Retinal Area Detector From Scanning Laser Ophthalmoscope (SLO) Images for Diagnosing Retinal Diseases

Muhammad Salman Haleem, Liangxiu Han, Jano van Hemert, Baihua Li, and Alan Fleming

Abstract—Scanning laser ophthalmoscopes (SLOs) can be used for early detection of retinal diseases. With the advent of latest screening technology, the advantage of using SLO is its wide field of view, which can image a large part of the retina for better diagnosis of the retinal diseases. On the other hand, during the imaging process, artefacts such as eyelashes and eyelids are also imaged along with the retinal area. This brings a big challenge on how to exclude these artefacts. In this paper, we propose a novel approach to automatically extract out true retinal area from an SLO image based on image processing and machine learning approaches. To reduce the complexity of image processing tasks and provide a convenient primitive image pattern, we have grouped pixels into different regions based on the regional size and compactness, called superpixels. The framework then calculates image based features reflecting textural and structural information and classifies between retinal area and artefacts. The experimental evaluation results have shown good performance with an overall accuracy of 92%.

Index Terms—Feature selection, retinal artefacts extraction, retinal image analysis, scanning laser ophthalmoscope (SLO).

I. INTRODUCTION

EARLY detection and treatment of retinal eye diseases is critical to avoid preventable vision loss. Conventionally, retinal disease identification techniques are based on manual observations. Optometrists and ophthalmologists often rely on image operations such as change of contrast and zooming to interpret these images and diagnose results based on their own experience and domain knowledge. These diagnostic techniques are time consuming. Automated analysis of retinal images has the potential to reduce the time, which clinicians need to look at the images, which can expect more patients to be screened and more consistent diagnoses can be given in a time efficient manner [1].

The 2-D retinal scans obtained from imaging instruments [e.g., fundus camera, scanning laser ophthalmoscope (SLO)] may contain structures other than the retinal area; collectively regarded as artefacts. Exclusion of artefacts is important as a

Fig. 1.   Example of (a) a fundus image and (b) an SLO image annotated with true retinal area and ONH.

preprocessing step before automated detection of features of retinal diseases. In a retinal scan, extraneous objects such as the eyelashes, eyelids, and dust on optical surfaces may appear bright and in focus. Therefore, automatic segmentation of these artefacts from an imaged retina is not a trivial task. The purpose of performing this study is to develop a method that can exclude artefacts from retinal scans so as to improve automatic detection of disease features from the retinal scans.

To the best of our knowledge, there is no existing work related to differentiation between the true retinal area and the artefacts for retinal area detection in an SLO image. The SLO manufactured by Optos [2] produces images of the retina with a width of up to $200°$ (measured from the centre of the eye). This compares to $45° - 60°$ achievable in a single fundus photograph. Examples of retinal imaging using fundus camera and SLO are shown in Fig. 1. Due to the wide field of view (FOV) of SLO images, structures such as eyelashes and eyelids are also imaged along with the retina. If these structures are removed, this will not only facilitate the effective analysis of retinal area, but also enable to register multiview images into a montage, resulting in a completely visible retina for disease diagnosis.

In this study, we have constructed a novel framework for the extraction of retinal area in SLO images. The three main steps for constructing our framework include:

1) determination of features that can be used to distinguish between the retinal area and the artefacts;
2) selection of features which are most relevant to the classification;
3) construction of the classifier which can classify out the retinal area from SLO images.

For differentiating between the retinal area and the artefacts, we have determined different image-based features which reflect grayscale, textural, and structural information at multiple

resolutions. Then, we have selected the features among the large feature set, which are relevant to the classification. The feature selection process improves the classifier performance in terms of computational time. Finally, we have constructed the classifier for discriminating between the retinal area and the artefacts. Our prototype has achieved average classification accuracy of 92% on the dataset having healthy as well as diseased retinal images.

The rest of this paper is organised as follows. Section II introduces the previous work for feature determination and classification. Section III discusses our proposed method. Section IV provides the quantitative and visual results of our proposed method. Section V summarizes and concludes the method with future work.

II. LITERATURE SURVEY

Our literature survey is initiated with the methods for detection and segmentation of eyelids and eyelashes applied on images of the front of the eye, which contains the pupil, eyelids, and eyelashes. On such an image, the eyelashes are usually in the form of lines or bunch of lines grouped together. Therefore, the first step of detecting them was the application of edge detection techniques such as Sobel, Prewitt, Canny, Hough Transform [3], and Wavelet transform [4]. The eyelashes on the iris were then removed by applying nonlinear filtering on the suspected eyelash areas [5]. Since eyelashes can be in either separable form or in the form of multiple eyelashes grouped together, Gaussian filter and Variance filter were applied in order to distinguish among both forms of eyelashes [6]. The experiment showed that separable forms of eyelashes were most likely detected by applying Gaussian filter, whereas Variance filters are more preferable for multiple eyelash segmentation [7]. Initially, the eyelash candidates were localized using active shape modeling, and then, eight-directional filter bank was applied on the possible eyelash candidates. Kang and Park [8] used focus score in order to vary the size of convolution kernels for eyelash detection. The size variation of the convolution kernels also differentiated between separable eyelashes and multiple eyelashes. Min and Park [9] determined the features based on intensity and local standard variation in order to determine eyelashes. They were thresholded using Otsu's method, which is an automatic threshold selection method based on particular assumptions about intensity distribution. All of these methods have been applied on CASIA database [10], which is an online database of Iris images. In an image obtained from SLO, the eyelashes show as either dark or bright region compared to retinal area depending upon how laser beam is focused as it passes the eyelashes. The eyelids show as reflectance region with greater reflectance response compared to retinal area. Therefore, we need to find out features, which can differentiate among true retinal area and the artefacts in SLO retinal scans. After visual observation in Fig. 1(b), the features reflecting the textural and structural difference could have been the suggested choice. These features have been calculated for different regions in fundus images, mostly for quality analysis.

The characterisation of retinal images were performed in terms of image features such as intensity, skewness, textural

analysis, histogram analysis, sharpness, etc., [1], [11], [12]. Dias *et al.* [13] determined four different classifiers using four types of features. They were analyzed for the retinal area including colour, focus, contrast, and illumination. The output of these classifiers were concatenated for quality classification. For classification, the classifiers such as partial least square (PLS) [14] and support vector machines (SVMs) [15] were used. PLS selects the most relevant features required for classification. Apart from calculating image features for whole image, grid analysis containing small patches of the image has also been proposed for reducing computational complexity [11]. For determining image quality, the features of region of interest of anatomical structures such as optic nerve head (ONH) and Fovea have also been analyzed [16]. The features included structural similarity index, area, and visual descriptor etc. Some of the above mentioned techniques suggest the use of grid analysis, which can be an time effective method to generate features of particular region rather than each pixel. But grid analysis might not be an accurate way to represent irregular regions in the image. Therefore, we decided the use of superpixels [17]–[20], which group pixels into different regions depending upon their regional size and compactness.

Our methodology is based on analyzing the SLO image-based features, which are calculated for a small region in the retinal image called superpixels. The determination of feature vector for each superpixel is computationally efficient as compared to feature vector determination for each pixel. The superpixels from the images in the training set are assigned the class of either retinal area or artefacts depending upon the majority of pixels in the superpixel belonging to the particular class. The classification is performed after ranking and selection of features in terms of effectiveness in classification. The details of the methods are discussed in the following section.

III. METHODOLOGY

The block diagram of the retina detector framework is shown in Fig. 2. The framework has been divided into three stages, namely training stage, testing and evaluation stage, and deployment stage. The training stage is concerned with building of classification model based on training images and the annotations reflecting the boundary around retinal area. In the testing and evaluation stages, the automatic annotations are performed on the "test set" of images and the classifier performance is evaluated against the manual annotations for the determination of accuracy. Finally, the deployment stage performs the automatic extraction of retinal area.

The subtasks for training, testing, and deployment stages are briefly described as follows:

1) *Image Data Integration:* It involves the integration of image data with their manual annotations around true retinal area.
2) *Image Preprocessing:* Images are then preprocessed in order to bring the intensity values of each image into a particular range.
3) *Generation of Superpixels:* The training images after preprocessing are represented by small regions called
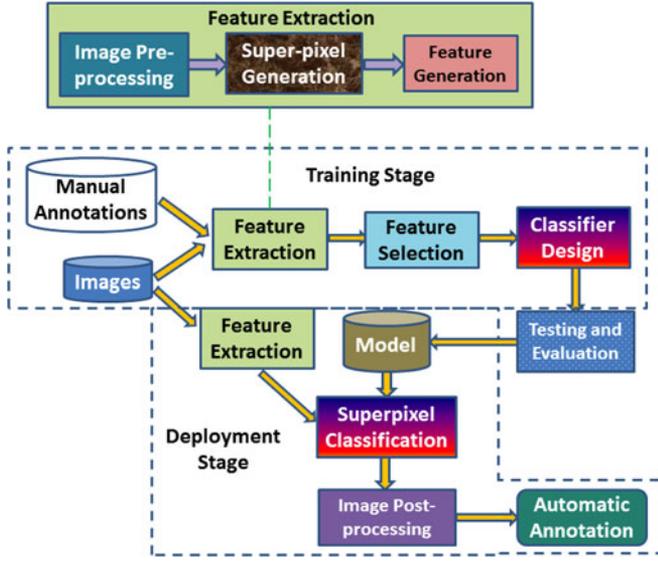
Fig. 2.   Block diagram of retina detector framework.

superpixels. The generation of the feature vector for each superpixel makes the process computationally efficient as compared to feature vector generation for each pixel.

4) *Feature Generation:* We generate image-based features which are used to distinguish between the retinal area and the artefacts. The image-based features reflect textural, grayscale, or regional information and they were calculated for each superpixel of the image present in the training set. In testing stage, only those features will be generated which are selected by feature selection process.

5) *Feature Selection:* Due to a large number of features, the feature array needs to be reduced before classifier construction. This involves features selection of the most significant features for classification.

6) *Classifier Construction:* In conjunction with manual annotations, the selected features are then used to construct the binary classifier. The result of such a classifier is the superpixel representing either the "true retinal area" or the "artefacts."

7) *Image Postprocessing:* Image postprocessing is performed by morphological filtering so as to determine the retinal area boundary using superpixels classified by the classification model.

The elements of our detection framework are discussed as follows.

A. Image Preprocessing

Images were normalized by applying a Gamma (γ) adjustment to bring the mean image intensity to a target value. γ was calculated using

$$\gamma = \frac{\log_{10}(\mu_{\text{target}}) - \log_{10}(255)}{\log_{10}(\mu_{\text{orig}}) - \log_{10}(255)} \qquad (1)$$

where μ_{orig} is the mean intensity of the original image and μ_{target} is the mean intensity of the target image. For image

visualization, μ_{target} is set to 80. Finally, the Gamma adjustment of the image is given as

$$I_{\text{norm}} = \left(\frac{I}{255}\right)^{\gamma}. \qquad (2)$$

B. Generation of Superpixels

The superpixel algorithm groups pixels into different regions, which can be used to calculate image features while reducing the complexity of subsequent image processing tasks. Superpixels capture image redundancy and provide a convenient primitive image pattern. As far as fundus retinal images are concerned, the superpixels have been generated for analyzing anatomical structures [21] and retinal hemorrhage detection [22]. For retinal hemorrhage detection, the superpixels were generated using watershed approach but the number of superpixels generated in our case need to be controlled. The watershed approach sometimes generates number of superpixels of the artefacts more than desired. The superpixel generation method used in our retina detector framework is simple linear iterative clustering [17], which was shown to be efficient not only in terms of computational time, but also in terms of region compactness and adherence. The algorithm is initialized by defining number of superpixels to be generated. The value was set to 5000 as a compromise between computational stability and prediction accuracy.

C. Feature Generation

After the generation of superpixels, the next step is to determine their features. We intend to differentiate between the retinal area and artefacts using textural, grayscale gradient, and regional based features. Textural and gradient based features are calculated from red and green channels on different Gaussian blurring scales, also known as smoothing scales [23]. In SLO images, the blue channel is set to zero; therefore, no feature was calculated for the blue channel. The regional features are determined for the image irrespective of the colour channel. The details of these features are described as follows

1) Textural Features: Texture can be analyzed using Haralick features [24] by gray level co-occurrence matrix (GLCM) analysis. GLCM determines how often a pixel of a gray scale value i occurs adjacent to a pixel of the value j. Four angles for observing the pixel adjacency, i.e., $\theta = 0°, 45°, 90°, 135°$ are used. These directions are shown in Fig. 3(a). GLCM also needs an offset value **D**, which defines pixel adjacency by certain distance. In our case, offset value is set to 1. Fig. 3(b) illustrates the process of creating GLCM using the image **I**. The features, which are calculated using GLCM matrix are summarized in Table I. The mean value in each direction was taken for each Haralick feature and they were calculated from both red and green channels.

2) Gradient Features: The reason for including gradient features was illumination nonuniformity of the artefacts. In order to calculate these features, the response from Gaussian filter bank [23] is calculated. The Gaussian filter bank includes Gaussian $\mathcal{N}(\sigma)$, its two first-order derivatives $\mathcal{N}_x(\sigma)$ and $\mathcal{N}_y(\sigma)$ and three second-order derivatives $\mathcal{N}_{xx}(\sigma)$, $\mathcal{N}_{xy}(\sigma)$, and $\mathcal{N}_{yy}(\sigma)$
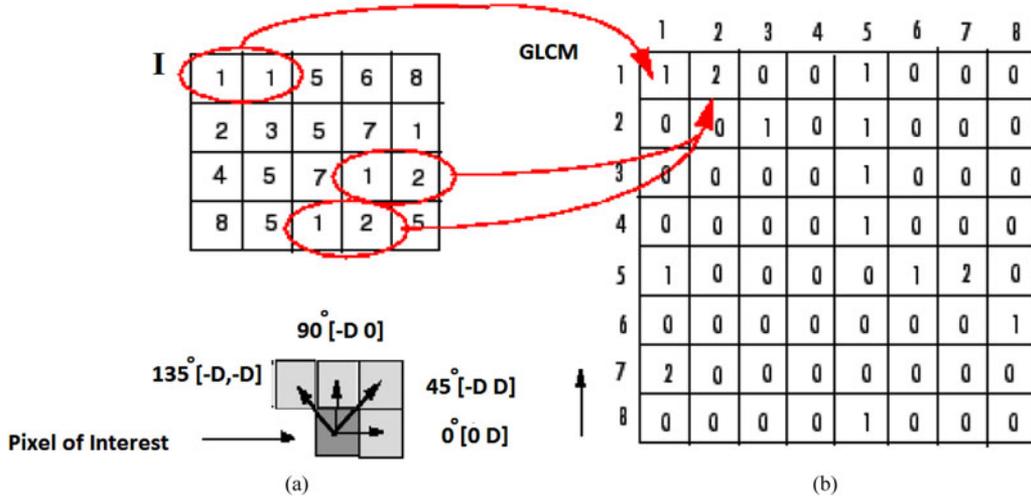
Fig. 3. (a) GLCM directions and offset. (b) GLCM process using image **I** [25].

in horizontal (x) and vertical (y) directions. After convolving the image with the filter bank at a particular channel, the mean value is taken over of each filter response over all pixels of each superpixel.

3) Regional Features: The features used to define regional attributes were included because superpixels belonging artefacts have irregular shape compared to those belonging the retinal area in an SLO image. Table II represents the features describing regional attributes.

The image features are calculated for each superpixel of the images present in the training set and they form a matrix of the form as

$$FM = \begin{bmatrix} A^{tex}{}_R & A^{tex}{}_G & A^{g}{}_R & A^{g}{}_G & A^{re} \\ B^{tex}{}_R & B^{tex}{}_G & B^{g}{}_R & B^{g}{}_G & B^{re} \end{bmatrix} \quad (3)$$

where A and B represent class of true retinal area and class of artefacts, superscripts tex, re, g represent textural features, regional features, and gradient based features, respectively, and subscript R and G represent the red and green channel, respectively. For determining features at different smoothing scales, both red and green channels of images are convolved with the Gaussian [23] at scales $\sigma = 1, 2, 4, 8, 16$. The textural features are calculated at the original scale, as well as at five different smoothing scales so as to accommodate their image response in the training set after blurring. In this way, the total number of columns in both channels of A^{tex} and B^{tex} will be 114 making it 228 altogether. The gradient features has six columns in each scale making 30 columns for each channel of A^g and B^g so 60 columns in total for each superpixel. As far as regional features are concerned, except I_μ, they are independent of channel variation. Therefore, they are calculated only once for the superpixel so seven columns for A^{re} and B^{re} (I_μ is calculated for both red and green channels). In this way, there are the total number of 295 features in the feature matrix for each superpixel of the image present in the training set. Each column of the feature matrix calculated for the particular image is normalized using

z-score normalization [26]. Z-score normalization returns the scores of the column with zero mean and unit variance.

D. Feature Selection

The main purposes for feature selection are reducing execution time, determination of features most relevant to the classification and dimensionality reduction. For feature selection, we have selected sequential forward selection (SFS) approach.

In the "SFS approach," the interaction among features is taken into account. From the available set of features, the feature with the highest *area under the curve* (AUC) [27] is selected. The next feature is chosen in such a way that when it is used along with the first selected feature, it will give the highest AUC compared to other nonselected features. The process is repeated until ten features were selected, since a higher number of features resulted in a very small improvement in AUC.

The performance of the SFS has been compared against other feature selection approaches such as "Filter approach" and "Filter and SFS" approach. In the filter approach, the features are ranked with respect to their effectiveness in classification and higher ranked features are thresholded out. In order to determine most relevant features, an independent evaluation criterion for binary classification is used [28] and AUC is selected as its evaluation measure [27]. The features with higher AUC are ranked higher, and the features with AUC greater than 0.9 are selected. In this way, 33 features are selected for classifier construction. The "Filter and SFS" approach is similar to SFS approach except that it is applied on the filtered feature set rather than complete feature set.

The individual and collective performance of the features selected in the feature sets from the above mentioned approaches are shown in Figs. 4 and 5. The axis of "Feature Index" in Fig. 4 is ordered according to descending independent evaluation criterion. The axis of "Number of Selected Features" in Fig. 5 represents the order with which the features are selected using the SFS approach. We have not applied SFS on the "Filter approach;" therefore, axis of "Number of Selected Features"

TABLE I
TEXTURAL FEATURES EXTRACTED USING GLCM

Feature Name	Equation	Definition
Autocorrelation	$\mathrm{acorr} = \sum_i \sum_j ij p(i,j)$	Linear dependence in GLCM between same index
Cluster Shade	$C_{\mathrm{shade}} = \sum_i \sum_j (i + j - \mu_x - \mu_y)^3 p(i,j)$	Measure of skewness or non-symmetry
Cluster Prominence	$C_{\mathrm{prom}} = \sum_i \sum_j (i + j - \mu_x - \mu_y)^4 p(i,j)$	Show peak in GLCM around the mean for non-symmetry
Contrast	$\mathrm{con} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i-j|^2 p(i,j)$	Local variations to show the texture fineness.
Correlation	$\mathrm{corr} = \frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$	Linear dependence in GLCM between different index
Difference Entropy	$H_{\mathrm{diff}} = - \sum_{i=0}^{N_g-1} p_{x-y} \log(p_{x-y}(i))$	Higher weight on higher difference of index entropy value
Dissimilarity	$\mathrm{diss} = \sum_i \sum_j |i-j| p(i,j)$	Higher weights of GLCM probabilities away from the diagonal
Energy	$E = \sum_i \sum_j p(i,j)^2$	Returns the sum of squared elements in the GLCM
Entropy	$H = - \sum_i \sum_j p(i,j) \log(p(i,j))$	Texture randomness producing a low value for an irregular GLCM
Homogeneity	$\mathrm{homom} = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$	Closeness of the element distribution in GLCM to its diagonal
Information Measures 1	$\mathrm{IM}_1 = (1 - \exp[-2.0(H_{xy} - H)])^{0.5}$	Entropy measures
Information Measures 2	$\mathrm{IM}_2 = \frac{\mathrm{Entropy} - H_{xy2}}{\mathrm{MAX}(H_x, H_y)}$	Entropy measures
Inverse Difference Normalized	$\mathrm{IDN} = \sum_i \sum_j \frac{p(i,j)}{1 + \frac{|i-j|}{N_g}}$	Inverse contrast normalized
Inverse Difference Moment Normalized	$\mathrm{IDMN} = \sum_i \sum_j \frac{p(i,j)}{1 + \frac{(i-j)^2}{N_g}}$	Homogeneity normalized
Maximum Probability	$Pr_{\mathrm{max}} = \mathrm{MAX}_{(x,y)} p(i,j)$	Maximum value of GLCM
Sum average	$\mu_{\mathrm{sum}} = \sum_{i=2}^{2N_g} i p_{x+y}(i)$	Higher weights to higher index of marginal GLCM
Sum Entropy	$H_{\mathrm{sum}} = - \sum_{i=2}^{2N_g} p_{x+y} \log(p_{x+y}(i))$	Higher weight on higher sum of index entropy value
Sum of Squares: Variance	$\sigma_{\mathrm{sos}} = \sum_i \sum_j (i - \mu)^2 p(i,j)$	Higher weights that differ from average value of GLCM
Sum of Variance	$\sigma_{\mathrm{sum}} = \sum_{i=2}^{2N_g} (i - H_{\mathrm{sum}}) p_{x+y}(i)$	Higher weights that differ from entropy value of marginal GLCM

(i,j) represent rows and columns, respectively; N_g is the number of distinct gray levels in the quantized image; $p(i,j)$ is the element from normalized GLCM matrix; $p_x(i)$ and $p_y(j)$ are the marginal probabilities of matrix obtained by summing rows and columns of GLCM, respectively, i.e., $p_x(i) = \sum_{j=1}^{N_g} p(i,j), p_y(j) = \sum_{i=1}^{N_g} p(i,j), p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j), k = i + j - 1 = 1, 2, 3, \ldots, 2N_g$ and $p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j), k = |i-j| + 1 = 1, \ldots, N_g$; H_x and H_y are entropies of p_x and p_y, respectively, $H_{xy} = - \sum_i \sum_j p_x(i) p_y(j) \log(p_x(i) p_y(j))$, and $H_{xy2} = - \sum_i \sum_j p(i,j) \log(p_x(i) p_y(j))$.

for "Filter Approach" would be same as that of "Feature Index" in Fig. 4. The features represented by "Feature Index" and "Number of Selected Features" are shown in Table III. SFS is computationally intensive as it required 5 min/feature on filtered feature set and 30 min/feature on complete feature set. But the results show that the SFS approach performed better compared to other two approaches despite of the fact that the feature set also consists of those features which ranked low in independent evaluation criterion. The Table IV represents the percentage of different types of features selected in each feature set. The table shows clear dominance of textural features compared to gradient features and regional features.

E. Classifier Construction

The classifier is constructed in order to determine the different classes in a test image. In our case, it is a two class problem:

true retinal area and artefacts. We have applied Artificial Neural Networks (ANNs). The ANN is the classification algorithm that is inspired by human and animal brain. It is composed of many interconnected units called artificial neurons. ANN takes training samples as input and determines the model that best fits to the training samples using nonlinear regression. Consider the Fig. 6 which shows three basic blocks of ANN, i.e., input, hidden layer (used for recoding or providing representation for input), and output layer. More than one hidden layer can be used but in our case, there is only one hidden layer with ten neurons. The output of each layer is in the form of matrix of floating values, which can be obtained by sigmoid function as

$$h_W(x) = \frac{1}{1 + \exp(-W^T x + b)} \quad (4)$$

TABLE II
REGIONAL FEATURES

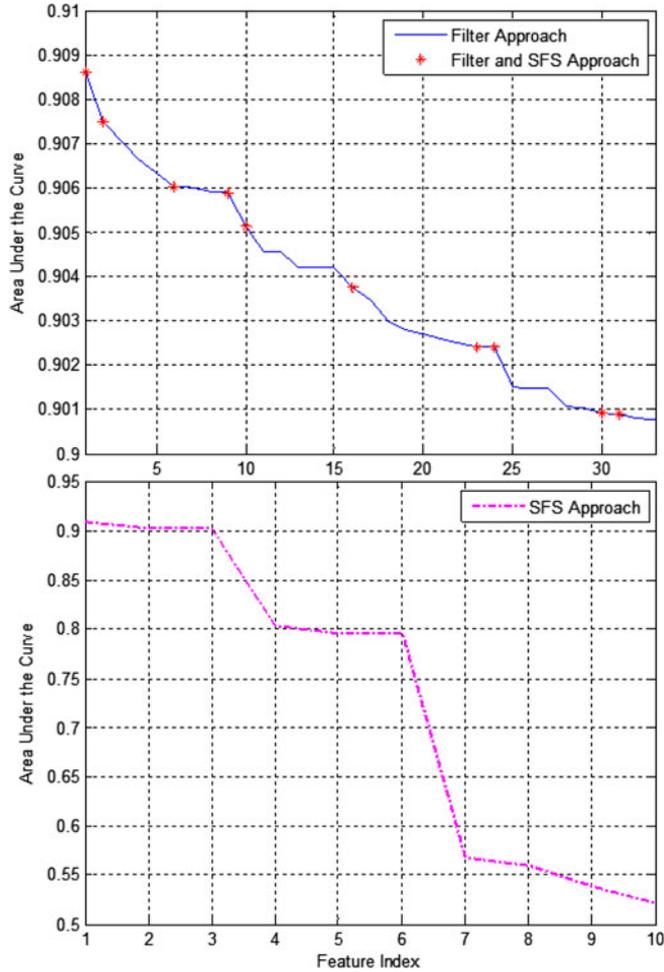| Feature Name	Equation	Definition
Mean Intensity	$I_\mu = \dfrac{\sum_i \sum_j I_s(i,j)}{N_s}$	Mean value of superpixel
Area	N_s	Number of pixels in Superpixel
Convex Area	N_{sc}	Number of pixels in convex area of superpixel
Extent	$\text{Ext} = \frac{N_s}{N_{sb}}$	Ratio of area to number of pixels in the bounding box
Orientation	θ_s	Superpixel angle with respect to *x*-axis
Solidity	$\text{Sol} = \frac{N_s}{N_{sc}}$	Ratio of area to convex area

Fig. 4. Plot of independent evaluation criterion. The features are ranked in descending order of independent evaluation criterion value. In top figure, red dots for "Filter and SFS approach" represent the ten features selected by applying SFS on "Filter approach" set. By applying SFS on complete feature set, ten out of 295 features have been selected as shown in bottom figure ("SFS approach").

where b is the bias value and W are the weights of input x. These weights can be determined by *backpropagation algorithm,* which tends to minimize mean square error value between desired output and actual output as

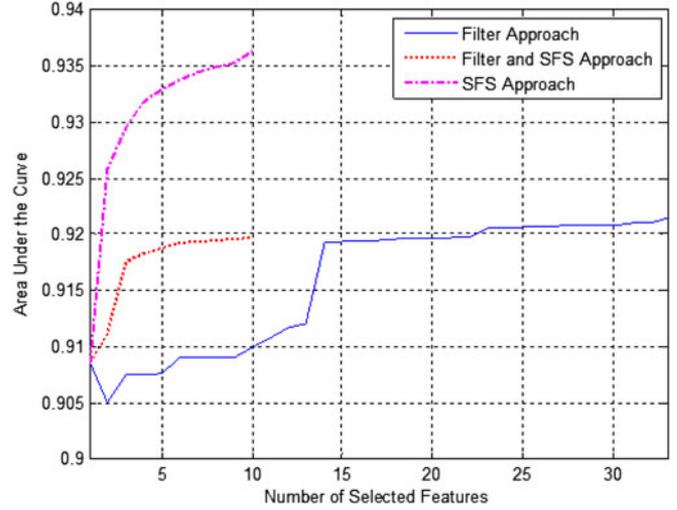$$\text{err} = \frac{1}{2}(t - y)^2 \tag{5}$$

Fig. 5. Plot of AUC by selecting the features one by one in different feature set.

TABLE III
FEATURE SETS OBTAINED USING DIFFERENT FEATURE SELECTION
APPROACHES

Feature Selection Method	Feature Symbols
Filter Approach (feature index and number of selected features)	$\mu_{\text{sum}R}(16), \sigma_{\text{sum}R}(16), \mu_{\text{sum}R}(8),$ $\mathcal{N}_R(16), \mu_{\text{sum}R}(4), \sigma_{\text{sum}R}(8),$ $\mu_{\text{sum}R}(2), \mu_{\text{sum}R}(1), \mu_{\text{sum}R},$ $\sigma_{\text{sum}R}(4), \sigma_{\text{sum}R}(2), \mathcal{N}_R(8),$ $\sigma_{\text{sum}R}(1), \text{acorr}_R(16), \sigma_{\text{sos}R}(16),$ $\sigma_{\text{sum}R}, \mathcal{N}_R(4), \mathcal{N}_R(2), \mathcal{N}_R(1),$ $\mathcal{N}_{yyR}(1), I_{\mu R}, \mathcal{N}_{xxR}(1), \text{acorr}_R(8),$ $\sigma_{\text{sos}R}(8), \text{acorr}_R(4), \sigma_{\text{sos}R}(4),$ $\mathcal{N}_{yyR}(2), \text{acorr}_R(2), \sigma_{\text{sos}R}(2),$ $\text{acorr}_R(1), \sigma_{\text{sos}R}(1), \text{acorr}_R, \sigma_{\text{sos}R}$
Filter and SFS Approach (feature index)	$\mu_{\text{sum}R}(16), \sigma_{\text{sum}R}(16), \sigma_{\text{sum}R}(8),$ $\mu_{\text{sum}R}, \sigma_{\text{sum}R}(4), \sigma_{\text{sum}R},$ $\text{acorr}_R(8), \sigma_{\text{sos}R}(8), \text{acorr}_R(1),$ $\sigma_{\text{sos}R}(1)$
Filter and SFS Approach (number of selected features)	$\mu_{\text{sum}R}(16), \sigma_{\text{sos}R}(1), \sigma_{\text{sum}R}(8),$ $\sigma_{\text{sos}R}(8), \sigma_{\text{sum}R}(16), \mu_{\text{sum}R},$ $\sigma_{\text{sum}R}, \text{acorr}_R(8), \text{acorr}_R(1),$ $\sigma_{\text{sum}R}(4)$
SFS Approach (feature index)	$\mu_{\text{sum}R}(16), \text{acorr}_R(8), \sigma_{\text{sos}R}(8),$ $\sigma_{\text{sum}G}, \text{acorr}_G, \sigma_{\text{sos}G}, H_G(8),$ $\mathcal{N}_{yR}(16), H_G(1), H_{\text{diff}G}(1)$
SFS Approach (Number of Selected Features)	$\mu_{\text{sum}R}(16), \sigma_{\text{sos}G}, H_G(8),$ $\mathcal{N}_{yR}(16), \sigma_{\text{sos}R}(8), H_{\text{diff}G}(1),$ $\text{acorr}_G, \text{acorr}_R(8), \sigma_{\text{sum}G}, H_G(1)$

"Feature index" represents the order of highest independent evaluation criterion measure, and "number of selected features" represent the sequence of feature selection in the feature set. R and G subscripts represent red and green channel, respectively.

TABLE IV
PERCENTAGE OF DIFFERENT TYPES OF FEATURES ACROSS
DIFFERENT FEATURE SET

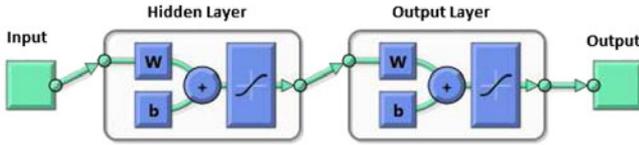| Feature Set	Textural Features	Gradient Features	Regional Features
SFS Approach	90%	10%	0%
Filter Approach	72.73%	24.24%	3.03%
Filter and SFS Approach	100%	0%	0%

Fig. 6.   AANs diagram.

where t and y represent the target output and actual output of the output layer. The minimization of (5) can be represented as

$$\frac{\partial \text{err}}{\partial W_i} = (y - t)y(1 - y)x_i. \tag{6}$$

Since it is an iterative process, therefore weights are updated by *delta rule* as

$$\Delta w_i = \alpha(t - y)x_i \tag{7}$$

α represents the step size. The weights were updated until 1000 iterations.

F. Image Postprocessing

After classification of the test image, the superpixels are refined using morphological operation [3], so as to remove misclassified isolated superpixels. The morphological closing was performed so as to remove small gaps among superpixels. The size of disk structuring element can be a smaller value, say 10. For better results, we can perform area opening so as to remove one or two misclassified isolated superpixels.

G. Comparison Study

After the construction of our classifier, we have compared its performance against different classifiers in terms of accuracy and computational time. The classifiers have been applied across different feature sets, which are obtained by using different feature selection procedures as mentioned in Section III-D. The classifiers we have selected for comparing the performance of our classifier are SVMs and *k* Nearest Neighbours (*k*NNs) [26].

The idea behind *k*NN method is to find out samples whose feature are similar to the classes to be detected. The function, which we are following in order to determine the similarity of the features with true retinal area is "Euclidean distance." SVM finds a separating hyperplane with the maximal margin in higher dimensional space. In our comparison study, we are using nonlinear SVM with radial-based function kernel with default parameter of (number of features)$^{-1}$ = 0.1 [29].

IV. EXPERIMENTAL EVALUATION

The images for training and testing have been obtained from Optos [2] and are acquired using their ultrawide field SLO. Each image has a FOV of up to 200° of the retina in a resolution of 14 μm. The device captures the retinal image without dilation, through a small pupil of 2 mm. The image has two channels: red and green. The green channel (wavelength: 532 nm) provides information about the sensory retina to retinal pigment epithelium, whereas the red channel (wavelengh: 633 nm) shows deeper structures of the retina toward the choroid. Each image has a dimension of 3900 × 3072 and each pixel is represented by 8-bit on both red and green channels. The dataset is composed of healthy and diseased retinal images; most of the diseased retinal images are from Diabetic Retinopathy patients. The system has been trained with 28 images and tested against 76 images.

Fig. 7 compares the classification power of different feature sets with the help of receiver operating characteristics (ROC). One of those feature sets include all features calculated. The rest of other feature sets include features selected by the approaches discussed in Section III-D. By using SFS approach, ten features out of 295 features have been selected and their calculation time is 25 s per image, whereas calculating the complete feature set can take around 10 min per image. The ROC curves and AUC values reveal that if the features are selected using the SFS approach, they can have a classification power almost similar to the complete feature set while reducing the computational time.

The visual results and the accuracies of different classifiers among different feature sets has been presented using Dice Coefficient as evaluation metric. The Dice Coefficient is the degree of overlap between the framework output and the benchmark obtained from the clinician. The Dice Coefficient is defined as

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|} \tag{8}$$

where A and B are the segmented images obtained from the framework and the benchmark, respectively, $|.|$ represents number of samples of the region, and $\cap$ denotes the intersection. Its value varies between 0 and 1, where a higher value indicates an increased degree of overlap. Let RA_1 and AR_1 represent samples from the retinal area and the artefact area obtained from the framework, respectively, and RA_2 and AR_2 be these samples from the benchmark. The class of superpixels in the benchmark was decided based on majority of pixels in the superpixel belonging to particular class. Also, $|RA_1| + |AR_1| = |RA_2| + |AR_2| = N_{\text{sample}}$, i.e., total number of samples (superpixels or pixels) in an image. If we calculate Dice Coefficient for the image, (8) can be deduced as

$$D_I = \frac{(|RA_1 \cap RA_2| + |AR_1 \cap AR_2|)}{N_{\text{sample}}}. \tag{9}$$

The Dice Coefficient for the retinal area D_R and artefacts D_A will be given as

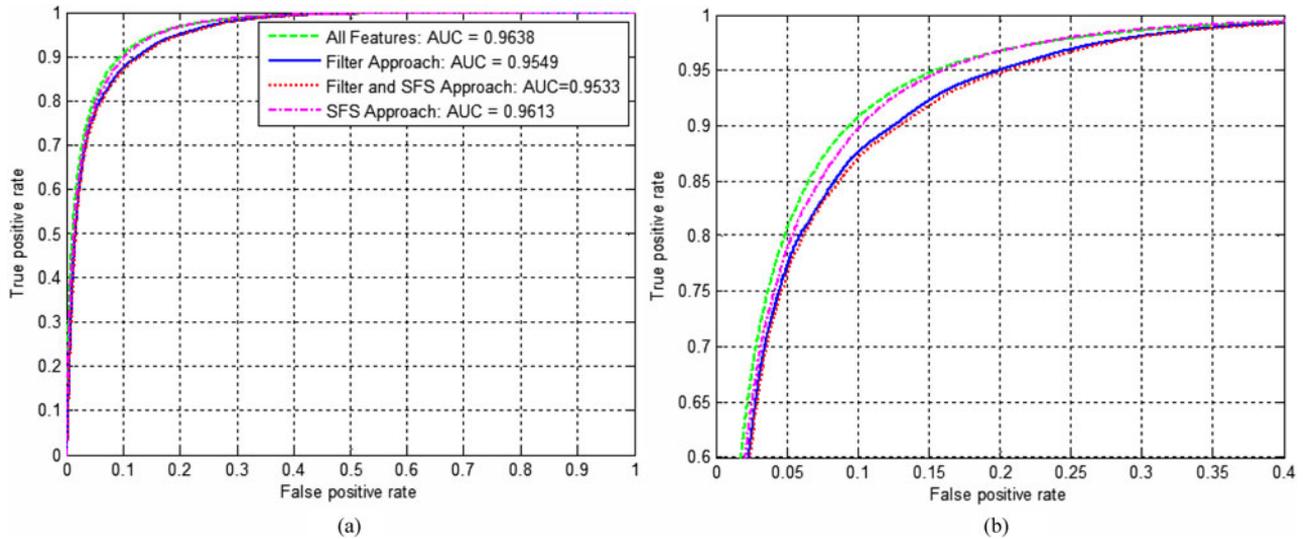$$D_R = \frac{2|RA_1 \cap RA_2|}{|RA_1| + |RA_2|}, D_A = \frac{2|AR_1 \cap AR_2|}{|AR_1| + |AR_2|}. \tag{10}$$

Fig. 7.   (a) ROC on the test sets. (b) Magnified version of (a).

TABLE V
AVERAGE CLASSIFICATION ACCURACY

Classifier	Filter Approach			Filter/SFS Approach			SFS Approach		
	D_I	D_R	D_A	D_I	D_R	D_A	D_I	D_R	D_A
ANN	89.36%	89.49%	89.22%	88.88%	89.00%	88.75%	90.48%	90.28%	90.68%
SVM	88.48%	88.48%	88.47%	88.41%	88.36%	88.46%	90.93%	90.89%	90.96%
kNN	88.35%	88.53%	88.17%	88.09%	88.24%	87.94%	90.34%	90.17%	90.52%

Degree of overlap has been calculated by taking superpixels as samples.

TABLE VI
COMPARISON OF FRAMEWORK OUTPUT PERFORMANCE
USING DIFFERENT CLASSIFIERS

Classifier	Training Time	Testing Time	D_I	D_R
ANN	30 min	0.013 s	91.93%	91.87%
SVM	12.5 min	8.5 s	92.00%	91.94%
kNN	1.45 s	2.05 s	91.43%	91.31%

The performance is compared with respect to computational time taken during training and testing and average accuracy. The training time is calculated for 28 images. Testing time shows the average time taken by the framework.

Table V compares the performance of different classifiers across different feature sets. As far as classification accuracy is concerned, there is a little difference among the outputs of different classifiers. The advantage of using ANN is its high-computational efficiency in terms of testing time as shown in Table VI. Although the training time of ANN is longer compared to its other two counterparts, the training time is once in a life-time process and once the model is deployed, it can process any image. Fig. 9 represents the total time taken by an image to be processed for automatic annotations. The block diagram and the Table VI shows that while using ANN, couple of seconds can be saved per image during automatic–annotation process. As shown in Table V, SVM although performed better on SFS feature set compared to ANN and kNN, ANN has the highest classification accuracy in other two feature sets. This shows that classification accuracy is highly dependent on type of features selected.

Fig. 8 shows superpixel classification results and final output after postprocessing of different examples of healthy and diseased retinal images. ANN is able to achieve the average accuracy nearer to that of other two classifiers, while saving significant computational time when processing millions of images for automatic annotations.

V. DISCUSSION AND CONCLUSION

Distinguishing true retinal area from artefacts in SLO images is a challenging task, which is also the first important step toward computer-aided disease diagnosis. In this study, we have proposed a novel framework for automatic detection of true retinal area in SLO images. We have used superpixels to represent different irregular regions in a compact way and reduce the computing cost. Feature selection enables the most significant features to be selected and, thus, reduces computing cost too. A classifier has been built based on selected features to extract out the retina area. It has been compared to other two classifiers and was compatible while saving the computational time. The experimental evaluation result shows that our proposed
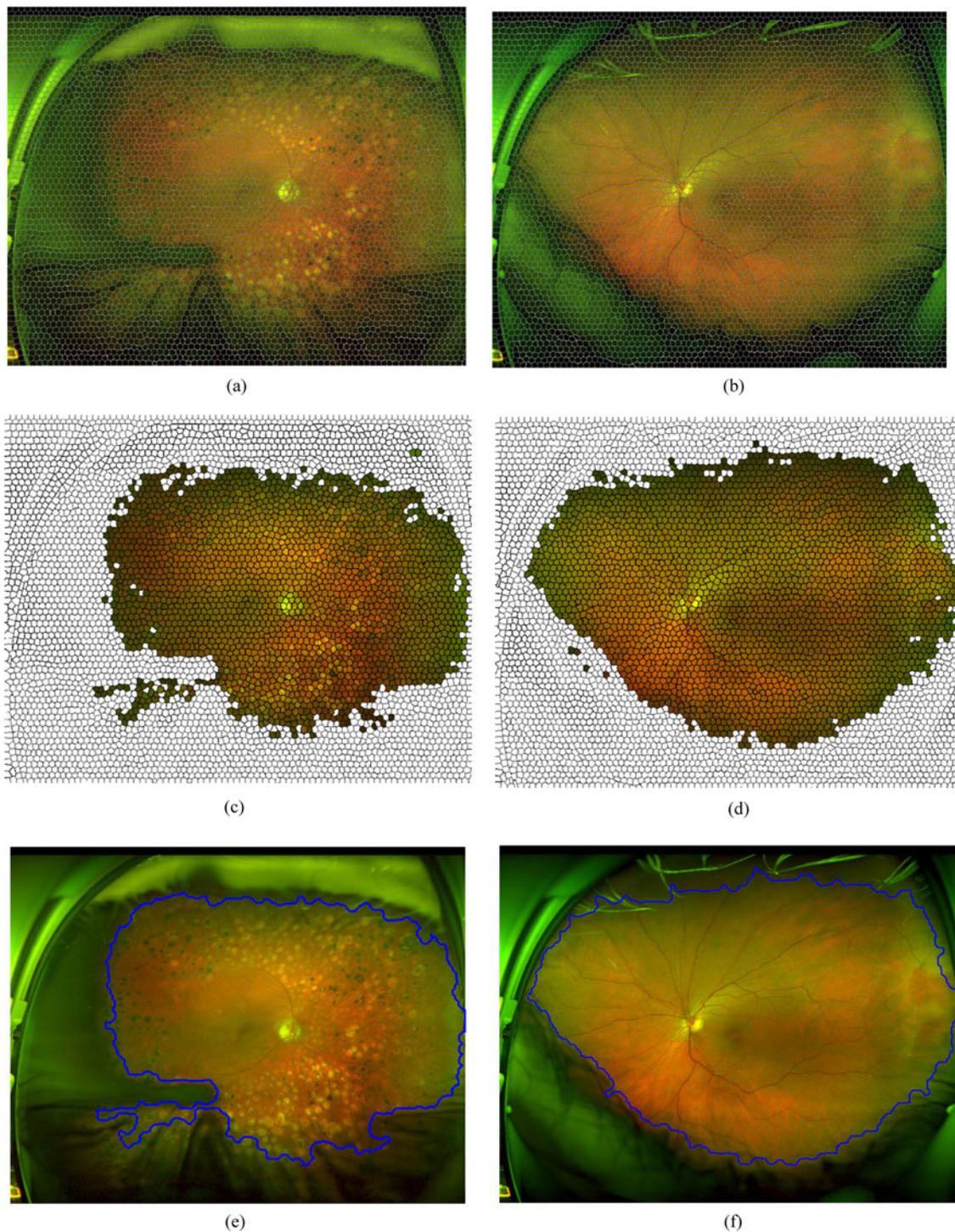
Fig. 8. Superpixel classification result of two examples of SLO images. Columns represent different examples of retinal images. Left column are retinal scans with lesions, whereas right column is the retinal scan from healthy subject. (a) and (b) represent the test images divided into superpixels. (c) and (d) represent superpixel classification results and (e) and (f) represent output after postprocessing.

framework can achieve an accuracy of 92% in segmentation of the true retinal area from an SLO image.

Feature selection is necessary so to reduce computational time during training and classification. Among different approaches used for feature selection, the performance of our feature selection approach surpassed the filter approach and "Filter and SFS" approaches in terms of classification power. The comparison of different feature selection approaches shows that selection of features based on their mutual interaction can provide the classification power close to that of feature set with
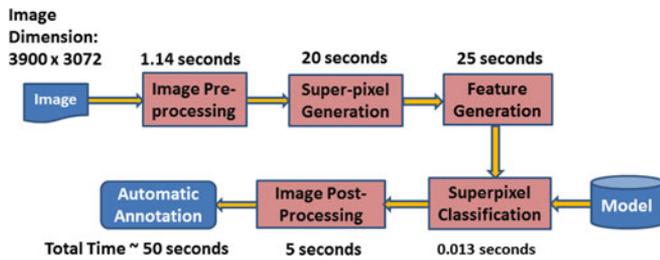
Fig. 9. Block diagram of deployment stage along with execution time of each block.

all features. Feature selection is once in a life-time process and we can compromise on computational time for feature selection on account of accuracy.

As far as the classifier is concerned, the testing time of ANN was the lowest compared to other two classifiers. Although the overall accuracy of SVM was the highest compared to other two classifiers, the training and testing time is quite long. Although *k*NN has the shortest training time, the testing time can be quite high compared to ANN while processing millions of images. Compared to SVM, we can tradeoff the overall accuracy of 0.1% on average while saving the testing time of 8 s per image. As far as images with lesions are concerned [see Fig. 8(a), (c), and (e)], ANN misclassified 1 or 2 superpixels at the corners, but they are corrected using morphological postprocessing as shown in Fig. 8(e).

Our retina detection framework serves as the first step toward the processing of ultrawidefield SLO images. A complete retinal scan is possible if the retina is imaged from different eye-steered angles using an ultrawidefield SLO and, then, montaging the resulting image. Montaging is possible only if the artefacts are removed before.

REFERENCES

[1] M. S. Haleem, L. Han, J. van Hemert, and B. Li, "Automatic extraction of retinal features from colour retinal images for glaucoma diagnosis: A review," *Comput. Med. Imag. Graph.*, vol. 37, pp. 581–596, 2013.

[2] Optos. (2014). [Online]. Available: www.optos.com

[3] R. C. Gonzalez and R. E. Woods, Eds., *Digital Image Processing*, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2006.

[4] M. J. Aligholizadeh, S. Javadi, R. S. Nadooshan, and K. Kangarloo, "Eyelid and eyelash segmentation based on wavelet transform for iris recognition," in *Proc. 4th Int. Congr. Image Signal Process.*, 2011, pp. 1231–1235.

[5] D. Zhang, D. Monro, and S. Rakshit, "Eyelash removal method for human iris recognition," in *Proc. IEEE Int. Conf. Image Process.*, 2006, pp. 285–288.

[6] A. V. Mire and B. L. Dhote, "Iris recognition system with accurate eyelash segmentation and improved FAR, FRR using textural and topological features," *Int. J. Comput. Appl.*, vol. 7, pp. 0975–8887, 2010.

[7] Y.-H. Li, M. Savvides, and T. Chen, "Investigating useful and distinguishing features around the eyelash region," in *Proc. 37th IEEE Workshop Appl. Imag. Pattern Recog.*, 2008, pp. 1–6.

[8] B. J. Kang and K. R. Park, "A robust eyelash detection based on iris focus assessment," *Pattern Recog. Lett.*, vol. 28, pp. 1630–1639, 2007.

[9] T. H. Min and R. H. Park, "Eyelid and eyelash detection method in the normalized iris image using the parabolic Hough model and Otsus thresholding method," *Pattern Recog. Lett.*, vol. 30, pp. 1138–1143, 2009.

[10] Iris database. (2005). [Online]. Available: http://www.cbsr.ia.ac.cn/IrisDatabase.htm

[11] H. Davis, S. Russell, E. Barriga, M. Abramoff, and P. Soliz, "Vision-based, real-time retinal image quality assessment," in *Proc. 22nd IEEE Int. Symp. Comput.-Based Med. Syst.*, 2009, pp. 1–6.

[12] H. Yu, C. Agurto, S. Barriga, S. C. Nemeth, P. Soliz, and G. Zamora, "Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation*, 2012, pp. 125–128.

[13] J. A. M. P. Dias, C. M. Oliveira, and L. A. d. S. Cruz, "Retinal image quality assessment using generic image quality indicators," *Inf. Fusion*, vol. 13, pp. 1–18, 2012.

[14] M. Barker and W. Rayens, "Partial least squares for discrimination," *J. Chemom.*, vol. 17, pp. 166–173, 2003.

[15] J. Paulus, J. Meier, R. Bock, J. Hornegger, and G. Michelson, "Automated quality assessment of retinal fundus photos," *Int. J. Comput. Assisted Radiol. Surg.*, vol. 5, pp. 557–564, 2010.

[16] R. Pires, H. Jelinek, J. Wainer, and A. Rocha, "Retinal image quality analysis for automatic diabetic retinopathy detection," in *Proc. 25th SIBGRAPI Conf. Graph., Patterns Images*, 2012, pp. 229–236.

[17] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[18] A. Moore, S. Prince, J. Warrell, U. Mohammed, and G. Jones, "Superpixel lattices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.

[19] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and supervoxels in an energy optimization framework," in *Proc. 11th Eur. Conf. Comput. Vis.*, 2010, pp. 211–224.

[20] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Learning*, vol. 13, no. 6, pp. 583–598, Jun. 1991.

[21] J. Cheng, J. Liu, Y. Xu, F. Yin, D. Wong, N.-M. Tan, D. Tao, C.-Y. Cheng, T. Aung, and T. Y. Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1019–1032, Jun. 2013.

[22] L. Tang, M. Niemeijer, J. Reinhardt, and M. Garvin, "Splat feature classification with application to retinal hemorrhage detection in fundus images," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 364–375, Feb. 2013.

[23] M. Abràmoff, W. Alward, E. Greenlee, L. Shuba, C. Kim, J. Fingert, and Y. Kwon, "Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features," *Invest. Ophthalmol. Vis. Sci.*, vol. 48, pp. 1665–1673, 2007.

[24] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[25] R. Haralick and L. Shapiro, Eds., *Computer and Robot Vision*. Reading, MA, USA: Addison-Wesley, 1991.

[26] R. O. Duda, P. E. Hart, and D. G. Stork, Eds., *Pattern Classification*. New York, NY, USA: Wiley-Interscience, 2000.

[27] A. J. Serrano, E. Soria, J. D. Martin, R. Magdalena, and J. Gomez, "Feature selection using ROC curves on classification problems," in *Proc. Int. Joint Conf. Neural Netw.*, 2010, pp. 1–6.

[28] H. Liu and H. Motoda, Eds., *Feature Selection for Knowledge Discovery and Data Mining*. Norwell, MA, USA: Kluwer, 1998.

[29] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," Dept. Comput. Sci., National Taiwan Univ., Taipei, Taiwan, 2010.

Muhammad Salman Haleem received the B.Eng. degree in electronic engineering from the NED University of Engineering and Technology, Karachi, Pakistan, in 2008, and the M.S. degree in electrical engineering from the Illinois Institute of Technology, Chicago, IL, USA, in 2011. He is currently working toward the Ph.D. degree as a Research Student at the School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester, U.K.

The title of his Ph.D. thesis is Automatic Detection of Features to Assist Diagnosis of Retinal Diseases. For this project, he received the prestigious Dorothy Hodgkin Postgraduate Award from the Engineering and Physical Sciences Research Council. His areas of research interests include computer vision, image processing, machine learning, and data mining.

Liangxiu Han received the Ph.D. degree in computer science from Fudan University, Shanghai, China, in 2002.

She is currently a Reader at the school of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Manchester, U.K., where he leads the Future Network and Distributed Systems Research Group. Her research interests include the development of novel architectures for large-scale networked distributed systems (e.g., cloud/grid/service-oriented computing/internet), large-scale data mining (application domains include web mining, biomedical images, environmental sensor data, network traffic data, cyber security, etc.), and knowledge engineering. As a Principal or Coprincipal Investigator, her research were funded by research councils, industries, and charity bodies, in her research areas. She is a Member of Engineering and Physical Sciences Research Council Peer Review College and as an Expert for Horizon 2020 proposal evaluation. She is also a regular Reviewer for several prestigious journals and international conferences in the field.

Baihua Li received the B.Sc. and M.Sc. degrees in electronic engineering from Tianjin University, Tianjin, China, and the Ph.D. degree in computer science from Aberystwyth University, Aberystwyth, U.K.

She is currently a Senior Lecturer in the School of Computing, Mathematics &; Digital Technology, Manchester Metropolitan University, Manchester, U.K. Her current research interests include computer vision, image processing, pattern recognition, advanced computer graphics, human motion analysis, and behavior understanding from multimodality imaging and sensory data. About 40 fully refereed research papers have been published in leading national/international journals and conferences, including IEEE Transactions, Pattern Recognition (PR) and Image and Vision Computing (IVC). She takes the role of a Reviewer and Program Committee Member for a number of high-quality journals and conferences.

Dr. Li is a Member of the British Machine Vision Association.

Jano van Hemert received the Ph.D. degree in mathematics and physical sciences from Leiden University, The Netherlands, in 2002.

He is currently the Imaging Research Manager and Academic Liaison at Optos, Dunfermline, U.K.; an innovative retinal imaging company with a vision to be the leading provider of retinal diagnostics. Since 2010, he is an Honorary Fellow of the University of Edinburgh. Since 2011, he is a Member of the Young Academy of the Royal Society of Edinburgh. From 2007 to 2010, he led the research of the U.K. National e-Science Centre, supported by an Engineering and Physical Sciences Research Council Platform Grant. He has held research positions at the Leiden University (NL), the Vienna University of Technology (AT), and the National Research Institute for Mathematics and Computer Science, NL.

Dr. Hemert received the Talented Young Researcher Fellowship by the Netherlands Organization for Scientific Research in 2004. In 2009, he was recognised as a promising Young Research Leader with a Scottish Crucible.

Alan Fleming received the Ph.D. degree from the Artificial Intelligence Department, University of Edinburgh, Edinburgh, U.K.

He is currently an Imaging Algorithms Developer and a Member of the Research Team at Optos, Dunfermline, U.K. He has held various academic and industrial positions in the fields of medical ultrasound and ophthalmic image analysis. He developed algorithms for the detection of diabetic retinopathy in retinal photographs and took part in clinical validations which led to their implementation within NHS Scotland's Diabetes Retinal Screening Programme.