We focus on accuracy and validation issues, and, importantly, the conditions for comparing meaningful results from different algorithms. This work is a part of VAMPIRE (Vasculature Assessment and Measurement Platform for Images of the REtina), which is an international collaboration growing a software suite for automatic morphometric measurements of the retinal vasculature.

Biomedical Image Understanding: Methods and Applications, First Edition.
Edited by Joo-Hwee Lim, Sim-Heng Ong, and Wei Xiong.

3.1 INTRODUCTION

Many eye and systemic diseases manifest themselves in the retinal vasculature and developments in retinal image processing (henceforth RIA) and computer-assisted diagnosis offer great potential for retinal fundus imaging to be used in large-scale screening programmes, promising significant resource savings, and freedom from observer bias [4]. An increasing number of RIA algorithms [6] for extracting the vascular structure and measuring different morphological parameters appears in the literature [7].

This chapter reports recent developments within the VAMPIRE project. VAMPIRE is an international collaboration of 10 clinical and image processing centers [8–12]. The VAMPIRE software suite aims to allow efficient quantification of morphological features of the retinal vasculature in large sets of fundus camera images, generating measurements suitable for biomarker discovery. *Biomarkers* are retinal vasculature (in our case) measurements that associate, in a statistical sense, with indicators of specific conditions [5, 13, 14]. Examples using VAMPIRE software include those used in References [15] (lacunar stroke) and [1] (cognitive ageing), and ongoing VAMPIRE studies address sarcopenia, schistosomiasis, and gene expression.

This chapter focuses on three important RIA topics: *vessel width estimation, artery–vein classification*, and *validation*. We aim to capture key concepts and challenges, describe briefly our solutions, and provide a reasonable spectrum of selected, representative further readings.

Any quantitative description of the retinal vasculature requires the location of the vascular network. This image segmentation task yields typically a *binary vessel map*, in which pixels are classified as vessel or not vessel. However, vessel width, an important parameter for clinical investigation [2, 3, 5], cannot be estimated reliably from the raw vessel map, in which width estimates present a relatively high standard deviation due to the jagged (pixelized) vessel contours, hence, the need and relevance of specialized algorithms for accurate width estimation (Section 3.2).

Vessel widths in the example shown vary between 4 and 16 pixels; in general, vessel width in pixel depends on the image resolution and the absolute size of the vessel itself. For reference on the absolute size, vessel diameters are reported by Ikram et al. [2] as between 92 and 238 μm for arterioles, and between 135 and 214 μm for venules. Pixel measurements vary with image resolution, optics, and other factors; in images acquired by state-of-the-art, commercial fundus cameras (approximately 3000 × 3000 pixels, 45° field of view), the largest vessels are up to about 30 pixels wide; the smallest ones discernible a few pixels only. As the relative error (estimated width over true width) of an automatic system increases with width, measurements using clinical studies are normally confined to the largest vessels around the optic disc.

Crucially, the vasculature changes that appear during the onset of a systemic disease *often affect arteries and veins differently*. For example, one of the early signs of retinopathy is generalized arteriolar narrowing in which the arteriolar-to-venular width ratio (AVR) decreases [3, 6]. There is also mounting evidence that narrowed retinal arterioles are associated with long-term risk of hypertension, while AVR is a well-established predictor of stroke and other cardiovascular events [2, 3, 5].

Hence, classifying vessels into arteries and veins is an essential component of RIA systems designed for both screening and computer-assisted diagnosis. This problem is only deceptively simple. We report an algorithm based on supervised learning and quadrant pairing, which proves effective in a simple four-dimensional feature space (Section 3.3).

Validation has been defined as the process of showing that an algorithm performs correctly by comparing its output with a reference standard [16]; that is, *the experimental process by which an RIA system is shown to achieve its purpose* (e.g., locating a lesion, or estimating the width of arteries in a specific region of a fundus color image) *to a certain quantitative extent*, established, for example, by receiver operating characteristic (ROC) analysis or confidence levels of statistical tests [6]. The main difficulty of validation stems from this definition: the need for large amounts of annotated data (ground truth) to guarantee sufficient power to statistical conclusions. There are, however, several other issues; for a detailed discussion, we refer the reader to the recent paper on RIA validation by an international, multicenter group [16]. In this chapter, we summarize the key concepts and challenges of RIA validation (Section 3.4).

3.2 ASSESSING VESSEL WIDTH

3.2.1 Previous Work

Several vessel segmentation algorithms have been reported in recent years for retinal fundus images [7, 12, 17, 18], as well as specialized algorithms to estimate vessel width accurately [19]. We sketch here a representative cross-section of recent papers addressing the topics of this chapter.

Coupled active contours are used in the Extraction of Segment Profiles (ESP) algorithm by al-Diri et al. [20], which simultaneously detects the vasculature and yields width estimates. This algorithm grows a "ribbon of twins" model, which captures pairs of contours and maintains width consistency. The algorithm is initialized with approximate vessels centerlines and the network topology is determined, resolving junction configurations. ESP performs well with blurred edges, close parallel vessels, light reflex, and thin vessels.

A graph-based approach has recently been reported by Xu et al. [21]. The problem of finding the two vessel boundaries is cast here as a two-slice, 3D surface segmentation problem, which is further converted into the problem of computing a minimum closed set in a node-weighted graph. An initial segmentation is generated from a vessel probability image. The authors report a study using the algorithm to explore the relationship between average vessel width and distance from the optic disc in 600 subjects.

Li et al. [22] report an algorithm within a study on AVR (arterio-venous ratio) estimation. Vessel starting points are detected using a matched Gaussian filter. Vessels are then traced by Kalman filtering. A modified Gaussian model is proposed to describe the vessel profile and account for central reflections. The width of a vessel is obtained

by data fitting. The authors report a 97.1% success rate (SR) in the identification of vessel starting points, and a 99.2% SR in the tracking of retinal vessels.

Fiorin et al. [23] improve vessel borders by means of spline interpolation, fitting a cubic spline to each of the two contours of a vessel independently. The authors report tests on a private set of 739 vessels with single-observer width annotations, with excellent correlation (0.97) between ground truth and estimated width. In the algorithm we report here, we too fit splines to vessel contours, but pairs of splines approximating the two contours of each vessel are coupled by a local parallelism constraint to promote locally consistent boundaries.

3.2.2 Our Method

This section focuses on width estimation of vessel contours obtained from binary maps. These maps can be generated by any vessel segmentation algorithms available [24]. Currently, the VAMPIRE software suite [9] employs a version of Soares's algorithm [17]. In essence, this well-known algorithm classifies each pixel as vessel or non-vessels using machine learning. The feature vector includes the pixel intensity and the response to two-dimensional Gabor wavelets at multiple scales. The classifier is a Bayesian classifier with class-conditional likelihoods, described as Gaussian mixtures. Probability distributions are estimated by training with a set of manually annotated images in which the vasculature has been traced (the DRIVE data set [25]). Very good detection performance has been reported given suitable training sets, which, ideally, ought to be sets of images consistent with the ones to be measured for the target application in terms of patient cohort, camera, acquisition protocol, quality, and any other relevant parameter. We are not aware of any *comprehensive* study on the variability of width estimates with different training sets.

Raw binary maps generated by vessel location algorithms identify each pixel as vessel or not. This results in pixelized, hence jagged contours, and in turn in significantly noisy, oscillating width estimates as the vessel is traversed along its centerline. For this reason, width estimation algorithms deploy considerable machinery to achieve accurate estimates (Section 3.2.1).

We present here an alternative approach: we refine the vessel contours of raw binary maps using a simple spline-fitting algorithm to interpolate the two contours of a vessel, augmented with a parallelism constraint promoting locally parallel contours. This simple approach brings about a large accuracy improvement when tested with the main public data set for width estimation, REVIEW [26]. Its relative simplicity, speed, and direct applicability to binary vessel maps constitute the main motivation for its inclusion in this chapter. The algorithm can be summarized in three key steps, as follows.

First, a temporary skeleton is obtained using morphological thinning on the binary mask; branching points are removed and a natural cubic spline is fitted to the thinned centerline. To fit a natural cubic spline to the thinned vessel centerline, we transform the reference frame into the principal directions of the vessel points. This guarantees that the centerline of vessel segments between junctions is well represented as a function mapping each x value to a single y value.

Second, two coupled cubic splines are fitted to the original (jagged) vessel contours. The splines are coupled by imposing a parallelism constraint, promoting locally parallel contours. This yields an overconstrained linear system, which can be written as follows:

$$\begin{cases} y_A = a_i(x - x_{A,i})^3 + b_i(x - x_{A,i})^2 + c_i(x - x_{A,i}) + d_i \\ y_B = \alpha_i(x - x_{B,i})^3 + \beta_i(x - x_{B,i})^2 + \gamma_i(x - x_{B,i}) + \delta_i \\ y'_A(x_{A,i+1}) = y'_B(x_{B,i+1}), \end{cases} \tag{3.1}$$

where the spline knots, $(x_{A,i}, y_{A,i})$ and $(x_{B,i}, y_{B,i})$, are n pairs of coupled contour points, and the last equation is the parallelism constraint. This system is overconstrained by the parallelism constraint at knots, and can be solved by least squares. We use singular value decomposition, which allows one to control easily the conditioning number of the system, hence the stability of the solution [27]. Our results suggest that more complex least-squares solving algorithms [28] are unnecessary.

Third and finally, given a point C_j lying on the spline-smoothed centerline, the vessel width w_j at C_j is estimated by the Euclidean distance between points D_j and E_j, intersections of the two refined contours with the direction, d_j, orthogonal to the centerline at C_j (see Fig. 3.1) [10].

3.2.3 Results

We tested our constrained spline fit, following the current literature, using the public standard REVIEW database [26]. The four REVIEW image sets offer a representative spectrum of vessel appearance in fundus images: high-resolution (HRIS data set), central light reflex (CLRIS data set), vascular diseases (VDIS data set), and kick-points (KPIS data set). Three experts (observers O_1, O_2, and O_3) marked manually vessel edge points and the average of the three width estimates is considered as the ground truth width ψ_i. REVIEW contains 5066 profiles. For comparison of different algorithms, the error χ_i is defined as $\chi_i = w_i - \psi_i$, where w_i is the width at the

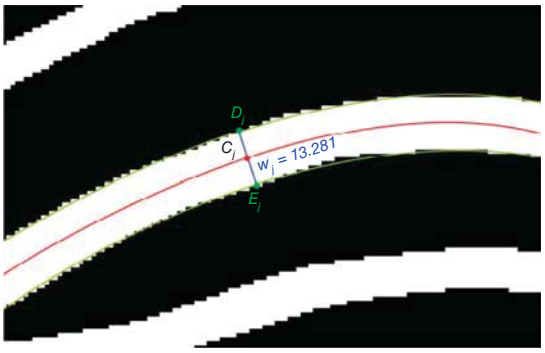

Figure 3.1 Vessel width at C_j estimated as the Euclidean distance w_j between D_j and E_j.

*i*th location estimated by the algorithm under examination, and ψ_i is the reference measurement (ground truth or different algorithm). The standard deviation of the error, σ_χ, is used to evaluate algorithm performance and considered more important than the mean [20]. A further useful parameter for performance evaluation is the SR, that is, the number of meaningful measurements returned by the algorithm over the total number of measurements attempted.

Table 3.1 reports the performance of our method and its comparison with two recent algorithms: ESP procedure [20] and Xu's graph-based method [21]. It must be noted that our method, as presented and tested, is not meant to and does not incorporate any way to detect and discard locations where binary maps may lead to severely inaccurate measurements, which other methods include. This is the reason behind the lower SR than other methods in the table. Considering accuracy and SR together, the overall performance achieved on the four REVIEW data sets by our simple algorithm is comparable, and sometimes better than that of specialized, sophisticated width estimation algorithms. For instance, with the HRIS data set our method has an accuracy comparable to the observers: $\sigma_\chi = 0.760$ pixels (2.75 times the mean of observers' σ_χ); Xu's graph-based method and ESP algorithm perform slightly better. With the CLRIS data set, our method yields the best accuracy, despite the presence of strong central light reflexes; the SR is lower than ESP and Xu's algorithms. With the noisy VDIS data set, the width estimates obtained by our method are the second best after ESP. Finally, with the KPIS data set, our method again provides the best accuracy, but it scores a SR that is slightly that is lower than that of other algorithms.

The performance of the proposed algorithm depends on the quality of the input binary mask. As our algorithm does not include any way to detect noisy locations in the vessel map, we have manually avoided such locations in our evaluation. Problem locations include missing or extremely noisy vessel segments, for which width estimates would be meaningless. To limit the effect of such errors in a completely automatic system, one can compare estimates with priors (ranges of expected values), and check consistency with several width estimates taken around the target position [29].

3.2.4 Discussion

Fully automatic systems estimating vessel width with the same accuracy as that of semiautomatic ones, in which estimates are revised and possibly corrected by an operator, are still outstanding. However, state-of-the-art estimates can be achieved with moderately complex algorithms, and we have presented one. It seems, however, plausible that current systems may be deployed in some cases. For instance, it seems reasonable to characterize the performance of a fully automatic system according to vessel width, as done, for instance, in Reference [29], and identify the range of widths leading to the minimum errors. If SR and accuracy are sufficiently high *in that range*, and the range is interesting for clinical purposes, the system could be used. We stress that, in general, the same error and accuracy figures may be acceptable or not; for instance, measurements directly used within a diagnostic process will require higher accuracy than measurements used in preliminary biomarker discovery studies.

TABLE 3.1 Performance Comparison of the Width Measurement Methods on the REVIEW Database

Method	Measurement		Error		SR%	Measurement		Error		SR%
	μ	σ	μ_χ	σ_χ		μ	σ	μ_χ	σ_χ	
	HRIS					CLRIS				
First observer: O_1	4.12	1.25	−0.23	0.288	100	13.19	4.01	−0.61	0.567	100
Second observer: O_2	4.35	1.35	0.002	0.256	100	13.69	4.22	−0.11	0.698	100
Third observer: O_3	4.58	1.26	0.23	0.285	100	14.52	4.26	0.72	0.566	100
Ground truth: O	4.35	1.26	—	—	100	13.80	4.12	—	—	100
ESP [20]	4.63	—	0.28	0.420	99.7	15.7	—	−1.90	1.469	93.0
Graph [21]	4.56	1.30	0.21	0.567	100	14.05	4.47	0.08	1.78	94.1
Proposed method	**3.93**	**1.40**	**−0.42**	**0.760**	**95.7**	**13.81**	**3.68**	**−0.16**	**1.229**	**90.2**
	VDIS					KPIS				
First observer: O_1	8.50	2.54	−0.35	0.543	100	7.97	0.47	0.45	0.233	100
Second observer: O_2	8.91	2.69	0.06	0.621	100	7.60	0.42	0.08	0.213	100
Third observer: O_3	9.15	2.67	0.30	0.669	100	7.00	0.52	−0.53	0.234	100
Ground truth: O	8.85	2.57	—	—	100	7.52	0.42	—	—	100
ESP [20]	8.80	—	−0.05	0.766	99.6	6.56	—	−0.96	0.328	100
Graph [21]	8.35	3.00	−0.53	1.43	96.0	6.38	0.59	−1.14	0.67	99.4
Proposed method	**8.17**	**2.82**	**−0.79**	**1.381**	**92.1**	**6.06**	**0.28**	**−1.32**	**0.319**	**93.9**

3.3 ARTERY OR VEIN?

3.3.1 Previous Work

Arteries and veins in fundus images are differentiated by visual inspection by several features. Following Reference [30]:

- Arteries appear brighter than veins.
- Arteries are normally thinner than neighboring veins.
- The central reflex (the light reflex of the inner parts of the vessels shown in Fig. 3.3) is usually wider in arteries and smaller in veins.
- Arteries and veins usually alternate near the optic disc before branching out; that is, an artery near the optic disc is usually next to two veins, and the other way round.

However, due to the many and variable factors influencing the appearance of fundus retinal images (including instruments, protocols, patient ethnicity, lesions, and image quality), the above-mentioned features cannot provide a complete guarantee. Highly reliable automatic classifiers still prove very difficult to develop.

Color and its representation and color spaces [31] play an important role in A/V classification and, in general, retinal image analysis. It is well known that some tasks such as segmentation and classification may be facilitated by transforming the red green blue (RGB) image into a suitable color space. It is still unclear, however, which space should considered the best one. Common choices include the green channel of the RGB image and the HIS space (hue, saturation, intensity). The green channel of RGB and the so-called red-free images are used often for vessel detection (see for instance, References [6, 25, 29] and references therein). Zamperini et al. [32] report a feature selection investigation to determine the most discriminative color features in RGB space for A/V classification. Yu et al. [33] compute color features in different color spaces simultaneously (RGB, CIElab, and YCbCr) before classifying the vessels. The HSI color space is another common choice, for example, to locate bright features. For instance, Osareh et al. [34] report experiments in which the optic disc is detected more reliably in HSI space, and extend their approach to the location of exudates [35]. Wang et al. [36] report spherical color coordinates, representing brightness and chromaticity, to detect bright lesions. Goatman et al. [37] investigate the effect of three color normalization algorithms to reduce background color variations in the framework of automatic population screening for diabetic retinopathy.

Various authors have reported supervised learning approaches to A/V classification [30, 38–40] using predominantly support vector machines or neural networks. Generating training data from large amounts of images is laborious and time consuming, but supervised vessel classification methods seem to yield higher classification rates than approaches based on manual models, as, for example, [41], where A/V classification is based on a rule-based method to propagate vessel labels through the vascular graph. Grisan and Ruggeri [13] proposed an unsupervised learning method performing quadrantwise A/V classification in a concentric zone around the optic

disc. They used fuzzy *c*-mean clustering on 443 vessels from 35 images, reporting 87.6% correct classification. Saez et al. [42] performed classification again in a concentric zone around the optic disc; quadrants were rotated in steps of 20° to include at least one artery and one vein in each. Classification was performed by *k*-means clustering on 58 images; positive and negative likelihood ratios were (7.2386, 4.2218) and (0.2445, 0.1528) for (arteries, veins), respectively. The authors also reported 87% and 90.08% correct classification before and after applying their vessel tracking method [43]. Joshi et al. [44] applied fuzzy *c*-mean clustering. The vessel tree was separated into a structurally mapped vessel network. The authors excluded centerline pixels that may be similar for both classes, treating them as noisy pixels. The proposed method was applied to 15 retinal color fundus images resulting in a classification accuracy of 88.28%.

3.3.2 Our Solution

3.3.2.1 Image Correction To counteract the effect of noisy variations within and across images, we first compensate for background illumination in the red, green, and hue channels. Background illumination is estimated in each channel using the method described in Reference [45]. Median filtering with a mask of size 100×100 pixels is performed. Then, correction coefficients are calculated by dividing the maximum image intensity by the intensity of each pixel. Finally, the corrected image is achieved by multiplying each channel by its correction coefficients. Fig. 3.2 shows an example. The illumination-corrected image is used to extract features for A/V vessel classification.

3.3.2.2 Centerline Location We then extract centerline pixels from Zone B (region between the blue concentric circles, Fig. 3.3(a)), as follows. Each vessel is tracked between two manually marked points, S and E; see Fig. 3.3(c) and (d). Then, the coordinates of the new point, P_{new}, 5 pixel ahead of S, are calculated (see Fig. 3.3(c)). At P_{new}, the intensity profile across the vessel centerline is computed (Fig. 3.3(b)). Point C, the approximate center of the vessel (red on the intensity profile in Fig. 3.3(b) and (c)) is then located by averaging the two local minima on the profile (green points on profile in Fig. 3.3(b)). Then, a new P_{new} is located (yellow point ahead of P_{new} in Fig. 3.3(c)). This procedure continues until the end point E is reached (blue lines in Fig. 3.3(d)). Next, vessel edges are located on each profile by Canny edge detection [46] (yellow in Fig. 3.3(d)). Finally, centerline pixels (pink in Fig. 3.3(d)) are estimated as the midpoints of pairs of edge points. Centerline pixels are extracted from vessels in each quadrant, yielding a set of n vessel segments $V_1, \ldots, V_n$, where each vessel is represented by the sequence of its centerline pixels. For each of these, we store coordinates and local vessel diameter.

3.3.2.3 Features The image is divided into four quadrants after locating the outer diameter (OD) and its approximate diameter [11]. Then, four color features, mean of red (MR), mean of green (MG), mean of hue (MH), and variance of red (VR) are

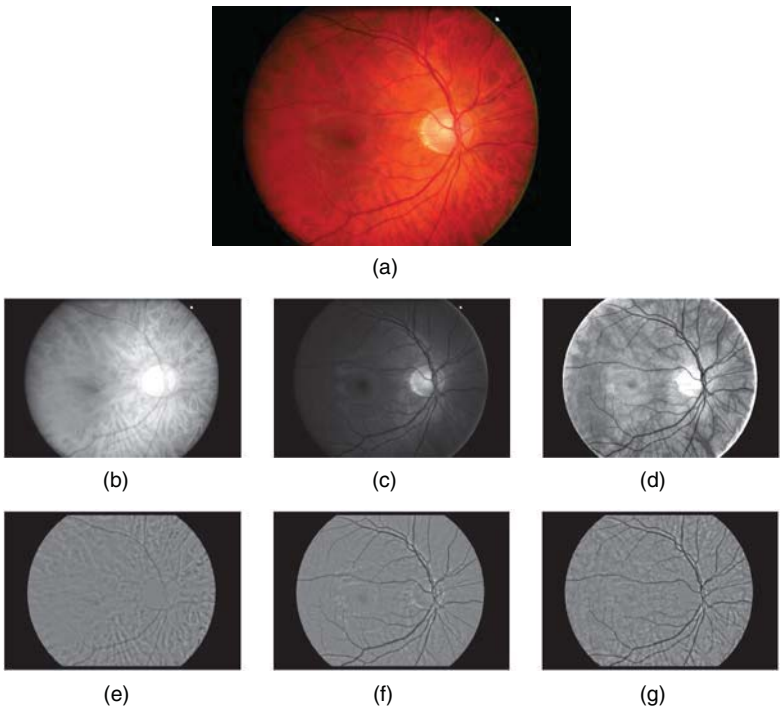

Figure 3.2 (a) Original image; (b), (c), and (d) uncorrected red, green, and contrast-adjusted hue channels; (e), (f), and (g) illumination-corrected channels.

computed, from the corrected channels, within a circular neighborhood around each centerline pixel, with diameter 60% of the mean vessel diameter. This is a small set of highly discriminant color features [13]. This yields four sets of feature vectors for each pair of adjacent quadrants clockwise, (I, II), (II, III), (III, IV), and (IV, I); see Fig. 3.3(a). Each set is represented by a $N_i \times 4$ matrix, where N_i is the number of pixels in a pair of quadrants.

3.3.2.4 Classifier Each feature vector is classified using a Gaussian mixture model-expectation maximization (GMM-EM) classifier. The classification is performed on pairs of adjacent quadrants. GM-EMM classifies the pixels into three clusters: *artery* (A), *vein* (V), and *not labeled* (N). The centroid of each cluster is associated with a vector of four mean values for the four color features. The two average values of the green channel intensity representing the centroids (i.e., for two clusters) are compared to determine the class. The cluster with higher mean green channel intensity at its centroid is labeled *artery* and the other *vein* [44]. As each quadrant is considered twice in all pairs, each pixel gets two labels. To improve the chances of correct classification, the quadrants are then rotated by 45° clockwise (white solid lines in Fig. 3.3(a)) and pixels classified again, generating two more

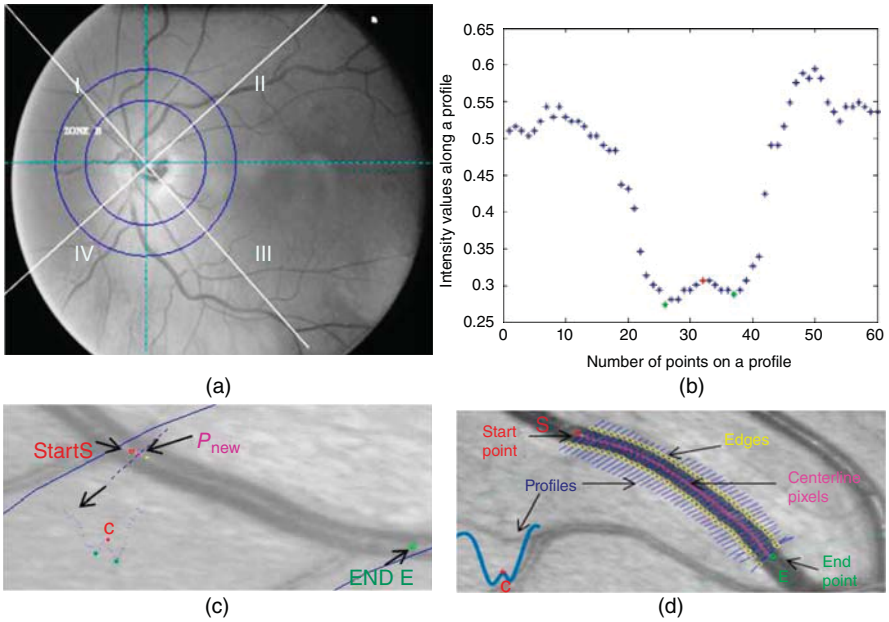

Figure 3.3 (a) Rotated quadrant by 45° (white lines) and measurement zone (Zone B, 0.5–1 disc diameter from the optic disc margin); (b) vessel profile; (c) and (d) centerline extraction. (*See insert for color representation of this figure.*)

labels for each pixel. Vessels are assigned a label based on the maximum number of labels of each kind of their pixels. The final label of each pixel is then decided based on following rules, where $n(x)$ indicates the number of instances of label x:

1. IF $n(A) > n(V)$ assign A.
2. IF $n(V) > n(A)$ assign V.
3. IF ($n(A) = n(V)$) OR ($n(N) \geq (n(A)$ OR $n(V)$)) assign N.

After each pixel has been assigned a final label, the vessel label is given by the pixel label occurring most often within the vessel itself. Table 3.2 shows an example of vessel assignment to A based on labels (maximum number A) in column III. Fig. 3.4 shows an example of an image with vessel classification complete.

3.3.3 Results

We used a total of 406 vessels from 35 color fundus images to test classification. The system did not assign a final label (artery or vein) to 55 vessels (13.5%). A total of 92% of the remaining 351 vessels were assigned an artery or vein label correctly.

Ground truth labels were provided by two trained observers. Observer 1 classified all vessels; Observer 2 did not classify 1.48% of vessels. Various performance

TABLE 3.2 Assigning a Final Label to a Vessel: Example for A

Pixels Belonging to a Vessel I	Four Labels per Pixel II	Final Label to Each Pixel (Artery) III
1	*a a a a*	A
2	*a a v a*	A
3	*a a v a*	A
4	*a a a a*	A
5	*a v v a*	N
6	*a n n a*	N
7	*a n a a*	A

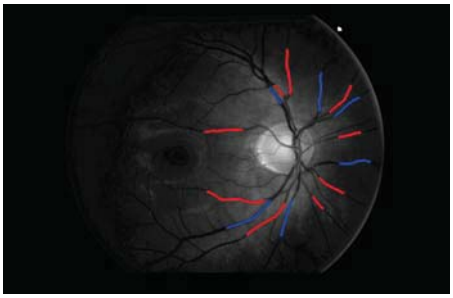

Figure 3.4 Final classification into arteries (red) and veins (blue). (*See insert for color representation of this figure.*)

measures based on ROC analysis were computed separately for arteries and veins. Results based on Observer 1 are given in Table 3.3. The sensitivity was 0.8181 for arteries and 0.7688 for veins, that is, the probability of an incorrect classification was 18.2% for arteries, and 23.1% for veins. The precision (positive predicted value) for both arteries and veins from our methods was higher than 0.8802 and 0.8118 for arteries and veins, respectively, compared to what reported in [42]. Likelihood ratios

TABLE 3.3 Classification Performance Parameters

Performance Measure	Arteries	Veins
Sensitivity	0.8181	0.7688
Specificity	0.8978	0.9590
Positive predicted value	0.9045	0.9408
Negative predicted value	0.8067	0.8307
Positive likelihood ratio	8.0095	18.7933
Negative likelihood ratio	0.2025	0.2410
Classification accuracy	0.8547	0.8719
Classification error rate	0.1453	0.1281

were also high. Results were slightly higher when compared with those for Observer 2. Overall, our approach led to 92% correct classification, compared with 87.6%, 90.08%, and 88.28% reported in References [42–44], respectively. Of course, firm comparative conclusions cannot be drawn given the relatively modest amounts of vessel analyzed and the different data sets used in different papers.

3.3.4 Discussion

Labeling large numbers of vessels manually is a time-consuming task. It is, therefore, important to include (semi-) automated classification with high sensitivity and specificity in a system meant for the analysis of large image data sets, such as VAMPIRE [8, 9]. We have proposed an unsupervised method using color features to classify arteries and veins. Our system does not require the presence of at least one artery and one vein per quadrant, and is applicable even if there is no vessel in a quadrant. At present, centerline pixels were extracted by manually marking the start and end points in a vessel. We note that, to make this algorithm fully automatic requires further refinement to computerize this step. In our tests, classification results were compared to manual labels and our system shows a low false positive rate, that is, (1-specificity) of 0.041 and 0.102 compared to 0.208 and 0.108 reported in Reference [42] for vein and artery, respectively. Furthermore, likelihood ratios, which confirm the high reliability of our proposed system was also greater. Our system results in positive likelihood ratio of (18.793, 8.009) and negative likelihood ratio of (0.2410, 0.2025) as compared to positive likelihood ratio of (4.2218, 7.2386) and negative likelihood ratio of (0.1528, 0.2445) for (veins, arteries), respectively reported in [42]. Also, the percentage of correct classification by our system was higher than those reported in References [13, 43, 44], respectively. We analyzed 35 color fundus images compared to 35, 58, and 15 images analyzed in References [13, 43, 44], respectively. In Reference [44], the images were obtained from diabetic subjects but there is no subject information in Reference [13, 43]. Also, the resolution of our images is 2048×3072, which is greater than 1300×1000, 768×576, and 800×1000 as used in References [13, 43, 44], respectively. Our proposed system has high classification accuracy and low classification error rate for both vein and artery in our test image data set. It should be pointed out that the resolution of our test images is higher than those used to test other classification algorithms reported in the literature and so our classification accuracy may differ with a different data set. Also, the classification result is highly dependent on segmentation results (for extracting centerline pixels). Moreover, choosing a different retinal zone, classifier, and framework would likely impact on the classification performance.

Several systems proposed in the literature adopt different classifiers, and it is difficult to establish which one works best for the problem at hand. We suggest that, for A/V classification, the choice of features, a robust way to compute them, and a suitable representation (feature vector) is more important than the specific classifier *per se*. Clustering algorithms were reported in References [13] (fuzzy *c*-mean clustering), [44] (fuzzy *k*-mean clustering), and [42] (*k*-mean clustering). Narasimha-Iyer et al. [38] tested four classifiers for A/V classification (SVM,

nearest neighbor, five-nearest neighbor, and Fisher linear discriminants) using 251 dual-band images (570 and 600 nm) to provide functional and structural features. The SVM provided the best performance, evaluated in terms of correct classification rate (97% V, 90% A). Niemejier et al. [39] used a different criterion, the area under the ROC curve, in a comparison of four classifiers: linear discriminant analysis, quadratic discriminant analysis, SVM, and *k*-nearest neighbor (kNN). Using 20 DRIVE images, they found the kNN to be the best performer. The diversity of criteria and data sets used makes a fair comparison clearly impossible, let alone deciding which classifier may be the most suitable.

We conclude that our system performance is very promising. Further tests with much larger data sets are needed to declare suitability to support A/V classification in biomarker research. If successful in such tests, the algorithm will become a central component to the VAMPIRE software suite for the automatic quantification of the retinal vasculature in fundus images.

3.4 ARE MY PROGRAM'S MEASUREMENTS ACCURATE?

Validation required three main components [47]: *standardization of validation methodology (protocols), design of public data sets*, and *standardization of validation metrics*. As any medical image processing system needs validation as an integral part of translation toward the clinic, and as there may be significant differences in validation requirements across clinical domains and applications (e.g., therapy, biomarkers, intervention, and screening), the literature is largely fragmented. Validation-specific paper collections include the 2002 and 2006 IEEE TMI special issues [47, 48]; forums include the working group on medical image processing within the European Federation for Medical Informatics, the Validation in Medical Image Processing initiative, and the Quantitative Imaging Network. All maintain easily located websites. An extensive picture of the RIA validation state-of-the-art, including surveys on the main clinical applications (diabetic retinopathy, retinopathy of prematurity, and glaucoma) is given in Reference [16].

The validation of RIA software introduces domain-specific issues, summarized hereafter [49].

1. *Variability of Expert Judgment.* This well-known fact is countered by *multiple-expert annotations*. This requires that variations among experts be characterized quantitatively, but there is no ultimate consensus on how to achieve this. Solutions include, depending on the nature of the variables at hand, averaging, consensus, and interrater reliability metrics such as AC1 or Kappa, Krippendorff α, histograms, and distributions.
2. *Annotation Protocols.* Annotating specific image elements is not performed by clinicians normally. To save clinicians' time, the requirements and protocols of RIA validation and clinical tasks should be aligned as much as possible. This would avoid asking clinicians to annotate explicitly anatomical structures. Protocols used to take photographs represent another source of variability.

3. *Generating Annotations Directly Comparable to Software Output.* As stated, most annotation tasks are not part of normal clinical practice (e.g., estimating accurately the width of blood vessels at many locations in a fundus images). Therefore, researchers have begun to explore alternative paradigms, for example, weak learning methods [50] (moving from algorithm-oriented annotations to the use of clinical notes directly) and STAPLE [51] (addressing the simultaneous reliability estimation of algorithm and reference standard from annotations by multiple experts).
4. *Outcome Point.* It is not always clear where to set the outcome for validation. In screening programs, a refer–no refer decision seems the obvious choice; other cases are not so clearcut.
5. *Physiological Short-Term Changes.* Recent studies have investigated the variation of retinal vessel width with pulse. If significant, taking photographs at random instants in the pulse cycle may result in unrecognized variations in the measurements. No firm conclusions seem possible from the few studies reported so far [52, 53].
6. *Different Imaging Instruments*. Algorithms suitable to one type of image may not be directly usable for a different type, for example, fundus images versus fluorescein angiography. But even within the same class of machines, instrument variations can have a large effect on algorithm's performance.
7. *Data and Image Quality.* Image quality depends on instrument characteristics, acquisition procedure, and target conditions. Quality definitions applied by experts are elusive to quantitative rules. In general, images deemed suitable for clinical analysis may not produce good results with RIA systems.
8. *Data Sets.* Different data sets may lead to somewhat inconsistent performance assessments, as preparation protocols may differ. The design of data sets for RIA validation is a crucial issue. Among the most popular, current public data sets with annotations for RIA, we mention STARE, DRIVE (vasculature detection), REVIEW (vessel width estimation), MESSIDOR, and the diabetic retinopathy online challenge (DR-specific lesion detection); all have easily located web sites. Further public data sets are reported in Reference [16].

The creation of substantial, structured, public data sets built and certified by large groups of RIA researchers and clinicians would be a substantial push toward the development of RIA software tools closer to translation. The international group authoring [16] lists a number of criteria agreed by the authors for the design of public data sets. Here, we present a selection of those criteria, thought for research groups who have access to local clinicians and need to instruct them on how to generate images and annotations to form test data sets for RIA tasks. Such data sets should

1. *include multiple image annotations*, providing the standard reference for comparison for the outcome stated, by as many clinicians as possible; each should ideally annotate the data set multiple times, to estimate intraobserver variability;

2. *be maintained regularly*, to manage distribution, additions, and potential obsolescence of data and annotations;
3. *include as many images as possible*, with the indicative orders of magnitude being hundreds of images for pilot studies, and thousands to tens of thousands for studies proper;
4. *use standardized, patient-friendly imaging protocols* allowing large numbers of patients to be imaged effectively;
5. *include metadata*, that is, nonimage data characterizing imaging instruments, patients, and disease;
6. *be organized by outcome*, which depends on the task at hand, in case multiple tasks are considered.

3.4.1 Discussion

There is no way one can overemphasize the importance of systematic, thorough validation in retinal, and, more generally, medical image analysis. A solid introduction to validation concepts and techniques should, in our opinion, be part of any course on medical image processing. Clinicians collaborating with image processing experts should also be conversant with the principles of validation, so that results and translation feasibility can be discussed meaningfully and on a common good-practice platform. Many statistical tools are shared with data analysis in clinical studies, and validation is the first and fundamental step between algorithm development and translation.

This chapter has focused on three important RIA topics: vessel width estimation, artery–vein classification, and validation. The discussion has been based on tools developed within the VAMPIRE project, but we have tried to give a larger, introductory picture by considering an array of papers representative of the current state-of-the-art in RIA.

The main conclusion, in our view, concerns validation and its role in translation. The obvious and well-posed question posed by novices and experienced researchers alike is, which algorithm works best? This question can only be answered experimentally, and experience and literature alike indicate that experiments on different data sets may lead to different conclusions. The volume of data and the computer power necessary to establish firm conclusions statistically is considerable; having both available in one site is not common. This is a key reason why the creation of publicly available data sets, with substantial numbers of annotated data, is crucial for the development and, ultimately, the real applicability of RIA software in clinical practice. Research groups without access to substantial hospital structures or unable to obtain images and annotations would be able to exercise their software development power in full; above all, prototype algorithms could be tested on internationally recognized data sets, as suggested in Section 3.4.

Finally, it is important to remind the reader, especially the ones with a background in the quantitative sciences and with no experience of translation, that the hurdles to be jumped to deploy RIA software in clinical applications are not purely technical. Obtaining substantial volumes of data from hospitals requires compliance with

a number of procedures (e.g., in the United Kingdom, at least ethics permissions, sponsorship, and Caldicott Guardian approval). Rules and regulations imposed by the relevant national agencies, devised in the interest of patient safety, make the journey from the laboratory to the clinic a long and tortuous one. That said, well-designed, efficient, robust, and properly validated software remains the center of the game.

Acknowledgments

We are indebted to VAMPIRE colleagues for useful discussions and insights, especially Bal Dhillon, Alex Doney, Jean Pierre Hubschman, Carmen Lupascu, Enrico Pellegrini, Ilaria Pieretti, Gavin Robertson, Domenico Tegolo, Peter Wilson, and Kris Zutis. Thanks to OPTOS plc for continuous support. The validation section owes much to the international coauthors of Reference [16] and their coworkers. Devanjali Relan is funded by Leverhulme project grant RPG-419.

REFERENCES

[1] N. Patton, T. Aslam, T. J. MacGillivray, A. Pattie, I. J. Deary, and B. Dhillon. Retinal vascular image analysis as a potential screening tool for cerebrovascular disease. *Journal of Anatomy*, 206:318–348, 2005.

[2] M. K. Ikram, J. C. M. Witteman, J. R. Vingerling, M. M. B. Breteler, A. Hofman, and P. T. V. M. de Jong. Retinal vessel diameters and risk of hypertension: the Rotterdam Study. *Hypertension*, 47(2):189–194, 2006.

[3] H. Leung. Relationships between age, blood pressure, and retinal vessel diameters in an older population. *Investigative Ophthalmology and Visual Science*, 44(7):2900–2904, 2003.

[4] N. Patton, T. M. Aslam, T. MacGillivray, I. J. Deary, B. Dhillon, R. H. Eikelboom, K. Yogesan, and I. J. Constable. Retinal image analysis: concepts, applications and potential. *Progress in Retinal and Eye Research*, 25(1):99–127, 2006.

[5] T. Y. Wong, M. D. Knudtson, R. Klein, B. E. K. Klein, S. M. Meuer, and L. D. Hubbard. Computer-assisted measurement of retinal vessel diameters in the Beaver Dam Eye Study: methodology, correlation between eyes, and effect of refractive errors. *Ophthalmology*, 111(6):1183–1190, 2004.

[6] M. D. Abràmoff, M. K. Garvin, and M. Sonka. Retinal imaging and image analysis. *IEEE Reviews in Biomedical Engineering*, 3:169–208, 2010.

[7] M. M. Fraz, S. A. Barman, P. Remagnino, A. Hoppe, A. Basit, B. Uyyanonvara, A. R. Rudnicka, and C. G. Owen. An approach to localize the retinal blood vessels using bit planes and centerline detection. *Computer Methods and Programs in Biomedicine*, 108(2):600–616, 2012.

[8] E. Trucco, L. Ballerini, D. Relan, A. Giachetti, T. MacGillivray, K. Zutis, C. Lupascu, D. Tegolo, E. Pellegrini, G. Robertson, P. Wilson, A. Doney, and B. Dhillon. Novel vampire algorithms for quantitative analysis of the retinal vasculature. In *4th IEEE Biosignals and Biorobotics Conference (ISSNIP/BRC)*, 2013.

[9] A. Perez-Rovira, T. MacGillivray, E. Trucco, K. S. Chin, K. Zutis, C. Lupascu, D. Tegolo, A. Giachetti, P. J. Wilson, A. Doney, and B. Dhillon. VAMPIRE: vessel assessment and measurement platform for images of the REtina. In *Proceeding of the 33rd IEEE EMBS International Conference on Engineering in Medicine and Biology*, pages 3391–3394, 2011.

[10] A. Cavinato, L. Ballerini, E. Trucco, and E. Grisan. Spline-based refinement of vessel contours in fundus retinal images for width estimation. In *Proceeding of the 10th IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 860–863, 2013.

[11] A. Giachetti, K. S. Chin, E. Trucco, C. Cobb, and P. J. Wilson. Multiresolution localization and segmentation of the optical disc in fundus images using inpainted background and vessel information. In *Proceeding of the IEEE International Conference on Image Processing*, pages 2145–2148, 2011.

[12] C. Lupascu and D. Tegolo. Automatic unsupervised segmentation of retinal vessels using self-organizing maps and K-means clustering. In *Computational Intelligence Methods for Bioinformatics and Biostatistics*, pages 263–274, 2011.

[13] E. Grisan and A. Ruggeri. A divide et impera strategy for automatic classification of retinal vessels into arteries and veins. In *Proceeding of the 25th IEEE EMBS International Conference on Engineering in Medicine and Biology*, pages 890–893, 2003.

[14] D. Fiorin and A. Ruggeri. Computerized analysis of narrow-field ROP images for the assessment of vessel caliber and tortuosity. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*, pages 2622–2625, 2011.

[15] F. N. Doubal, T. J. MacGillivray, P. E. Hokke, B. Dhillon, M. S. Dennis, and J. M. Wardlaw. Differences in retinal vessels support a distinct vasculopathy causing lacunar stroke. *Neurology*, 72:177–1778, 2009.

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

[17] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree. Retinal vessel segmentation using the 2-d Gabor wavelet and supervised classification. *IEEE Transactions on Medical Imaging*, 25:1214–1222, 2006.

[18] B. S. Y. Lam, Y. Gao, and A. W.-C. Liew. General retinal vessel segmentation using regularization-based multiconcavity modeling. *IEEE Transactions on Medical Imaging*, 29(7):1369–1381, 2010.

[19] J. Lowell, A. Hunter, D. Steel, A. Basu, R. Ryder, and R. L. Kennedy. Measurement of retinal vessel widths from fundus images based on 2-D modeling. *IEEE Transactions on Medical Imaging*, 23(10):1196–1204, 2004.

[20] B. Al-Diri, A. Hunter, and D. Steel. An active contour model for segmenting and measuring retinal vessels. *IEEE Transactions on Medical Imaging*, 28:1488–1497, 2009.

[21] X. Xu, M. Niemeijer, Q. Song, M. Sonka, M. K. Garvin, J. M. Reinhardt, and M. D. Abramoff. Vessel boundary delineation on fundus images using graph-based approach. *IEEE Transactions on Medical Imaging*, 30(6):1184–1191, 2011.

[22] H. Li, W. Hsu, M. L. Lee, and T. Y. Wong. Automatic grading of retinal vessel caliber. *IEEE Transactions on Biomedical Engineering*, 52(7):1352–1355, 2005.

[23] D. Fiorin, E. Poletti, E. Grisan, and A. Ruggeri. Fast adaptive axis-based segmentation of retinal vessels through matched filters. *World Congress on Medical Physics and Biomedical Engineering*, 25/11:145–148, 2009.

[24] M. M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A. R. Rudnicka, C. G. Owen, and S. A. Barman. Blood vessel segmentation methodologies in retinal images – a survey. *Computer Methods and Programs in Biomedicine*, 108(1):407–433, 2012.

[25] J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23:501–509, 2004.

[26] B. Al-Diri, A. Hunter, D. Steel, M. Habib, T. Hudaib, and S. Berry. REVIEW-A reference data set for retinal vessel profiles. In *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2262–2265, 2008.

[27] G. Strang. *Linear Algebra and its Applications*. Cengage, 2006.

[28] E. Trucco, A. Nayak, and N. Thacker. When are simple LS estimators enough? An empirical investigation of LS, TLS, and GTLS. *International Journal of Computer Vision*, 68(2):203–216, 2006.

[29] C. Lupascu, D. Tegolo, and E. Trucco. Accurate estimation of retinal vessel width using bagged decision trees and an extended multiresolution hermite model. *Medical Image Analysis*, 17(8):1164–1180, 2013, in press.

[30] C. Kondermann, D. Kondermann, and M. Yan. Blood vessel classification into arteries and veins in retinal images. *Proceedings of the SPIE*, 6512:651247–651249, 2007.

[31] R. Gonzalez and R. Woods. *Digital Image Processing*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 2002.

[32] A. Zamperini, A. Giachetti, E. Trucco, and K. S. Chin. Effective features for artery-vein classification in digital fundus images. In *2012 25th International Symposium on Computer-Based Medical Systems (CBMS)*, pages 1–6, 2012.

[33] H. Yu, S. Barriga, C. Agurto, S. Nemeth, W. Bauman, and P. Soliz. Automated retinal vessel type classification in color fundus images. *Proceedings of the SPIE*, 8670:86700P–86700P–8, 2013.

[34] A. Osareh, M. Mirmehdi, B. Thomas, and R. Markham. Comparison of colour spaces for optic disc localisation in retinal images. In *International Conference on Pattern Recognition*, volume 16, pages 743–746, 2002.

[35] A. Osareh, M. Mirmehdi, B. Thomas, and R. Markham. Automated identification of diabetic retinal exudates in digital colour images. *British Journal of Ophthalmology*, 87(10):1220–1223, 2003. PMID: 14507751.

[36] H. Wang, W. Hsu, K. G. Goh, and M. L. Lee. An effective approach to detect lesions in color retinal images. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition, 2000*, volume 2, pages 181–186, 2000.

[37] K. A. Goatman, A. D. Whitwam, A. Manivannan, J. A. Olson, and P. F. Sharp. Colour normalisation of retinal images. In *Proceedings of the Medical Image Understanding and Analysis*, pages 49–52, 2003.

[38] H. Narasimha-Iyer, J. M. Beach, B. Khoobehi, and B. Roysam. Automatic identification of retinal arteries and veins from dual-wavelength images using structural and functional features. *IEEE Transactions on Biomedical Engineering*, 54(8):1427–1435, 2007.

[39] M. Niemeijer, B. van Ginneken, and M. D. Abramoff. Automatic classification of retinal vessels into arteries and veins. *Proceedings of the SPIE*, 7260:72601F–8, 2009.

[40] C. Muramatsu, Y. Hatanaka, T. Iwase, T. Hara, and H. Fujita. Automated selection of major arteries and veins for measurement of arteriolar-to-venular diameter ratio on retinal fundus images. *Computerized Medical Imaging and Graphics*, 35(6):472–480, 2011.

[41] K. Rothaus, X. Jiang, and P. Rhiem. Separation of the retinal vascular graph in arteries and veins based upon structural knowledge. *Image and Vision Computing*, 27(7):864–875, 2009.

[42] M. Saez, S. González-Vázquez, M. González-Penedo, M. A. Barceló, M. Pena-Seijo, G. Coll de Tuero, and A. Pose-Reino. Development of an automated system to classify retinal vessels into arteries and veins. *Computer Methods and Programs in Biomedicine*, 108(1):267–376, 2012.

[43] S. G. Vazquez, B. Cancela, N. Barreira, M. G. Penedo, and M. Saez. On the automatic computation of the arterio-venous ratio in retinal images: using minimal paths for the artery/vein classification. In *International Conference on Digital Image Computing: Techniques and Applications*, pages 599–604, 2010.

[44] V. S. Joshi, M. K. Garvin, J. M. Reinhardt, and M. D. Abramoff. Automated artery-venous classification of retinal blood vessels based on structural mapping method. In *Proceeding of the SPIE Conference on Medical Imaging*, volume 8315, pages 83151–83158, 2012.

[45] R. Chrástek, M. Wolf, K. Donath, H. Niemann, D. Paulus, T. Hothorn, B. Lausen, R. Lämmer, C. Y. Mardin, and G. Michelson. Automated segmentation of the optic nerve head for diagnosis of glaucoma. *Medical Image Analysis*, 9(4):297–314, 2005.

[46] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[47] P. Jannin. Validation of medical image processing in image-guided therapy. *IEEE Transactions on Medical Imaging*, 21(12):1445–1449, 2002.

[48] P. Jannin, E. Krupinski, and E. Warfield. Validation in medical image processing. *IEEE Transactions on Medical Imaging*, 25(11):1405–1409, 2006.

[49] E. Trucco and A. Ruggeri. Towards a multi-site international data set for the validation of retinal image analysis software. In *Proceeding of the Annual International Conference of the IEEE EMBS*, Osaka, Japan, 2013.

[50] G. Quellec, M. Lamard, G. Cazuguel, M. D. Abràmoff, B. Cochener, and C. Roux. Weakly supervised classification of medical images. In *Proceedings of IEEE International Symposium on Biomedical Imaging*, 2012.

[51] O. Commonwick, H. Warfield, and D. Simon. Estimation of inferential uncertainty in assessing expert segmentation performance from STAPLE. *IEEE Transactions on Medical Imaging*, 29(3):771–780, 2010.

[52] F. Moret, C. Poloschek, W. Lagrèze, and M. Bach. Visualization of fundus vessel pulsation using principal component analysis. *Investigative Ophthalmology and Visual Science*, 52:5457–5464, 2011.

[53] D. K. Kumar, H. Hao, B. Aliahmad, T. Y. Wong, and R. Kawasaki. Does retinal vascular geometry vary with cardiac cycle? *Investigative Ophthalmology and Visual Science*, 53(9):5799–5805, 2012.