

Automatic Detection of Red Lesions in Digital Color Fundus Photographs

Meindert Niemeijer*, Bram van Ginneken, *Member, IEEE*, Joes Staal, *Member, IEEE*, Maria S. A. Suttorp-Schulten, and Michael D. Abràmoff, *Member, IEEE*

Abstract—The robust detection of red lesions in digital color fundus photographs is a critical step in the development of automated screening systems for diabetic retinopathy.

In this paper, a novel red lesion detection method is presented based on a hybrid approach, combining prior works by Spencer *et al.* (1996) and Frame *et al.* (1998) with two important new contributions. The first contribution is a new red lesion candidate detection system based on pixel classification. Using this technique, vasculature and red lesions are separated from the background of the image. After removal of the connected vasculature the remaining objects are considered possible red lesions. Second, an extensive number of new features are added to those proposed by Spencer–Frame. The detected candidate objects are classified using all features and a k-nearest neighbor classifier.

An extensive evaluation was performed on a test set composed of images representative of those normally found in a screening set. When determining whether an image contains red lesions the system achieves a sensitivity of 100% at a specificity of 87%. The method is compared with several different automatic systems and is shown to outperform them all. Performance is close to that of a human expert examining the images for the presence of red lesions.

Index Terms—Computer-aided diagnosis, fundus, microaneurysms, pixel classification, red lesions, retina, screening.

I. INTRODUCTION

DIABETIC RETINOPATHY (DR) is the leading cause of blindness in the working population of the western world. It is an eye disease which in some form afflicts 17% of diabetic subjects five years after diagnosis of diabetes and 97% 15 years after diagnosis [3]. Early diagnosis through regular screening

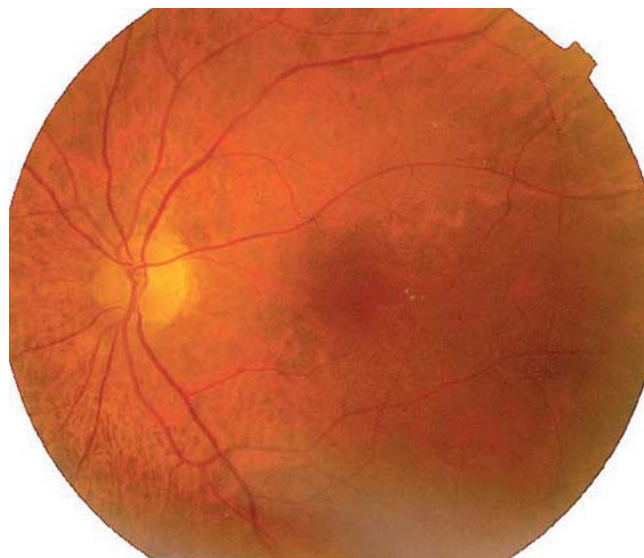


Fig. 1. A digital color fundus photograph containing red lesions. For a detailed analysis of this image see Fig. 5.

and timely treatment has been shown to prevent visual loss and blindness. Digital color fundus photography allows acquisition of fundus images (see Fig. 1) in a noninvasive manner which is a prerequisite for large scale screening.

In a DR screening program, the number of fundus images that need to be examined by ophthalmologists can be prohibitively large. The number of images without any sign of DR in a screening setting is typically over 90%. Therefore, an automated system that can decide whether or not any signs suspicious for DR are present in an image can improve efficiency; only those images deemed suspect by the system would require examination by an ophthalmologist. The method described in this paper is intended to be a first step toward such a prescreening system.

Signs of DR include red lesions such as microaneurysms and intraretinal hemorrhages, and white lesions, such as exudates and cottonwool spots. This paper concerns only the red lesions, which are among the first unequivocal signs of DR. Therefore, their detection is critical for a prescreening system.

Previously published methods for the detection of red lesions have focussed on detecting microaneurysms in fluorescein angiography images of the fundus [1], [2], [4], [5]. In this type of image the contrast between the microaneurysms and background is larger than in digital color photographs. However, a mortality of 1:222 000 associated with the intravenous use of fluorescein [6] prohibits the application of this technique for large-scale screening purposes.

Manuscript received April 29, 2004; revised December 24, 2004. The work of M. Niemeijer was supported in part by the Dutch Ministry of Economic Affairs under Grant IOP IBVA02016. The work of M. D. Abràmoff was supported by a K-12 Career Development Award from the University of Iowa College of Medicine, an unrestricted grant from Research to Prevent Blindness, New York, NY, and by the Netherlands Organization for Health Research and Development. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was E. Krupinski. Asterisk indicates corresponding author.

*M. Niemeijer is with Image Sciences Institute—Q0S.459, Heidelberglaan 100, 3584CX, Utrecht, The Netherlands (e-mail: meindert@isi.uu.nl).

B. van Ginneken and J. Staal are with the Image Sciences Institute, 3584CX, Utrecht, The Netherlands.

M. S. A. Suttorp-Schulten is with the Vrije Universiteit Medical Center, Department of Ophthalmology, De Boelelaan 1117, 1081HV, Amsterdam, The Netherlands.

M. D. Abràmoff is with the Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, Iowa City, IA 52242 USA and also with the Department of Veterans Affairs, Iowa City VA Medical Center, Iowa City, IA 55242 USA.

Digital Object Identifier 10.1109/TMI.2005.843738

The detection method described by Spencer, Cree, Frame, and co-workers [1], [2], [4], [5] employed a mathematical morphology technique that eliminates the vasculature from a fundus image yet left possible microaneurysm candidates untouched. This approach was first suggested by Baudoin *et al.* [7]. It has also been applied, in a modified version, to high-resolution red-free fundus photographs by Hipwell *et al.* [8]. These last authors present a sensitivity of 85% and a specificity of 76% on a per image basis.

A number of other approaches for the detection of red lesions in color fundus photographs have also been described. Gardner *et al.* [9] used a neural network to detect both hemorrhages and exudates. Each image was divided into 20×20 pixel grids, these were then individually classified. The per image results showed a sensitivity of 88.4% and a specificity of 83.5%.

Sinthanayothin *et al.* [10] applied a recursive regiongrowing procedure to segment both the vessels and red lesions in a fundus image. Next, a neural network was used to detect the vessels exclusively. The objects that remain after removal of the detected vasculature are labeled as microaneurysms. The evaluation was carried out on 10×10 pixel grids and not for individual images or lesions. A sensitivity of 77.5% and specificity of 88.7% was reported.

Results of a commercially available automatic red lesion detection system were given by Larsen *et al.* [11]. However, their method was not described. The system had a sensitivity of 93.1% and specificity of 71.4% on a per patient basis.

The aim of this paper is to develop an automated method which can detect images containing red lesions with a very high sensitivity and a reasonable specificity. This allows the method to be used in a prescreening setting. We address the problem of detecting red-lesions in three stages. First, each image is preprocessed; next, candidate objects that may represent red-lesions are extracted; and in the final stage the probability for each candidate to represent a red-lesion is estimated using a classifier and a large set of specifically designed features. To take advantage of prior works, and to enable comparisons between our system and previous approaches, the system of Spencer *et al.* [1] together with the extensions to this system proposed by Frame *et al.* [2] was taken as our starting point. The Spencer–Frame system employs a similar processing pipeline.

Two novel methodological contributions are made in this paper. First, a new candidate detection scheme based on pixel classification is proposed. The performance of this scheme is compared to the detection scheme based on mathematical morphology as used by Spencer–Frame. A hybrid scheme that combines both detection strategies is tested as well. Second, to improve the classification of the candidate objects into lesion and nonlesion, an extensive number of features has been added to the set proposed by Spencer–Frame. Several classifiers have been tested and best results were obtained with k-nearest-neighbors (kNN) classification.

The system is tuned and trained on a set of 50 photographs representative of those used in a screening setting, and tested on another, completely independent set of 50 photographs. An experienced ophthalmologist (MDA) carefully indicated all red-lesions in these images to provide a reference standard. A second experienced ophthalmologist (MS) indicated all

red-lesions in the test set to enable comparison between the automatic systems' and human performance.

The setup of the article is as follows. Section II describes the preprocessing stage. In Section III the candidate object detection systems are described. The mathematical morphology based method of Spencer–Frame is reviewed, the pixel classification based scheme is described and it is shown how to combine the output of both systems. This section also motivates the development of the pixel classification method and explains how it avoids some of the limitations of the other system. Section IV contains details regarding the candidate classification stage and discusses the classifier and the feature set we propose. A description of the used material is given in Section V. In Section VI the experiments and results are presented. We end with a discussion and conclusion in Section VII.

II. IMAGE PREPROCESSING

In order to prepare the fundus images for red lesion extraction some image preprocessing is performed on the green-plane I_{green} of the original RGB color image I_{org} . As red lesions have the highest contrast with the background in the green color plane [12] information from the red and blue color planes will not be used in this step.

In this and the following sections those parts of the system proposed by Spencer *et al.* and Frame *et al.* [1], [2] use the original parameter values wherever possible. However, because of differences in image modalities some of the parameters had to be adjusted. This is mentioned in the text where necessary.

A. Spencer–Frame

Fundus photos often contain an intensity variation in the background across the image, called “vignetting” [13] (see Fig. 1). This effect can deteriorate the performance of the candidate selection system. Any slow gradients in the background of I_{green} were removed, resulting in a “shade corrected” image I_{sc} . This was accomplished by estimating the background image I_{bg} and subtracting that from I_{green} . I_{bg} is produced by median filtering I_{green} with a 25×25 pixel kernel. The size of the median filter was chosen such that it is wider than the widest blood vessel in our dataset of images. A typical result of the operation is shown in Fig. 2(b).

B. Bright Lesion Removal

We extended the Spencer–Frame system in the following manner. As mentioned, bright pathologies can also appear in retinas affected by DR. These lesions can have well defined sharp edges. When they lie close together, small “islands” of normal retina are formed between them. These can be picked up as false positives (FPs) by the candidate object detection in the pixel classification based system in later steps. To prevent this, these bright pathologies were removed as follows.

The shade corrected image I_{sc} has negative values for all pixels which have an intensity lower than the background. By removing all pixels with a positive value from the image, bright pathologies no longer influence the later analysis. All pixels in

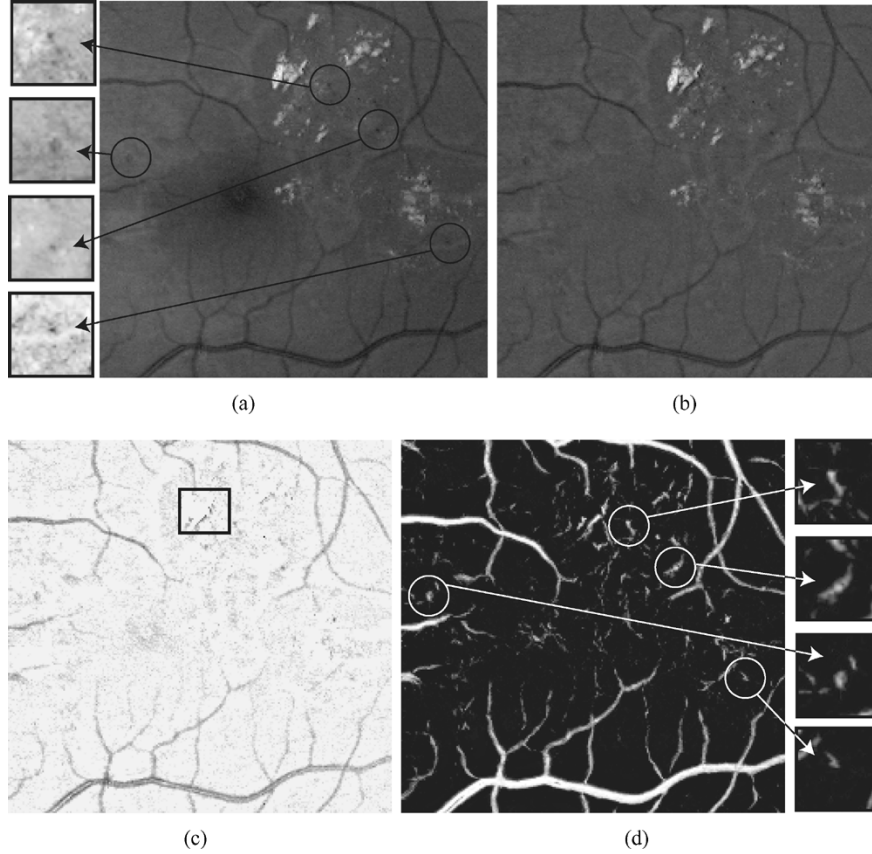


Fig. 2. (a) Part of the green color plane of a fundus image. Shown are pieces of vasculature and several red lesions. The bright lesions called exudates are also a symptom of DR. Circles mark the location of some of the red lesions in the image. (b) After subtracting a median filtered version of the green plane large background gradients are removed. (c) All pixels with a positive value are set to zero to eliminate the bright lesions in the image. Note that the exudates often partially occlude red lesions. The nonoccluded parts of the red lesions show up clearly in this image. An example of this is marked with a rectangle. (d) The pixel classification result produced by the contrast enhancement step. The nonoccluded parts of the hemorrhages are visible together with the vasculature and a number of red lesions.

I_{sc} with a positive value were set to zero resulting in the pre-processed image I_{pp} . The effect of this operation is shown in Fig. 2(c).

III. CANDIDATE EXTRACTION SYSTEM

Next, objects that are potential red lesions are extracted from the preprocessed images. These objects will be called candidate objects.

A. Mathematical Morphology Based Candidate Extraction

This system is described by Spencer *et al.* [1] and Frame *et al.* [2] and is briefly reviewed here. After the image preprocessing stage the actual candidate objects were extracted from I_{sc} . To discriminate between circular, nonconnected red lesions and the elongated vasculature a morphological top-hat transformation [7] was used. This operation is based on morphologically opening the image with a linear structuring element at different orientations. A total of twelve rotated structuring elements were used with a radial resolution of 15° . The length of a structuring element should be such that it is larger than the biggest red lesion present in the set. It was determined empirically that a length of 9 pixels gave the best balance between vessel segmentation and red lesion detection for our set. In each of the 12 opened images only those parts of the vasculature in which the linear structuring

element can fit remain. By taking the maximum pixel value at each pixel location in all 12 images a map of only the vasculature was obtained. This vasculature map was subtracted from I_{sc} , the result was an image I_{lesion} containing mainly nonelongated structures such as red lesions.

To enhance the contrast between background and red lesions in I_{lesion} a matched filter was used. The matched filter is a 2-D Gaussian with $\sigma = 1$ pixel and has a size of 11×11 pixels. Next the filtered image I_{match} was thresholded to produce a binary image I_{bin} . The threshold was fixed at a certain level above the modal value of the image. In this implementation 4 was used as the threshold. The extracted binary objects are not a good representation of the pathologies as found in I_{org} . A region growing procedure was used to grow back the original pathologies. The darkest pixel under each of the binary objects serves as the starting point.

The estimated background image I_{bg} obtained in the preprocessing step (see Section II) can be used to find the threshold t for the region growing procedure as follows:

$$t = i_{seed} - x(i_{seed} - i_{bg}) \quad (1)$$

where i_{seed} is the intensity at the starting position, i.e., the pixel under the binary object with the lowest gray value. Then i_{bg} is the intensity of the same pixel in the background image and

$x \in [0, 1]$. Here, $x = 0.5$, is similar to Spencer–Frame. Growing starts in the seed pixel and stops when no more connected pixels below the threshold can be found. The grown objects together form the final candidate object set.

B. Pixel Classification Based Candidate Extraction

A limitation of the previously described candidate extraction technique is that any red lesions which are larger than the linear structuring element cannot be detected. When the length of the structuring element is increased to be able to detect larger objects, the vessel segmentation deteriorates leading to more spurious candidate objects being detected on the vasculature. We have largely removed this limitation by using a pixel classification based method. This approach can detect larger candidate objects and reduces the number of spurious candidate objects on the vessels by integrating a vessel segmentation technique [14] with red lesion detection. Both vasculature and possible red lesions are extracted at once, and subsequently the vasculature is separated from the red lesion candidate objects.

The digital color photographs used in this research have a relatively low resolution, such that small red lesions typically have areas as low as four pixels, with relatively low contrast. Obviously this is not an ideal situation, but such is the image quality of data used routinely for DR screening. With pixel classification the contrast between foreground pixels (red lesions and vasculature) and pixels in the background can be improved such that a global threshold can be used to extract all relevant objects in the image.

Pixel classification is a supervised method, so it needs to be trained using example pixels. Each of the example pixels was extracted from a training set for which a labeled reference standard is available. All pixels in the training set are, thus, assigned a label and a feature vector. A classifier can now establish a decision boundary in the feature space which will optimally separate the two classes (i.e., foreground and background pixels).

A large number of sample pixels of red lesions, vasculature as well as background pixels are needed for training. A set of 40 fundus images (Set 1, see Section V), not used in the development or testing of any other part of the system, provided the training pixels. In all of the images the vasculature and red lesions were manually segmented. Each of these images was pre-processed as described in Section II before training the pixel classification procedure. This training procedure needs to be completed only once, after it has finished the system is ready to automatically process unseen images.

An important observation is the similarity, on a pixel level, in color and response to first and second order Gaussian filters for the red lesions and the vasculature. To obtain the feature vectors, filter outputs at each pixel location of I_{pp} were used. These filters consisted of all Gaussian derivatives up to and including second order at scales $\sigma = 1, 2, 4, 8, 16$ pixels. Also the pixel intensities taken from I_{pp} were added to the feature vector. Preliminary experiments comparing several classifiers showed that a kNN classifier [15] produced the best results. For all experiments, $k = 55$.

The kNN classifier (with $k > 1$) allows for soft-classification in which each pixel is assigned a probability of it being a foreground pixel. To accomplish this, all k neighbors in

the feature space of a query pixel were examined. When n neighbors were labeled as being a foreground pixel, the posterior probability that the query pixel is a foreground pixel itself p was determined by

$$p = \frac{n}{k}. \quad (2)$$

For the kNN classifier the optimized implementation as provided by Arya *et al.* [16] was used. Fig. 2(a) shows part of the green plane of a fundus image, its probabilistic map is shown in Fig. 2(d).

To extract possible candidate locations the probabilistic map produced by the pixel classification was thresholded. To determine this threshold the maximum pixel value in the probability map under each of the objects marked in the ground truth was examined. This gave an indication of the maximum amount of objects that could be detected at a certain threshold. Further experimentation showed that 0.4 gave a good balance between the number of detected spurious objects and the number of detected red lesions. After thresholding, a binary image containing any red lesions and a large part of the vascular tree was obtained. Because red lesions in general do not appear on larger (visible) vessels they are disconnected from the vasculature. To obtain possible candidate locations, connected component analysis was applied on the binary objects. Any object which was too large to be a red lesion was removed. To determine a size threshold for this operation a histogram was made of the sizes of all lesions in the reference standard segmentation of the training set. A threshold of 300 pixels was found to include 98% of all red lesions. Most of the vasculature is connected, forming objects larger than 300 pixels and will, thus, be removed by this step. What remains are a number of small vessel fragments and those red lesions not connected to the vasculature. In Fig. 3, a binary example image and the remaining connected components are shown.

As with the mathematical morphology based approach the extracted objects are not a good representation of the actual pathologies and objects in the original image. Therefore, the same region growing technique as described by Spencer–Frame (see Section III-A) was used to determine the actual shape of the objects.

The objects which were produced by the region growing procedure are then postprocessed such that any holes in the grown objects are filled. This post processing was done to facilitate feature extraction. The post processed objects together form the final set of candidate objects.

C. Hybrid Candidate Object Extraction

A number of experiments were conducted showing similar performance of both the pixel classification and the mathematical morphology based system in terms of detected lesions. However, when the candidates of both systems were combined a higher number of true positive (TP) candidate objects was extracted. The combination was done by putting all candidate objects in a set. Then all objects were checked to see if they overlap with another object in the set. In case of overlap one of the objects is removed. When both candidate detection systems detect the same lesion it does not matter which object

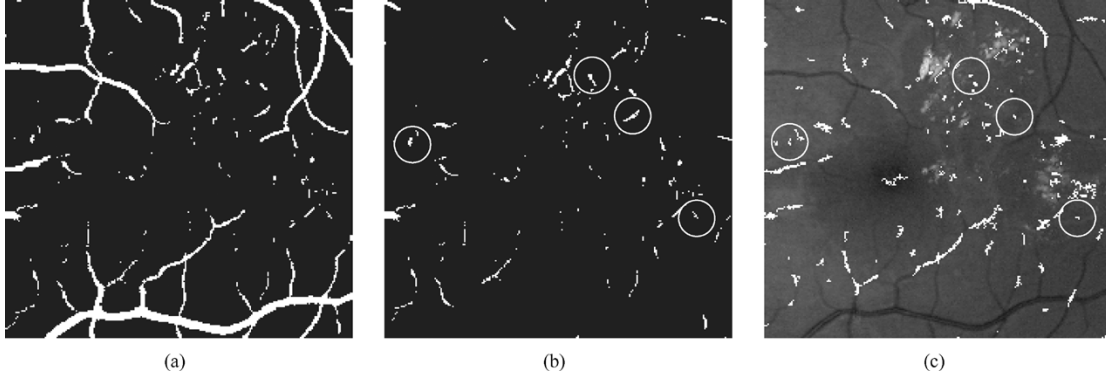


Fig. 3. (a) The thresholded probability map. (b) The remaining objects after connected component analysis and removal of the large vasculature. (c) The shape and size of the extracted objects in (b) does not correspond well with the actual shape and size of the objects in the original image. A final region growing procedure is used to grow back the actual objects in the original image which are shown here. In (b) and (c), the same red lesions as in Fig. 2(a) are indicated with a circle.

is removed because the segmentation method used in both approaches is the same. Overlap also often occurs when a region growing seed point is placed in a normal area of the retina (e.g., due to noise), the region growing procedure will grow into neighboring red lesions and vessels. To exclude these relatively large spurious objects we always remove the largest of two overlapping objects.

IV. CANDIDATE CLASSIFICATION SYSTEM

The candidate extraction step results in a set of candidate objects. The purpose of the candidate classification system is to classify each of these objects as either a red lesion or a non red lesion. In order to accomplish this another classifier needed to be trained using example objects. These example objects are first extracted from a training set (part of Set 2, see Section V) using the candidate object extraction systems described in the previous section. Using the reference standard segmentation of the training set each of the example objects is appropriately labeled. Preliminary experiments comparing a Linear Discriminant Classifier, a Quadratic Discriminant Classifier and a kNN classifier [15] were performed. The kNN classifier showed the best performance.

A. Features

A total of 13 different candidate object features were proposed by Frame *et al.* [2] for use with the Spencer–Frame system. The features can be divided into two groups: shape features 1)–4) and pixel intensity features 5)–13). Experiments showed (see Section VI) that with just these 13 features system performance was limited.

Several new features were added to the 13 in the Spencer–Frame feature set to improve performance. To determine what features should be added the behavior of the system using the Spencer–Frame feature set was observed in a number of experiments. These showed a significant number of FP objects on the vasculature. To counter this, Gaussian filterbank outputs [feature 19)] were added to the feature set as well as a feature which can help to determine if an object is on an elongated structure [feature 20)]. As fluorescein angiograms do not contain color, color features were not used in

the Spencer–Frame system. Since the color of the photographs is an extra source of information features 15)–18) were added. Feature 14 was chosen as an extra shape feature to help eliminate elongated structures. The last feature [feature 21)], was added to help improve the detection of larger circular objects. The total feature set contains 68 features, they are listed below. Some features are obtained at several image scales σ , these have not been individually listed.

- 1) The area $a = \sum_{j \in \Omega} 1$ where Ω is the set of pixels in the candidate.
- 2) The perimeter p which is approximated using the chain-codes [17] of the object. The number of odd n_o and even n_e chaincodes are counted then $p = n_o\sqrt{2} + n_e$.
- 3) The aspect ratio $r = l/w$. l is the length of the largest and w of the second largest eigenvector of the covariance matrix of the object.
- 4) The circularity $c = p^2/4\pi a$.
- 5) The total intensity of the object in the original green plane image $i_{\text{green}} = \sum_{j \in \Omega} g_j$ where g_j represents the j th pixel in the original green plane image.
- 6) The total intensity of the object in the shade corrected image $i_{\text{sc}} = \sum_{j \in \Omega} s_j$ where s_j represents the j th pixel in the shade corrected image.
- 7) The mean intensity under the object in the original green plane image $m_{\text{green}} = i_{\text{green}}/a$.
- 8) The mean intensity under the object in the shade corrected image $m_{\text{sc}} = i_{\text{sc}}/a$.
- 9) The normalized intensity in the original green plane image $NI_{\text{green}} = (1/\sigma)(i_{\text{green}} - \bar{x})$ where σ and \bar{x} are the standard deviation and average pixel value of I_{bg} .
- 10) The normalized intensity in the shade corrected image $NI_{\text{sc}} = (1/\sigma)i$.
- 11) The normalized mean intensity in the original image $NM_{\text{green}} = (1/\sigma)(m_{\text{green}} - \bar{x})$.
- 12) The normalized mean intensity in the shade corrected image $NM_{\text{sc}} = (1/\sigma)m_{\text{green}}$.
- 13) The intensity of the region growing seed in the match filtered image I_{match} .
- 14) The compactness $v = \sqrt{(\sum_{j=1}^n d_j - \bar{d})/n}$ where d_j is the distance from the centroid of the object to its j th boundary pixel and \bar{d} is the mean of all the distances

from the centroid to all the edge pixels. Here n is the number of edge pixels.

- 15) The difference between the mean pixel values inside the object and mean values in a circular region centered on the object in the red plane from the RGB color space $rd(j) = (1/n_\Omega)\sum_{j \in \Omega} R_j - (1/n_\Theta)\sum_{j \in \Theta} R_j$. Here Ω is the set of pixels in the candidate and Θ is the set of pixels included in the circular region excluding the pixels in Ω . R_j is the j th pixel in the red image plane. The diameter of the circular region is determined by taking the maximum distance from the object centroid to any of the edges of an axis aligned bounding box and adding three. n_Ω and n_Θ are the number of pixels in Ω and Θ respectively.
- 16) As item 15 but extracted from the green image plane.
- 17) As item 15 but extracted from the blue image plane.
- 18) As item 15 but extracted from the hue image plane taken from the HSI color space.
- 19) The mean and standard deviation of filter outputs under the object. The filters consist of the Gaussian and its derivatives up to second order at scales $\sigma = 1, 2, 4, 8$ pixels. The total amount of features obtained in this way is $2(4 \times 6) = 48$.
- 20) The average value under the object of the absolute difference of the two largest eigenvalues of the Hessian tensor. The scale $\sigma = 2$ is used for the Gaussian partial derivatives that make up the Hessian.
- 21) The average output under the object of an iris filter as described in [18] used on I_{sc} with a minimum circle radius of 4, maximum radius of 12 and 8 directions.

In some cases the use of feature selection schemes can help improve the performance of a classifier. We have used several feature selection methods on the feature set defined above. Experiments with the obtained feature subsets showed no clear improvement in classification performance over using all features.

V. MATERIAL

A set of 40 images (Set 1) was used to train the kNN classifier in the first candidate object extraction step as described in Section III-B. A second set of 100 images (Set 2) was used to train and test the complete system. All of the images in both sets were acquired digitally.

A. Set 1

Set 1 is a publicly available database of 40 manually segmented fundus images. All the vasculature has been manually segmented. For this research all red lesions were manually segmented and added to the existing segmentation by the first author. A complete description of the database can be found in [14] and [19].

B. Set 2

Set 2 contains a total of 100 images and was used to train and test the presented system. This set of images ideally would be composed of just data taken from a screening program. However, because screening data typically contains less than 10%

pathological images, we were not able to assemble Set 2 using only screening images. Of the images in Set 2, 74 were obtained at a tertiary referral hospital. A team of two ophthalmologists (MDA and MS), both retinal disease specialists, selected 37 images with and 37 images without pathology. These images were captured using a Topcon TRC-50 at 45° field of view. The remaining 26 images were taken from a DR screening program in the Netherlands. These images were captured using a Canon CR5 nonmydriatic 3CCD camera at 45° field of view. Of these images 13 contained red lesions and 13 showed no visible red lesions. Some patients have two images included in the set but these were never of the same eye.

All images were JPEG compressed. This lossy image compression format is not ideal for image processing but is commonly used in a screening setting. The size of the images is 768×576 pixels and the field of view is approximately 540 pixels in diameter. All personally identifiable information was stripped from all the images so that only the raw image data was available for this research.

The team separately annotated the entire set of 100 images. The annotation was performed by manually marking all pixels of a red lesion for every possible red lesion in the set. All 100 images were annotated in random order. The annotation of the first ophthalmologist was taken as the reference standard. According to the reference standard, a total of 55 images from the entire set contained pathologies. In these 55 images a total of 858 red lesions were identified. Next the complete set was randomly split into a test and a training set of 50 images each. Both the test and the training set contained 12 screening and 37 clinical images. As for red lesions, the test set contained 27 images with a total of 348 lesions while the training set contained 28 images with a total of 510 red lesions. Care was taken that, if two images were obtained from different eyes of the same patient, both images were assigned to either the test or training set.

VI. EXPERIMENTS AND RESULTS

A number of different system setups were tested to determine which provides the best performance.

- The hybrid system (HS) uses both the hybrid object candidate extraction system (Section III-C) and the complete set of features as defined in Section IV.
- The pixel classification system (PC) uses the pixel classification based candidate extraction system (Section III-B) and the complete set of features except feature 13 which is unique for the mathematical morphology based system.
- The mathematical morphology based system (Section III) using the complete set of features (MMA).
- The mathematical morphology based system using the original 13 features as proposed by Spencer–Frame (MMO).

All systems used a kNN classifier for the classification of the candidate objects. As noted in the previous section, Set 2 was also annotated by a second ophthalmologist. This allowed us to compare the performance of the automatic systems with that of a human expert (HE).

TABLE I

	Hybrid System	Pixel Classification	Mathematical Morphology	Human Expert
Total nr. of extracted candidate objects	14906	13680	10209	478
Nr. of true lesions that were not extracted	33 (9.48%)	49 (14.1%)	49 (14.1%)	108 (31.0%)

The performance of the different candidate object extraction methods on the test set which contained 50 images and 348 lesions. The results of the human expert have been added for comparison.

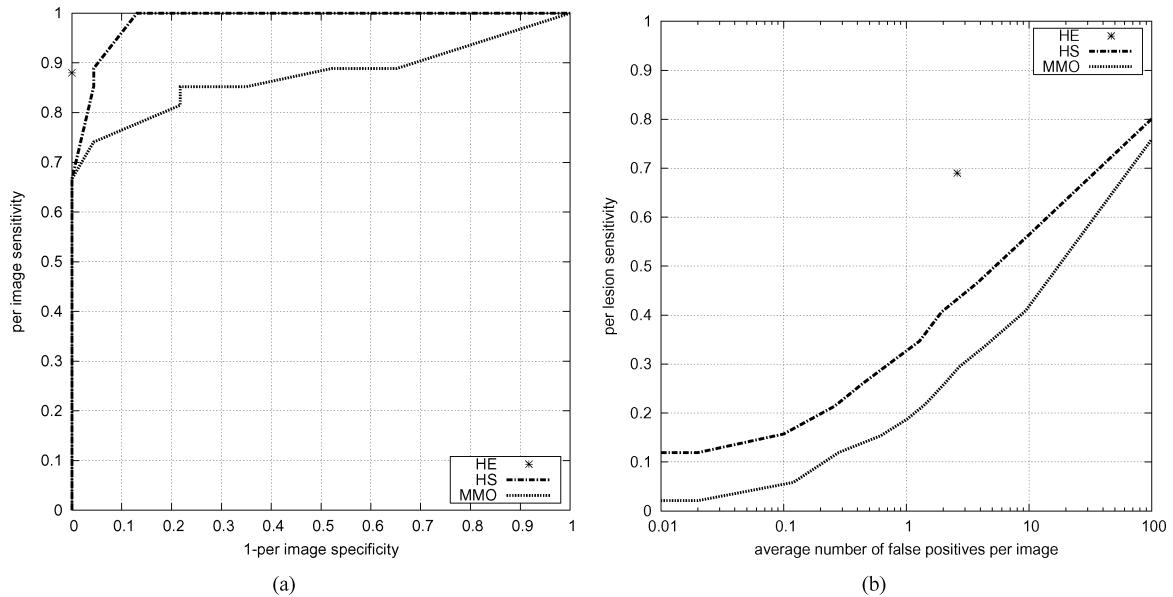


Fig. 4. (a) ROC curves of the HS, the MMO and the HE. The sensitivity and specificity are on a per image basis. (b) FROC curves of the HS, MMO and HE. The sensitivity of the detection is given on a per lesion basis. The average number of FPs per image are taken over all images in the test set. The horizontal axis has a logarithmic scale.

A. Settings

For each system the parameter settings as detailed in the previous sections were used.

For the candidate object classification step the optimal value of k for the kNN classifier was determined using the training set. For all systems an optimal value of $k = 11$ was found except for the MMA where $k = 9$.

B. Results

Table I shows an overview of the performance of all tested systems when extracting candidate red lesion objects from the test set. The number of true lesions that were not extracted indicates the number of lesions that were in the reference standard but were not extracted during the candidate object extraction step. These lesions are missed by the systems and are not present in the subsequent candidate classification step. For comparison the performance of the HE has been added.

How to evaluate the final classification performance of a red lesion detection system is application dependent. When using the method as a prescreening system as described in Section I, its purpose is to determine whether or not an image is normal, i.e., free of red lesions, or abnormal, i.e., containing one or more red lesions. Receiver operating characteristic (ROC) curve analysis

[20] can be used to evaluate this type of performance. The ROC curve plots the *sensitivity* against *1-specificity* of a system. It is an indication of the capability of the system to discriminate between those images containing red lesions and those that do not contain red lesions. Fig. 4(a) shows the ROC curves of the HS and the MMO. Because the HE provides one sensitivity/specificity pair its performance can only be represented by a single point in Fig. 4.

If the system were to be used as a tool to assist an HE in detecting (subtle) red lesions, performance can be measured using free-response operating characteristic (FROC) analysis [21]. When assisting a human, the goal of the system would be to detect as many lesions as possible while not generating an excessive number of FPs. Fig. 4(b) shows the FROC curves of both the HS and the MMO. An FROC curve plots the sensitivity of the system with regard to all the red lesions in the test set against the average number of FPs per image. This gives an indication of the actual red lesion detection performance of the systems. Please note that the horizontal axis in Fig. 4(b) has a logarithmic scale.

Table II shows the final performance of all systems at two important operating points. The first operating point has the same sensitivity per image as the HE, 0.89, and the second operating point is the point on the ROC curve where the per image sensitivity first reaches 1.00. Further, the per lesion sensitivity and the

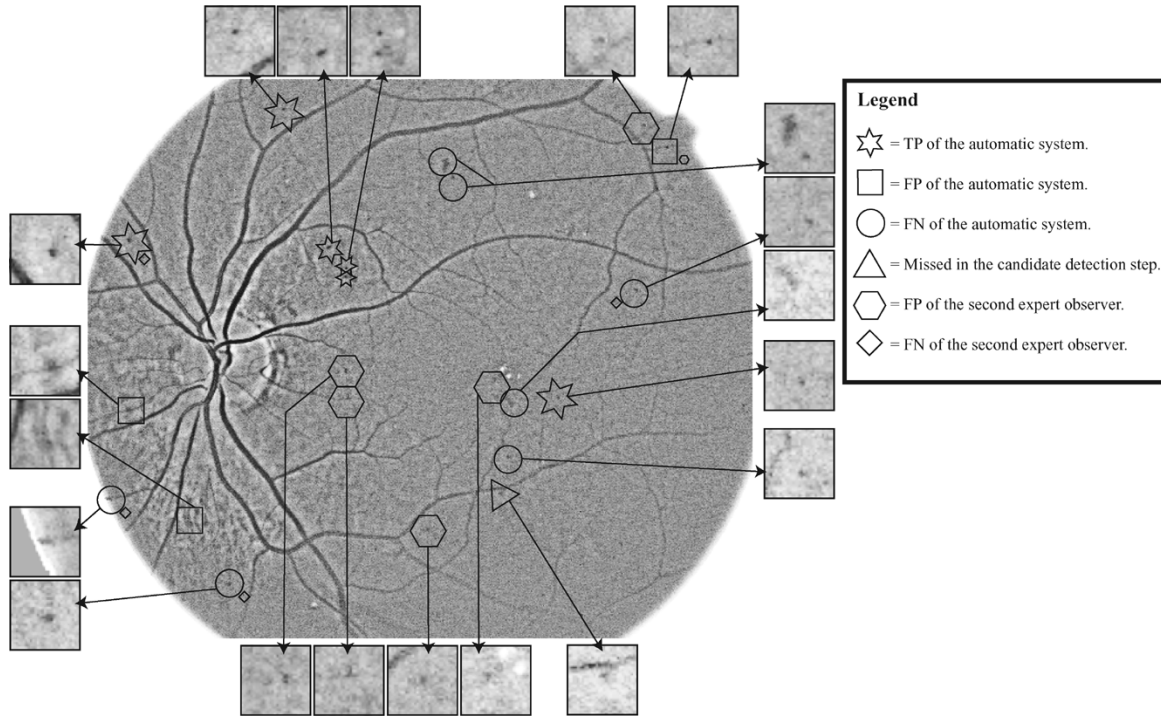


Fig. 5. An example of a final result of the hybrid red lesion detection system. The system was operating at a per image sensitivity/specificity of 1.00/0.87 and a per lesion sensitivity of 0.31 at 0.80 average FPs per image. This image contains 14 red lesions according to the reference standard. The second observer correctly identified 10 of those (71.4%), produced 6 FPs and 4 false negatives (FNs). The automatic system correctly identified 6 red lesions (42.9%), produced 3 FPs and 8 FNs. One of the FNs is a lesion that was missed in the first hybrid candidate detection step and, thus, the system will not be able to detect this lesion regardless of the operating point that is used. All detected objects have been marked on the image using different shapes. When an object belongs to multiple classes (e.g., is both a TP of the HS and a FN of the HE) the indicator of the second class is displayed smaller close to the first class. The contrast in this image has been adjusted for optimal display.

TABLE II

System	Sensitivity per image	Specificity per image	Sensitivity per lesion	Average nr. of FP per image
HS	0.89	0.96	0.26	0.44
	1.00	0.87	0.31	0.80
PC	0.89	0.96	0.21	0.36
	1.00	0.53	0.46	4.10
MMA	0.89	0.96	0.22	0.42
	1.00	0.60	0.42	2.54
MMO	0.89	0.48	0.34	4.38
	1.00	0.00	0.87	196
Human Expert	0.89	1.00	0.69	2.60

The final performance of the different red lesion detection methods on the test set at two operating points. The first point is at a per image sensitivity equal to that of the human expert. The second point is the operating point of a prescreening system

average number of FPs per image at these two operating points are shown. For comparison with the automatic systems the performance of the HE is also included.

VII. DISCUSSION AND CONCLUSION

The results show that the HS described in this paper, when used as a prescreening tool can operate at a sensitivity of 100% combined with a 87% specificity. This means a large decrease in images which have to be screened by an ophthalmologist. The performance of the system on a per image basis is close to that of a HE. Using the previously published system in a prescreening setup is not practical because it is not able to attain 100% sensitivity without its specificity dropping to 0%.

The first candidate detection step of the HS combines the detected candidates of the mathematical morphology with the candidates of the pixel classification based system. It is important to miss as few red lesions as possible in the first candidate extraction step because any lesions lost in this step cannot be retrieved later. Therefore, each system has been adjusted for maximum sensitivity in the first step, assuring that the largest number of lesions are extracted. This leads to the detection of a large number of spurious candidate objects as well. Table I shows that the HS extracted more red lesions than the other systems but it also extracted the largest number of candidate objects overall. However, 33 lesions are still not extracted by the hybrid detection system. Approximately half of them are very close to or connected with the vasculature, both the mathematical morphology as well as the pixel classification based approach have problems extracting these. An example of this type of missed lesions is shown in Fig. 5. The other half of the missed lesions is extremely subtle or blurred. Some of these candidate objects are not detected by neither the pixel classification nor the mathematical morphology based candidate extraction procedure. If they are detected, the region growing procedure often fails to find the object border in subtle cases and grows into nearby objects.

As Fig. 4(a), (b) and Table II show, the HE did not detect every abnormal image in the test set nor found all red lesions. This disagreement with the reference standard has several possible reasons. Because the images have been JPEG compressed and many red lesions are very small it is easy for some to be missed. Further, it is possible that the second expert (indeliberately) read the images differently than the first. When an ophthalmologist

performs a clinical reading of a color fundus photograph, the precise number of lesions is less important than their location. Observation of a number of cases showed that a significant number of FN lesions of the second observer were located outside of the vascular arch. Lesions in that part of the fundus are clinically less relevant. It can be that the second HE interpreted the images more in terms of clinical relevance of the lesions than the expert who defined the reference standard even though both were given the same task description. This effect is visible in Fig. 5. Most of the FN of the HE are located outside of the vascular arch while most TP and FP are located inside the vascular arch.

The HS reaches 100% sensitivity on a per image basis when the sensitivity on a per lesion basis is at 30%. The probable reason for this sensitivity difference is that some of the images contain more red lesions than others and, thus, the system does not need a very high per lesion sensitivity as long as it is able to detect at least one of the lesions in each of the images. On average, the 27 positive images in the test set contained 12.9 lesions with a standard deviation of 9.47. The high standard deviation indicates the diversity of the test set, in which both images containing a large number of lesions (5 images with more than 20 red lesions) and subtle images (9 images with less than 7 lesions) are represented. Generally the more lesions an image contains, the easier it will be for the automatic system to detect it as a positive. The test set used in this paper partly consisted of images obtained in a tertiary referral hospital in addition to images obtained in a screening program. When performance was assessed on both sets separately, results were very similar with slightly better performance in the screening images.

The focus of this paper is on the development of a pre-screening system. However, as the HS is able to detect individual red lesions, it could be used in the role of a CAD system assisting an ophthalmologist in the detection of red lesions. When the HS is operating at an average of 2 FPs per image, which may be an acceptable number of FPs for a CAD system, it detects 21 red lesions which were not found by the HE who operated at an average of 2.6 FPs per image.

Through the use of advanced pattern recognition techniques some parts of the system may be optimized further. One possibility would be to use more advanced classifiers such as support vector machines or techniques such as boosting [15]. By increasing the amount of training data for the candidate object classification, overall system performance may be further improved. Large red lesions are not always completely segmented by the region growing procedure, by replacing it with a different segmentation method this could be improved.

The total time required by the fully automatic HS to process a single image is approximately 15 minutes. All experiments were conducted using a Intel Pentium IV based computer running at 1.7 GHz. Most of this time is spent during the initial candidate detection step, that is, the pixel classification procedure. However, our implementation is experimental and performance could be improved.

Summarizing, we have attained the aim as set out in Section I. The system presented in this paper detected every positive image

in the test set while classifying only a small number of negative images as positive.

REFERENCES

- [1] 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.
- [2] 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.
- [3] R. Klein, B. Klein, S. Moss, M. Davis, and D. DeMets, "The Wisconsin epidemiologic study of diabetic retinopathy. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years," *Arch. Ophthalmol.*, vol. 102, no. 4, pp. 520–526, 1984.
- [4] T. Spencer, R. P. Phillips, P. Sharp, and J. Forrester, "Automated detection and quantification of microaneurysms in fluorescein angiograms," *Graefes's Arch. Clin. Exp. Ophthalmol.*, vol. 230, pp. 36–41, 1992.
- [5] 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.
- [6] L. Yannuzzi, K. Rohrer, and L. Tindel, "Fluorescein angiography complication survey," *Ophthalmology*, vol. 93, pp. 611–617, 1986.
- [7] C. Baudoin, B. Lay, and J. Klein, "Automatic detection of microaneurysms in diabetic fluorescein angiographies," *Revue D'Épidémiologie et de Santé Publique*, vol. 32, pp. 254–261, 1984.
- [8] J. Hipwell, F. Strachant, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool," *Diabetic Med.*, vol. 17, pp. 588–594, 2000.
- [9] G. Gardner, D. Keating, T. Williamson, and A. Elliot, "Detection of diabetic retinopathy using neural network analysis of fundus images," *Br. J. Ophthalmol.*, vol. 80, no. 11, pp. 937–948, 1996.
- [10] C. Sinthanayothin, J. Boyce, T. Williamson, H. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Med.*, vol. 19, pp. 105–112, 2002.
- [11] M. Larsen, J. Godt, N. Larsen, H. Lund-Andersen, A. Sjølie, E. Agardh, H. Kalm, M. Grunkin, and D. Owens, "Automated detection of fundus photographic red lesions in diabetic retinopathy," *Investigat. Ophthalmol. Vis. Sci.*, vol. 44, no. 2, pp. 761–766, 2003.
- [12] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.
- [13] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 951–958, Aug. 2003.
- [14] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. Abràmoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," *Proc. SPIE (Medical Imaging 2004: Image Processing)*, pp. 648–657, 2004.
- [15] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2001.
- [16] S. Arya, D. Mount, N. Netanyahu, R. Silverman, and A. Wu, "An optimal algorithm for approximate nearest neighbor searching in fixed dimensions," *J. ACM*, vol. 45, no. 6, pp. 891–923, 1998.
- [17] H. Freeman, "On encoding arbitrary geometric configurations," *IRE Trans. Electron. Comput.*, vol. 10, pp. 260–268, 1961.
- [18] H. Kobatake and Y. Yoshinaga, "Detection of spicules on mammogram based on skeleton analysis," *IEEE Trans. Med. Imag.*, vol. 15, no. 3, pp. 235–245, Jun. 1996.
- [19] J. Staal, M. Abràmoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
- [20] C. Metz, "ROC methodology in radiologic imaging," *Investigat. Radiol.*, vol. 21, no. 9, 1986.
- [21] P. Bunch, J. Hamilton, G. Sanderson, and A. Simmons, "A free-response approach to the measurement and characterization of radiographic-observer performance," *J. Appl. Photographic Eng.*, vol. 4, pp. 166–171, 1978.