

A Successive Clutter-Rejection based Approach for Early Detection of Diabetic Retinopathy

Keerthi Ram, Gopal Datt Joshi, *Member, IEEE*, and Jayanthi Sivaswamy, *Member, IEEE*

Abstract—The presence of microaneurysms (MAs) is usually an early sign of diabetic retinopathy (DR) and their automatic detection from color retinal images is of clinical interest. In this paper, we present a new approach for automatic MA detection from digital colour fundus images. We formulate MA detection as a problem of target detection from clutter, where the probability of occurrence of target is considerably smaller compared to the clutter. A successive rejection-based strategy is proposed to progressively lower the number of clutter responses. The processing stages are designed to reject specific classes of clutter while passing majority of true MAs, using a set of specialized features. The true positives that remain after the final rejector are assigned a score which is based on its similarity to a true MA. Results of extensive evaluation of the proposed approach on three different retinal image datasets is reported, and are used to highlight the promise in the presented strategy.

Index Terms—Diabetic retinopathy, Microaneurysm, Clutter-rejection, Retinal image

I. INTRODUCTION

DIABETIC Retinopathy (DR) is a major public health issue since it can lead to blindness in patients with diabetes. Microaneurysms (MAs) are usually the first clinical symptom of DR. They are swellings of capillaries caused by a weakening of the vessel wall [1]. Their sizes range from $10\mu m$ to $125\mu m$ [2]. In the clinical scenario, experts rely either on direct manual examination or fluorescein fundus angiography where MAs appear with high contrast as bright white spots. Given the high cost and the cumbersome requirement of intravenous injection of a dye for this type of imaging, interest in the recent past has been on detecting MAs from a colour fundus/retinal image (CFI). In CFIs, MAs appear as tiny, reddish isolated dots. Automatic detection of MAs from digital CFIs can play an important role in DR screening at a large scale [3][4]. It can significantly reduce the workload of the ophthalmologists and the health costs in the DR screening [3].

Early published work attempted to address the problem of MA detection in fluorescein angiogram images of the retina [5][6][7][8][9]. Lay et al., [5] presented the first MA detection method for angiograms. In this method, MA candidates were obtained using top-hat transformation which eliminates the

vasculature structure from the image leaving possible MA candidates untouched. Spencer et al., [7] presented a shade correction technique and a candidate detection method using matched filtering. However, potential mortality associated with the intravenous use of fluorescein [10][4] prohibits the application of this technique for large-scale screening purposes. Instead, colour fundus imaging has emerged as a preferred modality due to its non-invasive nature [10]. Extensive clinical studies show the effectiveness of CFI for large-scale DR screening [3].

Existing methods for MA detection generally consist of two-stages where, the first stage is aimed at obtaining potential MA candidates while the second stage is used to assign MA or non-MA category to the candidate using features computed around the candidate location. The main processing components include 1) pre-processing; selection of candidate MA and 2) feature extraction; classification. The focus of the early methods has been on pre-processing and candidates selection steps. Later methods focus more on designing new sets of features and choosing of classifiers. Recently published work have re-examined the individual processing components and presented improvements on certain aspects.

Numerous algorithms have been proposed to detect early signs of DR (MAs) from CFI. The first such method was presented by Oien et al. [11]. The pre-processing used here is similar to the approach used by [5]. In later methods, a rule-based classification was added to the processing pipeline [6][8][12][13]. Usher et al., [14] employed a neural network based classification on the candidate regions obtained using recursive region growing and adaptive intensity thresholding.

Huang et al., [2] presented a local adaptive approach to extract candidates, where multiple subregions of each image were automatically analyzed to adapt to local intensity variation and properties. Niemeijer et al., [4] presented a supervised, pixel classification technique to extract red lesions to get MA candidates. A large set of features was added to the original feature set used in [6]. A knn classifier was used for MA recognition. Fleming et al., [1] presented a local image contrast normalization technique to get more discriminative features for MA. A vessel-free region is obtained around each detected candidate using watershed segmentation which is then used to enhance the contrast of candidate. A parametric model of a paraboloid is used for the MA and fitted on a set of pixels obtained by applying region growing on the candidate location. The model parameters are used to derive a new set of features for the candidate and finally classified using a knn classifier.

Walter et al., [15] used a morphological (diameter) closing technique for detecting candidates. A supervised density-based

Manuscript received May 25, 2010; revised Aug 12, 2010. This work was supported partly by the Department of Science and Technology, Govt. of India, under Grant SR/S3/EECE/17/2005.

All authors are with the Centre for Visual Information Technology, IIT Hyderabad, India, 500032. e-mail: keerthiram@research.iit.ac.in, gopal@research.iit.ac.in and jsivaswamy@iit.ac.in

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org

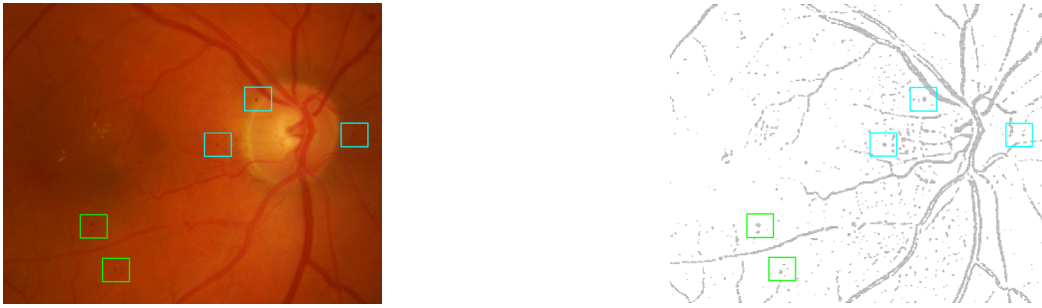


Fig. 1. Left Image: A sample region of a CFI. Green box highlights the true MA locations and magenta box shows the similar looking image noise. Right Image: A sample result for MA candidate.

classifier is used for MA classification. Quellec et al. [16] presented a method based on template matching with a generalized Gaussian template. The matching is performed in the wavelet domain to obtain MA candidates. The classification stage optimizes the selection of wavelet sub-bands in which maximum discriminative information exists for MAs versus non-MA regions.

The existing approaches have primarily given importance to robust modelling of the MA class and seek to suppress the false MA class by explicit segmentation of optic disk and vessels. Complex modeling of MA structure for candidate detection [17][16], local enhancement for illumination invariant MA features [1], use of local context/statistics and color information [1][4][15] are few attempts in this direction. This approach has, we believe, reached a saturation point as indicated by the lack of significant improvement in the detection performance [18] [19]. We think part of the reason for this scenario is the strategy followed for false MA suppression via optic disk (OD) and vessel segmentation. These segmentation problems are well researched, however the highest precision reported in literature is still inadequate for the suppression task. Consequently, there is a trade-off between detection of thin vessels and dark objects such as MAs/noise [20][21]. Alternatively it might be interesting to shift the focus to non-MA structures (clutter) and hence, we propose a strategy that gives prime importance to clutter rejection rather than signal (MA) detection.

The paper is organized as follows. Section 2 gives the motivation for a new approach and Section 3 conceptualizes it. Section 4 illustrates a system developed upon the proposed approach. Section 6 analyzes the results and draw some conclusions.

II. APPROACH FORMULATION

MAs appear as tiny, reddish isolated dots, subject to small intensity- or structure-based transformations. As mentioned above, detection of MAs is compounded by the presence of similar looking structures or image noise, leading to high number of false positives. Figure 1 shows a sample result for MA candidates. In the sample considered, many false candidates occur at vessel structures and general image background. Some unknown structures also contribute to false alarms, those highlighted in cyan arise due to camera noise.

If we consider true MAs and non-MAs (similar structures) as two classes, in a given image, the probability that a

candidate belongs to the true MA (P_T) class is substantially smaller, compared to that of belonging to non-MA class (P_C). Here, we formulate the MA detection problem as a problem of detecting a target embedded in a background clutter, where the target occurs with a much lower probability compared to the clutter ($P_T \ll P_C$). From this formulation point of view, the earlier methods can be viewed as attempts towards getting better characterization of target class followed by a classification stage.

We are interested in exploring whether knowledge of the clutter class can play a positive role in MA detection. Thus, instead of the earlier formulations where MA is the *only* object of interest, we wish to gain better understanding of objects in the clutter class, in addition to the target class.

Here, we propose to model the clutter, attempting to address the discrimination aspect early, and postpone the target modeling. Such a strategy that aims at very early clutter labeling, can be beneficial to the overall detection as this can facilitate progressive rejection of clutter responses (using many rejectors sequentially), and target recognition may be performed when fewer clutter responses remain.

In each rejection stage, responses classified as clutter can be removed from further consideration, retaining the remaining responses as putative targets. These are to be passed on to the subsequent rejector for further examination. The objective of such a cascade of rejectors is to reduce P_C while maintaining P_T .

III. PROPOSED APPROACH

Fig. 2 illustrates the proposed method where the strategy is to get a set of candidate MAs using a simple threshold, from a pre-processed image, and then culling the clutter among the candidates using a set of rejectors in cascade. Since the clutter class has multiple objects with different characteristics, the known and frequently occurring clutter objects are rejected first, and a second stage is designed to discriminate the remaining class of (largely unknown) clutter objects. In the final stage, the candidates are assigned a similarity score based on their similarity to true MAs.

The candidate selection method is kept simple since current focus is on rejection of false positives rather than acquiring good candidates. The first rejection stage is aimed at eliminating candidates originating from dark structures like vessels and hemorrhages. Candidates occurring on such structures can be well-characterized using local morphological information

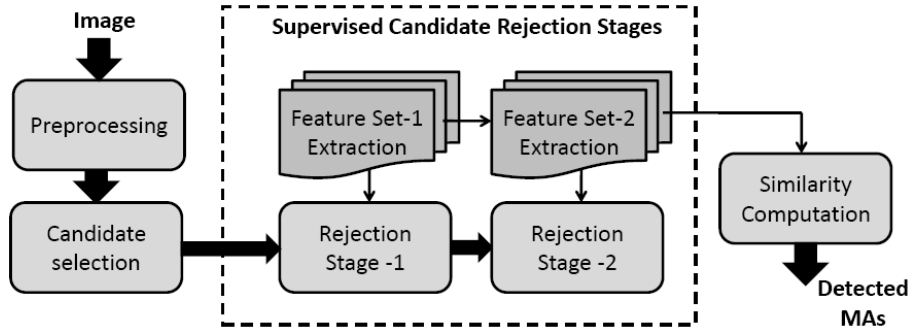


Fig. 2. Flow Diagram of the proposed approach

(for example, elongated structure co-located with the candidate indicates the possibility of the candidate occurring on a vessel). Thus a set of shape-based features and a two-class classifier are used to eliminate such clutter candidates.

The sources of remaining non-MA candidates could be due to a variety of reasons including local minima formed by image noise, region between two bright regions, optic disk, etc. Locally, these candidates have shape profile similar to a MA however can be discriminable if surrounding background region is considered. Since, it is difficult to obtain a common characterisation for this range of clutter candidates, a target-oriented rejection strategy is employed. A coarse model for MA candidates is learnt using context sensitive local features. The candidates found to be inconsistent with this model are classified into the clutter and are eliminated.

Culling of clutter by two stages results in a significant reduction in the number of reported candidates. In the final stage, we compute the degree of similarity of each remaining candidate to a true MA profile, and assign a score which ranges from $[0 - 1]$. This stage uses a composite set of features capturing true MA profile based on morphological and appearance based information. A final set of MA points can be obtained by applying a threshold on the similarity score. In the forthcoming sections, each of the processing stages is elaborated in detail.

IV. PRE-PROCESSING AND CANDIDATE SELECTION

To minimise the effect of intensity variation in the background across the image, preprocessing is performed on the green colour plane of CFI I_g . The green colour plane is chosen because it provides highest MA contrast compared to other colour planes [1] [4][16]. The background is normalised by subtracting an estimate of the background from I_g . The estimate of the background (called I_{bg}) is obtained by median-filtering I_g with a pre-defined sized kernel whose size is chosen such that it can include the widest blood vessel in the dataset.

$$I_{fg} = I_g - I_{bg} \quad (1)$$

In I_{fg} , the background is normalized and of high intensity, whereas dark structures of I_g , such as vessels and MAs have low value. This is followed by morphological bottom-hat enhancement using a disk for the structuring element, to enhance small size structures like MA, resulting in I_{bothat} .

Next, a morphology-based approach presented by [6] is used to extract linear structures in various orientations. The suprema of morphological openings (obtained with linear structuring elements of different orientations [4]) is used as the marker, and with I_{bothat} as the mask, we perform morphological reconstruction, to get I_{recon} . The final preprocessed image I_{pp} is obtained by subtracting I_{recon} from I_{bothat} , thereby suppressing linear structures. The potential candidate locations in I_{pp} have a high intensity. Since, MAs have high values in I_{pp} , an empirically chosen threshold is applied to get candidate regions. The local minima in the green colour plane of each candidate regions are used as candidates (designated as C_0) in further stages.

V. REJECTION STAGE-1 (RS_1)

The objective of RS_1 is to identify from C_0 , the known class of clutter namely, candidates on vessels, hemorrhages, vessel junctions etc. A set of features are designed to capture information and aid in discriminating above clutter candidates from true-MA. The information about the location of occurrence is extracted from each candidate using some specially designed filters, and scale-specific statistics, as explained below.

Anisotropic Filters: Vessel fragments can be modeled as elongated structures. A set of oriented filters are obtained by rotating a second-derivative of Gaussian filter [22]. The analytical expression for the second derivative in x-direction based on 1-dimensional kernels and separability, is derived using the following relationships:

$$\begin{aligned} g_\sigma(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \\ g'_\sigma(x) &= -g_\sigma(x) \times \frac{x}{\sigma^2} \\ g''_\sigma(x) &= g_\sigma(x) \times \frac{x^2 - \sigma^2}{\sigma^4} \end{aligned} \quad (2)$$

A smoothed anisotropic Gaussian second derivative filter g_{xx} is constructed using separability as:

$$g_{xx}(x, y, \sigma) = g''_\sigma(x)g_{\sigma_c}(y), \quad (4)$$

where σ_c is the standard deviation of a static 1-dimensional smoothing Gaussian function with a value of 9 pixels. Responses obtained from these filters help in discriminating

between false candidates on vessels and true MAs by way of high response to the former and low response to the latter.

A bank of filters at 6 equi-spaced orientations and 3 different scales are used at the output of which the maximum (r_m), variance (r_v) and sum (r_s) of the responses are computed. The following features are then derived for each candidate at each scale: a) ($r_s - r_m$): this difference is high for true MA locations which are characterised by high r_s (about 6 times that of r_m) compared to clutter located on vessels and b) r_v : this is expected to be low at true MA locations, and high at vessel and junction locations.

Scaled difference-of-Gaussians: A difference of Gaussian (DoG) filter acts as a blob detector, giving a high response to dark, isotropic structures. We introduce a variant of DoG, given by

$$f_d = \alpha g(\sigma_2) - g(\sigma_1) \quad (5)$$

where $\sigma_1 < \sigma_2$ and $\alpha > 0$ is a parameter controlling the height of the rim, σ_2 controls the width of the rim. At a candidate resembling a well-defined MA, this filter's response r_d is high. If a candidate lies on a vessel, r_d is low value (going negative if the vessel is thick). This is hence an informative feature for discrimination.

Inverted Gaussians: While the first filter responses help in discriminating candidates on vasculature, a second type of clutter structure that is similar to MAs are hemorrhages. In order to capture these, inverted Gaussian filters at high scale are used. These filters will maximally respond to larger objects such as hemorrhages and thick vessels in contrast to well-defined MA.

A. Feature Set-1

- $r_s - r_m$: Difference between sum and max of responses from rotated $g_{xx}(2^{(i/2)})$ for $i = 3, 4, 5$ pixels
- r_v : Variance of responses from rotated $g_{xx}(2^{(i/2)})$ for $i = 3, 4, 5$ pixels
- r_d : Response to DoG filter at $\sigma_2 = 2, 4, 6$ pixels
- $r_g(\sigma)$: Response to inverted Gaussian at $\sigma = 2, 4, 6$ pixels

B. Classifier-1

In the above feature space, true samples occupy the positive (first) hyper-quadrant and are agglomerated near the coordinate origin (have low positive values). In contrast, false samples are scattered away from the origin. We use the nearest-mean classifier, which computes the mean of the true and false training samples, and stores them as prototypes. A random selection of the vessel points from the segmented vessel map obtained by [21] is chosen as training samples while true MA samples are taken from the ground truth.

A new sample x_q is labeled by considering the distance to the prototypes and assigning the label of the nearest prototype to the new sample:

$$l_q = \arg \min(\|x_q - \mu_i\|), i = true, false \quad (6)$$

where μ_i is the prototype of class i in the training set. Figure 3 illustrates sample on OD rejected by RS_1 .

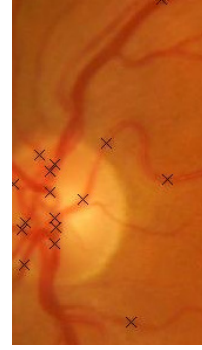


Fig. 3. Sub-image around optic disk indicating candidates rejected by RS_1

VI. REJECTION STAGE-2 (RS_2)

The function of this rejector is to identify from the candidates C_1 passed by RS_1 , the remaining class of clutter objects. These clutter arise due to a variety of reasons including image noise, poor image resolution etc. and are difficult to model. Therefore, we employ different rejection strategy where clutter candidate is defined as a sample which appears to be inconsistent with the target [23] (or abnormal). Here, we coarsely learn a model for target class using features representing isotropic nature and absolute topography of a candidate in addition to the appearance-based feature derived in RS_1 . Outliers to this model can be isolated as clutter, and rejected. Next, features used for target modeling are elaborated in detail.

Distance feature: In RS_1 , the distance between a sample x_p and the true-sample prototype (denoted as $d_{true} = \|x_p - \mu_{true}\|$) encodes a *condensed* appearance-based information about the sample. The value of d_{true} is low for candidates that are similar in appearance to well-defined MAs. It is thus carried forward to RS_2 as a feature.

Correlation features: MA structure is typically found to be isotropic in nature. A set of features to capture this information would be the correlation between a local neighborhood containing the structure, with itself after rotation. A high correlation at several orientations indicates a highly isotropic structure. The set of such correlation values is used to quantify the isotropy of the candidate.

The features are computed by correlating a square window (from I_g) around the candidate (larger than the expected size of the lesion), with rotated versions of the window. The rotation is performed about the minima of the candidate. A total of 5 equally-spaced orientations (each $\pi/5$ radians apart) is used to get 5 features. These features are denoted as $\mathbb{R}_{\theta\theta}$. This feature only encodes isotropy of a candidate and not capable of discriminating between large and small objects. Next, we derived a set of features capturing absolute topography around a candidate.

Level cuts: The local gray-scale topography around a candidate can be represented using iso-contours or level-curves of the local neighborhood considering it as a height map. Since, MAs are local minima in I_g , the level curves at a MA-like candidate can be expected to be closed curves, making it possible to perform filling within each level-curve, to obtain a finite area. We call this area a *level-cut*.

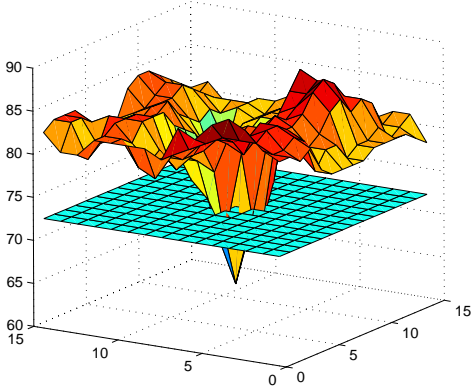


Fig. 4. Surface profile of a MA candidate with a plane at level $l = 72$ sectioning the surface.

Fig. 4 depicts the topographic surface obtained by visualizing the local gray-scale neighborhood of a candidate as a height map. A plane parallel to the ground plane (at level l) when intersecting with the surface, sections it and the intersection points define the level-curve (shown in Fig. 4). A level cut is the closed area bound by a level-curve, containing within it the coordinate of the minima. The area of a level-cut at level l_i is taken to be the number of pixels in the level cut and is denoted as $A(l_i)$.

At each candidate, the lowest and highest relevant levels, denoted as l_{min} and l_{max} , are found from the minimum and maximum gray values within a window (of radius 5) centered at the candidate minimum. M equi-spaced level cuts are chosen between these extrema and the area $A(l_i)$; $i = 1, 2, \dots, M$ of each level cut is determined and used to derive the following features:

$d_1 = l_{max} - l_{min}$: the estimated depth of the candidate grayscale topography

$l_c = \arg \max\{A(l_{i+1})/A(l_i); i = 1, 2, \dots, M\}$: this denotes the level at which the level-cut area changes significantly (by 15 pixels) at the next level.

ν : the ratio of volume of the candidate, to the volume of an inverted cone with base area $A(l_c)$, and height h :

$$\nu = \frac{V_c}{A(l_c)h/3} \quad (7)$$

$$\text{where } V_c = \sum_{i=0}^{l_c} A(l_i), \quad h = d_1 l_c/M$$

A. Feature Set-2

- d_{true} : distance of sample from μ_{true} of FS_1
- $\mathbb{R}_{\theta\theta}$: correlation of candidate with small window at 5 angles of rotation (36°)
- m/d_1 : depth of the candidate
- $A(l_1)$: area at the first level above l_0
- m/l_c : where l_c is the ‘‘rim-level’’ of the candidate
- $A(l_c)$: area of the candidate
- Γ : measure of ‘‘jump’’ defined as $\Gamma = \frac{A(l_c+1)}{A(l_c)}$

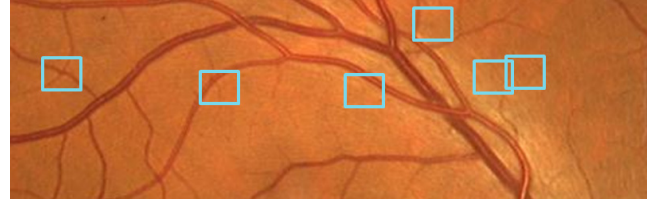


Fig. 5. Sub-image indicating candidates rejected by RS_2 .

- Ω : measure of ‘‘overflow’’ defined as $\Omega = \left. \frac{\partial A}{\partial l} \right|_{l_c} = A(l_c + 1) - A(l_c)$
- ν : Volume of the lesion relative to volume of cone of similar dimensions

B. Classifier-II

In this second feature space, the MA samples are designed to agglomerate near the origin, and clutter samples are ideally scattered. The clutter samples are thus amenable to discrimination as outliers to a model dictated by the distribution of true samples in the feature space.

We model a hyper-cuboid H around the true samples, defined by the range occupied in each feature dimension for the true samples. The true samples are taken from the ground truth available with the training images. The true samples ideally have a limited range and enclose the samples within H near the origin. Clutter samples lie outside the hyper-cuboid obtained. The dimensions of the model H are stored. A candidate obtained from RS_1 is rejected if it does not lie within H . Figure 5 shows some candidates rejected by RS_2 .

VII. SIMILARITY MEASURE COMPUTATION (L)

The rejector cascade outputs a set C_2 of candidates which are likely to be true MAs. This final module assigns a numerical similarity score to each sample in C_2 , indicating the chance of it being a true lesion. We choose to perform the score assignment by considering the signed distance of a sample from the optimal hyperplane of a two-class SVM, in feature space. A complete representation for true MA is obtained by considering features from the previous rejection stages in addition to the features encoding context and structure symmetry information which are explained next.

Context features: The following set of context features are derived, which considers the pixels within the candidate, and a context surrounding it.

- Difference in mean value of the candidate region and its surround of size (49 X 49) pixels computed in 4 spectral bands: red, green, blue and hue. $msd_j = mean_j(cand) - mean_j(surround)$, where $j = \{red, green, blue, hue\}$
- The response of the candidate to a center-surround binary filter [24] with off-center. This is used as a rough descriptor of candidate computed on a image patch centered at local minima.
- The perimeter p of the candidate, found as the number of pixels in the level curve at l_c (defined in FS_2)
- Mean response and standard deviation of derivative of Gaussian filter bank: $g_x, g_y, g_{xx}, g_{yy}, g_{xy}$ at pixels within

the candidate (5 filters at 4 scales each, resulting in 20 features; scales used are $\sigma = \{1, 2, 4, 8\}$) pixels

Symmetry features: A set of 8 features is obtained at each candidate by filtering with rotated Haar-like wavelets [25]. The vertical 2-dimensional non-standard Haar wavelet is rotated in 16 orientations (each separated by $\pi/8$) to get 16 filters. The axially anti-symmetric feature pairs capture symmetry of the candidate along different axes, and the ratio of the pair responses is used as features.

A. Feature Set-3

- $d_{true}, A(l_1), A(l_c), \Gamma, \Omega, \nu$: Taken from feature set-2.
- md_r, md_g, md_b : Difference in mean values within the candidate and its surrounding region for red, green and blue colour plane.
- md_h : Same as previous, in hue plane
- $c.s$: Response of center-surround binary filter [24]
- p : Perimeter of the candidate
- $mr(\sigma), sd(\sigma)$: Mean and standard deviation of response to Gaussian derivative filters $g_x, g_y, g_{xx}, g_{xy}, g_{yy}$ within the candidate. $\sigma = 1, 2, 4, 8$ pixels
- $s.f$: Symmetry features from non-standard Haar wavelet

B. Classifier-3

Here, we estimate the similarity score for a sample based on its distance from the optimal hyperplane of a support vector machine (SVM). A strength of SVM is its ability to handle imbalanced distributions of true and false samples [26]. Additionally, it permits the use of non-linear kernel transformations, to overcome hyperplane linearity assumption. The similarity score ψ (a function of x) obtained is such that it models a posterior probability of the two-class SVM assigning a label “true-MA” to x_q , given its feature values, i.e., $\psi(x_q) = p(y_q \leftarrow true|x_q)$. In our experimentation, we parameterised this probability score to get detection sensitivities at different false positive per image (fppi) rates.

VIII. DATASETS

For the purpose of evaluation three datasets were considered: two are the publicly available datasets namely, the *DIARETDB1* [27] and *ROC_a* [18] datasets; a custom-built dataset called *CRIAS*. Images in each dataset are divided into training and testing sets. Images in each dataset gives a range of image sizes (768 X 586 to 1500 X 1100), resolution, etc. The detailed specifications of the selected datasets is given next.

DIARETDB1 consists of 89 images, of which 5 images do not contain any DR-indicative lesions. The images were collected from a screening program and taken under a fixed imaging protocol. The images were selected by the medical experts, but their distribution does not correspond to any typical population. The ground truth supplied with this dataset is a soft map consisting of regions indicating expert consensus level information averaged from multiple experts. A bright region thus indicates high consensus about the presence of MA. According to the guidelines given with the dataset, evaluation of the presented method is done on a 75% consensus level

(relative to maximum) as the ground truth. A total of 182 MAs are obtained at 75% consensus level (i.e. majority vote between 4 experts). A test set of 68 images is formed and remaining 21 images were used in training.

ROC_a consists of 50 training images with associated ground truth, and a test set of 50 images whose ground truth is retained by the organizers of a competition [18]. The images are taken from a DR screening program across multiple sites, and hence captured with different cameras, fields of view and resolution. The images in this set are relatively heterogeneous [18]. The number of MAs in the training set is 336 and 343 in the test set [18].

CRIAS consists of 288 images collected by us from a local hospital. These images are mainly collected for clinical documentation and patient profiling. These images are of diabetic patients who already have been diagnosed with DR. Therefore, these images have high pathology occurrence and laser marks. Ground truth on these images was obtained from two experts and merged using the OR rule: a location is considered to have an MA if at least one expert has marked it. This dataset contains a total of 1436 MAs based on the above criteria which is far higher compared to the two public datasets. A training set of 100 images and test set of 188 images were created. The detailed specifications and other variability occurring in each of the selected datasets is summarized in Table I.

IX. EXPERIMENTS AND RESULTS

Various experiments were conducted with the following objectives: To assess the detection performance 1) at the lesion-level and image-level, 2) across datasets which present different kinds of challenges and 3) against other approaches, to assess the potential of the underlying approach. The second objective is motivated by the fact that all existing approaches report only on one dataset and hence no attempt has been made to address the question as to whether performances can be generalised across datasets.

A. Settings

We accommodate the variability in the given image resolution, which can affect feature values, by averaging the responses of filters with centers positioned at each of the 8-neighbors of the candidate location. The responses are given equal weight of 1 for the 8 neighbors, and 1.2 for the center.

To obtain optimal classifier settings, we use 90% MAs from the training images for training while holding out the rest for evaluation. We perform folded validation with 8 folds, and the best performing classifier setting during this evaluation was retained and used for testing. *RS₁* and *L* use two-class classifiers which require negative samples during training phase. False samples for training were generated via a random selection of the background and vessel regions of training images. Since, *RS₁* mainly focuses on outlier modeling, the ratio of true- to false-samples used for training was higher 1 : 15 compared to the *L* stage 1 : 5. For the L-stage, a radial-basis kernel was chosen: $\kappa(x_1, x_2) = \exp(-\gamma||x_1 - x_2||^2)$, and a L2-soft-margin kernel-SVM (with slack coefficient=10) [26] was trained.

TABLE I

DATASET SPECIFICATIONS. ABBREVIATIONS USED: IVW: ILLUMINATION VARIATION WITHIN IMAGES; IVA: ILLUMINATION VARIATION ACROSS IMAGES; BLA: BLURRING AND LIGHTING ARTIFACTS; CP: IMAGES TAKEN UNDER A COMMON PROTOCOL; ICT: IMAGE COMPRESSION TYPE[UC:UNCOMPRESSED, C: COMPRESSED]

	No. of Images (Training/Testing)	Imaging Factors						
		Cameras	FOV	IVW	IVA	CP	Image resolution	Mydriatic
<i>DIARETDB1</i>	89 (21/68)	fixed	50°	high	low	yes	fixed	no
<i>ROC_d</i>	100 (50/50)	varying	45°	low	medium	no	mixed	no
<i>CRIAS</i>	288 (100/188)	fixed	30° – 45°	low	high	yes	fixed	yes
	Image Quality				Pathological Category		Ground Truth	
	Clarity	Contrast	ICT	BLA	High	Mild	Type	Number of experts
<i>DIARETDB1</i>	low	low	PNG (UC)	low	no	yes	soft	multiple
<i>ROC_d</i>	medium	medium	JPG (C)	low	yes	yes	hard	4
<i>CRIAS</i>	high	high	TIF (UC)	high	yes	yes	hard	2

TABLE II

SENSITIVITIES OF THE PROPOSED METHOD AT DIFFERENT FPPI.

Dataset	FPPI						
	1	2	4	8	12	16	20
<i>DIARETDB1</i>	0.71	0.74	0.78	0.83	0.85	0.87	0.88
<i>ROC_d</i>	0.24	0.33	0.40	0.50	0.53	0.55	0.57
<i>CRIAS</i>	0.10	0.15	0.24	0.36	0.44	0.49	0.51

B. Assessments

First, the evaluation is done using both a free-response operating characteristic (FROC) and a receiver operating characteristic (ROC) as they help assess the potential of the proposed solution for two different type of applications: as an assistive MA detection tool and as a decision support tool to decide if an image is normal, respectively.

Next, an assessment of the proposed solution is done against 8 different methods which were recently analysed on the *ROC_d* dataset [18]. Dataset specifications and evaluation methodology presented in [18] are adopted to facilitate an objective comparison. Various operating points are considered to capture our method's performance at different sets of fpfi.

C. Results

Lesion-level detection: Figure 6 shows the FROC curves for our method obtained on three different datasets. These curves indicate that the best and worst performance is on *DIARETDB1* and *CRIAS* respectively. A comparable performance is achieved on *ROC_d* and *CRIAS* beyond 30 fpfi. The sensitivity is good on *ROC_d* relative to *CRIAS*, even for lower values of fpfi. The maximum sensitivity achieved varies with the dataset due to the limitation imposed by the candidate selection scheme. As mentioned in section VIII, *DIARETDB1* contains images of equal size and resolution with minimal inter-image variability unlike the *ROC_d*. Hence, there is only a small variation among true MAs in this dataset. This facilitates better candidate selection, therefore highest sensitivity S_{max} of 88.46% (at 18.02 fpfi) was achieved for *DIARETDB1*. Whereas, *CRIAS* dataset contains images from the patients having confirmed DR. It has only small variations in imaging, but poses challenges in terms of the high number of hemorrhages of varying sizes (some being similar to MAs) and other lesions in the images. The first row in fig.8 shows a false negative example missed by the

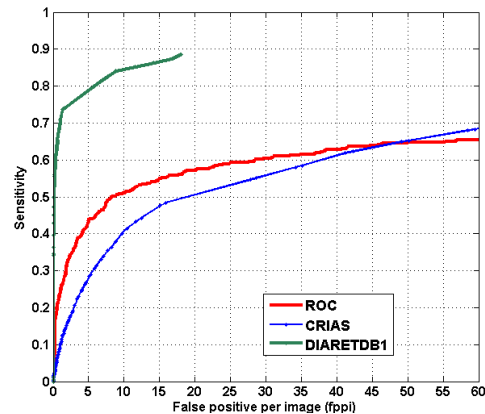


Fig. 6. FROC curves: Lesion-level performance on 3 datasets.

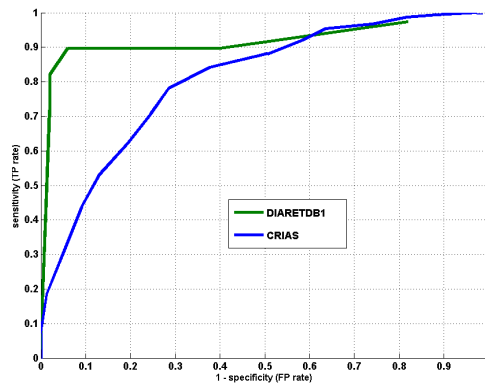


Fig. 7. ROC curves: Image-level performance on *DIARETDB1* and *CRIAS*.

candidate detection stage, due to its size being larger than a typical MA.

For a closer look at the FROC curves, sensitivities obtained at a set of particular fpfi rates are given in Table II.

The end-to-end performances on the individual dataset can be understood by examining the performance at individual stages. In *DIARETDB1*, the candidate selection stage has sensitivity above 0.90 and detects, on average, 80 candidates per image. Both RS1 and RS2 altogether reject 60% of the candidates and retain an average sensitivity of 0.88. The L-stage receives, on average, 48 candidates per image. Comparatively,

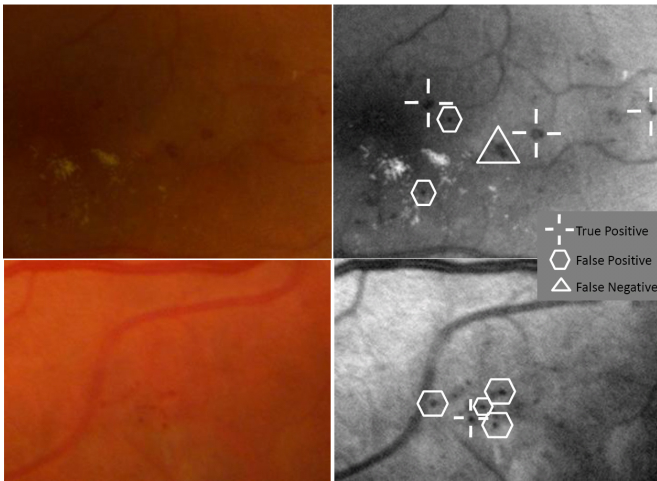


Fig. 8. Sample detection results with ground truth. Column 1 shows original colour images and column 2 highlights detection on normalised green colour plane.

the overall rejection achieved on ROC_d and $CRIAS$ was poor. In $CRIAS$, for example, the candidate selection stage has a sensitivity of 0.79 and produces 157 candidates per image. Both RS stages reject 45% candidates while maintaining a sensitivity 0.75, passing, on average, 87 candidates per image to the L-stage. Thus, due to poor clutter suppression in ROC_d and $CRIAS$, the L-stage received twice as many candidates (as in $DIARETDB1$) and incurs a higher cost in terms of fppi, for attaining a similar sensitivity as $DIARETDB1$. This manifests itself in the form of a slow rise in the FROC curves for ROC_d and $CRIAS$.

The higher cost incurred is due to the following. By design, the false candidates passed on to L stage will be hard to classify, since easier false candidates (such as local intensity minima formed due to hard exudate clusters and discontinuity in a vessel segment) would have been rejected in the earlier stages. The second row of Fig. 8 shows instances of ambiguous false positives. The ambiguity is in the image characteristics, based on which it is difficult to distinguish between a false and true MA. This distinction is possible only by incorporating high-order information about the local surround like the presence of exudate cluster or part of a discontinued vessel segment.

Image-level detection was carried out as follows: An image is declared abnormal if it contains at least one MA. Figure 7 shows the ROC curves obtained on two datasets. It can be seen that the proposed method achieves sensitivity of 0.90 on $DIARETDB1$ and 0.45 on $CRIAS$ at specificity of 0.90. The difference in lesion- and image-level sensitivity is due to the fact that the method does not require detecting every lesion in an image having multiple instances of MAs. The image-level performance is not reported on ROC_d since ground truth is not available on the test set.

Comparison against other methods: A detailed comparative analysis of 8 different detection methods including ours (IRIA-Group), is reported in [28]. The evaluation criterion used is a score obtained by averaging the sensitivity values at a set of fppi rates. Likewise, we consider two sets of fppi: the

TABLE III
SCORES OBTAINED ON ROC_d BY DIFFERENT METHODS [18] [28] AT TWO SETS OF FPPI RATES.

Method	At OP_1	At OP_2
Niemeijer et al. [4]	0.395	0.558
LaTIM [16]	0.381	0.565
GIB Valladolid [29]	0.322	0.514
OKmedical [30]	0.357	0.502
Proposed Method	0.264	0.503
Fujita Lab [31]	0.310	0.468
ISMV [32]	0.256	0.438
Waikato [28]	0.206	0.355

first set (OP-1:fppi: [0.125, 0.25, 0.5, 1, 2, 4, 8]) to capture the performance in the rising part of the FROC curve [18] while the second set (OP-2: fppi: [2, 4, 8, 12, 16, 20, 24]) to largely capture the plateau region.

Table III shows the scores reported for 8 different methods [18]. The methods are ordered according to the obtained average performance on OP-1 and OP-2. It can be seen that for OP-1, the score obtained by the proposed method is low due to its slow rising behavior as mentioned earlier. Whereas, for OP-2, the score is competitive. A poor score at OP-1 is mainly due to the inability of the L stage to handle the ambiguity between the target and the clutter.

Recently, one of the 8 methods (namely [30][33]) included in Table III has also been tested on $DIARETDB1$. A proper comparison of our method with this report is difficult since the results have been reported in [30] only on 11 images which are said to be randomly selected out of the full set of 89 images in this dataset. The average sensitivity obtained at fppi: [0.5, 1, 2] is reported to be 0.713 for these 11 images. In comparison, the proposed method has an average sensitivity of 0.706 on similar fppi rates however tested on a much larger set of 68 images (test set). It is noteworthy that the scores of the proposed method and [30] differed more significantly (score of 0.264 compared to 0.357) on ROC_d .

From the above results, we conclude that the performance achieved by the proposed method on $DIARETDB1$ indicates a promise in the approach. Further, the lesion-level performance of a method on a dataset may be an insufficient basis to predict its general performance across multiple datasets.

X. CONCLUSIONS

In this work, we formulated MA detection as a *target detection in clutter* problem and proposed a successive clutter-rejection solution for MA detection. The rejection stages are formulated based on the occurrence frequency and discriminability of the underlying clutter. A new set of morphological and appearance-based features are introduced to characterize the clutter and MA structures. In earlier approaches, a single classification step is employed to distinguish between the target and multiple sub-groups of clutter class together which demands design complexity in feature space and underlying classifier. The proposed processing pipeline separates both classes and further allows various sub-groups of clutter class to be handled through a cascade solution. This gives flexibility to achieve independent optimal solution for individual sub-tasks

and helps simplify the final classification step (L stage) by reducing the number of clutter sub-groups to be handled.

The assessment of the proposed method demonstrated that it performs well in image-level detection of MAs across 2 datasets. At the lesion-level, the performance is variable with the dataset. Experiments performed at lesion-level detection indicate scope for improvement in performance across datasets and also highlight the difficulties that arise in candidate selection due to the variability found across datasets. Advanced machine learning approaches such as kernel-based approaches and training regimes with larger training dataset are some solutions that can be explored.

Overall, our experimental evaluation on three different datasets offers some insights about challenges in developing automated solutions for MA detection: a) In the current scenario, automated screening solution is easier to realise if a fixed protocol is used to acquire images as in *DIARETDB1*. This however, may be practically difficult; b) In order to develop robust solutions (and obtain a predictably consistent performance for an MA detection algorithm), a large, heterogeneous dataset is needed. Hence, more effort needs to be directed towards building large benchmark datasets composed of images acquired in different settings such as different cameras, imaging protocol, population etc. This will give a common platform to evaluate methods and make evaluation close to a real screening scenario [34].

REFERENCES

- [1] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Trans. Medical Imaging*, vol. 25, no. 9, pp. 1223–1232, 2006.
- [2] K. Huang and M. Yan, "A local adaptive algorithm for microaneurysms detection in digital fundus images," in *Proc. CVBIA*, 2005, pp. 103–113.
- [3] M. Abramoff, M. Niemeijer, M. Suttorp-Schulten, M. Viergever, S. Russell, and B. van Ginneken, "Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes," *Diabetes Care*, vol. 31, no. 2, pp. 193–198, 2008.
- [4] M. Niemeijer, B. van Ginneken, J. Staal, M.S.A. Suttorp-Schulten, and M.D. Abramoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Trans. Medical Imaging*, vol. 24, no. 5, pp. 584–592, 2005.
- [5] B. Lay, "Analyse automatique des images angiofluorographiques au cours de la rétinopathie diabétique," *Ph.D. Thesis, Centre of Mathematical Morphology, Paris School of Mines, June*, 1983.
- [6] T. Spencer, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "An image processing strategy for the segmentation and quantification in fluorescein angiograms of the ocular fundus," *Comput. Biomed. Res.*, vol. 29, pp. 284–302, 1996.
- [7] T. Spencer, R. P. Phillips, P. Sharp, and J. Forrester, "Automated detection and quantification of microaneurysms in fluorescein angiograms," *Graefes' Archive for Clinical and Experimental Ophthalmology*, vol. 230, pp. 36–41, 1991.
- [8] A. Frame, P. Undrill, M. Cree, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms," *Comput. Biol. Med.*, vol. 28, pp. 225–238, 1998.
- [9] M. Cree, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "A fully automated comparative microaneurysm digital detection system," *Eye*, vol. 11, pp. 622–628, 1997.
- [10] L. Yannuzzi, K. Rohrer, and L. Tindel, "Fluorescein angiography complication survey," *Ophthalmology*, vol. 93, pp. 611–617, 1986.
- [11] G. E. Oien and P. Osnes, "Diabetic retinopathy: Automatic detection of early symptoms from retinal images," in *Proc. Norwegian Signal Process. Symp.*, 1995, pp. 135–140.
- [12] A. M. Mendonca, A. J. Campilho, and J. M. Nunes, "Automatic segmentation of microaneurysms in retinal angiograms of diabetic patient," in *Proc. Int. Conf. Image Anal. Process.*, 1999, pp. 728–733.
- [13] I. Autio, J. Borra, I. Immonen, P. Jalli, and E. Ukkonen, "A voting margin approach for the detection of retinal microaneurysms," in *Proc. Visualization, Imaging and Image Processing*, 2005.
- [14] D. Usher, M. Dumskyj, M. Himaga, T. Williamson, S. Nussey, and J. Boyce, "Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening," *Diabet. Med.*, vol. 21, pp. 84–90, 2004.
- [15] T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, and J. Klein, "Automatic detection of microaneurysms in color fundus images," *Medical Image Analysis*, vol. 11, no. 6, pp. 555–566, 2007.
- [16] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microaneurysms in retinaphotographs," *IEEE Trans. Medical Imaging*, vol. 27, no. 9, pp. 1230–1241, 2008.
- [17] A. Bhalerao, A. Patanaik, S. Anand, and P. Saravanan, "Robust detection of microaneurysms for sight threatening retinopathy screening," in *Proc. ICVGIP*, 2008, pp. 520–527.
- [18] M. Niemeijer, B. van Ginneken, M. Cree, A. Mizutani, G. Quellec, C.I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. M. Q. Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garcia, H. Fujita, and M. Abramoff, "Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs," *IEEE Trans Med Imaging*, vol. 1, no. 29, pp. 185–195, 2010.
- [19] M. D. Abramoff, M. Niemeijer, and S. R. Russell, "Automated detection of diabetic retinopathy: barriers to translation into clinical practice," *Expert Review of Medical Devices*, vol. 7, no. 2, pp. 287–296, 2010.
- [20] 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 Trans Med Imaging*, vol. 23, no. 4, pp. 501–509, 2004.
- [21] S. Garg, J. Sivaswamy, and S. Chandra, "Unsupervised curvature-based retinal vessel segmentation," *Proc. ISBI*, pp. 344–347, 2007.
- [22] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International Journal of Computer Vision*, vol. 43, no. 1, pp. 29–44, 2001.
- [23] V. Barnett and T. Louis, *Outliers in statistical data*. John Wiley & Sons, 1994.
- [24] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proc. Int. Conf. Image Processing*, vol. 1, 2002, pp. I: 900–903.
- [25] C. Papageorgiou, M. Oren, and T. Poggio, "A general framework for object detection," in *Proc. Int. Conf. Computer Vision*, 1998, pp. 555–562.
- [26] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [27] T. Kauppi, V. Kalesnykiene, J. K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Klviinen, and J. Pietil, "DIARETDB1 diabetic retinopathy database and evaluation protocol," *Proc. Medical Image Understanding and Analysis*, no. 2, 2007.
- [28] *Retinopathy Online Challenge Website*. The University of Iowa and the ROC organizers; <http://roc.healthcare.uiowa.edu/results.php>.
- [29] C. I. Sanchez, R. Hornero, A. Mayo, and M. Garcia, "Mixture model-based clustering and logistic regression for automatic detection of microaneurysms in retinal images," in *Proc. SPIE Medical Imaging 2009: Computer-Aided Diagnosis*, vol. 7260, 2009.
- [30] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, "Detection of microaneurysms using multi-scale correlation coefficients," *Pattern Recognition*, vol. 43, pp. 2237–2248, 2010.
- [31] A. Mizutani, C. Muramatsu, Y. Hatanaka, S. Suemori, T. Hara, and H. Fujita, "Automated microaneurysm detection method based on double ring filter in retinal fundus images," in *Proc. SPIE Medical Imaging 2009: Computer-Aided Diagnosis*, vol. 7260, 2009.
- [32] L. Giancardo, F. Meriaudeau, T. P. Karnowski, K. Tobin, Y. Lic, and E. C. MD, "Microaneurysms detection with the radon cliff operator in retinal fundus images," *Proc. SPIE Medical Imaging*, vol. 7623, p. 29, 2010.
- [33] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, "Hierarchical detection of red lesions in retinal images by multiscale correlation filtering," in *Proc. SPIE Medical Imaging 2009: Computer-Aided Diagnosis*, vol. 7260, 2009.
- [34] R. Winder, P. Morrow, I. McRitchie, J. Bailie, and P. Hart, "Algorithms for digital image processing in diabetic retinopathy," *Comput Med Imaging Graph.*, 2009.