

Discriminative Vessel Segmentation in Retinal Images by Fusing Context-Aware Hybrid Features

Erkang Cheng · Liang Du · Yi Wu · Ying J. Zhu · Vasileios Megalooikonomou · Haibin Ling

Received: date / Accepted: date

Abstract Vessel segmentation is an important problem in medical image analysis and is often challenging due to large variations in vessel appearance and profiles, as well as image noises. To address these challenges, we propose a solution by combining heterogeneous context-aware features with a discriminative learning framework. Our solution is characterized by three key ingredients: First, we design a hybrid feature pool containing recently invented descriptors including the stroke width transform (SWT) and Weber's local descriptors (WLD), as well as classical local features including intensity values, Gabor responses and vesselness measurements. Second, we encode context information by sampling the hybrid features from an orientation invariant local context. Third, we treat pixel-level vessel segmentation as a discriminative classification problem, and use a random forest to fuse the rich information encoded in the hybrid context-aware features. For evaluation, the proposed method is applied to retinal vessel segmentation using three publicly available benchmark datasets. On the DRIVE and STARE datasets, our approach achieves average classification accuracies of 0.9474 and 0.9633, respectively. On the

high resolution dataset HRFID, our approach achieves average classification accuracies of 0.9647, 0.9561 and 0.9634 on three different categories, respectively. Experiments are also conducted to validate the superiority of hybrid feature fusion over each individual component.

Keywords Vessel segmentation · Random forest · Stroke width transform · Weber's local descriptors

1 Introduction

Assessment of the characteristics of blood vessels plays an important role in automatic systems for detecting diseases such as diabetes, hypertension, and arteriosclerosis. As an initial step, accurate vessel segmentation lays down a critical basis for subsequent operations and therefore has been attracting a great amount of research attention.

Existing vessel segmentation methods can be roughly divided into two groups: rule-based methods and learning-based ones. Rule-based methods find the vessel locations using presumed rules for vessels. For example, vessel tracking methods utilize a profile model to incrementally trace along and finally segment a vessel [4, 15]. Mathematical morphology is used to explore vasculature shape features such as piecewise linearity and connectivity as prior [20]. The *match filters* (MF) are designed to simulate the profile of the cross section of a blood vessel for vessel localization [5, 33]. A likelihood ratio-based approach is introduced in [29] that fuses matched filter responses, confidence measures and vessel boundary measures for the extraction of vessel centerlines. *Vesselness* is another popular approach that utilizes eigenvalues of the Hessian matrix to enhance all vascular structures including vessel bifurcations and to suppress non-vessel structures [10]. Learning-based methods, in contrast, turn the segmentation problem into a vessel/non-vessel classification task. Classifiers are trained from a set of features

Erkang Cheng · Liang Du · Haibin Ling
Department of Computer and Information Science, Temple University, Philadelphia, PA, 19122 E-mail: erkang.cheng, liang.du, hbling@temple.edu

Vasileios Megalooikonomou
Department of Computer and Information Science, Temple University, Philadelphia, PA, 19122 & Computer Engineering and Informatics Department, University of Patras, Greece, 26500 E-mail: vasilis@temple.edu

Ying J. Zhu
Electrical & Computer Engineering Department, Temple University, Philadelphia, PA, 19122 E-mail: yingzhu97@gmail.com

Yi Wu
School of Information and Control Engineering, Nanjing University of Information Science and Technology, Nanjing, China, 210044. E-mail: ywu.china@gmail.com

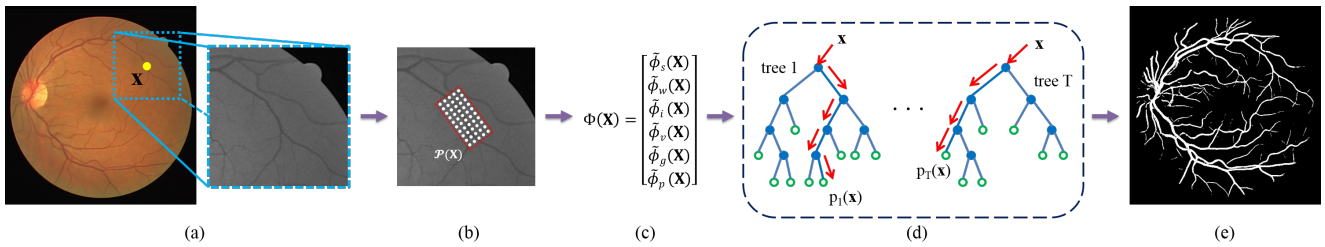


Fig. 1 An overview of our proposed vessel segmentation algorithm. (a) An input image I and a pixel \mathbf{x} (in yellow) to be labeled as vessel or non-vessel. Patch zoomed in is a part of green channel of I centered at \mathbf{x} . (b) Detected local context $\mathcal{P}(\mathbf{x})$ (red rectangle) of \mathbf{x} . (c) Extracted hybrid feature $\Phi(\mathbf{x})$. (d) Random forest classifier learned from annotation. (e) Segmentation result of I .

along with labels [26]. Features like pixel intensities, edges and Gabor wavelets are fed to different classifiers such as multilayer neural network [27], Bayesian inference [28], random forest [7], graph cut [23] and support vector machine (SVM) [19, 26]. These methods, by introducing discriminative power in modeling, often outperform traditional rule-based solutions [19].

Despite many existing approaches, automatic vessel segmentation remains a challenging task due to challenges arising from various sources such as the variation in vessel appearance, shape and orientations, the low contrast between vasculature and background, the presence of noise, large abnormal regions due to the presence of lesions, exudates, and other pathological regions [26]. For instance, the variation in vessel shape and orientation can mislead the match filter to respond to non-vessel edges [33]; the low contrast areas in retinal images can cause similar problems. It is hard to use a single type of feature to address all these challenges. A typical strategy is to employ additional post-processing steps in vessel segmentation (e.g. [19]).

In this paper, we treat vessel segmentation as a pixel-wise classification problem and present a solution that fuses hybrid discriminative features with context modeling. There are three key ingredients in our solution: *hybrid features* which bring diverse types of discriminative information, *local context* that provides clues to distinguish vessel pixels from similar clutters, and a *discriminative learning framework* that automatically leverages and fuses all information encoded in the feature pool. The three ingredients, illustrated in Figure 1, are briefly summarized below:

- **Hybrid features.** We use a heterogeneous set of features to form a rich feature pool. We first introduce two recently invented image features, the *stroke width transform* (SWT) [9] and the *Weber’s local descriptor* (WLD) [6], both of which capture line-like structures to some extent and have never been used for vessel segmentation. We also include intensity, Gabor responses and vesseness, all of which are previously used for vessel segmentation. Furthermore, position information is encoded in the feature pool.

- **Local context.** For each pixel, we extract a local context patch which is an oriented rectangle center at the pixel. The orientation is estimated from local intensity pattern to achieve rotation invariance. Then, the hybrid features are extracted on sampled position inside the local context to encode the context information.
- **Discriminative learning.** For fusing hybrid features extracted from context patch of a pixel, we build a classifier using the *random forest* (RF) [3] by taking advantage of its strong discriminative power and its flexibility of fusing heterogeneous features.

We apply the proposed method to retinal vessel segmentation and evaluate it using three publicly available datasets: the DRIVE dataset [30], the STARE dataset [14], and the High-Resolution Fundus Image Database (HRFID) [24]. Quantitative evaluation results demonstrate the effectiveness of the proposed approach. Furthermore, we investigate the performances of individual features and their behavior in the experimental results. These studies confirm that the proposed features are complimentary to each other and fusing them by the random forest framework boosts the performance. Since our method uses a general learning framework that automatically exploits feature pools, it can be easily generalized to other similar tasks for different clinical applications.

The rest of this paper is organized as follows: In the next section, we describe the proposed vessel segmentation method. After that, in Section 3, the experimental results are presented and discussed. Finally, conclusions are drawn in Section 4.

2 Method

2.1 Overview of the proposed method

We formulate vessel segmentation as a pixel-level classification problem. Now we briefly summarize the proposed approach. Given an image I and a pixel $\mathbf{x} \in \mathbb{R}^2$ to be classified, we first calculate an orientation invariant local context $\mathcal{P}(\mathbf{x})$. Then, a feature representation, denoted as $\Phi(\mathbf{x})$, is generated by extracting heterogeneous features from $\mathcal{P}(\mathbf{x})$.

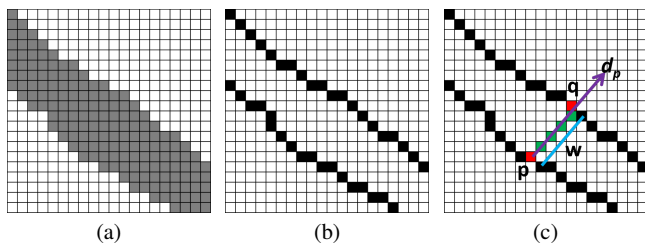


Fig. 3 Stroke width transform. (a) Example of an image patch. (b) Edge extraction of (a). (c) Stroke width calculation at point \mathbf{p} .

After that, $\Phi(\mathbf{x})$ is fed into a random forest classifier to determine whether \mathbf{x} is a vessel pixel. Figure 1 gives the overview of our proposed algorithm.

In the following subsections, we first describe the hybrid feature set, then detail the extraction of local context, and present the random forest algorithm used in our approach.

2.2 Hybrid Features

To address the diverse range of challenges in vessel segmentation, we use a hybrid set of features to capture discriminative information from various aspects. Our feature pool consists of recently proposed descriptors such as the stroke width transform and the Weber’s local descriptors, the classical vessel features including pixel intensity, vesselness and Gabor responses, as well as position information of each pixel. Hybrid features are extracted from the green channel of retinal image because it presents the largest contrast between vessels and the background. Figure 2 shows some feature components used in our method.

Stroke width transform (ϕ_s). One difficulty in modeling vessels lies in the large variance of vessel profile, including vessel width, vessel orientation and local vessel shape. Having observed that vessels are continuous parts of an image with nearly constant width within a limited length, we propose to use stroke width transform (SWT) to extract vessel width features. SWT is a local descriptor that computes per pixel the width of the most likely stroke-like structure containing the pixel [9]. The output of SWT is a 2D matrix with the same size as the input image. Each matrix element contains the width of the stroke passing the corresponding pixel. Another desirable property of SWT for vessel feature extraction is its flexibility in extracting vessel features at arbitrary orientations.

Figure 3 illustrates the calculation of SWT. The initial SWT value for each pixel is set to ∞ . First, the edge of the original image is computed using the Canny edge detector (Figure 3 (b)). Then, for each edge pixel \mathbf{p} , a ray starts from \mathbf{p} along its gradient direction \mathbf{d}_p , i.e.,

$$\mathbf{y} = \mathbf{p} + r\mathbf{d}_p, \quad r > 0 \quad (1)$$

till reaching another edge pixel \mathbf{q} . The length of the ray $w = |\mathbf{p} - \mathbf{q}|$ is assigned as the stroke width to all pixels lying between \mathbf{p} and \mathbf{q} (the blue line in Figure 3(c)). If several rays intersect at a pixel, then the smallest w will be assigned to that pixel. For a detailed explanation of SWT, one can refer [9].

Intuitively, if \mathbf{p} lies on a vessel boundary, \mathbf{d}_p should be roughly perpendicular to the orientation of the vessel and w approximates the width of the vessel. For this reason, SWT was originally proposed for text detection in natural scenes.

Ideally, SWT generates meaningful responses only on edge pixels and leaves non-vessel areas untouched. In practice, we set the default SWT response values to 100, which distinguishes most non-vessel pixels from vessel pixels. Some example results are shown in Figure 2 (d). We denote the SWT at a pixel \mathbf{x} as $\phi_s(\mathbf{x})$.

Weber’s local descriptors (ϕ_w). Another challenge for vessel detection in retinal images comes from the large low contrast areas. We propose using the Weber’s local descriptors (WLD) [6] to alleviate this problem. WLD is a simple, yet very powerful and robust local descriptor inspired by the Weber’s Law. Specifically, we use the differential excitation component in the original WLD described below.

For a pixel \mathbf{x} in a vessel image I , its differential excitation component, denoted as $\phi_w(\mathbf{x})$, measures the relative intensity differences of \mathbf{x} against its neighbors:

$$\phi_w(\mathbf{x}) = \arctan \left(\sum_{\mathbf{z} \in \mathcal{N}(\mathbf{x})} \frac{I(\mathbf{z}) - I(\mathbf{x})}{I(\mathbf{x})} \right), \quad (2)$$

where $\mathcal{N}(\mathbf{x})$ is the set of pixels in \mathbf{x} ’s neighborhood. The arc-tangent function is used to prevent the output from increasing or decreasing too quickly.

Figure 2 (e) shows two example WLD results. We can see that the low contrast areas are enhanced.

Intensity feature (ϕ_i). To a certain degree, vessel pixels often distinguish themselves from background ones by intensity. For example, vessels in retinal images often appear darker than other areas as shown in Figure 2 (b). Inspired by this observation, for each pixel \mathbf{x} in a vessel image I , we include its intensity $I(\mathbf{x})$, denoted as $\phi_i(\mathbf{x})$, in our hybrid feature pool.

Gabor feature (ϕ_g). Gabor wavelet is known to be very effective for texture representation [28]. The center symmetric Gabor wavelet filter used in our method is expressed as:

$$\begin{aligned} g(\mathbf{x}; \lambda, \theta, \sigma, \gamma) &= \exp \left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} \right) \exp \left(i2\pi \frac{x'}{\lambda} \right) \\ x' &= x \cos(\theta) + y \sin(\theta) \\ y' &= -x \sin(\theta) + y \cos(\theta), \end{aligned} \quad (3)$$

where $\mathbf{x} = (x, y)$ is a 2D point, λ represents the wavelength of the sinusoidal factor, θ represents the orientation of the

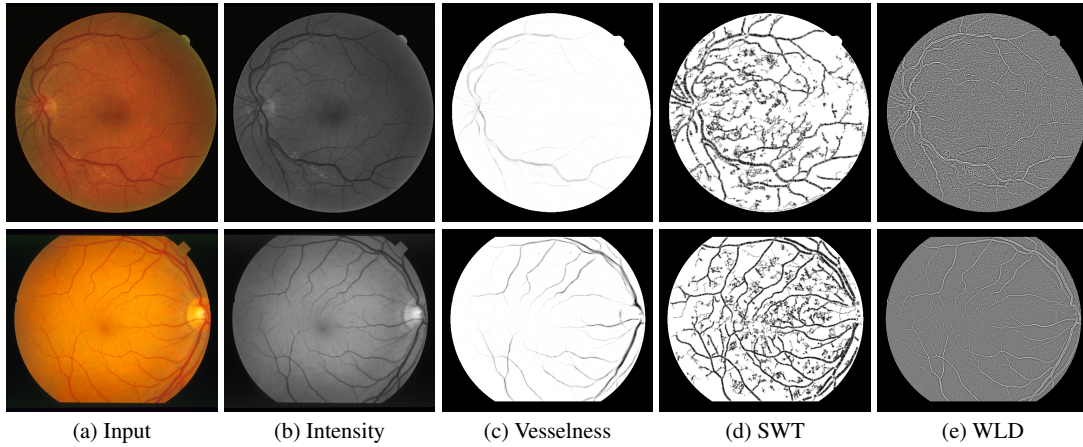


Fig. 2 Components of hybrid feature used in the proposed method. (a) An input image. (b) Green channel intensity. (c) Vesselness enhancement. (d) Stroke width transform. (e) Weber's local descriptor. The original images of the first row and the second row are from the DRIVE and STARE datasets, respectively.

normal to the parallel of the Gabor function, σ is the sigma of the Gaussian envelope and γ is the spatial aspect ratio which specifies the ellipticity of the support of the Gabor function. The bandwidth of Gabor filter b is related to the ratio σ/λ . Then, the Gabor response is parameterized by $G(\mathbf{x}; \lambda, \theta, b, \gamma)$. For each pixel position and considered scale value (λ, b, γ) , we are interested in the maximum responses of real, imaginary and magnitude over all orientations, denoted as,

$$M^c(\mathbf{x}; \lambda, b, \gamma) = \arg \max_{\theta \in \Theta} \|G^c(\mathbf{x}; \lambda, \theta, b, \gamma)\|, \quad (4)$$

where $\Theta = \{k\pi/18, k = 0, \dots, 17\}$, $c \in \{r, i, m\}$, and the corresponding G^r, G^i, G^m respectively denote the real, imaginary and magnitude components of $G(\mathbf{x}; \lambda, \theta, b, \gamma)$. Thus for specific parameters (λ, b, γ) , three response values are extracted. A filter bank consisting of Gabor filters with multiple scales is applied to a vessel image, and the responses of pixel \mathbf{x} form the Gabor feature vector $\phi_g(\mathbf{x})$.

In our implementation, the parameters are set as $b \in \{1, 2\}$, $\lambda \in \{2, 4, 6, 8\}$ and $\gamma \in \{0.25, 0.5\}$. This generates $2 \times 4 \times 2 = 16$ different configurations of Gabor filters. Therefore, the Gabor feature $\phi_g(\mathbf{x})$ is a $16 \times 3 = 48$ dimension vector.

Vesselness measurement (ϕ_v). Based on the observation that the eigenvalues of Hessian matrices capture vessel like structures, Frangi et al. proposed vesselness measurement [10] for vessel representation. Vesselness has been popularly used for its effectiveness and computational efficiency. Examples of vesselness measurements of given retinal images are provided in Figure 2 (c). We denote the vesselness of a pixel \mathbf{x} as $\phi_v(\mathbf{x})$.

Position information (ϕ_p). In some anatomic structures such as retinal vessels, different locations in the FOV (field-of-view) have different vessel-related priors. For example, the

distribution of vessel pixels is much sparser in peripheral regions than around the center. Moreover, an optic disk is unlikely located at the center of an image. Such information can help detect vessel pixels at certain areas of an image. Motivated by the observation, we include position-related features in our hybrid feature set, denoted as $\phi_p(\mathbf{x})$ at pixel \mathbf{x} . In particular, let \mathbf{c} be the center of the image, we have

$$\phi_p(\mathbf{x}) = (\|\mathbf{x} - \mathbf{c}\|, \sin(\theta_p), \cos(\theta_p), d_m(\mathbf{x}))^\top, \quad (5)$$

where $d_m(\mathbf{x})$ is the distance from \mathbf{x} to the boundary of the FOV and θ_p is the angle of vector $\overrightarrow{\mathbf{x}\mathbf{c}}$.

2.3 Feature Extraction from Orientation Invariant Local Context

In practice, the information restricted on a single pixel is often insufficient to accurately determine its label (i.e., vessel or not). Many previous approaches rely on post-processing to aggregate context information from neighborhood, often in an ad hoc fashion. We instead address this issue by explicitly modeling context information. Note that this approach will result in a high dimensional feature space, which fortunately poses no problem in our flexible learning framework.

Estimation of Local Context. For a pixel \mathbf{x} of a given image I , we define its local context $\mathcal{P}(\mathbf{x})$ as an oriented rectangular region centered at \mathbf{x} with fixed width and height. To achieve insensitivity against rotation, we estimate the orientation of \mathbf{x} , denoted by $\theta(\mathbf{x})$, by a simple searching procedure

$$\theta(\mathbf{x}) = \arg \min_{\theta \in \Theta} \sum_{\mathbf{p} \in R(\theta, \mathbf{x})} I(\mathbf{p}), \quad (6)$$

where $R(\theta, \mathbf{x})$ is a rectangle centered at \mathbf{x} with orientation θ , $\Theta = \{k\pi/12, k = 0, \dots, 11\}$ is the set of candidate orientations. Note that there are many other ways for estimating

local patch orientation, while the proposed strategy particularly captures the property that vessel pixels are in general darker than other pixels. The procedure is illustrated in Figure 4 (a-b). In our implementation, we fix the size of $R(\theta, \mathbf{x})$ as 3×11 pixels.

Once $\theta(\mathbf{x})$ is ready, the local context $\mathcal{P}(\mathbf{x})$ is formally defined as a set of sample points uniformly sampled in the rectangle centered at \mathbf{x} and oriented at $\theta(\mathbf{x})$. In practice we sample $w = 5$ pixels along one dimension and $h = 11$ along the other, resulting $n_p = w \times h = 55$ sample points, i.e., $\mathcal{P}(\mathbf{x}) = \{\mathbf{p}_k, k = 1, 2, \dots, n_p\}$. An example is shown in Figure 4 (c).

Feature Aggregation. In order to use the context information, we aggregate features extracted from all sample points in $\mathcal{P}(\mathbf{x})$ for pixel \mathbf{x} . In particular, for a feature function $\phi \in \{\phi_s, \phi_w, \phi_i, \phi_v\}$, we define the aggregated feature extraction as

$$\tilde{\phi}(\mathbf{x}) = (\phi(\mathbf{p}_1), \phi(\mathbf{p}_2), \dots, \phi(\mathbf{p}_{n_p}), \mu(\mathbf{x}), \sigma(\mathbf{x}))^\top, \quad (7)$$

where $\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ denote respectively the mean and standard variation of $\{\phi(\mathbf{p}_1), \dots, \phi(\mathbf{p}_{n_p})\}$, and $\mathbf{p}_k \in \mathcal{P}(\mathbf{x}), k = 1, \dots, n_p$. The calculation results a $n_p + 2 = 57$ dimensional feature vector.

Note that we do not apply the context aggregation for all hybrid features described in Section 2.2. This is because context information has been implicitly captured by Gabor responses ϕ_g and is not applicable to position information ϕ_p .

Now we can formally define the feature pool $\Phi(\mathbf{x})$ for pixel \mathbf{x} by combining all features, either with or without context, as

$$\Phi(\mathbf{x}) = (\tilde{\phi}_s(\mathbf{x})^\top, \tilde{\phi}_w(\mathbf{x})^\top, \tilde{\phi}_i(\mathbf{x})^\top, \tilde{\phi}_v(\mathbf{x})^\top, \phi_g(\mathbf{x})^\top, \phi_p(\mathbf{x})^\top)^\top. \quad (8)$$

2.4 Discriminative learning framework

The proposed hybrid features form a large feature pool. In addition, we have a large number of training samples which are essentially all pixels from the training images. To handle the large heterogeneous feature pool and the large training set, we choose the random forest framework [3] due to its flexibility and robustness demonstrated in similar medical image analysis tasks [8].

A random forest is an ensemble classifier that consists of decision trees and each tree is constructed via some randomized configuration. The randomization allows the flexibility to explore a large feature space effectively because it only considers a subset of features in each tree node. Also, it has the ability to handle huge training samples since each tree is only fed with a random subset of the whole training data. As illustrated in Figure 5, a leaf node encodes the class distribution for samples that reach it. An internal node instead

performs a binary test to split the samples to its child nodes. The splitting terminates when a leaf node is reached. The posterior probability at each leaf node is learned as the proportion of the training samples labeled as vessels at the given leaf node. Node optimization is the key to select a best feature while guiding the split. A stump is applied in our experiments for the task. Specifically, for the input feature values Φ^i of samples S , a stump is used to select a best threshold to split samples in S to minimize the mis-classification error. More details can be found in [3] and [8].

In the testing phase, the feature $\Phi(\mathbf{x})$ of a pixel \mathbf{x} is first fed into the root of each tree and then it follows the splitting rule till it reaches a leaf (red paths in Figure 5). Each tree returns a posterior probability that \mathbf{x} belongs to a vessel. The mean of leaf distributions from all trees is used for final decision. Specifically, the probability that \mathbf{x} is a vessel pixel is estimated by

$$\Pr(\text{vessel}|\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T p_t(\text{vessel}|\Phi(\mathbf{x})), \quad (9)$$

where $p_t(\cdot)$ is the output from the t -th tree and T is the number of trees in the forest.

3 Experiments

In this section, we evaluate our algorithm on two groups of datasets. The first group includes two publicly available datasets: the DRIVE dataset [30] and the STARE dataset [14], both containing low resolution images. The second group includes the High-Resolution Fundus Image Database, which contains high resolution images. Experimental results on both low and high resolution datasets show the effectiveness of our algorithm. In the first group of datasets, we also conduct detailed comparisons of the hybrid features with individual components. The results demonstrate that the hybrid features outperform the individual ones.

3.1 Low resolution dataset

3.1.1 Implementation details

Datasets. Two public datasets, the DRIVE dataset [30] and the STARE dataset [14], are used in our evaluation. The DRIVE dataset consists of 40 images. The images are divided into a training set and a testing set. Ground truth of vessel segmentation is available for all the images in DRIVE dataset. FOV binary masks are also provided for all the training and testing images in the DRIVE dataset. For the images in the testing set, a second independent manual segmentation is also given.

The STARE dataset has 20 images, 10 of them are from healthy ocular fundi and the other 10 from unhealthy ones.

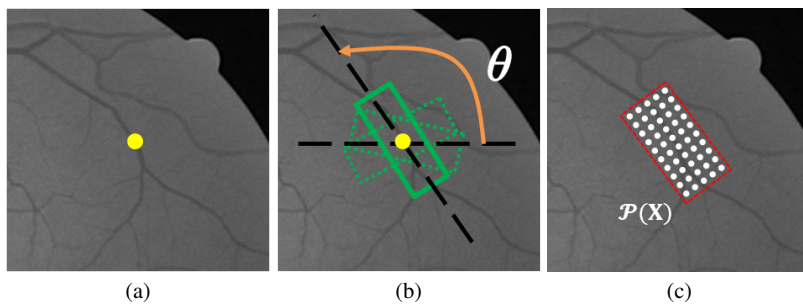


Fig. 4 Orientation invariant local context. (a) An input image I and a pixel \mathbf{x} (yellow circle) whose local context is to be estimated. (b) Orientation approximation: Rectangles are aligned with different orientation candidates. The solid rectangle reflects the true vessel orientation θ . (c) Local context $\mathcal{P}(\mathbf{x})$ with sample points.

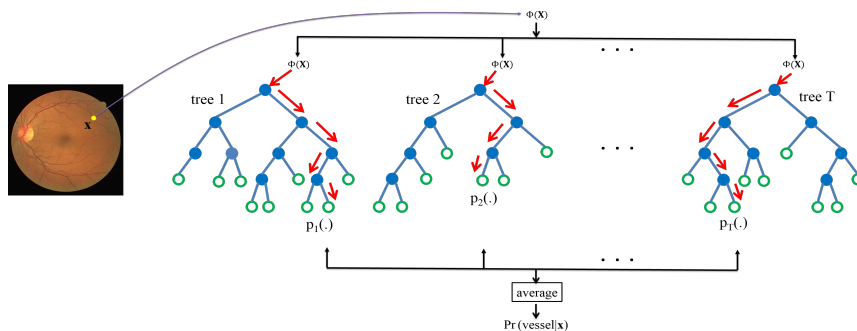


Fig. 5 A random forest consists of T individual decision trees. The internal nodes (in blue) are stump classifiers and the leaves (green circles) store an estimated vessel distribution $p(\text{vessel})$. A query pixel \mathbf{x} with computed feature $\Phi(\mathbf{x})$ traverses each tree (the red paths) to reach leaf nodes. Then, $p_t(\cdot)$ of the visited leaf node is the result of each tree. The final output is the average of all the trees.

Manually labeled vessel groundtruth is also provided for the STARE dataset. The STARE dataset does not have separate training and testing sets. We follow the approach in [28] to create FOV binary mask for each image in the STARE dataset.

Experiment protocol. For the DRIVE dataset, the classifier is built on training set. To train each tree, 10,000 positive (vessel) and 10,000 negative (non-vessel) samples are randomly selected from each training image. The results are collected on the separate testing set.

Due to that there are no available labeled training images for the STARE dataset. Soares et al. [28] and Staal et al. [30] performed leave-one-out tests on this database (i.e., every image is classified by using samples from the other 19 images), while Ricci et al. [26] and Marin et al. [19] built their classifier by using a training set comprising samples randomly extracted from test images.

We follow the experiment setup in [26] and [19] on the STARE dataset. To train each tree, 10,000 positive (vessel) and 10,000 negative (non-vessel) samples are randomly selected from each image. The results are evaluated on the whole dataset. Note that, due to the sampling, only around 6% of the pixels are involved when training a tree of the random forest.

There is no post-processing in the experiments for both DRIVE and STARE datasets.

Parameters. For a pixel \mathbf{x} in an image I , we use 5×11 samples around it to extract context information for SWT, WLD, intensity and vesseness features, resulting in $4 \times (55 + 2)$ feature values. In addition, we have a 48 dimensional Gabor feature and a 4 dimensional position feature. So the final hybrid feature vector, i.e., $\Phi(\mathbf{x})$, has the dimension of $57 \times 4 + 48 + 4 = 280$. Details can be found in Sections 2.2 and 2.3.

For the learning framework, we construct a random forest using 100 decision trees. Each decision tree is of depth 15 and built in a parallel fashion. Stump classifiers are used for internal tree nodes. Finally, 100 features are randomly selected from the 280-dimensional feature pool to train an internal node.

Quantitative evaluation. The initial output of our segmentation system is a probability map obtained by the voting of all the trees (e.g., Figure 8 and 9). The vessel segmentation is achieved by applying a threshold ($0 \leq \theta_{th} \leq 1$) to the probability map.

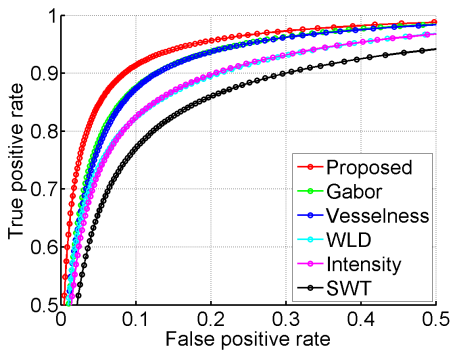
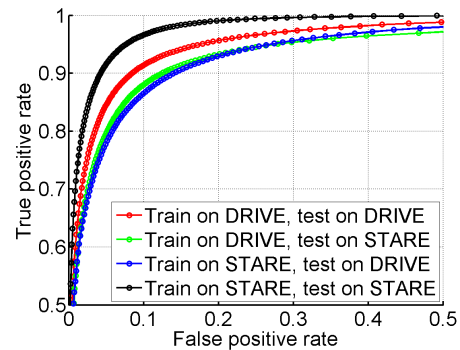
For quantitative evaluation of the segmentation results, we follow previous studies [19] by using *sensitivity* (SE), *specificity* (SP) and *classification accuracy* (ACC). In addition, the performance is also evaluated by receiver operating characteristic (ROC) curves. Figure 6 shows the ROC curves using different types of features on the DRIVE dataset. Figure 7 gives the ROC curves of different experimental proto-

Table 1 Comparison with state-of-the-art methods on the DRIVE dataset. (ACC-CT denotes classification accuracy in cross training mode.)

Type	Methods	SE	SP	ACC	AUC	ACC-CT
	2nd Human Observer	0.7796	0.9717	0.9470	–	–
Unsupervised	Chaudhuri et al. [5]	–	–	0.9284	–	–
	Zana et al. [32]	0.7796	0.9717	0.9470	–	–
	Jiang et al. [15]	0.6971	–	0.9377	0.8984	–
	Mendonca et al. [20]	–	–	0.9212	0.9114	–
	Perfetti et al. [25]	–	–	0.0.9261	–	–
	Al-Diri et al. [2]	0.7344	0.9764	0.9452	–	–
	Lam et al. [16]	–	–	0.9472	0.9614	–
	Zhang et al. [33]	–	–	0.9382	–	–
	Miri et al. [21]	0.7352	0.9795	0.9458	–	–
	Fraz et al. [11]	0.7152	0.9759	0.9430	–	–
You et al. [31]	0.7410	0.9751	0.9434	–	–	
Supervised	Niemeijer et al. [22]	–	–	0.9416	0.9294	–
	Soares et al. [28]	0.7332	0.9782	0.9461	0.9614	0.9397
	Staal et al. [30]	–	–	0.9441	0.9520	–
	Ricci et al. [26]	–	–	0.9595	0.9558	0.9266
	Lupascu et al. [18]	0.7200	–	0.9597	0.9561	–
	Fraz et al. [13]	0.7406	0.9807	0.9480	0.9747	0.9456
	Marin et al. [19]	0.7067	0.9801	0.9452	0.9588	0.9448
	Proposed	0.7252	0.9798	0.9474	0.9648	0.9384

Table 2 Comparison with state-of-the-art methods on the STARE dataset. (ACC-CT denotes classification accuracy in cross training mode.)

Type	Methods	SE	SP	ACC	AUC	ACC-CT
	2nd Human Observer	0.8951	0.9384	0.9348	–	–
Unsupervised	Chaudhuri et al. [5]	–	–	0.9276	–	–
	Hoover et al. [14]	0.6747	0.9565	0.9264	–	–
	Jiang et al. [15]	–	–	0.9009	–	–
	Mendonca et al. [20]	0.6996	0.9730	0.9440	–	–
	Lam et al. [17]	–	–	0.9474	0.9392	–
	Al-Diri et al. [2]	0.7521	0.9681	–	–	–
	Lam et al. [16]	–	–	0.9567	0.9739	–
	Fraz et al. [11]	0.7311	0.9680	0.9442	–	–
	You et al. [31]	0.7260	0.9756	0.9497	–	–
Supervised	Staal et al. [30]	–	–	0.9516	0.9614	–
	Soares et al. [28]	0.7207	0.9747	0.9479	0.9671	0.9327
	Ricci et al. [26]	–	–	0.9584	0.9602	0.9464
	Fraz et al. [13]	0.7548	0.9763	0.9534	0.9768	0.9493
	Marin et al. [19]	0.6944	0.9819	0.9526	0.9769	0.9526
	Proposed	0.7813	0.9843	0.9633	0.9844	0.9476

**Fig. 6** ROC curves of different features on the DRIVE dataset.**Fig. 7** ROC curves of different protocols on the DRIVE and STARE datasets.

cols on the DRIVE and STARE datasets. The area under the curve (AUC) is also calculated for evaluation.

For each dataset, the performance results are obtained considering the same threshold value θ_{th} for all the images

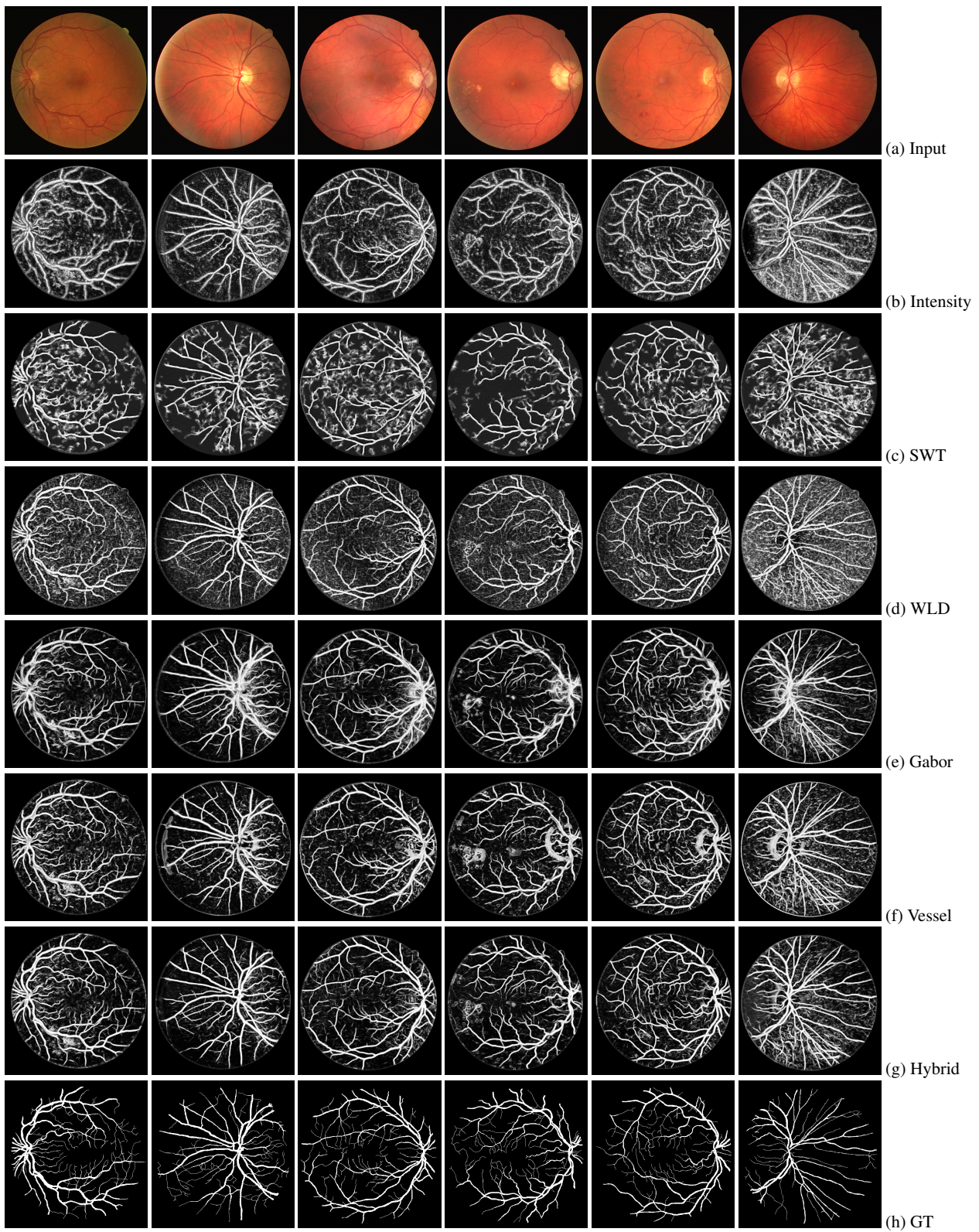


Fig. 8 Example results from the DRIVE dataset. From top to bottom: input image (a), probability maps using intensity (b), SWT (c), WLD (d), Gabor (e), Vesselness (f), the proposed hybrid combination (g), and the ground truth (GT) (h).

Table 3 Performance of individual feature components on the DRIVE dataset.

Methods	SE	SP	ACC	AUC
2nd Human Observer	0.7796	0.9717	0.9470	–
Intensity	0.6158	0.9739	0.9284	0.9303
Gabor	0.6723	0.9736	0.9353	0.9514
SWT	0.5526	0.9718	0.9184	0.9303
WLD	0.6198	0.9782	0.9326	0.9320
Vesselness	0.6468	0.9755	0.9337	0.9504
Proposed	0.7252	0.9798	0.9474	0.9648

in the same dataset. θ_{th} is taken to provide the maximum average classification accuracy. For a detailed explanation of θ_{th} , one can refer to [19]. The optimal threshold value for both the DRIVE and STARE datasets is $\theta_{th} = 0.84$.

3.1.2 Experimental results

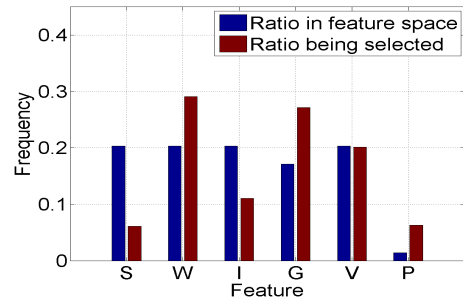
Tables 1 and 2 list the performance of our approach along with the results reported in previous studies for DRIVE and STARE datasets. Results of previous methods are reproduced from [12, 13]. From the tables we see that the proposed method produces excellent results and outperforms previous state-of-the-art solutions and even the 2nd human observer. In addition, we also conduct a cross training experiment on both DRIVE and STARE datasets. In particular, the classifier trained on images in DRIVE (or STARE) is tested on images in STARE (or DRIVE). The accuracies are listed in the last column (“ACC-CT”) of Tables 1 and 2. Evaluation results show that our method performs similarly with the state-of-the-art methods.

To further study the effects of individual feature components in our method, we have also conducted segmentation experiments using each individual feature component. In particular, a series of segmentation algorithms that are similar to the proposed solution. The only difference is that, instead of using the hybrid features, each algorithm here uses a specific type of feature, and the configuration of the random forest is adjusted accordingly. To simplify the experiments, all the classifiers are trained on training images in the DRIVE dataset. Tables 3 and 4 summarize the results in comparison with the proposed solution. Figures 8 and 9 show some example results on the two datasets for qualitative evaluation. The figures show that the probability maps from our method resemble very much the ground truth.

There are mainly two pieces of conclusion from the evaluations. First, the proposed method, by combining hybrid features, context information and a discriminative learning framework, effectively handles challenges in vessel segmentation. Second, the heterogeneous types of features act complementarity in the proposed approach.

Table 4 Performance of individual feature components on the STARE dataset.

Methods	SE	SP	ACC	AUC
2nd Human Observer	0.8951	0.9384	0.9348	–
Intensity	0.4795	0.9809	0.9289	0.8943
Gabor	0.4978	0.9781	0.9284	0.9194
SWT	0.5224	0.9759	0.9289	0.9082
WLD	0.5475	0.9783	0.9336	0.9280
Vesselness	0.5491	0.9787	0.9342	0.9342
Proposed	0.6340	0.9839	0.9476	0.9465

**Fig. 10** Feature importance in the random forest. (S: STW; W: WLD; I: Intensity; G: Gabor; V: Vesselness; P: Position.)

3.1.3 Feature importance analysis

To further understand the roles of individual features in the proposed method, we explore the discriminability of each feature by calculating the frequency by which it is being selected in the forest. When training the random forest, each node chooses one optimal feature from 100 randomly sampled feature subset. The relative frequency of a type of feature being selected reflects its importance in the classifier learned. In particular, for each feature type, we are interested in the ratio of the number of features from this type to the total number of features (i.e., ratio in feature space), and the ratio of the number of selected features from this type to the total number of selected features (i.e., ratio being selected). The results are provided in Figure 10. It is clear that all the features make significant contributions to the final inference, which is consistent with our intuition. In addition, the results show that the WLD feature is indeed the most effective one, followed by Gabor, Vesselness, intensity, SWT and the position feature.

3.1.4 Computation time analysis

The computation time of the proposed method for segmenting an input image is composed of two main parts: feature extraction and random forest inference. The computation for feature extraction involves mainly four components: Gabor responses, vesselness, SWT, and WLD. Techniques such as direct Gabor response index and integral images have been implemented for improving the efficiency. For random for-

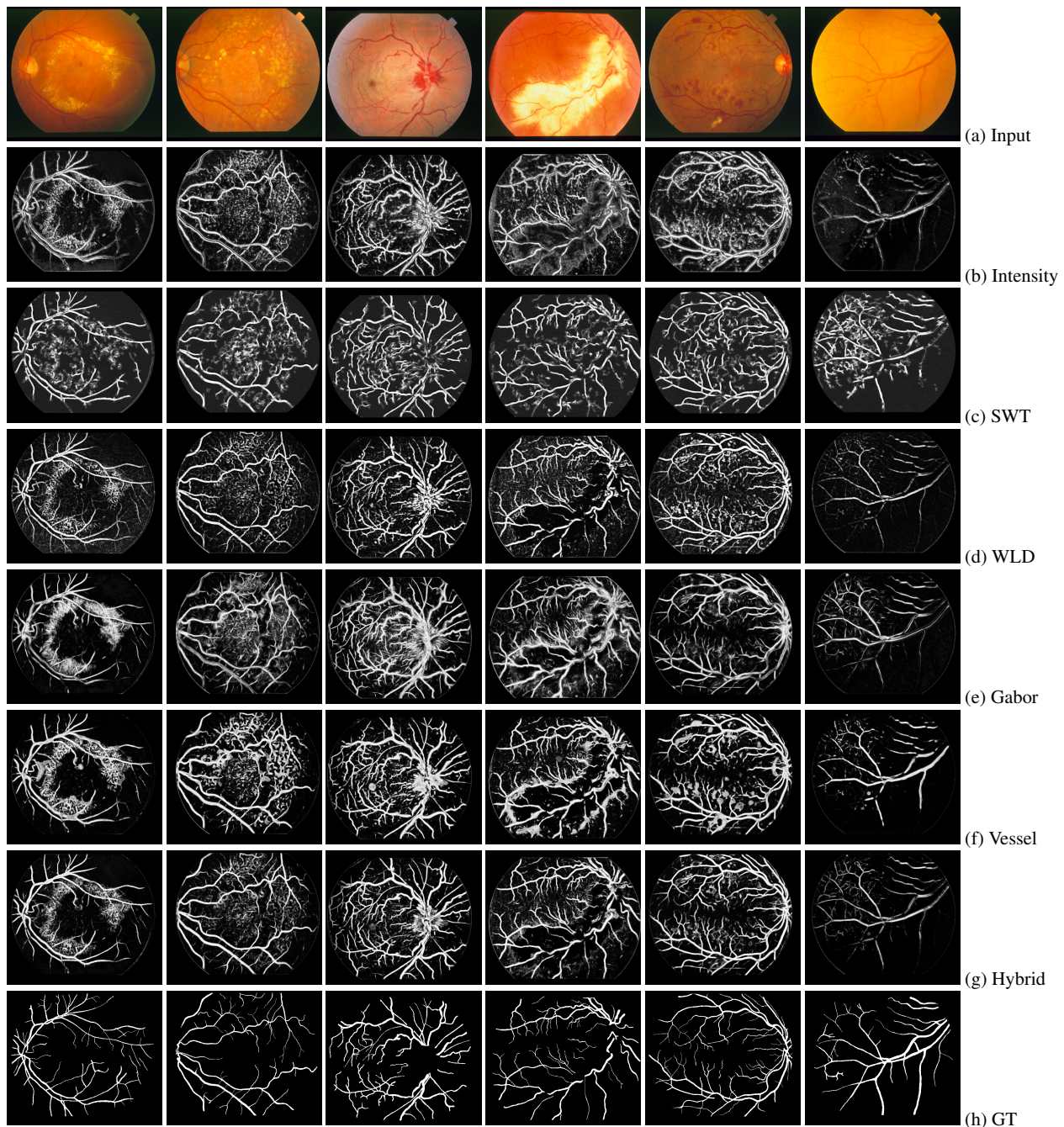


Fig. 9 Example results from the STARE dataset. From top to bottom: input image (a), probability maps using intensity (b), SWT (c), WLD (d), Gabor (e), Vesselness (f), the proposed hybrid combination (g), and the ground truth (GT) (h).

est inference, the calculation mainly lies in the traverse of decision trees of depth 15, and at each node only one feature needs to be calculated. In our experiments, the average time for segmenting an image is less than 1 minute.

3.1.5 Failure analysis

To analyze weaknesses of the proposed method, failure analysis of one sample image is shown in Figure 11. In this

figure (a) is an input image from DRIVE dataset, and (b) is the probabilistic map calculated by our approach. Vessel segmentation result is then achieved by thresholding ($\theta_{th} = 0.84$) of the probabilistic map. The segmentation results are color-coded in (c): true positives are indicated in red, false positives in green, and missing vessels in blue. Figure 11 demonstrates that our proposed method sometimes fails to capture thin vessel structures.

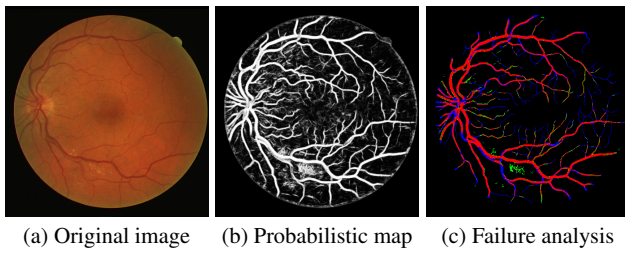


Fig. 11 Failure analysis of the proposed method. (a) Original image; (b) Probabilistic map computed by the proposed method; (c) Failure analysis of vessel segmentation by thresholding (b) with $\theta_{th} = 0.84$. True positives are shown in red; false positives in green; and missing vessels in blue.

3.2 High resolution dataset

3.2.1 Implementation details

Datasets. The high resolution images used in our evaluation is from the High-Resolution Fundus Image Database [24] (HRFID). The dataset has 45 retinal fundus images of size 2336×3504 . These images are divided into three categories: healthy images (HI), diabetic retinopathy images (DRI) and glaucomatous images (GI). Each category has 15 images. Ground truth for vessel segmentation and FOV binary masks are provided for all the images in this dataset.

Experiment protocol. Since there is no previous protocol for splitting the dataset to training and testing sets, we use a three-fold cross validation for the evaluation. Each fold is composed of 15 images, 5 from each of the three categories. Therefore, for each run, we have 30 training images and 15 testing ones. In the training of each of the three runs, 10,000 positive (vessel) and 10,000 negative (non-vessel) samples are randomly selected to train a decision tree. The testing results are collected from the random forest results without further post-processing.

Parameters. For a pixel \mathbf{x} in an image I , we use a sampled pattern of size 21×45 around it to extract context information for SWT, WLD, intensity and vesselness features. The sampling step size is set to 4, resulting in $4 \times (72 + 2)$ feature values. In addition, we have a 48 dimensional Gabor feature. So the final hybrid feature vector, i.e., $\Phi(\mathbf{x})$, has a dimension of $74 \times 4 + 48 = 344$. We construct a random forest using 50 decision trees. The other parameters of random forest are the same as in the experiments with the low resolution datasets.

3.2.2 Experimental results

Similar to the experiment on low resolution images, we report *sensitivity* (SE), *specificity* (SP) and the *classification accuracy* (ACC) as evaluation metric. The comparison with

the state-of-the-art method¹ is presented in Table 5. Figure 12 shows some example results on three categories in this dataset for qualitative evaluation. It can be seen that the probability maps from our method are very close to the ground truth.

4 Conclusions

In this paper, we proposed a method for discriminative vessel segmentation by fusing hybrid features with local context modeling. In particular, new vessel features as well as traditional filter-based features extracted from orientation invariant local context are fed into a random forest framework for pixel level vessel determination. The effectiveness of the proposed method is validated through experiments on three public benchmark datasets, where our approach outperforms other state-of-the-art solutions.

In the future, we plan to investigate the use of high level semantic priors of vessel structures for further improvement. In addition, we are also interested in extending the proposed method for 3D vessel segmentation.

Acknowledgment

The authors would like to thank the reviewers for valuable suggestions to improve the paper. The work is supported in part by the NSF Grants IIS-1407156 and IIS-1350521.

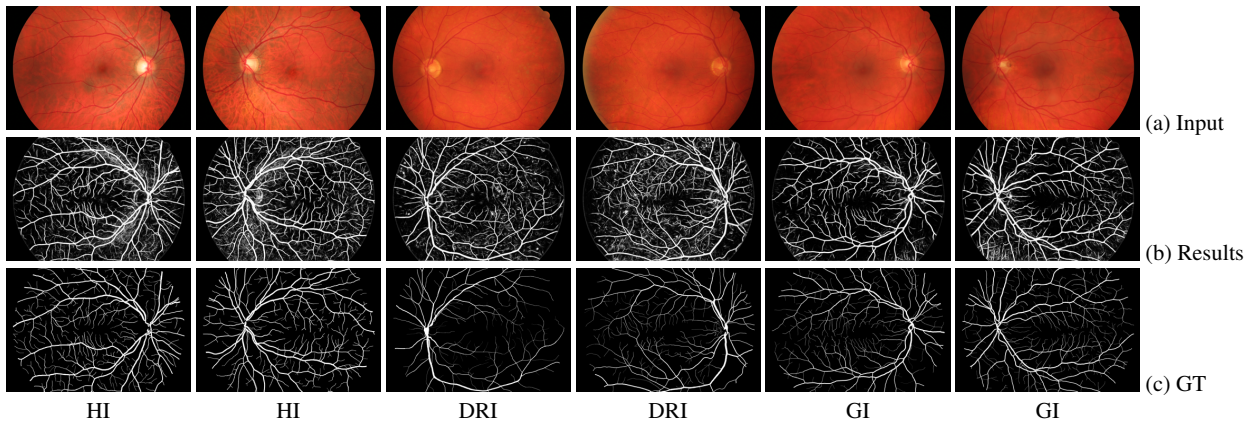
References

1. <https://www5.cs.fau.de/research/data/fundus-images/>
2. Al-Diri, B., Hunter, A., Steel, D.: An active contour model for segmenting and measuring retinal vessels. *Medical Imaging, IEEE Transactions on* **28**(9), 1488–1497 (2009)
3. Breiman, L.: Random forests. *Machine Learning* **45**(1), 5–32 (2001)
4. Cai, W., Chung, A.: Multi-resolution vessel segmentation using normalized cuts in retinal images. *Medical Image Computing and Computer-Assisted Intervention* pp. 928–936 (2006)
5. Chaudhuri, S., Chatterjee, S., Katz, N., Nelson, M., Goldbaum, M.: Detection of blood vessels in retinal images using two-dimensional matched filters. *Medical Imaging, IEEE Transactions on* **8**(3), 263–269 (1989)
6. Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., Gao, W.: Wld: a robust local image descriptor. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **32**(9), 1705–1720 (2010)
7. Cheng, E., McLaughlin, S., Megalooikonomou, V., Bakic, P., Maidment, A., Ling, H.: Learning-based vessel segmentation in mammographic images. In: *Proc. of the IEEE Int’l Conf. on Healthcare Informatics, Imaging and Systems Biology* (2011)
8. Criminisi, A., Shotton, J., Bucciarelli, S.: Decision forests with long-range spatial context for organ localization in ct volumes. In: *MICCAI Workshop on Probabilistic Models for Medical Image Analysis* (2009)

¹ Results of the state-of-the-art method are taken from [1] using the method reported in [24].

Table 5 Performance on high resolution dataset (HRFID, [24]).

	Healthy images			Diabetic retinopathy images			Glaucomatous images		
	SE	SP	ACC	SE	SP	ACC	SE	SP	ACC
Odstrcilik et al. [1, 24]	0.7861	0.9750	0.9539	0.7463	0.9619	0.9445	0.7900	0.9638	0.9497
Proposed	0.7889	0.9865	0.9647	0.6313	0.9851	0.9561	0.6921	0.9875	0.9634

**Fig. 12** Example results of the HRFID dataset. From top to bottom: input image (a), probability maps produced by our algorithm (b) and the ground truth (GT) (c).

9. Epshtein, B., Ofek, E., Wexler, Y.: Detecting text in natural scenes with stroke width transform. In: Computer Vision and Pattern Recognition, IEEE Conference on, pp. 2963–2970. IEEE (2010)
10. Frangi, A., Niessen, W., Vincken, K., Viergever, M.: Multiscale vessel enhancement filtering. Medical Image Computing and Computer-Assisted Intervention pp. 130–137 (1998)
11. Fraz, M., Barman, S., Remagnino, P., Hoppe, A., Basit, A., Uyyanonvara, B., Rudnicka, A., Owen, C.: An approach to localize the retinal blood vessels using bit planes and centerline detection. Computer methods and programs in biomedicine (2011)
12. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: Blood vessel segmentation methodologies in retinal images—a survey. Computer methods and programs in biomedicine **108**(1), 407–433 (2012)
13. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: An ensemble classification-based approach applied to retinal blood vessel segmentation. Biomedical Engineering, IEEE Transactions on **59**(9), 2538–2548 (2012)
14. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. Medical Imaging, IEEE Transactions on **19**(3), 203–210 (2000)
15. Jiang, X., Mojon, D.: Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. Pattern Analysis and Machine Intelligence, IEEE Transactions on **25**(1), 131–137 (2003)
16. Lam, B.S., Gao, Y., Liew, A.C.: General retinal vessel segmentation using regularization-based multiconcavity modeling. Medical Imaging, IEEE Transactions on **29**(7), 1369–1381 (2010)
17. Lam, B.Y., Yan, H.: A novel vessel segmentation algorithm for pathological retina images based on the divergence of vector fields. Medical Imaging, IEEE Transactions on **27**(2), 237–246 (2008)
18. Lupascu, C.A., Tegolo, D., Trucco, E.: Fabc: retinal vessel segmentation using adaboost. Information Technology in Biomedicine, IEEE Transactions on **14**(5), 1267–1274 (2010)
19. Marin, D., Aquino, A., Gegúndez-Arias, M., Bravo, J.: A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. Medical Imaging, IEEE Transactions on **30**(1), 146–158 (2011)
20. Mendonca, A.M., Campilho, A.: Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. Medical Imaging, IEEE Transactions on **25**(9), 1200–1213 (2006)
21. Miri, M.S., Mahloojifar, A.: Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction. Biomedical Engineering, IEEE Transactions on **58**(5), 1183–1192 (2011)
22. Niemeijer, M., Staal, J., van Ginneken, B., Loog, M., Abramoff, M.D.: Comparative study of retinal vessel segmentation methods on a new publicly available database. In: Medical Imaging 2004, pp. 648–656. International Society for Optics and Photonics (2004)
23. Nuzhnaya, T., Cheng, E., Ling, H., Kontos, D., Bakic, P., Megalooikonomou, V.: Segmentation of anatomical branching structures based on texture features and graph cut. In: Proc. of the IEEE International Symposium on Biomedical Imaging (2011)
24. Odstrčilík, J., Jan, J., Gazárek, J., Kolář, R.: Improvement of vessel segmentation by matched filtering in colour retinal images. In: World Congress on Medical Physics and Biomedical Engineering, September 7–12, 2009, Munich, Germany, pp. 327–330. Springer (2009)
25. Perfetti, R., Ricci, E., Casali, D., Costantini, G.: Cellular neural networks with virtual template expansion for retinal vessel segmentation. Circuits and Systems II: Express Briefs, IEEE Transactions on **54**(2), 141–145 (2007)
26. Ricci, E., Perfetti, R.: Retinal blood vessel segmentation using line operators and support vector classification. Medical Imaging, IEEE Transactions on **26**(10), 1357–1365 (2007)
27. Sinthanayothin, C., Boyce, J., Cook, H., Williamson, T.: Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. British Journal of Ophthalmology **83**(8), 902–910 (1999)
28. Soares, J.V., Leandro, J.J., Cesar, R.M., Jelinek, H.F., Cree, M.J.: Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. Medical Imaging, IEEE Transactions on **25**(9), 1214–1222 (2006)

29. Sofka, M., Stewart, C.V.: Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures. *Medical Imaging, IEEE Transactions on* **25**(12), 1531–1546 (2006)
30. Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. *Medical Imaging, IEEE Transactions on* **23**(4), 501–509 (2004)
31. You, X., Peng, Q., Yuan, Y., Cheung, Y.m., Lei, J.: Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. *Pattern Recognition* **44**(10), 2314–2324 (2011)
32. Zana, F., Klein, J.: Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *Image Processing, IEEE Transactions on* **10**(7), 1010–1019 (2001)
33. Zhang, B., Zhang, L., Zhang, L., Karray, F.: Retinal vessel extraction by matched filter with first-order derivative of gaussian. *Computers in Biology and Medicine* **40**(4), 438–445 (2010)