

Splat Feature Classification With Application to Retinal Hemorrhage Detection in Fundus Images

Li Tang, Meindert Niemeijer, Joseph M. Reinhardt, *Senior Member, IEEE*, Mona K. Garvin, *Member, IEEE*, and Michael D. Abramoff*, *Senior Member, IEEE*

Abstract—A novel splat feature classification method is presented with application to retinal hemorrhage detection in fundus images. Reliable detection of retinal hemorrhages is important in the development of automated screening systems which can be translated into practice. Under our supervised approach, retinal color images are partitioned into nonoverlapping segments covering the entire image. Each segment, i.e., splat, contains pixels with similar color and spatial location. A set of features is extracted from each splat to describe its characteristics relative to its surroundings, employing responses from a variety of filter bank, interactions with neighboring splats, and shape and texture information. An optimal subset of splat features is selected by a filter approach followed by a wrapper approach. A classifier is trained with splat-based expert annotations and evaluated on the publicly available Messidor dataset. An area under the receiver operating characteristic curve of 0.96 is achieved at the splat level and 0.87 at the image level. While we are focused on retinal hemorrhage detection, our approach has potential to be applied to other object detection tasks.

Index Terms—Diabetic retinopathy (DR), fundus image, retinal hemorrhage, splat feature classification.

I. INTRODUCTION

AUTOMATED detection of diabetic retinopathy (DR), as used in screening systems, is important for allowing timely treatment [1], and thereby increasing accessibility to and productivity of eye care providers. Because of its cost-effectiveness and patient friendliness, digital color fundus photography is a prerequisite for automated DR detection [2]. Patients with images that are likely to contain DR are detected and referred for further management by eye care providers.

Manuscript received August 12, 2012; revised October 10, 2012; accepted October 21, 2012. Date of publication November 16, 2012; date of current version January 30, 2013. This work was supported in part by the National Institutes of Health (EY017066) and in part by the U.S. Department of Veterans Affairs (VA 5101CX000119, 11K2RX000728). Asterisk indicates corresponding author.

L. Tang and M. Niemeijer are with the Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA 52242 USA (e-mail: li-tang-1@uiowa.edu; meindertn@gmail.com).

J. M. Reinhardt is with the Department of Biomedical Engineering, University of Iowa, Iowa City, IA 52242 USA (e-mail: joe-reinhardt@uiowa.edu).

M. K. Garvin is with the Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA and also with the Department of Veterans Affairs, Iowa City VA Medical Center, Iowa City, IA 55242 USA (e-mail: mona-garvin@uiowa.edu).

*M. D. Abramoff is with the Department of Ophthalmology and Visual Sciences, Department of Biomedical Engineering, Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 USA and also with the Department of Veterans Affairs, Iowa City VA Medical Center, Iowa City, IA 55242 USA (e-mail: michael-abramoff@uiowa.edu).

Digital Object Identifier 10.1109/TMI.2012.2227119

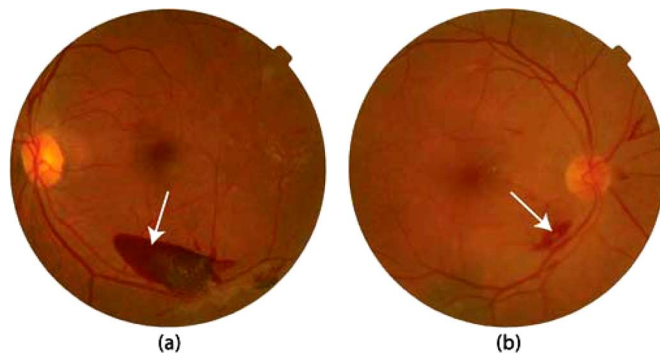


Fig. 1. Examples of retinal hemorrhages with different shapes and appearances.

The most common signs of DR are microaneurysms, small hemorrhages, exudates, drusen, and cottonwool spots. Because of the variability in appearance of these lesions, different techniques have been designed to detect each type of these lesions separately in DR detection systems. Retinal hemorrhages are caused by retinal ischemia and primarily caused by abnormally fragile blood vessels in hypertension, malaria and primarily, preproliferative and proliferative DR. Large hemorrhages are asymptomatic except when they are located in the center of the macula. Two examples of large retinal hemorrhages are demonstrated in Fig. 1. Compared with anatomical structures, such as optic disc, fovea and blood vessels, the shape and appearance of hemorrhages show substantial variability.

Our work on evaluation of automated DR detection systems shows that an important cause of false negatives, as high as 50%, is formed by images that contain only large hemorrhages [3]. Large hemorrhages indicate more severe disease, and improved detection of such lesions will lead to elimination of more severe false negatives.

Small hemorrhages are regular in shape and many systems have been developed by us and others to detect them [4]–[6]. A review of most recent work on hemorrhage detection can be found in [7]. They primarily fall into three categories: pixel-based approaches, lesion-based approaches, and image-based approaches. Pixel-based approaches focus on the location of hemorrhages on the retina. Lesion-based approaches use morphological operations to define candidate lesions and count them. Image-based approaches are aimed at detecting images or eyes with hemorrhages. However, the size of the lesion is yet another important factor to consider in decision making processes of DR detection systems, which is closely related to the severity of disease that need timely treatment. Large hemorrhages occur infrequently, and their

appearance is highly variable, making their shape modeling and automated detection a challenge.

Detecting DR lesions is often accomplished by supervised classification [3], which involves training of classifiers using expert labeled target objects at pixel level. Features are extracted from each pixel and soft labels are assigned accordingly, indicating the probability of the pixel being one or part of a target object. Abnormal pixels are then combined into objects [8]. While their performance approaches that of clinicians [1], problems still exist.

- It is expensive to acquire expert labeled reference standards for training and evaluation. Designing such systems requires substantial work by clinicians to define the reference standard, which is expensive and prone to error.
- Ideally training samples are intended to be both informative to the classification model and diverse so that information provided by individual samples overlaps as little as possible [9]. However, often in a single training image, there can be a huge number of very similar pixel samples.
- Large hemorrhages occur infrequently, have nonregular shape and can occur without accompanying other signs of DR, such as microaneurysms or small hemorrhages. They will thus be missed by systems designed to detect the regular DR lesions [8]. Because of their low occurrence, sensitivity for detection of large hemorrhages has negligible effect on unweighted performance metrics such as AUC.

We address these problems using a higher level entity—the splat, which is a collection of pixels with similar color and spatial location [10]. As hemorrhages consist of blood, they share appearance features with intravascular blood. That makes it difficult to differentiate these from retinal vessels using low level pixel features. On the contrary, by upgrading samples for classification from pixel level to splat level, information is encoded at the splat level, with less disturbance from pixel level noise.

The purpose of this study is to present a supervised classification algorithm to detect large, irregular retinal hemorrhages. Reference standard hemorrhage locations were delineated by a retinal specialist (MDA) using splat-based image representation. Supervised classification predicts the likelihood of splats being hemorrhages with the optimal feature subset selected in a two-step feature selection process. From the resulting *hemorrhageness map*, a hemorrhage index is assigned as the image level output.

The remainder of the paper is organized as follows. Section II describes splat-based image segmentation and representation. Section III explores splat feature space. In Section IV splat feature extraction and classification are presented. The proposed algorithm is verified in experiments and results are given in Section V. We end with discussion and conclusion in Section VI.

II. SPLAT SEGMENTATION

Based on the assumption that pixels that are part of the same object or structure share similar color, intensity and spatial location, the image is partitioned into nonoverlapping splats of similar intensity covering the entire image [10].

Splat-based representation is an image re-sampling strategy onto an irregular grid. Background regions, with gradual variations in appearance, tend to consist of fewer large splats while

foreground regions consist of a larger number of smaller splats. At pixel level, the distributions of hemorrhage pixels and non-hemorrhage pixels are imbalanced [11], since hemorrhages usually account for a small fraction of the entire image. Instead of including only a subset of background pixels for training, as many resampling methods do, a splat-based approach maximizes the diversity of training samples by retaining all important samples.

A. Scale-Specific Image Over-Segmentation

Splats are created by over-segmenting images using watershed or toboggan algorithms [10]. Conventional image over-segmentation on a regular grid generates so called “superpixels” [12], [13], a similar concept to “splats.” But superpixels are roughly homogeneous in size and shape, resulting in a lattice pattern [14]. In contrast, a splat-based approach divides images into an irregular grid, depending on properties of target objects to be detected.

To create splats which preserve desired boundaries precisely, i.e., boundaries separating hemorrhages from retinal background, we perform a scale-specific image over-segmentation in two steps. Due to the variability in the appearance of hemorrhages, we firstly aggregate gradient magnitudes of the contrast enhanced dark-bright opponency image at a range of scales for localization of contrast boundaries separating blood and retinal background. Next, the maximum of these gradients over scale-of-interest (SOI) is taken in performing a watershed segmentation [15].

Assuming that we establish a scale-space representation of image $I(x, y; s)$ with Gaussian kernels G_s at SOI $s \in s_1, \dots, s_n$, the gradient magnitude $|\nabla I(x, y; s)|$ is computed from its horizontal and vertical derivatives

$$\begin{aligned}
 |\nabla I(x, y; s)| &= \sqrt{I_x(x, y; s)^2 + I_y(x, y; s)^2} \\
 &= \sqrt{\left[\frac{\partial}{\partial x} (G_s * I(x, y)) \right]^2 + \left[\frac{\partial}{\partial y} (G_s * I(x, y)) \right]^2} \\
 &= \sqrt{\left[\frac{\partial G_s}{\partial x} * I(x, y) \right]^2 + \left[\frac{\partial G_s}{\partial y} * I(x, y) \right]^2}
 \end{aligned}
 \quad s = s_1, \dots, s_n \quad (1)$$

where symbol $*$ represents convolution and $(\partial G_s)/(\partial x)$, $(\partial G_s)/(\partial y)$ are the first order derivatives of Gaussian at scale s along the horizontal and vertical direction. The maximum of the gradient magnitude aggregated over the scale band $|\nabla I(x, y)|$ is

$$|\nabla I(x, y)| = \max_i |\nabla I(x, y; s_i)|. \quad (2)$$

The application of gradient magnitude from a maximum pooling operation $|\nabla I(x, y)|$ across certain scales as the topographic surface in watershed segmentation [15] is important to obtain meaningful splats preserving hemorrhage boundaries precisely. A comparison of it with the original intensity image and gradient images outside SOI as the topographic surface for splat creation is given in Fig. 2. Each image in this figure contains a similar number of splats generated by

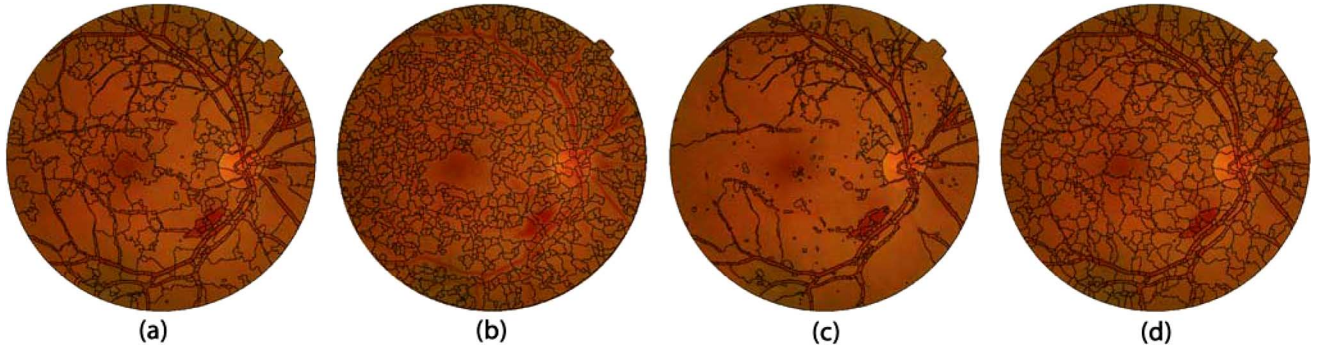


Fig. 2. Scale-specific image over-segmentation create splats preserving hemorrhage boundaries. Watershed algorithm with topographic surface obtained as (a) maximum of gradient magnitude from desired scale band; (b) pixel intensity; (c) gradient magnitude from a fine scale outside scale-of-interest; (d) gradient magnitude from a coarse scale outside scale-of-interest.

the same watershed algorithm. It reveals that splat distribution depends closely upon the method of acquiring a gradient image [16]. In Fig. 2(a), high-frequency noise at finer scales such as those in Fig. 2(c) is removed and splats boundaries are accurately aligned with boundaries of blood regions instead of being trapped at other contrast magnitudes as those shown in Fig. 2(d). Details related with blood regions are represented by small splats and retinal background consists of much larger splats in Fig. 2(a). Given similar number of splats, those in Fig. 2(d) have more uniform sizes. The scale in Fig. 2(d) may be more appropriate in dealing with structures such as optic disc while that in Fig. 2(c) may be more appropriate in dealing with microaneurysms, drusen, etc. The scale band controls the very type of intensity transitions or the amount of details to be captured using splat-based image representation.

The number of splats in each image is set to be within a limit, which is achieved by thresholding the topographic surface iteratively. The threshold is increased by a constant step until the number of splats is lower than a predefined upper bound. This is essentially a compromise between accuracy and efficiency. The larger the number of splats, the closer the splat-based representation approaches the original image.

B. Splat-Based Reference Standard Acquisition

Supervised algorithms require labeled samples by experts, but it is expensive to acquire such data [17], because substantial time is required to delineate irregular boundaries of hemorrhages. Any misalignment with true boundaries introduces noise at the training stage. Given the limited number of training samples, it has considerable impact on system performances. This problem may be simplified by splat-based image formulation. We compared both a pixel-based approach and a splat-based approach in Fig. 3 using “Truthmarker”—an iPad app developed to provide a convenient user interface for clinicians to perform reference standard annotation [18].

For pixel level annotation illustrated in Fig. 3(a), we allow two types of annotations by expert: large hemorrhages and small hemorrhages. Large hemorrhages are indicated by a few points along the boundaries (shown as small circles) and then spline fitting is applied to connect those discrete points as enclosed curves shown in cyan. Small hemorrhages are indicated by a single point shown as a green dot. Thus considerable

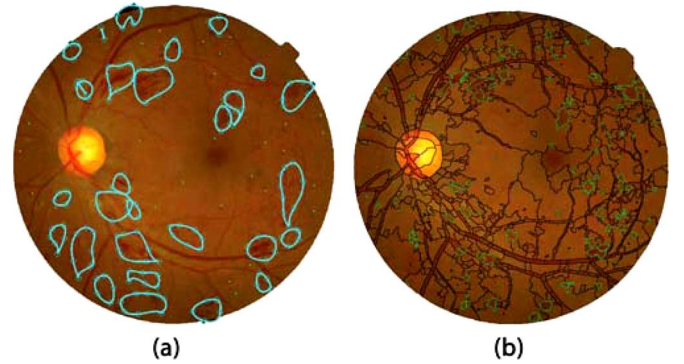


Fig. 3. Sample labellings acquired from expert annotation with: (a) pixel-based approach; (b) splat-based approach.

noise is introduced and the time costs of experts are still high. For splat level annotation shown in Fig. 3(b), this process is simplified substantially. Experts only perform a single click in a splat to indicate a hemorrhage splat. As splats preserve hemorrhage boundaries, the resulting reference standard is less noisy.

To produce an image level reference standard, images with splat-based annotation from the expert, i.e., images containing hemorrhages, are given labels of “1” and the rest are given labels of “0” as they contain no hemorrhage splats.

C. Edge Effect Removal

As a preprocessing step, edge effects due to limited field of view (FOV) and vignetting [19] in fundus photographs have to be addressed to suppress irrelevant responses during feature extraction. This effect is visible in Fig. 4. It is conventionally performed in two ways [20]. One is to fill the region outside FOV with the mean color of the region within FOV. The other possibility is to mirror the FOV outside the FOV. In Fig. 4(a), clear edges still exist as the mean color is not necessary to blend seamlessly with the color on boundaries of FOV [20]. In Fig. 4(b) bright strips are visible on the left and dark strips on the right due to imperfections of illumination or reflection during imaging process. If these artifacts were not completely eliminated, they would interfere with features to be identified. This problem can be easily handled with splat-based image representation as is shown in Fig. 4(c). While features are extracted

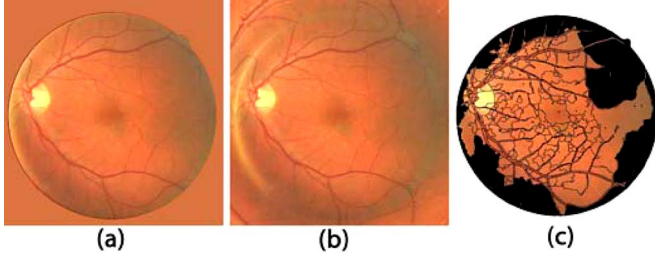


Fig. 4. Splats located on field of view boundaries are excluded to eliminate edge artifacts: (a) mean color background; (b) mirror circular reflection; (c) valid splat coverage.

from all of splats, those containing pixels on the circular boundaries of FOV are excluded from further processing. This avoids abrupt intensity changes across splat boundaries and enables the retention of only splats formed by the real content of the image.

III. SPLAT FEATURE EXTRACTION

Given splats with their associated feature vectors and reference standard labels, a classifier can then be trained to detect target objects. In this study, two categories of features are extracted for splat-based hemorrhage detection as follows: 1) splat features aggregated from pixel-based responses; 2) splat wise features (no aggregation is required).

A. Pixel-Based Feature Responses

1) *Color Channel and Opponency Images*: Color within each splat is extracted in RGB color space and dark-bright (db), red-green (rg), and blue-yellow (by) opponency images [21], which comprise six color components in splat feature space.

To accommodate color variations across the dataset, we normalize each image according to its dominant pixel values at three color channels, which means most frequent pixel values present in the image are shifted to the origin of RGB color space. No separate rescaling is performed in order to preserve the ratio between color components.

2) *Characteristics of Boundaries Across Neighboring Splats*: Factors such as limitations on the amount of illumination required for patient safety, vignetting artifacts and considerable intensity variations across the dataset, can deteriorate color-based classification. An example of intensity variation of the same structure is shown in Fig. 5 using a typical vasculature with retinal background masked out as black pixels. The histograms of the vasculature and the masked background are compared in Fig. 5(c) and (d) for the three color channels. Hemorrhages suffer from exactly the same problem—the absolute color of blood regions shows great variability while blood regions and the background exhibit similar color distributions with substantial overlap of their histograms. Therefore, using the absolute color of splat alone may still lead to a low separability between classes.

To distinguish a structure or object from its surroundings or background, it is crucial to differentiate distinct boundaries formed by neighboring splats, such as well defined sharp boundaries resulting from abrupt intensity transition, or blurred soft boundaries resulting from gradual intensity transition.

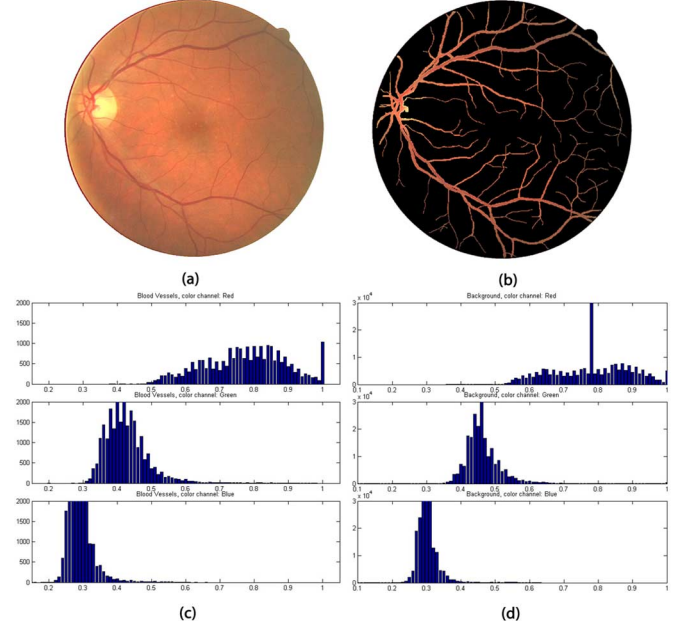


Fig. 5. Variability of color distributions of blood regions (left panel) and background (right panel) among three color channels: top: red channel; middle: green channel; bottom: blue channel. (a) One example image. (b) Vasculature. (c) Histogram of blood vessels. (d) Histogram of retinal background.

What also matters is if the transition is from dark to bright or from bright to dark towards inside/outside of the splat. The idea is similar to pixel feature classification where derivative kernels are used to characterize features of a pixel with respect to its surroundings. Different from pixels, which are on an orthogonal grid, splats are on a nonorthogonal grid, which makes extraction of regular derivative features harder. An alternative is to take advantage of the representation of Gaussian scale space [22]. Both sharp edges and soft edges are relative with respect to their underlying scale. The high intensity points and low intensity points evolve towards different directions across the scale space produced by Gaussian kernels with different σ [22].

Difference of Gaussian (DoG) kernels are applied at five different smoothing scales ($\sigma = 1, 2, 4, 8, 16$) and one baseline scale $\sigma_0 = 0.5$ to extract such features, which is expected to span potential bandwidth of boundaries present in fundus images [23].

3) *Responses From Other Filter Bank*: Besides DoG, responses from other filter bank are also computed in order to extract splat features. A Gaussian filter bank [21], which include Gaussian $G(\sigma)$, its first order derivatives at two orientations $G'_x(\sigma)$ and $G'_y(\sigma)$, its second-order derivatives at three orientations $G''_{xx}(\sigma)$, $G''_{xy}(\sigma)$ and $G''_{yy}(\sigma)$, at scales $\sigma = 1, 2, 4, 8, 16$, are applied to the green channel. A Schmid filter bank [24] consisting of 13 rotationally invariant kernels, is applied to dark-bright opponency images. Local texture filters include local range filter, local standard deviation filter and local entropy filter, which compute the intensity range, standard deviation and entropy of one pixel in a given neighborhood or region [25]. Dominant orientation and strength from steerable filters are applied to dark-bright opponency images [26].

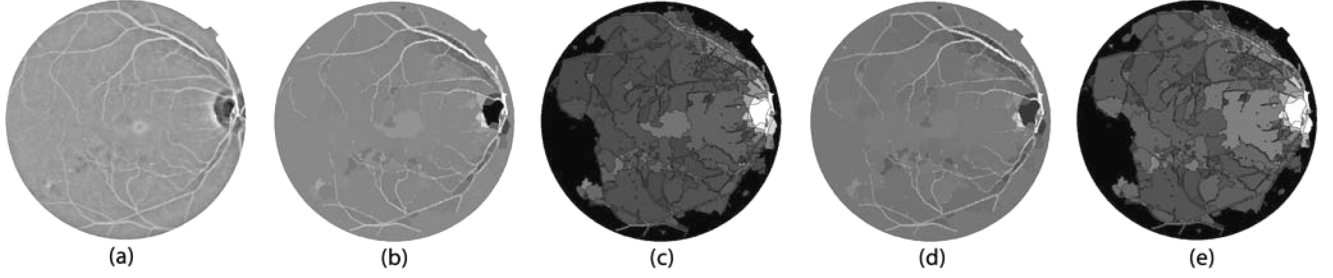


Fig. 6. Strategies to aggregate pixel responses within splats: (a) difference of Gaussian response image from green channel with $\sigma = 16$ and $\sigma_0 = 0.5$; (b) mean response within splats; (c) standard deviation within splats; (d) mean response along splat boundaries; (e) standard deviation along splat boundaries.

B. Aggregation of Pixel-Based Responses

Similar to the way splats are created so that hemorrhage boundaries are preserved precisely, splat features are more meaningful when response images exhibit high intra-splat similarity and low inter-splat similarity between target classes. To find the optimal strategy to aggregate pixel responses within each splat and associate it with a single feature value, two approaches are used, resulting in four sets of features.

Firstly, the mean and standard deviation (SD) of filtering response within splat p are computed. Taking the above DoG responses R_{DoG} for example

$$\begin{aligned} f_1^\sigma &= \frac{1}{a_p} \sum_{(x,y) \in \Omega_p} [G_\sigma * I(x,y) - G_{\sigma_0} * I(x,y)] \\ &= \frac{1}{a_p} \sum_{(x,y) \in \Omega_p} R_{DoG}^\sigma(x,y) \end{aligned} \quad (3)$$

$$f_2^\sigma = \left\{ \frac{1}{a_p} \sum_{(x,y) \in \Omega_p} [R_{DoG}^\sigma(x,y) - f_1^\sigma]^2 \right\}^{1/2} \quad (4)$$

where Ω_p represents the set of pixels within splat p with area a_p , G_σ represents Gaussian kernel at scale σ and $\sigma_0 = 0.5$.

Secondly, the mean and SD of filtering responses along boundaries of splat p are calculated as additional features of that splat

$$f_3^\sigma = \frac{1}{l_p} \sum_{(x,y) \in \omega_p} R_{DoG}^\sigma(x,y) \quad (5)$$

$$f_4^\sigma = \left\{ \frac{1}{l_p} \sum_{(x,y) \in \omega_p} [R_{DoG}^\sigma(x,y) - f_3^\sigma]^2 \right\}^{1/2} \quad (6)$$

where ω_p represents the set of pixels along the boundary of splat p with length l_p .

Fig. 6 shows one example of four feature images corresponding to a DoG filtering response [Fig. 6(a)]. In the feature images, each splat is assigned a single value represented as the intensity of that splat. These four features are all listed as candidates to identify hemorrhage splats.

It shows that aggregation over all pixels within splats [Fig. 6(b) and (c)] or over only pixels along splat boundaries [Fig. 6(d) and (e)] generate similar features. The mean splat responses from DoG filter bank show good separability between blood and nonblood splats. The SD splat responses show good separability between optic disc and the rest of the retina.

C. Splat-Wise Features

In addition to splat features aggregated from pixel-based responses, we also extract splat wise features which do not need to be aggregated. Shape features, such as splat area, extent, orientation and solidity, are derived based on individual splat distribution. Texture features are extracted according to the statistics of gray-level co-occurrence matrix (GLCM) [25] and Tamura signatures [27]. Other splat wise features are summarized in Section III-D.

D. List of Splat Features

In summary, we end up with 357 splat features, as listed in Table I. Each feature is normalized for all samples to have zero mean and unit variance.

To build an effective classifier, it is crucial to extract discriminative features tailored to target objects [5]. The point of the proposed feature extraction process is to include as many potentially relevant features as possible. We leave it for feature selection to decide which features or which aggregation strategy are most discriminative in identifying hemorrhage splats.

IV. SPLAT FEATURE SELECTION AND CLASSIFICATION

A. Two-Step Splat Feature Selection

Feature selection reduces the dimensionality of feature space by identifying relevant features and ignoring those *irrelevant* or *redundant* ones [30], which is particularly important to a higher separability between classes. There are two major approaches for feature selection: the filter approach and the wrapper approach [30]. The filter approach is fast, enabling their practical use on high dimensional feature spaces. It assesses individual feature separately without considering their interactions. The wrapper approach assesses different combinations of feature subsets tailored to a particular classification algorithm at the cost of longer computation time. To take advantage of both approaches, we use a two-step feature selection process—a filter approach followed by a wrapper approach.

The dataset is partitioned into a training set and a testing set. Feature selection evaluates discrimination power of candidate features according to reference standard labels of training set and comes up with an optimal subset as the input to a classifier.

1) *Preliminary Feature Selection With a Filter Approach:* The goal of preliminary feature selection is to exclude those *individual* features that are not effective or irrelevant in separating

TABLE I
LIST OF SPLAT FEATURES

Feature	Number	Description
Color	6×4	RGB and dark-bright (db), red-green (rg), blue-yellow (by) opponency [21], *
DoG filter bank	30×4	RGB and db, rg, by opponency [23], $\sigma = 1, 2, 4, 8, 16$, $\sigma_0 = 0.5$, *
Gaussian filter bank	30×4	$G(\sigma)$, $G'_x(\sigma)$, $G'_y(\sigma)$, $G''_{xx}(\sigma)$, $G''_{xy}(\sigma)$, $G''_{yy}(\sigma)$ [21], green channel, $\sigma = 1, 2, 4, 8, 16$, *
Schmid filter bank	13×4	13 kernels [24], db opponency, *
Local texture filter	3×4	Local range, standard deviation and entropy of one pixel in a given neighborhood [25], db opponency, *
Steerable filter	2×4	Dominant orientation and strength [26], db opponency, *
Splat area	1	Number of pixels in splat
Splat extent	1	Proportion of pixels in bounding box that are also in splat
Splat orientation	1	Angle between horizontal axis and major axis of the ellipse that has the same second-moments as splat
Splat solidity	1	Proportion of pixels in convex hull that are also in splat
Texture	4	Statistics of gray-level co-occurrence matrix (GLCM): contrast, correlation, energy and homogeneity [25]
Tamura signatures	3	Coarseness, directionality and contrast of db opponency patches associated with each splat [27]
Edge strength ratio	6	Ratios of the maximum and minimum edge strengths between neighboring splats, db, rg, by opponency
Vignetting artifacts	1	Closest distance of splat centroids to boundaries of FOV
Vessel probability	1	Vessel probability map averaged within splats derived from vessel segmentation algorithm [28]
Impact of major anatomical structures	2	Distances of automatically detected optic disc and fovea to splat centroids [29]

* Mean and SD aggregated both within splats and along their boundaries

hemorrhage/nonhemorrhage splats. It relies on general characteristics of the data to evaluate and select relevant feature subsets without involving any chosen induction algorithms [30].

The training set is further partitioned into a training subset and a testing subset. Given reference standard labels, splats in the training subset are grouped into hemorrhage splats and non-hemorrhage splats. The *t*-test is applied to each feature of the two groups. The *p*-values sorted in ascending order are taken as measures of how effective those features are in predicting the correct labels of splats.

The appropriate number of features to be retained is determined by inspecting how it varies with the misclassification error (MCE) using cross-validation. Classification is carried out using quadratic discriminant analysis (QDA), which performs likelihood ratio test under the assumption of multivariate normal distributions [31]. The percentages of misclassified splats on the training subset and the testing subset are plotted as a function of increasing numbers of sorted features. Overfitting occurs where the error on the testing subset increases while the error on the training subset decreases. The appropriate number of features is chosen according to the turning point where the smallest MCE on the test set is reached right before overfitting begins to occur.

2) *Feature Selection With a Wrapper Approach*: After preliminary selection, irrelevant features are removed. By taking interactions among features into account, a wrapper approach selects optimal combinations of relevant features with their redundancy minimized. Potential combinations are evaluated depending upon certain classification algorithms. A *k*-nearest neighbor (kNN) classifier is used for this purpose—the same as what we use for testing in the following sections.

Sorted relevant features identified from the filter approach are applied to sequential forward feature selection (SFS), which attempts to select a feature subset that maximizes area under the receiver operating characteristic (ROC) curve (AUC) of the classification system. The accuracy of splat labels predicted by kNN classifier is assessed using leave-one-out cross-validation.

B. *K-Nearest Neighbor (kNN) Classification*

After feature selection, a trained kNN classifier is set up in a “calibrated” feature space with a set of discriminative features and a set of labeled instances. The kNN classifier assigns

soft class labels to query splats based on the labels of their *k* nearest neighbors in the feature space, i.e., those instances in the training set. When *n* neighbors were labeled as being a hemorrhage splat, the posterior probability that the query splat comes from hemorrhage itself *p* was determined by $p = n/k$. The distance for finding the nearest neighbors is measured with Euclidean metric in the optimized feature space [31]. At the testing stage, the system is fully automatic.

The *k* nearest neighbor rule attempts to estimate the *a posteriori* probabilities from labeled training samples [31]. A large value of *k* is desirable to obtain reliable estimates. But only when all of the *k* nearest neighbors are close enough to the query sample, its *a posteriori* probability can be approximated by the majority labels of its neighbors. Therefore, a compromise has to be made so that the value of *k* accounts for only a small fraction of the training samples [31].

C. *Postprocessing: Assign Image Level Hemorrhage Index*

The ultimate goal of splat feature classification is to develop a hemorrhage detector, indicating whether or not an image is normal, i.e., free of hemorrhages, or abnormal, i.e., containing one or more hemorrhages. When the *a posteriori* probability of each splat being hemorrhage is determined, a *hemorrhageness map* can be created for each testing image. It is then upgraded to a single hemorrhage index as image level decision, which can be fused with results from other lesion detectors consisting a DR screening system [3].

To eliminate spurious responses in hemorrhageness map *h*, firstly low probability responses are suppressed

$$h(x, y) = \begin{cases} h(x, y) & \text{if } h(x, y) \geq h_0 \\ 0 & \text{if } h(x, y) < h_0 \end{cases} \quad (7)$$

where h_0 is a predefined threshold. The appropriate value for h_0 can be chosen according to the training set by collecting probabilities assigned to both hemorrhage and nonhemorrhage splats labeled by the reference standard. A desired separability between hemorrhage splats and the background retina can be reached by setting a threshold where a majority of hemorrhage splats receive higher probabilities than nonhemorrhage splats (see Section V-D).

Secondly, given a limited number of hemorrhage splats coming out from the first step, those neighboring ones are

merged together to form objects. Objects with small areas are removed because they are more likely to be red lesions or microaneurysms, which are supposed to be detected by separate detectors.

Thirdly, splats formed by the fovea, whose locations are detected automatically [29], are masked out to suppress potential false positives. Because detectors consisting a screening system attempt to help early detection of sight threatening diseases and prevent their progressions among a large populations unaware of any abnormalities for their vision, subjects would have noticed if there are any lesions present at the fovea.

The hemorrhage index assigned at image level is simply calculated as the probability summation of the consequently processed hemorrhageness map. Other alternatives would be to weight each hemorrhage suspicious by its area respectively, or weight each image by the area of the largest relevant objects detected.

V. EXPERIMENTS AND RESULTS

A. Data Collection

A set of 1200 fundus photographs from the publicly available Messidor database was used (<http://messidor.crihan.fr>), which were acquired by three ophthalmologic departments using a color video 3CCD camera on a Topcon TRC NW6 nonmydriatic retinograph with a 45° FOV. The images were captured using 8 bits per color plane at 1440×960 , 2240×1488 or 2304×1536 pixels. In a preprocessing step, the FOV was detected automatically and the images were rescaled to 1026×681 pixels with the FOV approximately 630 pixels in diameter across the entire dataset. All personally identifiable information was stripped so that only the raw image data was available.

A subset of the first 300 images in the Messidor dataset, not used in the development or testing of any other part of the system, provided the training splats. The testing set contains splats created from the rest of 900 images which is entirely independent of the training set.

The reference standard annotations were completed by a retinal specialist (MDA) under the splat-based formulation. As the number of large hemorrhages is far less than what required to train the classifier, the expert also annotated images which contain small or medium hemorrhages (see Figs. 8 and 12 for example). Overall, 138 of 1200 images were annotated by expert (27 from the training set and 111 from the testing set).

B. Splat Feature Subset Selection

From 300 training images, 233 776 splats are created and 708 of these splats are hemorrhage splats (0.30%). To alleviate the severely imbalanced data distribution between the two classes, only images containing more than five hemorrhage splats are included in the training process.

Fifty out of 357 features were selected by the filter approach and nineteen of those were selected by the wrapper approach:

- Mean of second-order Gaussian derivative from green channel at $s = 1, 2, 4$, orientation 1,2,3.
- Mean of second-order Gaussian derivative from green channel at $s = 8$, orientation 2,3.

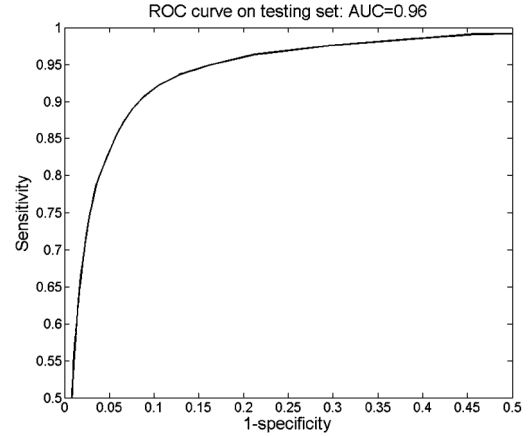


Fig. 7. Splat-based ROC curve with an AUC of 0.96 on the testing set. Note the limited range of both x and y axis.

- Mean of second-order Gaussian derivative from green channel at $s = 16$, orientation 3.
- Mean of DoG ($s2-s0.5$) from green channel.
- Mean of DoG ($s4-s0.5$) from db and rg opponency.
- Mean of DoG ($s8-s0.5$) from db opponency.
- Mean of Gaussian from green channels at $s = 8, 16$;
- Mean of response 11 of Schmid filter bank from db opponency.

These are all mean responses aggregated over splats. Twelve features are responses from second-order Gaussian derivatives at various scales and orientations, four from DoG filter, two from zero-order Gaussian kernel, and one from Schmid filter bank. Texture features are not selected, probably because interactions of splats with their surroundings are more informative than subtle complexion differences within splats themselves.

Responses from second-order Gaussian derivatives discern elongated structures such as blood vessels at different scales and orientations, while DoG filter responds mostly to blood splats with different sizes and shapes, including both vessels and hemorrhages. A combination of color channels and opponency images pick the right hue or contrast belonging to blood under various imaging conditions. Those 19 features coordinate to constitute a hemorrhage detector.

Most of the time in training the kNN classifier was devoted to the two-step feature selection, since the number of candidate features are as large as 357. It firstly took around 12 min to select 50 features from the initial feature set by the filter approach and then 45 min to select 19 of those by the wrapper approach. In classification phase, splat segmentation takes approximately 2 min and feature extraction from response images approximately 15 min. The extraction of the feature set is time consuming as the algorithm is not optimized for speed.

C. Hemorrhage Splat Classification

The kNN classifier is applied to splats created from 900 testing images. Each splat is represented as a normalized 19 dimensional feature vector. Different k were tested ranging from 21 to 151 for kNN classifier involved both in feature selection and classification. They gave close performances in terms of splat-based AUC on the training set. Based on the

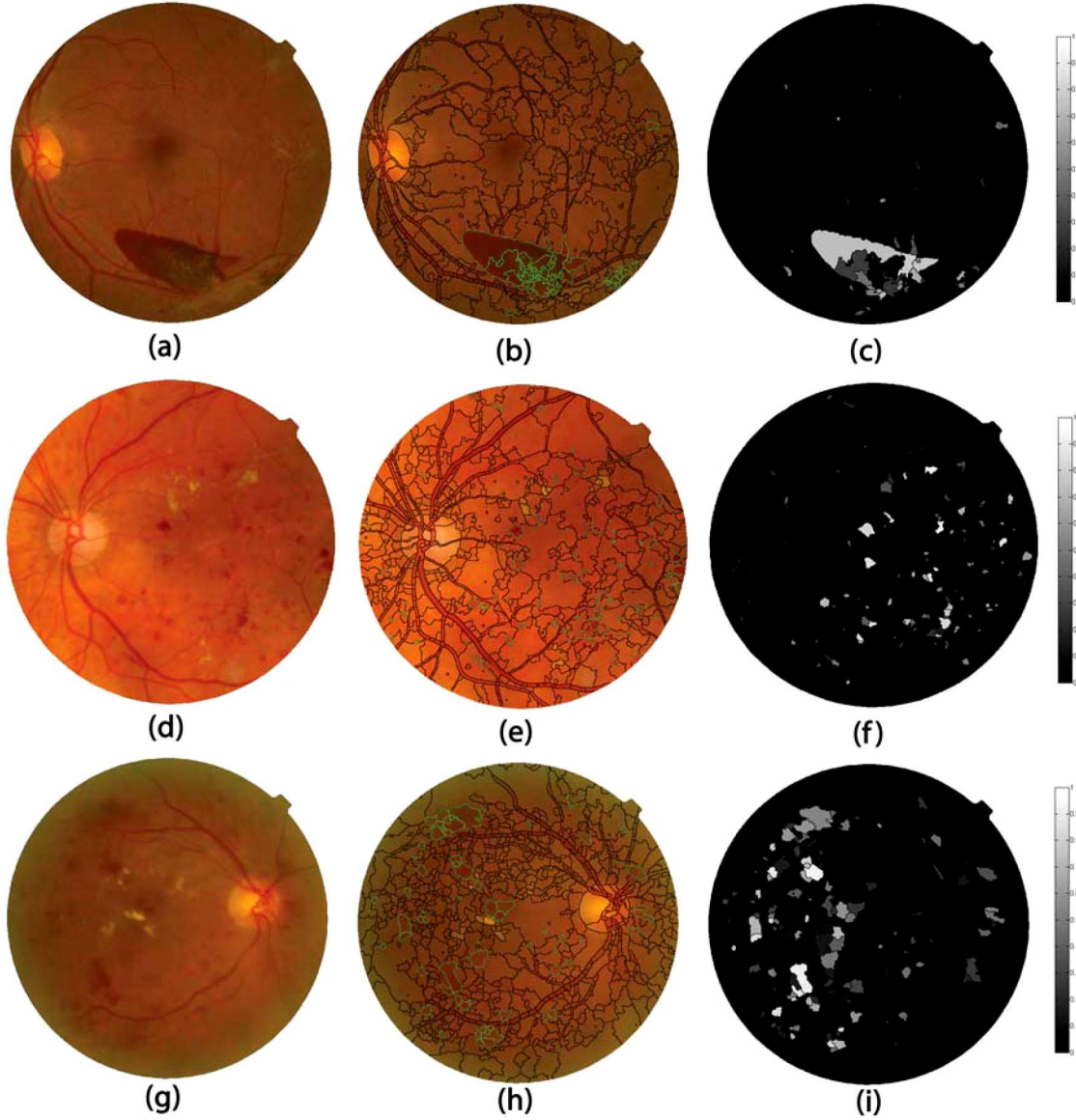


Fig. 8. Fundus images and their hemorrhageness maps from splat-based feature classification: (a) fundus image 1; (b) expert annotation 1; (c) hemorrhageness map 1; (d) fundus image 2; (e) expert annotation 2; (f) hemorrhageness map 2; (g) fundus image 3; (h) expert annotation 3; (i) hemorrhageness map 3.

sample numbers of two classes and our previous experiences, parameter k is chosen at 101 as a compromise between computational stability and prediction accuracy as discussed in Section IV-B. At splat level, an AUC of 0.96 was achieved on the testing set (Fig. 7).

Three hemorrhageness maps with their ground truth are shown in Fig. 8 as examples. Hemorrhage splats are delineated in green lines. Even though hemorrhages can have various sizes and shapes (such as large extensive hemorrhages from severe proliferative diabetic retinopathy and small red lesions), they typically receive high responses from the classifier. Given the advantage of scale invariance, splats are represented in feature space similarly regardless of their sizes. Therefore those more commonly seen small or medium hemorrhages can be used to train the classifier to identify unseen large hemorrhages. The boundaries of hemorrhages are well preserved by splat-based representation. These hemorrhageness maps, even

without any postprocessing, show a good separability between hemorrhage/nonhemorrhage splats, indicating splat features associated exclusively with hemorrhages were accurately captured by the selected feature subset.

The three corresponding splat-based ROC curves are plotted in Fig. 9. It shows possible variability in magnitude of the probability for different fundus images, which is closely related to the decision about how to set a universal threshold.

D. Image Level Performance Evaluation

To turn the hemorrhageness map shown in Fig. 8 into a single image level hemorrhage index, an appropriate threshold h_0 (7) is approximated from probability distributions of the two classes in the training set assigned by the same kNN classifier (Fig. 10). We sorted the two groups of splat probabilities in ascending order. Due to substantial variability of its appearance under various imaging conditions, probabilities assigned to hemorrhage

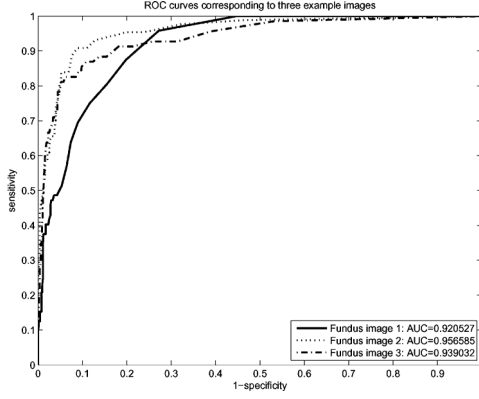


Fig. 9. Splat-based ROC curves corresponding to three example images shown in Fig. 8.

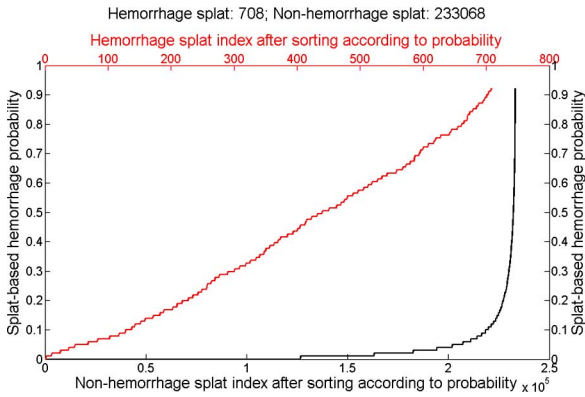


Fig. 10. Probabilities assigned to hemorrhage splats and nonhemorrhage splats in the training set.

splats, as indicated by the red curve, range from 0 to 1. Non-hemorrhage splats get probabilities shown by the black curve. Most of the nonhemorrhage splats are correctly classified with very low probabilities as desired, so only few nonhemorrhage splats get probabilities high enough so that they cannot be distinguished from hemorrhage splats without producing excessive false positives. The choice of threshold h_0 is application dependent. In the present application, the threshold $h_0 = 0.2$, where a majority of false positives would be suppressed.

By performing postprocessing as described in Section IV-C, image level ROC curve achieved an AUC of 0.87 on the testing set (Fig. 11). Plotted in the same figure on the right panel is the free-response operating characteristic (FROC) curve. When incorporated into existing DR screening systems, the goal of a hemorrhage detector would be to detect as many large hemorrhages as possible while not generating an excessive number of false positives. An FROC curve plots the sensitivity against the average number of false positives per image, which displays the effectiveness of the hemorrhage detector in improving performances of such systems when integrated as a complete component.

E. Analysis of Misclassification

Patient safety is the most important factor in designing and improving a system which can be translated into practice. Thus, top false negative (FN) cases, i.e., cases which do con-

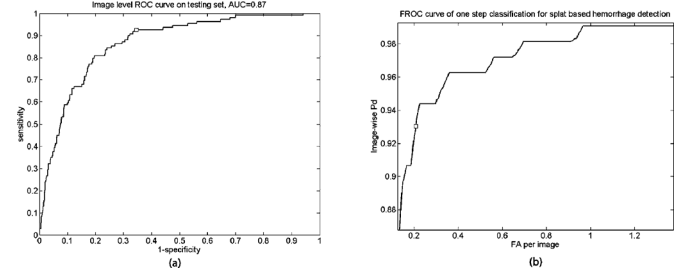


Fig. 11. Image level performance analysis on the testing set: (a) image level ROC curve with an AUC of 0.87; (b) FROC curve. The operating point gives a sensitivity of 93% and specificity of 66%, if the threshold is set to produce an average of 0.2 FPs on a per image basis.

tain hemorrhages but assigned very low hemorrhage indices, were reviewed and three of them shown in Fig. 12. Two cases [Fig. 12(a) and (g)] are caused by hemorrhages whose contrasts against the background are subtle. Their boundaries are preserved well by splat-based representation, but the classifier assigned them low probabilities, which remain challenging cases to be considered. Another case [Fig. 12(d)] actually contains small hemorrhages. The corresponding splats do get high probabilities in the hemorrhageness map. Its image level hemorrhage index is low because it is assigned according to the total area of hemorrhages present in that image. Other FNs include elongate hemorrhages present in the neighborhood or even connected with the vasculature.

We also reviewed top false positive (FP) cases, i.e., cases which actually contain no hemorrhages but assigned high hemorrhage indices, with examples shown in Fig. 13. Those FPs were found to display retinal membranes, extensive exudates and extensive atrophic laser scars, which were also labeled as abnormal by the original annotators of the Messidor dataset. The major number of misclassifications being nonhemorrhage pathology may not be a significant issue—the present hemorrhage detector can be integrated into the DR screening system to detect hemorrhages that would otherwise be missed by the classical systems [3] without introducing excessive false positives.

Major causes of misclassification (presence/absence) or discrepancy of the severity of hemorrhages (total area) include the presence of clearly visible choroid vessels [Fig. 13(c)], retinal background surrounded by a cluster of bright lesions [Fig. 13(a)], significantly different appearances between arteries and veins [Fig. 13(e)], or white reflections visible in young fundus [Fig. 13(e)]. Specifically, the misclassification shown in Fig. 13(a) is due to the fact that splat features essentially describe contrasts between splat and its neighbors. One of the predominant features of hemorrhage is it appears as a darker regions with brighter surroundings. Retinal background surrounded by a cluster of bright lesions happens to meet this criterion. Therefore such cases are misclassified.

VI. DISCUSSION AND CONCLUSION

In this study, we present a splat-based feature classification algorithm with application to large, irregular hemorrhage detection in fundus photographs. Neighboring pixels with similar intensity are grouped into nonoverlapping splats. A set of features

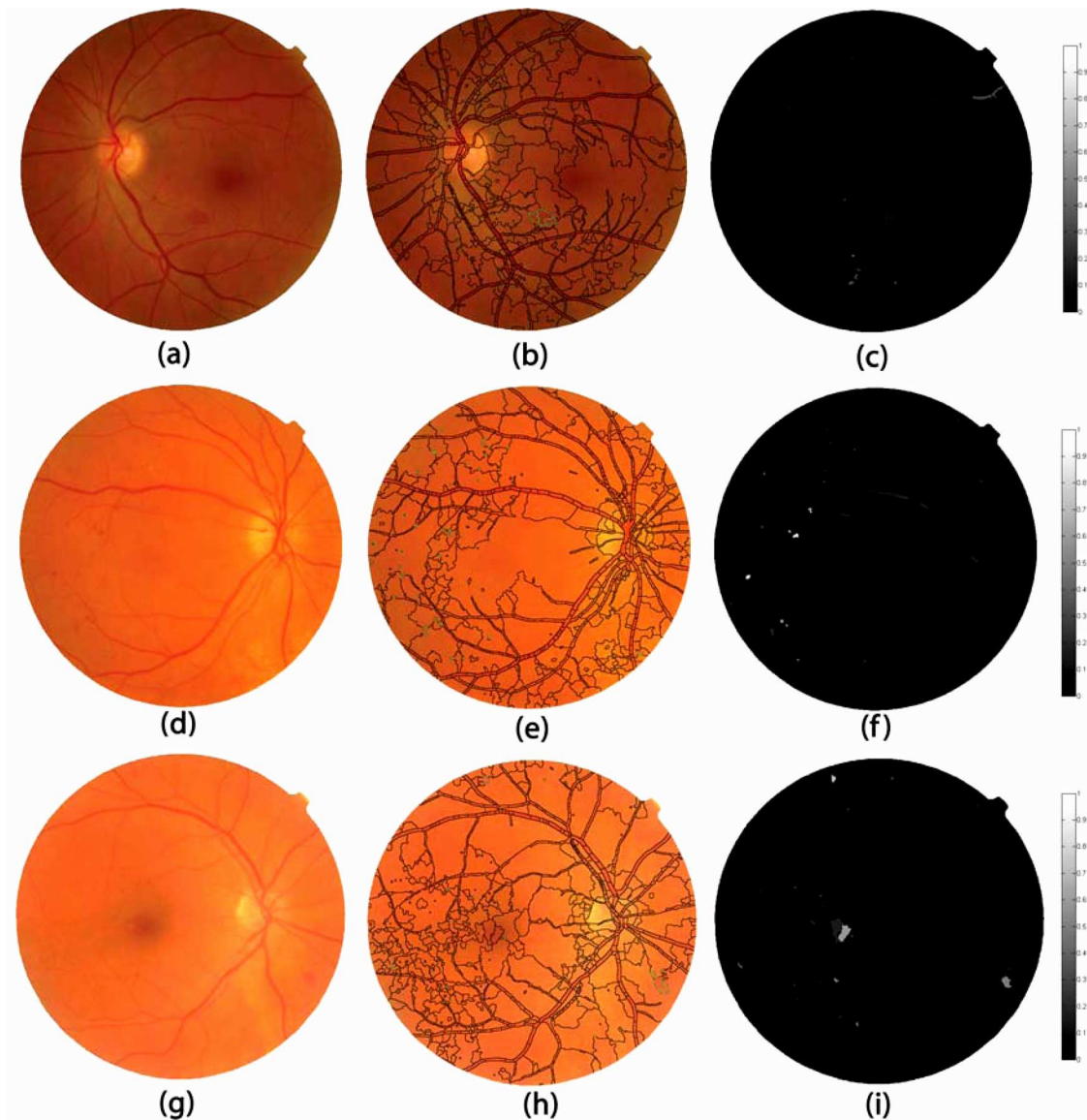


Fig. 12. Top false negatives produced by the proposed hemorrhage detector: (a) fundus image 1; (b) expert annotation 1; (c) hemorrhageness map 1; (d) fundus image 2; (e) expert annotation 2; (f) hemorrhageness map 2; (g) fundus image 3; (h) expert annotation 3; (i) hemorrhageness map 3.

is extracted from each splat to describe its characteristics. These splats are taken as samples for supervised classification in a selected feature space. The algorithm is validated on the publicly available Messidor dataset with an area under the ROC curve (AUC) of 0.96 at the splat level. At the image level, an AUC of 0.87 was achieved.

Splat-based image representation makes it easier for clinicians to annotate the boundaries of target objects, which may lower the cost of acquiring reference standard data for training. It also provides an efficient and natural way to model irregular shaped abnormalities in medical images. Aggregating features within splats improves their robustness and stability, as it is resistant to pixel level noise and intensity bias. Moreover, certain high level texture features are only meaningful when considering regions instead of pixels. Grouping of pixels into splats only depends on the attribute of neighboring pixels instead of the number of pixels contained in each splat. It results in splats on a nonorthogonal grid optimized for image homogeneity.

Sample size is decreased considerably in a splat-based framework, which is an image resampling method. For example, there are 200–300 K pixels within FOV while the average number of splats contained in an image in the Messidor dataset is only approximately 800–900. Decreased sample size leads to substantially less time for classification, which is desirable especially when dealing with large datasets as the one we used in experiments. After the training process, it takes the classifier no more than 15 s to assign hemorrhage/nonhemorrhage splat labels to one image on a computer equipped with a two-core Intel X9650 processor running at 3.00 GHz. On the other hand, experiments with large dataset presented in this study are supposed to reflect more accurately DR screening system performance.

Large, irregular hemorrhage detection is a challenging problem due to the fact that they are rare and irregular in shape with substantial variability in appearance. To assess performances of an automated system, we conducted image level ROC and FROC analysis, showing that the system can operate

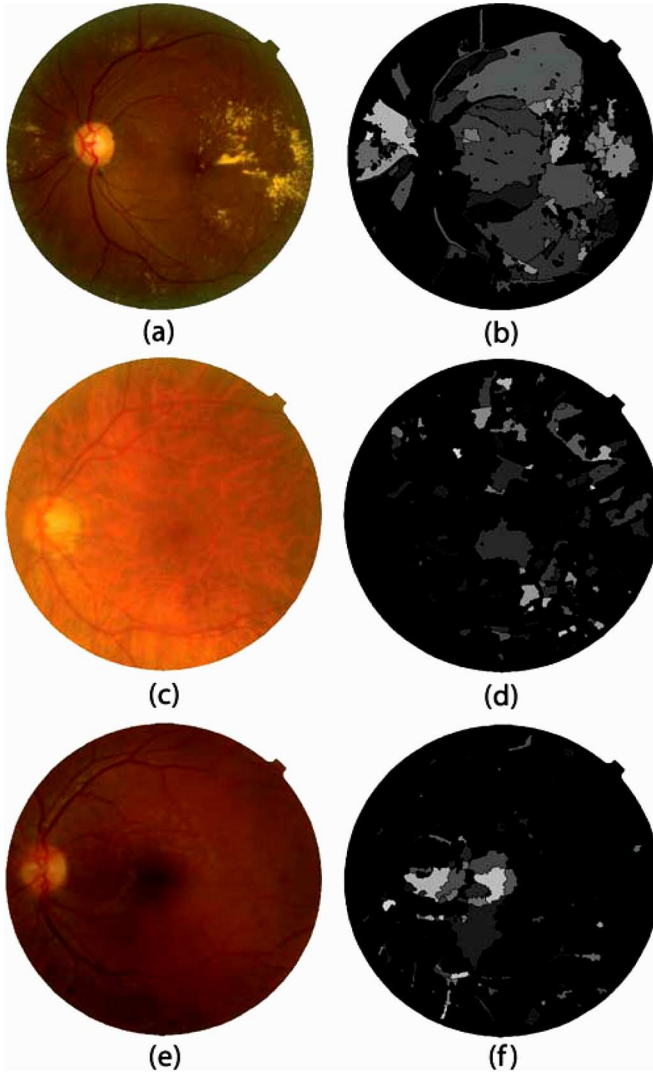


Fig. 13. Top false positives produced by the proposed hemorrhage detector: (a) fundus image 1; (b) hemorrhageness map 1; (c) fundus image 2; (d) hemorrhageness map 2; (e) fundus image 3; (f) hemorrhageness map 3.

at a sensitivity of 93% and specificity of 66%, if the threshold is set to produce an average of 0.2 FPs on a per image basis.

Initial evaluations on how the detection of rare large hemorrhages affects overall performance of a DR detection system, were encouraging. By integrating the present hemorrhage detector at an appropriate threshold, unweighted performance metrics such as AUC or sensitivity and specificity will not be affected, because the binary reference standard labels only indicate the presence or absence of DR. It will shift the type of false negatives, so that there are fewer false negatives with large hemorrhages, which are never missed by human readers, and more false negatives with retinopathy with only small hemorrhages, even though objectively the severity of diabetic retinopathy is the same for these missed cases.

An image level performance metric is appropriate when the goal is to identify abnormal images. In this scenario, the maximum number of possible FPs or FNs is limited to the number of images that were labeled by experts as normal or abnormal at the image level. However, given splat-based reference standard

annotation, experts can provide a binary label but also a quantitative indication of how abnormal those images are. That may lead to insight into the performance of the detector in a more meaningful way.

The hemorrhage index, produced as an image-level outcome by taking the total area of abnormal regions in one image, serves as a quantitative indication of the chance of the presence of retinal hemorrhages and more importantly, reflects how abnormal the image is. Images containing many hemorrhages or *any* large ones should not be missed. It can be compared with a similar measurement provided by human experts, by taking the summation of the annotated binary hemorrhage mask. A positive association between hemorrhage index and its reference standard reveals the ability of the detector in discriminating high risk images with higher priority.

Many of the hemorrhages are connected (continuous) with the retinal vessels. Because many of the false positives in our approach are parts of retinal vessel, an alternative approach would be to mask out all blood vessels using one of the common vessel segmentation methods [28]. However, preliminary studies not presented here show that such an approach, attractive at first consideration, also masked out many of the large hemorrhages we are trying to detect in the first place.

Another potential improvement is use of an active learning approach [9], [32]. As we mentioned earlier, one of the problems for supervised classification is its high cost in acquiring labeled data for training. If we design a classifier that can automatically choose examples with the highest classification uncertainty, i.e., at the decision surface boundary, for manual labeling during the learning process, human experts need to label as little data as possible to achieve the same classification confidence.

One of the limitations of the current study is that our reference standard was based on that of a single expert, reviewing only part of the dataset. With annotations from additional human experts, it would be possible to compare the variability of experts in interpreting the same set of images in terms of clinical relevance when given the same task description and to get a better definition of the reference standard.

To summarize, we present a splat-based feature classification algorithm with application of hemorrhage detection in fundus photographs. Splat-based feature classification is able to model shapes of various lesions efficiently regardless of their variability in appearance, texture or size. A variety of lesion detection tasks can therefore be generalized into exactly the same framework by training classifiers with optimal features learned from available examples projected onto a sub-feature space which maximizes the inter-class distances while minimizes the intra-class distance. The approach is validated on the Messidor dataset and achieved an area under splat wise ROC curve of 0.96 and an area under image wise ROC curve of 0.87. The hemorrhage detector could be integrated into comprehensive screening systems assisting ophthalmologists in the detection of diabetic retinopathy.

ACKNOWLEDGMENT

The authors would like to thank the Messidor program partners for kindly providing their database (see <http://messidor>).

crihan.fr). M. D. Abramoff and M. Niemeijer are listed as inventors on patents and patent applications that are not the subject of this study, but are related to the subject matter of this study. M. D. Abramoff is owner, and M. Niemeijer is employee of IDx LLC, the company that has licensed these inventions from the University of Iowa.

REFERENCES

- [1] M. D. Abramoff, J. M. Reinhardt, S. R. Russell, J. C. Folk, V. B. Mahajan, M. Niemeijer, and G. Quellec, "Automated early detection of diabetic retinopathy," *Ophthalmology*, no. 6, pp. 1147–1154, Apr. .
- [2] O. Faust, R. Acharya U., E. Y. K. Ng, K.-H. Ng, and J. S. Suri, "Algorithms for the automated detection of diabetic retinopathy using digital fundus images: A review," *J. Med. Syst.*, Apr. .
- [3] M. Niemeijer, M. D. Abramoff, and B. van Ginneken, "Information fusion for diabetic retinopathy CAD in digital color fundus photographs," *IEEE Trans. Med. Imag.*, no. 5, pp. 775–785, May .
- [4] M. Niemeijer, B. van Ginneken, J. Staal, M. S. A. Suttorp-Schulten, and M. D. Abramoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Trans. Med. Imag.*, vol. 24, no. 5, pp. 584–592, May 2005.
- [5] G. Quellec, S. Russell, and M. Abramoff, "Optimal filter framework for automated, instantaneous detection of lesions in retinal images," *IEEE Trans. Med. Imag.*, vol. 30, no. 2, pp. 523–533, Feb. 2011.
- [6] Y. Hatanaka, T. Nakagawa, Y. Hayashi, M. Kakogawa, A. Sawada, K. Kawase, T. Hara, and H. Fujita, "Improvement of automatic hemorrhages detection methods using brightness correction on fundus images," in *Proc. SPIE*, 2008, vol. 6915, pp. 69 153E-1–69 153E-10.
- [7] P. Jitpakdee, P. Aimmamee, and B. Uyyanonvara, "A survey on hemorrhage detection in diabetic retinopathy retinal images," in *Proc. 9th Int. Conf. Elect. Eng./Electron., Comput., Telecommun. Inf. Technol. (ECTI-CON)*, Bangkok, Thailand, 2012, pp. 1–4, vol. .
- [8] M. Abramoff, M. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010.
- [9] S. C. H. Hoi, R. Jin, J. Zhu, and M. R. Lyu, "Batch mode active learning and its application to medical image classification," in *Proc. ICML*, 2006, pp. 417–424.
- [10] J. Fairfield, "Toboggan contrast enhancement for contrast segmentation," in *Proc. Int. Conf. Pattern Recognit.*, 1990, vol. 1, pp. 712–716.
- [11] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Editorial: Special issue on learning from imbalanced data sets," *SIGKDD Explorations*, no. 1, pp. 1–6, 2004.
- [12] C. L. Zitnick and S. B. Kang, "Stereo for image-based rendering using image over-segmentation," *Int. J. Comput. Vis.*, no. 1, pp. 49–65, Feb. .
- [13] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Int. Conf. Comput. Vis.*, 2003, vol. 1, pp. 10–17.
- [14] A. Moore, S. Prince, J. Warrell, U. Mohammed, and G. Jones, "Super-pixel lattices," *Proc. Comput. Vis. Pattern Recognit.*, pp. 1–8, 2008.
- [15] Y.-C. Lin, Y.-P. Tsai, Y.-P. Hung, and Z.-C. Shih, "Comparison between immersion-based and toboggan-based watershed image segmentation," *IEEE Trans. Image Process.*, no. 3, pp. 632–40, Mar. .
- [16] G. Li, G. Liqun, P. Zhao-Yu, and W. Kun, "Image segmentation using multiscale gradient toboggan," in *Proc. 2nd IEEE Conf. Ind. Electron. Appl.*, May 2007, pp. 2206–2209.
- [17] G. Quellec, M. Lamard, M. Abramoff, E. Decencière, B. Lay, A. Erginay, B. Cochener, and G. Cazuguel, "A multiple-instance learning framework for diabetic retinopathy screening," *Med. Image Anal.*, vol. 16, no. 6, pp. 1228–40, 2012.
- [18] M. Christopher, D. C. Moga, S. R. Russell, J. C. Folk, T. Scheetz, and M. Abramoff, "Validation of tablet-based evaluation of color fundus images," *Retina*, vol. 32, no. 8, pp. 1629–35, 2012.
- [19] 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.
- [20] M. Niemeijer, M. D. Abramoff, and B. van Ginneken, "Fast detection of the optic disc and fovea in color fundus photographs," *Med. Image Anal.*, no. 6, pp. 859–870, Dec. .
- [21] M. D. Abramoff, W. L. M. Alward, E. C. Greenlee, L. Shuba, C. Y. Kim, J. H. Fingert, and Y. H. Kwon, "Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features," *Invest. Ophthalmol. Vis. Sci.*, vol. 48, no. 4, pp. 1665–1673, Apr. 2007.
- [22] B. M. T. H. Romeny, *Front-End Vision and Multi-Scale Image Analysis: Multi-Scale Computer Vision Theory and Applications, Written in Mathematica*. Berlin, Germany: Springer, 2003.
- [23] L. Tang, M. Niemeijer, and M. Abramoff, "Splat feature classification: Detection of the presence of large retinal hemorrhages," in *Proc. IEEE 8th Int. Symp. Biomed. Imag. (ISBI)*, 2011, pp. 681–684.
- [24] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *Int. J. Comput. Vis.*, vol. 62, no. 1–2, pp. 61–81, 2005.
- [25] T. Wagner, , B. Jahne, H. Haussecker, and P. Geissler, Eds., *Texture Analysis*. New York: Academic, 1999, vol. 2.
- [26] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.
- [27] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst., Man Cybern.*, vol. 8, no. 6, pp. 460–472, Jun. 1978.
- [28] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," in *Proc. SPIE*, pp. 648–656.
- [29] M. Niemeijer, M. D. Abramoff, and B. van Ginneken, "Segmentation of the optic disc, macula and vascular arch in fundus photographs," *IEEE Trans. Med. Imag.*, no. 1, pp. 116–127, Jan. .
- [30] R. Kohavi and G. John, "Wrappers for feature subset selection," *Artif. Intell.*, vol. 97, no. 1–2, pp. 272–324, 1997.
- [31] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York: Wiley-Interscience, 2000.
- [32] C. Sánchez, M. Niemeijer, M. Abramoff, and B. van Ginneken, "Active learning for an efficient training strategy of computer-aided diagnosis systems: Application to diabetic retinopathy screening," *Med. Image Comput. Comput. Assist. Intervent.*, vol. 13, no. 3, pp. 603–610, 2010.