Retinal image analysis based on mixture models to detect hard exudates

Clara I. Sánchez\(^a,b,*\), María García\(^a\), Agustín Mayo\(^c\), María I. López\(^b\), Roberto Hornero\(^a\)

\(^a\) Biomedical Engineering Group, E.T.S. Ingenieros de Telecomunicación, University of Valladolid, Campus Miguel Delibes s/n, Valladolid, Spain
\(^b\) Instituto de Oftalmobiología Aplicada (IOBA), University of Valladolid, Spain
\(^c\) Department of Statistics and Operative Investigation, Facultad de Ciencias, University of Valladolid, Spain

**Abstract**

Diabetic Retinopathy is one of the leading causes of blindness in developed countries. Hard exudates have been found to be one of the most prevalent earliest clinical signs of retinopathy. Thus, automatic detection of hard exudates from retinal images is clinically significant. In this study, an automatic method to detect hard exudates is proposed. The algorithm is based on mixture models to dynamically threshold the images in order to separate exudates from background. A postprocessing technique, based on edge detection, is applied to distinguish hard exudates from cotton wool spots and other artefacts. We prospectively assessed the algorithm performance using a database of 80 retinal images with variable colour, brightness, and quality. The algorithm obtained a sensitivity of 90.2% and a positive predictive value of 96.8%, using a lesion-based criterion. The image-based classification accuracy is also evaluated obtaining a sensitivity of 100% and a specificity of 90%.

© 2009 Elsevier B.V. All rights reserved.

**1. Introduction**

Retinal images are widely used by ophthalmologists and primary care physicians for the screening of epidemic eye diseases, such as Diabetic Retinopathy (DR). DR is one of the leading causes of blindness and vision defects in developed countries. Fundus images permit a high quality permanent record of retinal fundus for detecting early signs of DR and monitoring its progression. Moreover, their digital nature allows automatic analysis to reduce the workload of the ophthalmologists and the health costs in the screening of the disease.

Early detection of DR is crucial for the prevention of visual loss. Hard exudates (HEs) are one of the most prevalent signs in the earliest stages of the disease (Singer et al., 1992). Additionally, they represent the most specific marker for the presence of co-existent retinal oedema, the major cause of visual loss in the non-proliferative forms of DR (Singer et al., 1992). Automatic exudates detection is a difficult task due to the uneven illumination, poor contrast and colour variation of the retinal images. Several attempts have been made to segment this sign from the retinal background.

Grey level thresholding is the simplest exudate segmentation method (Kavitha and Devi, 2005; Liu et al., 1997; Philips et al., 1993; Ward et al., 1989). However, the automatic selection of a threshold is difficult due to the uneven illumination of the HEs. A global threshold is set manually for each image in (Ward et al., 1989), whereas a local threshold is used in (Philips et al., 1993) for different regions, which are selected by the user. A local dynamic thresholding algorithm is proposed in (Liu et al., 1997), based on the histogram shape but no results are reported. In another approach, a multilevel thresholding of the histogram is applied to segment exudates (Kavitha and Devi, 2005). Although a sensitivity of 100% and a mean number of 0.1 false positives per image were obtained, the resulting algorithm was tested using only ten images.

The grey level variation of the exudates using mathematical morphology is also exploited for HE detection (Walter et al., 2002), obtaining a sensitivity of 92.8% and a positive predictive value of 92.4% in a small dataset of 30 retinal images.

Region growing algorithms segment retinal images based on the homogeneity of the exudates illumination (Sinthanayothin et al., 2002) or combined with edge detection (Li and Chutatape, 2004). These techniques have the drawback of being computationally intensive.

Clustering algorithms (Hsu et al., 2001), statistical classification (Goh et al., 2000; Wang et al., 2000), supervised approaches (Niemeyer et al., 2007) and neural network (Gardner et al., 1996; Osareh, 2004; Zhang and Chutatape, 2005) have been also attempted to detect HE. They have used different exudate features (illumination, contrast, colour, etc.) guaranteeing the detection of exudates at the expense of simplicity.

Our purpose is to select a threshold that allows segmentation of the exudates from the background. A fixed value based on the exu-
dates grey level cannot be assigned for this threshold because of the wide variability in illumination and contrast from image to image, strongly correlated to subject’s intrinsic characteristics. We propose a innovative segmentation approach based on a statistical mixture model based clustering, which allows a robust segmentation of the image foreground in a totally unsupervised manner. In contrast to supervised algorithms (e.g. Niemeijer et al., 2007), there is no need of training phase or feature extraction procedure, reducing the computation costs. In other approaches (e.g. Osareh, 2004) the segmentation relies on the study of the histogram shape, which is highly influenced by outliers. The proposed method, by contrast, can deal with outlying observations, obtaining a robust separation of the foreground and background scenes and, specifically, a segmentation of hard exudates. A postprocessing technique, based on edge detection, is then applied to distinguish HEs from other bright lesions, such as cotton wool spots, and from other bright elements that are detected after the segmentation process, such as artefacts along large blood vessels due to light reflection. The resulting algorithm achieves a satisfactory performance with independence of the variable appearance of retinal fundus images, an important factor in clinical environment.

2. Mixture models

Mixture models (MMs) are a powerful semi-parametric statistical technique for estimating probability densities, such as histograms (McLachlan and Peel, 2000; Titterington et al., 1985). Due to their advantages over nonparametric and parametric methods, MMs are being increasingly exploited in medical image analysis to model pixel values as a combination of different populations mixed in varying proportions (Frosio et al., 2006; Noe and Gee, 2001; Osareh, 2004).

Let $Y_i$ be a $p$-dimensional random variable. Assuming that the probability density function $f(Y_i)$ can be written as

$$ f(Y_i; \Psi) = \sum_{i=1}^{g} \pi_i f_i(Y_i; \theta_i) $$

(1)

with

$$ \sum_{i=1}^{g} \pi_i = 1 \quad 0 \leq \pi_i \leq 1 $$

(2)

$$ \int f_i(x)dx = 1, $$

(3)

it is said that $Y_i$ follows a $g$-component MM distribution. $\pi_1, \ldots, \pi_g$ are the mixing parameters, $f_1(\cdot), \ldots, f_g(\cdot)$ are the component densities of the mixture, which belong to a known parametric family fully characterized by the parameter vector $\theta_i$, and $\Psi = (\pi_1, \ldots, \pi_g, \theta_1^1, \ldots, \theta_g^1)^T$ is the complete set of parameters needed to specify the mixture (McLachlan and Peel, 2000).

The component densities generally belong to the normal family. These normal mixtures can be written as

$$ f(Y_i; \Psi) = \sum_{i=1}^{g} \pi_i N(Y_i; \mu_i, \Sigma_i), $$

(4)

where

$$ N(Y_i; \mu_i, \Sigma_i) = (2\pi)^{-\frac{p}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (Y_i - \mu_i)^T \Sigma_i^{-1} (Y_i - \mu_i) \right\} $$

(5)

denotes the multivariate normal density with mean $\mu_i$ and covariance matrix $\Sigma_i$ (McLachlan and Peel, 2000). In this case, the vector $\Psi$ is defined as

$$ \Psi = (\pi_1, \ldots, \pi_g, \mu_1^T, \ldots, \mu_g^T, elements(\Sigma_1, \ldots, \Sigma_g))^T. $$

(6)

The common approach for the fitting of mixture distributions is the maximum likelihood (ML) estimation. An estimate $\hat{\Psi}$ can be obtained maximizing the likelihood function of the parameters, described by

$$ L(\Psi) = \prod_{i=1}^{N} f(Y_i; \Psi). $$

(7)

A closed-form solution for computing the parameters is not available, so iterative algorithms have been used. The common iterative method to obtain a ML estimate of the mixture parameters is the expectation–maximization (EM) algorithm (Dempster et al., 1977). This method produces the equations for updating the parameters in each iteration step until convergence (Dempster et al., 1977).

3. Methods

Exudates segmentation is obtained by five sequential steps. As stated in (Walter et al., 2002), the exudates appear more contrasted in the green component $I_g$ of the RGB colour model. First, $I_g$ is enhanced to obtain a histogram where HEs can be better segmented from the background. The histogram of the enhanced image is modelled using a MM and a dynamic threshold is set based on the statistical information obtained. Then, the optic disk is automatically localized and masked out. Finally, a postprocessing technique is then applied to distinguish HEs from other bright lesions, such as cotton wool spots, and from other bright elements that are detected after the segmentation process, such as artefacts along large blood vessels due to light reflection.

3.1. Preprocessing

The large luminosity and contrast variability of the retinal images, mainly due to retinal pigmentation and the acquisition process, affects seriously the diagnostic process of DR and the automatic detection of retinal lesions, especially HEs. Thus, a preprocessing step is essential to algorithm success, aimed at reducing the intra- and inter-image variability.

We followed the method proposed in (Foracchia et al., 2005) to normalize luminosity and contrast in retinal images based on a statistical model of the image. The method performs an estimation of the local luminosity and contrast using the mean and the standard deviation of pixels which are likely to belong to the background. The variability in luminosity and contrast across the image is then compensated in the whole image. A detailed description of this method can be found in (Foracchia et al., 2005).

The resulting image (Fig 1a and b) shows an improvement in the overall luminosity variation and in the contrast between lesions and background.

In order to better appreciate the enhancement achieved with the preprocessing step, we calculated a quantitative metric ($J$) to evaluate the separation between HEs and other regions in the image. The metric $J$ estimates the class separability of exudate and non-exudates pixels using within-class $S_w$ and between-class $S_b$ scatter matrices (Fukunaga, 1990):

$$ J = \text{trace}(\frac{S_b}{S_w}). $$

(8)

A higher value of $J$ indicates that the classes are more separated while members within each class are closer to each other. Thus, HEs can be more easily extracted from the background. The metric for $I_{enh}$ and for the colour components typically used in exudate detection are shown in Table 1. As the enhanced image presents the highest metric, $I_{enh}$ is the most suitable component for our retinal image analysis.
3.2. Mixture model for histogram modelling

The histogram of the enhanced image $I_{enh}$ usually has a unimodal shape, heavy tailed with different weight in the tails (see Fig. 2a), which comprises the vast majority of the pixels in the image. The peak corresponds roughly to the natural level of variation within homogeneous regions and represents the image background. The heavy tails are caused by foreground elements varying above the natural level, such as vessels, the optic disk and lesions. The left tail is associated with the pixels that belong to regions darker than the background, such as the blood vessels and haemorrhages; whereas the right tail corresponds to bright objects, such as optic disk, HEs and cotton wool spots. Several pixels are characterized by grey levels far removed from the majority distribution, which are considered as outliers.

Let $Y = Y_1, \ldots, Y_n$ be a finite set of pixels from a fundus image characterized by their intensity value, $Y_j$ in the enhanced image $I_{enh}$. Suppose that a pixel comes from one of the following classes:

- **class 1** (background elements),
- **class 2** (foreground elements, such as vessels, optic disk and lesions),
- **class 3** (outliers).

Visual analysis of the histograms of several retinal images (Fig. 2) has inspired us to assume that the distribution of grey levels for each class can be modelled by a normal distribution $N$ with mean $\mu_i$ and variance $\sigma_i^2$, with $i = 1, 2, 3$. Therefore, the overall normalized histogram of the image can be estimated as the following 3-component MM distribution ($\text{McLachlan and Peel, 2000}$):

$$f(y) = \sum_{i=1}^{3} p_i N(y; \mu_i, \sigma_i^2)$$

(9)

With $\Psi = \{\pi_1, \pi_2, \zeta\}$ and $\zeta = \{\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \mu_3, \sigma_3^2\}$.

We have introduced the third component to obtain a robust estimation of the MM, which tries to reduce the influence of the outlying observations ($\text{McLachlan and Peel, 2000}$).

Applying the EM algorithm, we can estimate the mixture parameters. The initialization of the parameters is crucial to the EM algorithm success because its convergence and accuracy of the final estimate depend on these initial parameters ($\text{McLachlan and Peel, 2000}$). The $K$-means clustering algorithm is used here to initialize these parameters. This unsupervised technique partitions the dataset into $K$ clusters in order to minimize an objective function, normally the squared error function. This simple algorithm converges fast and allows the initial MM parameters to near optimal values ($\text{McLachlan and Peel, 2000}$).

After the convergence of the mixture parameters (Fig. 2b), the two Gaussian components with higher mixing weights ($\pi_i$) are associated to the background pixels and the foreground features, such as vessels, optic disk and different lesions that can appear in a fundus image (haemorrhages, HEs, cotton wool spots, ...).
The component with the minimum mixing proportion presents a larger standard deviation than the others and it corresponds to grey levels that lie far away from the midline.

To evaluate the fitting process, we calculated the mean posterior probability of a dataset of 7800 pixels manually segmented from the training set. The dataset comprises of 2600 background pixels, 2600 foreground pixels belonging to dark elements (vessels and haemorrhages) and 2600 foreground pixels belonging to bright elements (optic disk, HEs and cotton wool spots). The mean posterior probability for the first component (background class) of the background pixels is 0.844 ± 0.367, mean ± standard deviation, whereas the mean posterior probability for the second component of the foreground pixels is 0.818 ± 0.362, mean ± standard deviation. Therefore, the estimated MM describes correctly the two image components: foreground and background.

### 3.3. Dynamic thresholding

An optimal threshold $\text{Th}_{\text{mm}}$ must be now set to allow the segmentation of HEs from background. As we can see in Fig. 2, a fixed value cannot be assigned to $\text{Th}_{\text{mm}}$ because the histogram varies from image to image, due to the large variation among images even after preprocessing stage. Therefore, we propose to dynamically select the threshold using the fitted MM. In order to avoid the influence of outliers, we set the threshold based on the two Gaussian components with higher mixing weights (McLachlan and Peel, 2000).

$$f_{\text{robust}}(y_j) = \frac{\pi_1}{\pi_1 + \pi_2} N(y_j; \mu_1, \sigma^2_1) + \frac{\pi_2}{\pi_1 + \pi_2} N(y_j; \mu_2, \sigma^2_2).$$

Due to the strong overlapping of the two components, the minimum error thresholding cannot be applied. As the point of discontinuity on the curve represents the intersection of two overlapping distributions (Sezgin and Sankur, 2004), we analyse the concavities of $f_{\text{robust}}$ to set the threshold. In order to get only the bright foreground elements, we set the threshold in the deepest concavity point that is found on the right tail of $f_{\text{robust}}$, as follows (Sezgin and Sankur, 2004):

$$\text{Th}_{\text{mm}} = \arg\max \left\{ f_{\text{robust}}(y_j) - \text{Hull}(f_{\text{robust}}(y_j)) \right\} \quad \forall y_j$$

where $\text{Hull}(\cdot)$ is the convex hull. Fig. 1c shows the result $I_{\text{enh}}$ after thresholding $I_{\text{enh}}$ with $\text{Th}_{\text{mm}}$.

### 3.4. Optic disk localization

The localization of the optic disk (OD) has a critical importance in retinal image analysis because it is used as a landmark for the other features in the fundus images. Moreover, its localization is indispensable in the detection of HEs because the OD has similar attributes in terms of brightness and contrast.

The OD regions belong to the bright foreground elements and they are mainly segmented after dynamic thresholding as we can see in Fig. 1c. Therefore, accurate OD detection is then necessary to remove these false positives from the final result.

The OD detection algorithm relies on two assumptions: the OD represents a bright region and the primary four vessels normally emanate near-vertically from it. The method is decomposed in two stages:

a. Candidate selection using mathematical morphology.

b. OD detection using Hough transform.

A group of OD candidates are found applying alternative sequential filters to the enhanced image to calculate the background approximation (Fig. 3a and b)

$$M = \text{ASF}(I_{\text{enh}}) = \psi^{(n\text{E})} \ldots \psi^{(2\text{E})} (\phi^{(\gamma)}(I_{\text{enh}})))$$

where $\phi$ and $\gamma$ refer to grey closing and opening, respectively, and $n\text{E}$ is an structuring element of size $n$ sufficiently large to remove bright elements but not the OD (Sollie, 1999). The centroids of the regional maxima of $M$, shown in Fig. 3b, represent the centres of potential ODs. In the second stage, a candidate is considered to be the OD if vertical vessels are found in its neighbourhood. First, vertical vessels are detected by matching $I_{\text{enh}}$ with a two-dimensional vertical-oriented filter characterized by a gaussian cross-profile section (Chaudhuri et al., 1989), as it is shown in Fig. 3c. Each detected vessel is modelled by a single line to obtain the image skeleton. Then, the Hough transform is applied to the neighbourhood of the candidate centroids. The OD is the candidate region with the maximum number of pixels which belongs to vertical lines passing through it (Fig. 3d).

In Fig. 1d, the OD regions are automatically excluded from the other segmented regions in $I_{\text{enh}}$, obtaining $I_{\text{enh}}$-od.

### 3.5. Postprocessing

After thresholding, other bright lesions, such as cotton wool spots, regions near the papillary region and artefacts that are characterized by a high grey level after the image enhancement may be erroneously detected as HEs. Therefore, they must be removed from the final result.

A significant attribute of HEs, in addition to high intensity, is the sharpness of their edges. This characteristic is not representative of other elements such as cotton wool spots or artefacts in the papillary regions. So, we make use of the edge strength of the HEs to eliminate these false positives.
As stated in (Walter et al., 2002), we use the green channel \( I_G \) of the RGB colour space to characterize the edge strength because the exudate boundaries appear sharper than in other colour components. Moreover, this channel presents less noise and artefacts in homogeneous regions than the enhanced image \( I_{\text{enh}} \), as it can be seen in Fig. 1e. In a first step, edges in the image are enhanced applying the Kirsch’s method to \( I_G \) (Jain, 1989):

\[
l_{\text{Kirsch}}(m,n) = \max_k \{ g_{mn} \cdot k_k \}
\]

where \( g_{mn} \) is a vector which represents a \( 3 \times 3 \) subimage of \( I_G \) centered in the pixel \((m,n)\) and \( k_k \) with \( k = 1, \ldots, 8 \) are the group of masks which form the Kirsch operator. Using the Kirsch operator, we can delineate the boundaries giving them a value (see Fig. 1f), which depends on the strength of the edge (Jain, 1989). To avoid the influence of the vessel edges, we remove these contours from \( I_{\text{Kirsch}} \). We set the elements of \( I_{\text{th}-\text{od}} \) to 0 in the green channel (Fig. 1g)

\[
I_0 = \begin{cases} I_{\text{C}}, & \text{if } I_{\text{th}-\text{od}} = 0 \\ 0, & \text{if } I_{\text{th}-\text{od}} \neq 0 \end{cases}
\]

and we then perform morphological reconstruction of the image \( I_0 \) under the mask \( I_{\text{C}} \) (Walter et al., 2002). In that way, all the bright regions detected are removed from \( I_{\text{C}} \) (Fig. 1b) and, after applying the Kirsch operator to the result, only the contours of the blood vessels and dark elements are detected. Calculating the difference between this image and \( I_{\text{Kirsch}} \), we obtain an edge strength map \( I_{\text{edge}} \) of the detected bright elements, as it is shown in Fig. 1i.

We define the edge strength \( \zeta \) of each candidate object \( e_i \) in \( I_{\text{th}-\text{od}} \) as the mean intensity under the object in the edge-enhanced image \( I_{\text{edge}} \):

\[
\zeta(e_i) = \frac{\sum_{j \in \Omega} I_{\text{edge}}(j)}{\sum_{j \in \Omega} 1}
\]

where \( \Omega \) is the set of pixels in the candidate object \( e_i \) using 8-connected neighbourhood. An candidate \( e_i \) is considered HE if \( \zeta(e_i) > Th_{\text{pp}} \). \( Th_{\text{pp}} \) represents an algorithm parameter. It determines the value an edge must have to be considered a sharp boundary. If it is chosen low, more HEs are detected, while the number of false positives increases. The final result, defined as the union of the elements with edge strength higher than \( Th_{\text{pp}} \), is shown in Fig. 4.

4. Retinal images and system performance evaluation

We assessed the performance of our algorithm on retinal images provided by the Instituto de Oftalmobiología Aplicada at University of Valladolid, Spain. The image data set consists of 106 images taken with a TopCon TRCNW6S non-mydratic retinal camera at 45° field of view. Image resolution was 576 x 768 pixels in 24 bit JPEG format. The images came from a clinical set of diabetic patients who were referred to the ophthalmologist for further examination. Retinopathy grading was performed on these images by an experienced ophthalmologist. According to the expert, the images belonged to patients who suffered from mild to moderate non-proliferative DR (NPDR). The hard exudates were prominent in images with moderate NPDR. Drusen were not present in any of the images. Other bright features such as internal limiting mem-
brane and circular scars resulting from panretinal photoagulation treatment were also presented in a few images. Images with media opacities were excluded from the study.

We divided our dataset into a training set and a test set. Twenty-six retinal images form the training set and were used to develop the algorithm. A test set of 80 unseen images were selected to assess the complete algorithm. From the test set, 40 of these images were identified by an ophthalmologist to have no signs of DR. The other 40 pathological images present HEs as well as other visible lesions of DR, such as haemorrhages or cotton wool spots. In order to validate the algorithm prospectively against expert annotated detection, one specialist performed manual annotations marking HEs in the 40 pathological images. This specialist performed the analysis on two separated occasion to assess intra-observer variation. Performance is computed with the segmentations of the first session as ground truth.

We have first evaluated the performance of the mixture model calculating the average absolute error between the observed histogram and the modelled one. In that way, we can measure quantitatively the similarity between the two probability density functions.

To allow a comparative study between our algorithm and alternative proposed HE detection methods, two different criteria have been used to assess the algorithm performance: lesion-based criterion and image-based criterion (Osareh, 2004). Using the image-based criterion, we evaluate the ability of the algorithm to the ability to classify an image without HE as a normal image and an image with HEs as a pathological one (Osareh, 2004). In this work, an image is considered pathological if it presents one or more HEs. In lesion-based criterion, the performance of the system is assessed based on the number of exudate lesions correctly detected in the image. A true exudate was considered found if the automatically detected cluster overlapped at least 50% of the area manually annotated by the expert but less than 150%. All findings outside this criterion were considered to be false detections.

The algorithm has a parameter, \( Th_{pp} \), which determines the value to consider an edge as a sharp boundary. In order to study the influence of this parameter on algorithm performance, the number of true positives (TPs), true negatives (TNs), false positives (FPs) and false negatives (FNs) was determined for each image comparing the regions detected by the algorithm with the ground truth. Different trade-offs between sensitivity and specificity were achieved, while the parameter was varied as follows:

\[
Th_{pp} \in \{0.02; 2\}.
\]  

To compare the results with other approaches that used pixel-based criterion, the ratio \( TP/(TP + FP) \), which is the probability that a cluster, which has been classified as exudate, is really an exudate (positive predictive value), was used instead of specificity. In that way, the sensitivity rate can be plotted as a function of the predictive value, varying the algorithm parameter (Walter et al., 2002). The different curves obtained were compared using the area under the curve (AUC). The Free-Response Operating Characteristic (FROC) curve was also provided to analyse the sensitivity of the system against the average number of false positives per image.

We have applied also other dynamic thresholding methods, such as Otsu’s approach (Sezgin and Sankur, 2004) and fuzzy c-means (Osareh, 2004), in order to compare the benefits of applying a dynamic threshold to the fitted mixture model over simply applying to the image pixel values.

5. Results

For each retinal images of the database, we modelled the histogram with a 3-component MM obtaining an average absolute error of 5.1 ± 2.9%.

We tested the complete algorithm using a lesion-based criterion over the test set. Fig. 5a shows the sensitivity–predictivity curve varying \( Th_{pp} \). The best overall performance was obtained at \( Th_{pp} = 0.12 \) with a mean sensitivity of 90.2% and a mean predictive value of 96.8%. The effect of varying \( Th_{pp} \) on the algorithm performance can be analyzed from this graph: the lower \( Th_{pp} \), the higher sensitivity and lower predictivity are obtained. In the second reading by the observer, a sensitivity of 82.7% and a predictive value of 100% were achieved. The FROC curve is depicted in Fig. 5b. For a sensitivity of 90.2%, the average number of false positives per image was 0.63.

In 13 of the 40 fundus images of diabetic subjects with HEs, a total of 30 cotton wool spots (CWs) were present. 27 of them were removed using the postprocessing step. Thus, the success rate of CWs removal processing in the optimum operation point (\( Th_{pp} = 0.12 \)) was 90%. The other 3 CWs could not be eliminated because they appear well contrasted in the original images.

Using the better performance operation point of the proposed algorithm (\( Th_{pp} = 0.12 \)), we evaluated also the algorithm performance to separate images with the presence of exudates from images of healthy retinal. The presence of exudates was successfully detected in all the images from diabetic subjects. In 36 of the 40 images from healthy retinas no exudates were found by our algorithm. In the other images, only a few FP were found. Therefore, a sensitivity of 100% and a specificity of 90% were obtained in terms of image-based classification accuracy.

The accuracy of other dynamic thresholding method was also assessed varying \( Th_{pp} \) on the same test set. The results obtained using the Otsu’s method (Sezgin and Sankur, 2004) and the fuzzy c-means approach (Osareh, 2004) are summarized in Table 2.

6. Discussion

6.1. Design considerations

The proposed exudate detection algorithm using MMs is an innovative thresholding method which sets the threshold based on the robust estimation of the histogram. Modelling each histogram with a different MM, we obtain a dynamic threshold for each image.

Preprocessing stage is crucial for the algorithm success due to the variability in colour, illumination and contrast of the retinal images. The proposed preprocessing approach obtains grey level normalization at the same time as contrast enhancement using luminosity and contrast model of the background. Moreover, the class separability of exudate and non-exudate pixels in this image is higher compared with other colour models.

After visual inspection, we modelled the retinal image histogram with three different grey level components using the MM of three single Gaussian distributions. We obtained a good fitting with an average absolute error of 5.1 ± 2.9%. The third component is associated with outliers in order to obtain a robust estimation of the histogram and to reduce the influence of the outlying observations.

Due to the large variation among images even after preprocessing stage (Fig. 2), it is difficult to set an appropriate threshold for all the images. Therefore, we proposed a dynamic thresholding technique based on the estimated mixture model. As the outliers have a strong influence on the threshold selection, we set the threshold using only the two Gaussian components with higher mixing weights.

The proposed segmentation process is generic, i.e., widely applicable to any preprocessing step. Fig. 6 shows the results using a preprocessing step based on the subtraction of a background esti-
mation, which was obtained by means of mean filtering. The statistical characteristics of the mixture model based clustering makes the method independent of the preprocessing step and, consequently, of the variable appearance.

We evaluated the algorithm on an independent database of retinal images with variable characteristics to investigate its robustness. The images from diabetic patients present different clinical signs, as exudates, haemorrhages, microaneurysms or cotton wool spots. In that way, we evaluated the robustness of the algorithm with the presence of other lesions. The approach is also effective in removing other yellow lesions, as cotton wool spots (removal rate of 90%).

Table 2
Comparative study of different thresholding methods.

<table>
<thead>
<tr>
<th>Thresholding method</th>
<th>Lesion-based</th>
<th>Image-based</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sens. (%)</td>
<td>Pred. (%)</td>
</tr>
<tr>
<td>Proposed dynamic thresholding</td>
<td>90.2</td>
<td>96.8</td>
</tr>
<tr>
<td>Otsu’s thresholding</td>
<td>79.3</td>
<td>84.9</td>
</tr>
<tr>
<td>Fuzzy c-means clustering</td>
<td>62.4</td>
<td>69.1</td>
</tr>
</tbody>
</table>

Sens: sensitivity; Pred: predictivity, Spec: specificity, AUC: area under the curve.

Fig. 5. (a) Sensitivity–predictivity curves varying $Th_{pp}$. The star (*) represents the performance obtained by the same observer in the second reading. (b) FROC curve varying $Th_{pp}$.

Fig. 6. Proposed segmentation process using a different preprocessing step. (a) Original image; (b) image after preprocessing based on the substraction of a background estimation; (c) segmentation result.
We have compared the algorithm performance with that of a human expert. The algorithm outperformed the sensitivity obtained in the second reading by the same observer. During the annotation of the images in the second session, the observer detected lesions with higher clinical relevance missing small subtle elements. This difference between the gold standard and the second reading proves the intra-observer variability in retinal image annotations.

Two different criteria were also used to assess the algorithm performance: lesion-based criterion and image-based criterion, in order to allow a comparative study with other exudate detection methods. The overall results were compared with the accuracy obtained using other dynamic threshold methods, showing that the proposed algorithm achieves significantly higher accuracies. Whereas the other methods use the whole image pixel values, the proposed algorithm is more robust in that potential outliers do not contribute to the threshold selection, setting a more accurate value for all the images.

6.3. Algorithm limitations

Despite the enhanced appearance of HEs, brought about by the preprocessing techniques, their diversity of brightness and size makes it difficult to detect them all. HES usually appear in groups and therefore missing some very faint exudates is not very important. However, when there are only a few very faint HE in the retina, the method may fail in the identification task.

Among retinal abnormalities, we have not handled the presence of drusen, the main confounding feature for HE detection. The system performance has been evaluated on a small clinical set of diabetic patients who were referred to the ophthalmologist for further examination. No retinal image in the database presents this abnormality and, therefore, we could not assess its influence on the algorithm performance.

A second reading by the same observer was provided in order to analyse the intra-observer variability and to compare the performance with that of a human expert. However, to assess the inter-observer variability, a second independent reading made by a different observer would be needed.

Finally, the real impact of this system in the daily use of physicians and its suitability for a clinical environment has to be assessed. The obtained trade-off between sensitivity and predictivity values need to be appropriately verified with a higher number of fundus images. Thus, it will be interesting to evaluate the algorithm in a clinical context using a dataset with more fundus images in future works.

References