Optimal Filter Framework for Automated, Instantaneous Detection of Lesions in Retinal Images

Gwénolé Quellec*, Stephen R. Russell, and Michael D. Abràmoff, *Senior Member, IEEE*

***Abstract*—Automated detection of lesions in retinal images is a crucial step towards efficient early detection, or screening, of large at-risk populations. In particular, the detection of microaneurysms, usually the first sign of diabetic retinopathy (DR), and the detection of drusen, the hallmark of age-related macular degeneration (AMD), are of primary importance. In spite of substantial progress made, detection algorithms still produce 1) false positives—target lesions are mixed up with other normal or abnormal structures in the eye, and 2) false negatives—the large variability in the appearance of the lesions causes a subset of these target lesions to be missed. We propose a general framework for detecting and characterizing target lesions almost instantaneously. This framework relies on a feature space automatically derived from a set of reference image samples representing target lesions, including atypical target lesions, and those eye structures that are similar looking but are not target lesions. The reference image samples are obtained either from an expert- or a data-driven approach. Factor analysis is used to derive the filters generating this feature space from reference samples. Previously unseen image samples are then classified in this feature space. We tested this approach by training it to detect microaneurysms. On a set of images from 2739 patients including 67 with referable DR, DR detection area under the receiver-operating characteristic curve (AUC) was comparable ($\mathrm{AUC} = 0.927$) to our previously published red lesion detection algorithm ($\mathrm{AUC} = 0.929$). We also tested the approach on the detection of AMD, by training it to differentiate drusen from Stargardt's disease lesions, and achieved an $\mathrm{AUC} = 0.850$ on a set of 300 manually detected drusen and 300 manually detected flecks. The entire image processing sequence takes less than a second on a standard PC compared to minutes in our previous approach, allowing instantaneous detection. Free-response receiver-operating characteristic analysis showed the superiority of this approach over a framework where false positives and the atypical lesions are not explicitly modeled. A greater performance was achieved by the expert-driven approach for DR detection, where the designer had sound expert knowledge. However, for both problems, a comparable performance was obtained for both expert- and data-driven approaches. This indicates that annotation of a limited number of lesions suffices for building a detection system for any type of lesion in retinal images, if no expert-knowledge is available. We are studying whether the optimal filter framework also generalizes to the detection of any structure in other domains.**

***Index Terms*—Drusen, factor analysis, lesion detection, microaneurysms, retinal diseases.**

I. Introduction

AUTOMATED detection of lesions caused by disorders of the eye, especially diabetic retinopathy (DR) and age-related macular degeneration (AMD), has rapidly expanded in the last decade. One important motivation is the need for reproducible, efficient, early detection, or screening, of large at-risk populations. In particular, there is a preponderance of evidence that early detection and timely treatment of DR, the most common cause of blindness in the working age population of the United States and of the European Union [1], can prevent visual loss and blindness in patients with diabetes [2]–[4]. Several projects aiming at identifying automatically the most at-risk patients, who should be seen by an expert, have thus been initiated (EyeCheck, Amsterdam, The Netherlands; Ophdiat/TeleOphta, Paris, France; etc.).

Most efforts have focused on automatically detecting the lesions in color fundus photographs, as this is a noninvasive and cost-effective modality [5]–[7]. One study of a large screening dataset has shown that, even though some automated detection algorithms achieve a detection performance comparable to that of human experts, further improvements are desirable before their translation into clinical practice [8]. A primary step in automated detection of DR is the detection of microaneurysms [9]–[21], which are highly pathognomic and often the first sign of DR, although it has been shown that detecting other lesions can improve DR detection performance [22], [23]. In order to compare the algorithms from various groups around the world, the first automated microaneurysm detection competition [the 2008 Retinopathy Online Challenge (ROC)] was recently held, and showed that human experts are still performing better at the lesion level [24]. In particular, all participating methods failed to sufficiently differentiate lesions from retinal blood vessels: when a more sensitive setting was used, false positives occurred on the vessels, while when a specific setting was used, lesions connected to or close to the vasculature were missed.

Several automated detection algorithms for detecting the lesions of AMD, the most common macular disease associated with aging [25] and the most important cause of blindness and visual loss in the developed world, have also been published. In

Manuscript received March 20, 2010; revised August 08, 2010, October 13, 2010; accepted October 14, 2010. Date of publication October 21, 2010; date of current version February 02, 2011. This work was supported in part by the National Institutes of Health (EY017066) and in part by the U.S. Department of Veterans Affairs. *Asterisk indicates corresponding author*.

*G. Quellec is with the Department of Ophthalmology and Visual Sciences and the Department of Biomedical Engineering, University of Iowa, Iowa City, IA 52242 USA (e-mail: gwenole-quellec@uiowa.edu).

S. R. Russell is with the Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA 52242 USA and with the Department of Veterans Affairs, Iowa City VA Medical Center, Iowa City, IA 55242 USA (e-mail: steve-russell@uiowa.edu).

M. D. Abràmoff is with the Department of Ophthalmology and Visual Sciences and 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: michael-abramoff@uiowa.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2010.2089383

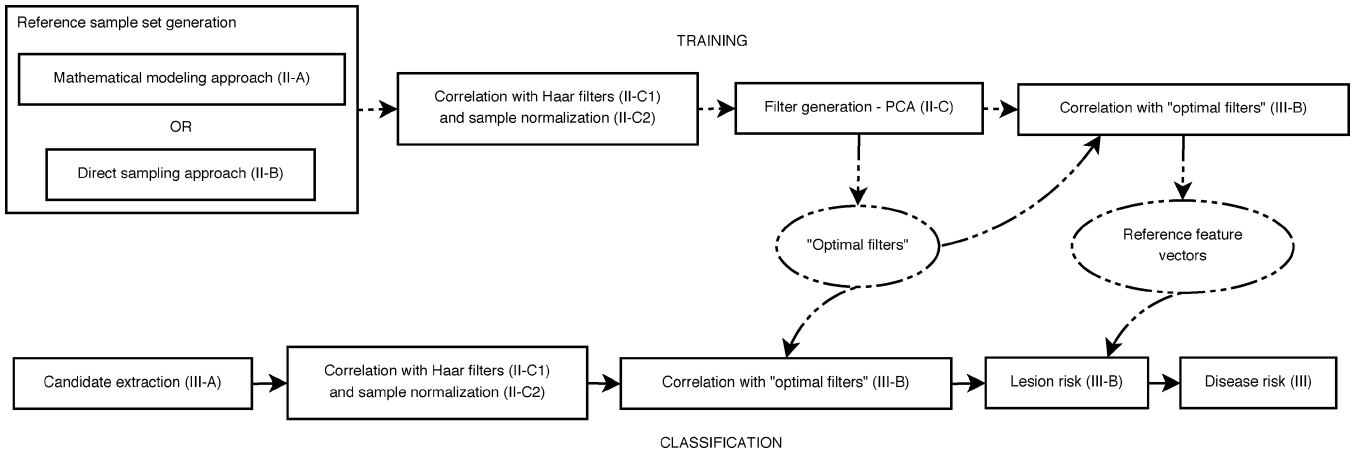

Fig. 1. Schematic view of the proposed approach. The offline training processing sequence is summarized by the dashed line path on the top row. The processing sequence to classify unseen image samples is summarized by the solid line path on the bottom row.

particular, these efforts have focused on drusen, the hallmark of AMD [26]–[30]. One of the main challenges in detecting drusen is the ability to differentiate them from bright lesions caused by other diseases, in particular, DR's exudates and cotton wool spots, as well as subtle, drusen-like lesions such as the *flecks* related to Stargardt's disease [31]. AMD tends to occur in older patients than Stargardt's disease, but both the individual drusen and *flecks* as well as their distribution are very similar.

We call *negative lesion confounders* those structures that are similar looking, but not the target lesion: vessel portions in the microaneurysm case, and exudates, cotton-wool spots and Stargardt flecks in the drusen case. We call *positive lesion confounders* that subset of target lesions that are easily missed because they have specific properties that are substantially different from the standard case, but should still be detected: in the case of microaneurysm detection, these are microaneurysms close to major vessels. In other words, negative (resp. positive) lesion confounders are the false positives (resp. false negatives) of simple detectors. The receiver-operating characteristic (ROC) experience shows that these lesions are hardly detected if they are not specifically modeled. In both cases, lesions that are very proximate to other lesions, including being fused with these other lesions, also are positive lesion confounders.

We propose in this paper an optimal filter framework to solve any lesion segmentation and differentiation problems in retinal images, including but not limited to microaneurysm detection and drusen versus flecks differentiation. The presented method is fast. A fast algorithm has advantages: it allows instantaneous feedback to the camera operator and to the patient. Moreover, fast algorithms allow simpler and therefore cheaper devices to be used. How these algorithms are translated into clinical practice, and in particular, whether these algorithms are used in autonomous mode without human review of every image, are still open questions, and are beyond the scope of this paper. We simply note that many blood tests are also never reviewed by a human expert, except for quality control purposes, and that the discussion has only started [32]–[34].

Previous studies have always dealt with the confounder problem with confounder-specific procedures [16] or confounder-specific features [15], [29], [30]. Instead, our optimal filter framework automatically generates features adapted for differentiating target lesions, including atypical target lesions, from negative lesion confounders. The proposed framework relies on the generation of a set of filters matched both to the lesions and to common positive as well as negative lesion confounders. The *design* of these filters is defined as either one of the following.

- 1) Expert-driven: a mathematical model is designed by a human modeler for both lesions and lesion confounders, and the model parameter space is subsequently traversed to generate (different) artificial image samples
- 2) Data-driven: a set of annotated image samples is used.

The set of filters then emerges from the samples obtained under 1) or 2) [35]. The framework has been generalized to the design of emergent color filters [36], [37]. On one hand, modeling the lesions mathematically is convenient in the sense that only a couple of representative images are required for training: the modeler (usually an experienced programmer) is able to generalize out of verbal descriptions by clinicians and a small number of samples. Moreover, no additional annotations are required each time the method is applied to a new dataset. On the other hand, deciding how to model the target lesions is subjective and biased, and a good description may not (yet) be available. Therefore, a design dependent only on manual annotations is attractive, as it makes the proposed approach immediately applicable to a wider range of lesion detection problems, and any modeling bias is avoided.

The purpose of this paper is to illustrate the proposed optimal filter framework and perform numerical tests of detection performance of detecting microaneurysms and drusen in retinal images.

II. Filter Generation

The proposed approach, summarized in Fig. 1, relies on the generation of an optimal set of filters to differentiate lesions, as well as positive and negative lesion confounders. An optimal filter framework for detecting target lesions involves obtaining a representative set of samples from these three categories of objects, and deriving from these samples a set of filters, or a basis, that optimally represents them.

Ideally, the representative set of samples would uniformly describe the space of all possible lesion samples, as well as the space of possible negative lesion confounders. However, we cannot sample all the possible lesions from all possible images that might be erroneously classified as target lesion: therefore we sample the negative lesion confounders whose erroneous detection as positive lesions significantly impacts the classification performance. In other words, we only sample the most common lesion confounders. As explained above, both an expert driven and data-driven design approach is used in order to obtain samples representative of the lesion of interest, as well as samples from positive lesions confounders, negative lesions confounders or any combination of these.

The expert-driven approach, as discussed in the introduction is presented in Section II-A, while the data-driven approach is presented in Section II-B. Both result in representative samples, and optimal filters are then generated as described in Section II-C, regardless of the approach used to generate the samples.

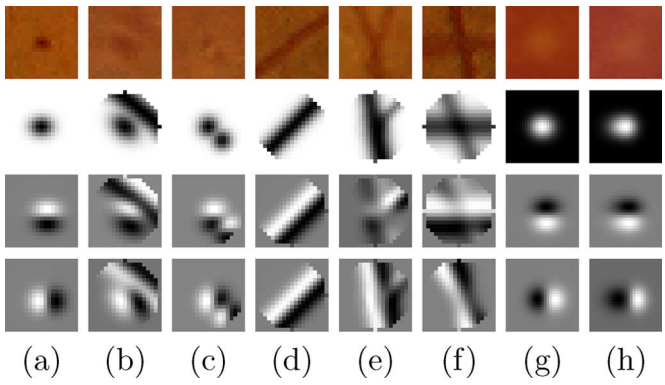

Fig. 2. Examples of image samples (first row) and similar samples derived from mathematical models in the spatial domain (second row) and after correlation with Haar filters (third row: vertical Haar filter, fourth row: horizontal Haar filter). The first six samples come from the microaneurysm detection problem: a typical lesion (a), positive lesion confounders (b), (c) and negative lesion confounders (d), (e), (f). The last two samples come from the drusen (g) versus flecks (h) classification problem.

A. Mathematical Model Based Sample Generation

In the expert-driven approach, a modeler translates verbal descriptions with clinicians, as well as intensity distributions in a limited number of samples selected by the clinicians, into mathematical equations.

The modeler is able to generalize out of a few samples. For example, if he/she sees linear vessel portions with a few different orientations, he/she can guess that in other images, linear vessel portions with slightly different orientations and slightly different sizes may also occur, and are thus also vessels. The modeler can also discriminate relevant information from noise, such as background intensity variation, compression artifacts, high frequency noise, so *a priori* fewer samples are required to obtain a representative set.

For the segmentation and differentiation problems studied in this paper (microaneurysm and drusen detection), the mathematical models were defined by one of the authors (GQ), a nonclinician, for typical lesions, for positive lesion confounders, if any were identified, and for negative lesion confounders, if any were identified. They rely on a continuous model of the intensity distribution of the lesions. In these models, pixel intensities along a line segment were modeled with a generalized Gaussian function with scale parameter α. Let R denote the distance to the pixel with maximal intensity in a line segment and let $r = R/\alpha$ denote the normalized distance. Pixel intensity in the line segment is given by

$$\text{intensity}(r;\beta,\delta) = \delta e^{-r^{\beta}}. \tag{1}$$

Note that asymmetric objects can be modeled with this formula: the value for α simply needs to be direction-dependent. β characterizes the sharpness of the contours; more or less sharp contours can be modeled by varying β: small values of β generate peaks, $\beta = 2$ generates Gaussian profiles and high values of β generate plateaus, which may be relevant for modeling exudates for instance. δ is the amplitude ($\delta = 1$ for drusen, $\delta = -1$ for microaneurysms).

For instance, the typical intensity distribution for both microaneurysms and drusen is modeled as

$$\text{typical}(x,y;\alpha_1,\alpha_2,\beta,\theta) = \text{intensity}(r;\beta,\delta) \tag{2}$$

$$u = x\cos(\theta) + y\sin(\theta) \tag{3}$$

$$v = -x\sin(\theta) + y\cos(\theta) \tag{4}$$

$$r = \sqrt{\left(\frac{u}{\alpha_1}\right)^2 + \left(\frac{v}{\alpha_2}\right)^2}. \tag{5}$$

The complete list of intensity distribution models for lesion confounders is described in Sections II-A1 and II-A2 for microaneurysm detection and in Sections II-A3 and II-A4 for drusen detection. Examples of image samples and samples derived from these models are given in Fig. 2 for illustration.

1) Common Negative Lesion Confounders for Microaneurysm Detection: The first negative microaneurysm confounder is a linear vessel portion

$$\text{linearVessel}(x,y;\alpha,\beta,\theta) = \text{intensity}\left(\left|\frac{v}{\alpha}\right|;\beta,-1\right). \tag{6}$$

The second lesion confounder is a vessel branch: three linear segments, modeled by (6), meet at the center of the model I, where the three intensity distributions, I_1, I_2 and I_3, are fused. It is empirically expressed as $I = (I_1+1)(I_2+1)(I_3+1)-1$.

The last lesion confounder is a vessel crossing: two vessels, modeled by (6), meet at the center of the model, where the two intensity distributions, I_1 and I_2, are fused. Since the two vessels overlap, the intensity is darker at the center of the model I. It is empirically expressed as: $I = -\sqrt{I_1^2 + I_2^2}$.

2) Common Positive Lesion Confounders for Microaneurysm Detection: We modeled two types of positive lesion confounders: a microaneurysm either close to a vessel or connected to it and a microaneurysm either close to or joined to another microaneurysms. Both models are combinations of (2) and (6), possibly involving a translation; the amplitude and angle of the translation are parameters of the model.

3) Common Negative Lesion Confounders for Drusen Detection: For drusen detection, we modeled one type of lesion confounders: Stargardt's disease *flecks*, with a pisciform intensity distribution [31]

$$\text{fleck}_{|u>0}(x,y;\alpha_1,\alpha_2,\beta,\theta) = \text{intensity}(r1;\beta,\delta) \quad (7)$$

$$\text{fleck}_{|u\leq 0}(x,y;\alpha_1,\alpha_2,\beta,\theta) = \text{intensity}(r2;\beta,\delta) \quad (8)$$

$$r1 = \sqrt{\left(\frac{u}{\alpha_1}\right)^2 + \left(\frac{v}{\alpha_2}\right)^2} \quad (9)$$

$$r2 = \sqrt{\frac{u^2+v^2}{\alpha_2^2}} \quad (10)$$

where $\text{fleck}_{|u>0}$ (resp. $\text{fleck}_{|u\leq 0}$) is the tail (resp. head) of the goldfish.

4) Common Positive Lesion Confounders for Drusen Detection: We modeled one type of positive lesion confounders: drusen close to, or fused to, other drusen.

Once the intensity distribution of all the lesions and lesion confounders were modeled, samples were generated from each intensity distribution model as follows.

- The scale for the generated samples is selected based on the average size of the lesions and lesions confounders in a particular dataset. The generated image samples are defined as circular image patches of radius R times as wide as the average lesion size (we chose $R = 3$ in this experiment). If the scale range for the target lesion is very important, then several classifiers can be trained, as described in this paper, using a different value for R in each classifier. It is important to chose a wide enough ratio R, since, by definition, most negative lesion confounders look similar to the lesions at the center of the patch. What is most likely to discriminate lesions from negative lesion confounders is the periphery of the patch. Typically, for microaneurysm detection, vessel crossings may look similar to a microaneurysm at the center, but, as opposed to microaneurysms, there will be linear structures, oriented toward the center of the patch, at the periphery.
- For each parameter π_i of a model, a range $[\pi_i^{\min};\pi_i^{\max}]$ of variation is manually measured by the modeler from a couple of representative images. For instance, to generate vessel branchings, there would be nine parameters: $\pi_1 = \alpha(I_1)$, $\pi_2 = \beta(I_1)$, $\pi_3 = \theta(I_1)$, $\pi_4 = \alpha(I_2)$, $\pi_5 = \beta(I_2)$, $\pi_6 = \theta(I_2)$, $\pi_7 = \alpha(I_3)$, $\pi_8 = \beta(I_3)$, $\pi_9 = \theta(I_3)$, where $\alpha(I_j)$, $\beta(I_j)$ and $\theta(I_j)$ are the α, β and θ in (6) corresponding to the jth vessel branch, with intensity distribution I_j.
- Image patches are generated from each model by sampling the product space $\prod_i[\pi_i^{\min};\pi_i^{\max}]$ uniformly. The more reference samples we want to generate, the smaller the sampling step. Note that, although many parameters might be normally distributed across actual lesions, a uniform sampling of $\prod_i[\pi_i^{\min};\pi_i^{\max}]$ ensures that rare lesions are equally well characterized.

A color model was used to improve further the presented gray-level modeling of the lesions: the gray levels in the generated samples are mapped to colors, using a look-up table.

- Intensity 0 is mapped to the average background color in the representative set of images $(R_{\text{bg}}, G_{\text{bg}}, B_{\text{bg}})$.
- Intensity δ [$\delta = 1$ or $\delta = -1$, see (1)] is mapped to the average color at the center of the target lesions in the representative set of images $(R_{\text{lc}}, G_{\text{lc}}, B_{\text{lc}})$.
- Intensity i, between 0 and δ, is mapped to $(R_{\text{bg}} - i\delta(R_{\text{bg}} - R_{\text{lc}}), G_{\text{bg}} - i\delta(G_{\text{bg}} - G_{\text{lc}}), B_{\text{bg}} - i\delta(B_{\text{bg}} - B_{\text{lc}}))$.

B. Direct Image Sampling

Many aspects of the mathematical models described above are *ad-hoc*. Also, modeling is challenging, in particular if color is involved: a direct sampling color model simply requires sampling in the entire color space, instead of complex modeling of the interaction of color channel intensities. Hence, a data-driven or direct sampling approach has many advantages.

In direct sampling, the target lesions—both typical lesions and positive lesion confounders—are annotated by a human expert on a training dataset; the expert either indicates the center of the lesions, or segments the lesions. In the latter case, the candidate lesion detector, described in Section III-A, is used to find the center of the lesion within the segmented region. The set of sample images should represent all images and all local image variations that the proposed filter framework may encounter. The expert may not know what the negative lesions confounders are: the candidate lesion detector is used to identify them. In this study, $\tilde{N}$ negative lesion confounders were selected at random among the false positives of the candidate detector, where $\tilde{N}$ is the number of annotated (positive) samples.

C. Postprocessing the Reference Samples

To maximize the performance of the proposed framework, the set of reference image samples for the typical lesions and the common positive or negative lesion confounders, collected as above, has to be as representative of all unseen samples as possible. In order to increase the representativeness of the reference samples, two initial steps are performed. First, the variability due to the noise potentially embedded in the reference samples is reduced (see Section II-C1). Second, the generality of the reference samples is increased by a normalization step (see Section II-C2).

1) Transformed Space With Increased Signal-to-Noise Ratio: The presence of low-frequency noise (slow background intensity variation) and high-frequency noise (salt-and-pepper noise, compression artifacts, etc.) is a major problem for characterizing the appearance of lesions in retinal images. We project both the noise-free samples S and noisy samples $S+\mathcal{N}$ into a transformed space where the signal-to-noise ratio is high, i.e., such as $T(S) \approx T(S+\mathcal{N}) \approx T(S+\mathcal{N}')$, where T is the transform operator, and $\mathcal{N}$ and $\mathcal{N}'$ are two realizations of a random noise. Such a transform has been described previously for retinal images [18]: image samples are correlated with horizontal and vertical Haar filters, dilated by a factor s to match the typical scale of the lesions ($s = 2$ in our applications). In this transformed space, most of the high- and low-frequency noise has been removed.

2) Sample Normalization: The representative samples (see Sections II-A and II-B) characterize the range of variations for the shape, the intensity and the color of the target lesions and

lesion confounders. One of these parameters, the intensity of the object can easily be normalized across samples. After normalizing intensity, two objects that are similar except for intensity (e.g., they come from differently-contrasted image areas) become similar. Without normalization, we would need to collect N samples with intensity I_1 plus N samples with intensity I_2 plus N samples with intensity I_3, etc., to represent the variability of image samples in a given image dataset. Thanks to normalization, we can simply collect N samples with varying intensities and normalize them. One may think that losing the intensity information would imply a loss of classification performance: our experience shows that it actually increases classification performance (see Fig. 5). Normalizing the intensity of samples generated by a mathematical model is trivial: the intensity at the center of the model simply needs to be set to one. Normalizing the intensity of samples directly extracted from images, on the other hand, is not: for a reliable estimation of the lesion intensity, the intensity distribution in the target sample is matched to the intensity distribution in the average sample as explained hereafter. Let S (resp. S_0) denote the target sample (resp. the average sample) projected in the transformed space of high signal-to-noise ratio (see Section II-C1), and $S(k)$ (resp. $S_0(k)$) denote its kth element in lexicographical order. The coefficient δ minimizing the sum of the squared errors between $\delta \cdot S$ and S_0 is found using the least mean squared algorithm

$$\delta = \frac{\sum_k S(k)S_0(k)}{\sum_k S_0(k)^2}. \tag{11}$$

Finally, each coefficient $S(k)$ is divided by δ.

D. Generating Optimal Filters From Maximally Representative Image Samples

The reference samples for lesions and lesion confounders, projected in the transformed space of high signal-to-noise ratio and subsequently normalized, are now maximally representative of all lesions and lesion confounders, provided that they are also projected into the transformed space and normalized. In the remainder of this section, "sample" stands for "sample projected in the transformed space of high signal-to-noise ratio and subsequently normalized."

Definition (α-Density): $\mathbb{S}$, the set of representative samples, is said to be α-dense in some set of samples $\mathbb{T}$ if, for any sample $S \in \mathbb{T}$, the distance from S to its nearest neighbor in $\mathbb{S}$, in the least square sense (i.e., the sample in $\mathbb{S}$ minimizing the L2-distance to S), is smaller than $\alpha (\alpha > 0)$.

Let $\tilde{\alpha} \geq 0$ be the constant such that $\mathbb{S}$ is $\tilde{\alpha}$-dense in $\tilde{\mathbb{T}}$, the set of all possible target lesions, including its inner border (the set of positive confounders) and its outer border (the set of negative confounders).

By definition of the outer border of a region, any path from one point outside the region to one point inside the region crosses the outer border before entering the region. It implies that the nearest neighbor in $\tilde{\mathbb{T}}$ of any negative sample S lies on the outer border (where reference samples are negative samples), when $\tilde{\alpha}$ tends towards 0. And by definition of inner border, the nearest neighbor in $\mathbb{S}$ of any positive sample S lies within the inner border (where reference samples are positive samples—either typical lesions or positive lesion confounders), when $\tilde{\alpha}$ tends towards 0.

As a consequence, in the absence of time constraints, classifying a new sample S comes down to finding its most similar samples in $\mathbb{S}$, in the least square sense, and classification performance tends towards optimality as $\tilde{\alpha}$ tends towards 0. In other words, we could use the normalized transformed space described in Section II-C as feature space for classification with a k-NN classifier.

However, we would like to achieve close to instantaneous detection, as a consequence the dimensionality of the feature space has to be as low as possible. So we need to compress the sample space to a space with lower dimension M. The optimal compression framework, i.e., the one that minimizes the distortion due to compression, in the least square sense, for any dimension M, and therefore minimizes the error in the L2-distance between any sample S and its *true* nearest neighbor (i.e., its nearest neighbor in the original space—Section II-C), is principal component analysis (PCA) [38]. PCA is a well-known orthogonal linear transformation that transforms the set of reference samples into a new coordinate system such that the greatest variance, by any projection of the data, comes to lie on the first coordinate, the second greatest variance on the second coordinate, etc.

As a conclusion, as $\tilde{\alpha}$ tends towards 0 and as the percentage of variance explained by the first M components tends towards 100%, the proposed framework tends towards optimality: the filters generating the "optimal feature space" can thus be derived through PCA; they are the first M principal components.

In this study, the number M of components is chosen so that 98% of the variance in the reference samples is explained by these M components ($M = 67$ for microaneurysm detection, $M = 32$ for drusen versus flecks differentiation). Independent component analysis [35] may also be used to generate an "optimal feature space." Note, however, that others found no difference with PCA, in terms of performance, in related applications [39].

III. Proposed Lesion Detector

Given a set of optimal filters, previously unseen images and image regions can be classified within this optimal feature space. A so-called *risk of presence* of the target lesion, as defined in Section III-B, is derived for each of the image samples; this risk can subsequently be used to segment the lesions in the image, or, to derive an automated probabilistic diagnosis for the entire image. Images are processed as follows.

1) A preprocessing step, described in Section III-A, is used to identify and normalize candidate target lesions in an image.
2) Each sample selected in the preprocessing step is input to the classifier after normalization. The risk of presence of the target lesion is computed by the classifier for that sample: a lesion risk map is thus obtained.
3) If the lesions need to be segmented, the risk map defined in step 2) is thresholded, and the connected foreground samples are identified using morphological labeling. Each connected component is regarded as an automatically detected lesion, and this lesion is assigned a risk value defined as the

maximal samplewise risk of presence within the connected component.

4) If we need a probabilistic diagnosis for the entire image, the risks assigned to each automatically detected lesions are combined [22]: the resulting combined risk is referred to as the disease risk (see Fig. 1, bottom right box). When a single type of lesions is detected, the probabilistic diagnosis for the image can simply be defined as the maximum risk of presence of a target lesion: this is what was done in this study.

A. Candidate Extraction

A candidate extraction step is required for both selecting reference negative lesion confounders in the direct sampling approach (see Section II-B) and selecting the samples that should be fed to the classifier when processing unseen images.

The first step to extract candidates in an input image is to transform the image as described in Section II-C1. For each pixel $p_{i,j}$ in the transformed image, the intensity of the potential lesion, within sample $S_{i,j}$ with central pixel $p_{i,j}$, is then estimated as described in Section II-C2. If this intensity is outside the normal range of intensities for the target lesions—the range observed on the training set plus a margin, $p_{i,j}$ is rejected as a potential lesion center. Indeed, it is preferable to discard candidates with low intensity (the background): by chance, magnified background pixels might look similar to the target lesions; similarly, attenuated pigment clumps or attenuated small nevi might look like lesions, so candidates with an unusually high intensity should be discarded. If the lesion intensity is within that range, sample $S_{i,j}$ is normalized (i.e., divided by the estimated intensity) before being further classified.

The specificity of this candidate extraction step needs to be arbitrarily high when it is used to select reference negative lesion confounders in the direct sampling approach (see Section II-B). Therefore, the candidates samples $S_{i,j}$ are further selected using a manually tuned constant threshold τ on the average pixel-to-pixel squared distance to the average transformed and normalized positive lesion sample ($\tau = 0.3$). Note that this candidate extraction step is related to the classifier described in [18].

B. Defining Risk of Presence for a Lesion

Since the set of filters generating the "optimal feature space" (see Section II-D) was derived from a representative set of lesions and lesion confounders, a straightforward approach to classify new image samples within the optimal feature space is to find the most similar representative samples within this space. A *risk of presence* $\rho(S)$ is computed for each new sample S (S has been obtained as described in Section II-C). $\rho(S)$ derives from the distance between S and the representative samples. To compute the distance between two samples S and R, S (resp. R) is projected on each filter in the optimal set of filters (i.e., on each PCA component): a M-dimensional vector $c(S)$ [resp. $c(R)$] is obtained. The distance between S and R is the Euclidean distance between $c(S)$ and $c(R)$. Approximate nearest neighbor search, based on k-d trees, was applied to find the most similar reference samples [40].

Once the k most similar samples $(R_i)_{i=1..k}$ with label $(l(R_i))_{i=1..k}$ have been found, $\rho(S)$ is defined as follows:

$$\rho(S) = \frac{1}{k}\sum_{i=1}^{k} l(R_i)\exp(-\|R_i - S\|) \tag{12}$$

where $l(R_i) = 1$ if R_i is a target lesion (typical lesion or common positive lesion confounder) and $l(R_i) = -1$ if R_i is a common negative lesion confounder. This *risk of presence* ranges from -1 (high risk of being negative) to 1 (high risk of being positive). If the new sample is neither a typical lesion nor a lesion confounder, then the distance to the most similar reference samples is likely to be high. As a consequence, a risk close to zero should be measured. In particular, if background pixels are not discarded by the candidate selection step (described in Section III-A), it is likely that a risk close to zero is measured.

C. Trading Off Performance and Complexity

If the set of reference samples is truly representative of the negative (resp. positive) lesion confounders, then by definition, false positives (resp. false negatives) should only occur for rare negative (resp. positive) lesion confounders. This implies that we can trade-off system performance and the system complexity by adjusting our definition of "common" and "rare" lesion confounders: as the number of modeled (expert-driven approach) or selected (data-driven approach) lesion confounders increases, the number of false positives and false negatives decreases, but computation time increases (see Fig. 6).

IV. Experiments

The proposed system was tested on microaneurysm detection, for DR screening, in Section IV-A and drusen detection, for AMD detection, in Section IV-B. If not mentioned otherwise, the computations are performed in the green channel of fundus images.

A. Microaneurysm Detection

The performance of the proposed framework at detecting microaneurysms was first evaluated at a lesion level, on a set of 100 deidentified images manually segmented by a clinician expert (MDA). Twenty-six of these images were taken from a DR screening program in the Netherlands [5]; these images were originally captured using a Canon CR5 nonmydriatic 3CCD camera at 45° field of view; and 74 were obtained at Abràmoff's retinal referral clinic; these images were captured using a Topcon TRC-50 at 45° field of view. The size of the images is 768×576 pixels and the field of view is approximately 540 pixels in diameter. 55/100 of the images contained lesions. In these images, a total of 858 red lesions were identified. The dataset is similar to the second set described in [15].

Half of the images in this dataset were used for training and the other half for testing. In order to validate the optimal fitting of the filters to the reference samples, the optimal filtering framework (including confounders) was also compared to an approach were analytical filters are selected to maximize the

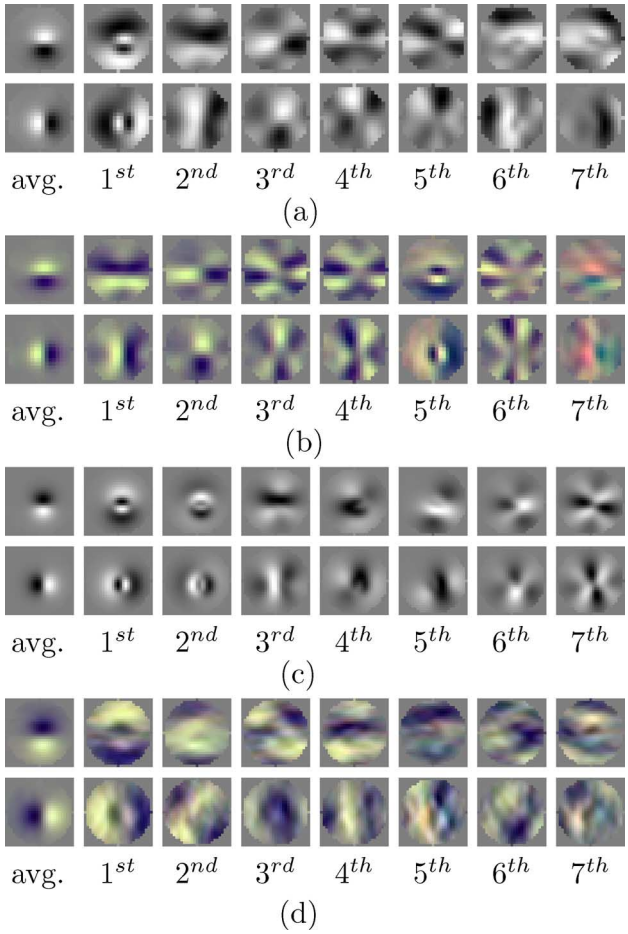

Fig. 3. Optimal filters found for microaneurysm detection (a), (b) and for drusen versus flecks classification (c), (d). The upper part (resp. the lower part) of each filter is correlated with vertical (resp. horizontal) Haar filters dilated by a factor of 2. This figure shows that similar sets of filters are obtained whatever method is used to collect the set of reference samples. (a) MA: mathematical model approach. (b) MA: Sampling approach. (c) Drusen: Mathematical model approach. (d) Drusen: sampling approach.

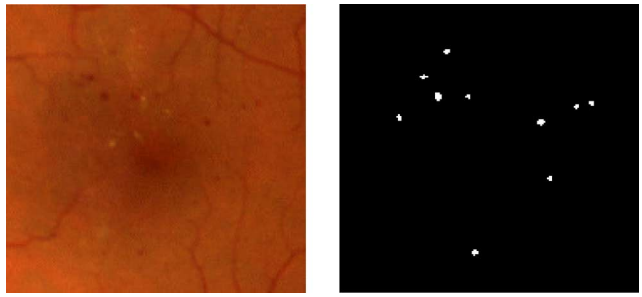

Fig. 4. Example of microaneurysm detection by the proposed mathematical model based method in the macular region of a JPEG-compressed image used in the EyeCheck screening program.

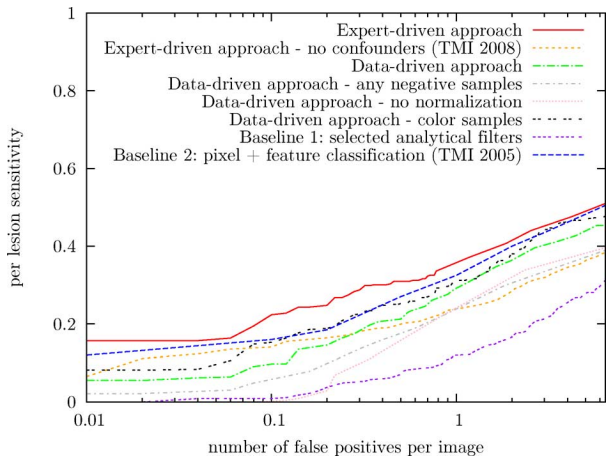

Fig. 5. FROC of microaneurysm detection in logarithmic scale. "Expert-driven approach—no confounders" shows the performance obtained with a simplified version of the proposed method [18] in which no lesion confounders were included in the set of reference samples; there were only two reference samples: a small and a large typical lesion [see (2)]. "Data-driven approach—any negative samples" shows the performance obtained if the negative reference samples are selected at random outside of the manually segmented regions, instead of being selected among the false positives of candidate selection (see Section II-B). "Data-driven approach—no normalization" shows the performance obtained if the normalization step described in Section II-C2 is omitted both for training and testing.

classification accuracy, by cross-validation, on the set of reference samples ("training set"). The analytical filters used in this experiment were typical: Gaussian derivatives and Gabor wavelets in color opponency space (see our use of these analytical filters for example in [37], [15]). All other aspects of the method (correlation and normalization of the data, k-NN classifier with the same risk) remained unchanged. The radius of the filters was set to 9 pixels for this application. The optimal filters are reported in Fig. 3(a) and (b) and a detection example is given in Fig. 4. The free-response receiver-operating characteristic (FROC) [41] curves of the different methods on the testing set are reported in Fig. 5.

The time complexity assessment of the method is presented in Fig. 6; the reported times are for a whole image. For each image, one sample was extracted around each pixel within the field of view, and about 2% of the samples were selected by the candidate extraction step (that is 4500 samples per image for microaneurysm detection). For convenience, we studied the time complexity of the mathematical model based approach (it is easier to increase the number of generated samples); similar computations times were obtained with the direct sampling approach.

Once the proposed method was calibrated at a lesion level, the performance of the best performing version of the proposed algorithm (the mathematical-model approach) was compared at an examination level to a re-implementation of Niemeijer's method [15]. This experiment was carried out on a set of images from the same screening program [8], [5]. The size of images and that of the field of view are similar to the previous set. 2739 consecutive examinations from diabetic patients on their first visit were selected from a set of first visits in 2008. Each examination consists of two images per eye (a macula-centered and an optic disc centered image). Sixty-seven of them were diagnosed with DR by a single retinal specialist, and the maximal microaneurysm presence risk measured within this set of four images was used as DR risk (see Section III). The ROC [42] of these two methods are reported in Fig. 7. Note that, in a screening context, all patients are expected to be asymptomatic, including those with referable DR. DR is referable when a couple of lesions are seen by the expert, not when a single subtle lesion is

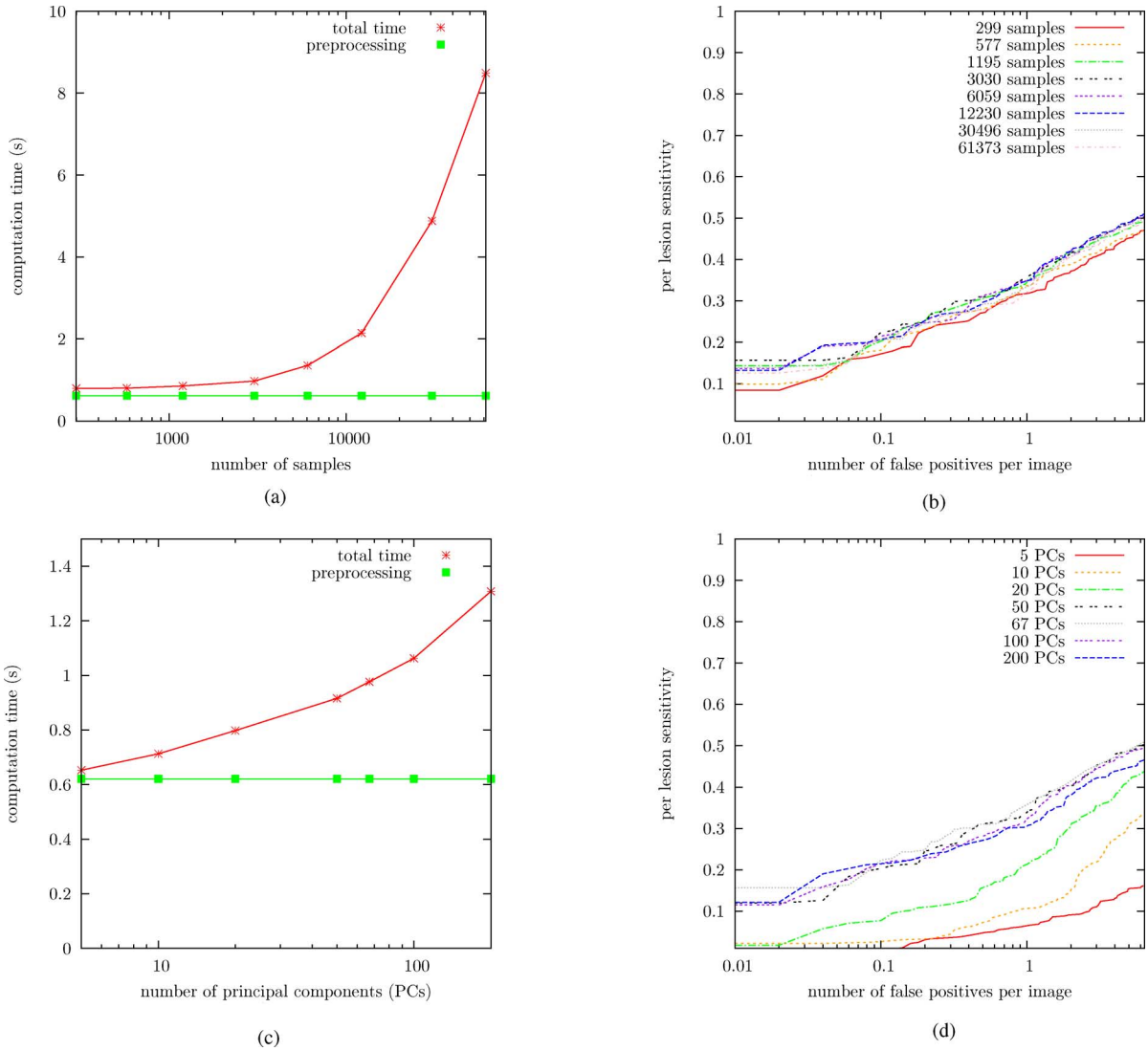

Fig. 6. Time complexity assessment of the method on the microaneurysm detection problem—mathematical model based approach. The computation times have been obtained on a standard personal computer running at 2.26 GHz, using a varying number of principal components (PCs) and reference samples. Each vertex in (a) [resp. (c)], from left to right, corresponds to one curve in (b) [resp. (d)] from "299 samples" (resp. "5 PCs") to "61373 samples" (resp. "200 PCs"). Sixty-seven PCs were used in (a) and (b) and 3030 samples were used in (c) and (d). (a) Number of samples/time complexity. (b) Number of samples/performance. (c) Size of optimal filter set (Number of PCs)/time complexity. (d) Size of optimal filter set (Number of PCs)/performance.

present; lesions missed by the expert, as well as by the algorithm, are usually subtle.

B. Drusen Detection

In the second application of the proposed framework, drusen are differentiated from Stargardt's disease flecks. Thus, the lesions and lesion confounders appear in different image sets, and the problem is to differentiate the two types of lesions as to whether they are typical for AMD (drusen) or Stargardt's disease (flecks). A set of 15 images of patients with AMD, all containing drusen, as confirmed by a clinician expert (MDA), and 15 images containing flecks as confirmed to have Stargardt's disease by positive genetic testing for the ABCA4 haplotype were selected. Images were acquired with a Zeiss FF3 30° camera and recorded on Kodachrome film (ASA 100). Film-based images were scanned at 3456×2304 pixels, 12 bit depth, using a three color channel CCD scanner. The field of view is approximately 930 pixels in diameter. A set of 300 drusen and 300 flecks were manually annotated in these images and used as reference samples for the sampling approach. A leave-one-image-out cross-validation was performed to assess the sampling approach. To validate the mathematical model approach, the 600 samples were used as a testing set. The radius of the filters was set to 13 pixels for this application. The optimal filters are reported in Fig. 3(c) and (d). The ROCs are reported in Fig. 8.

V. Discussion and Conclusion

We presented and validated a novel, general, optimal filter framework for detecting target lesions in retinal images. It relies on the design of a set of filters optimally representing the typical target lesions, as well as the *negative lesion confounders*, i.e., lesions that are similar looking but not the target lesions, and the *positive lesion confounders*, i.e., lesions that are easily missed because they have specific properties that are substantially different from the standard case. In order to design the op-

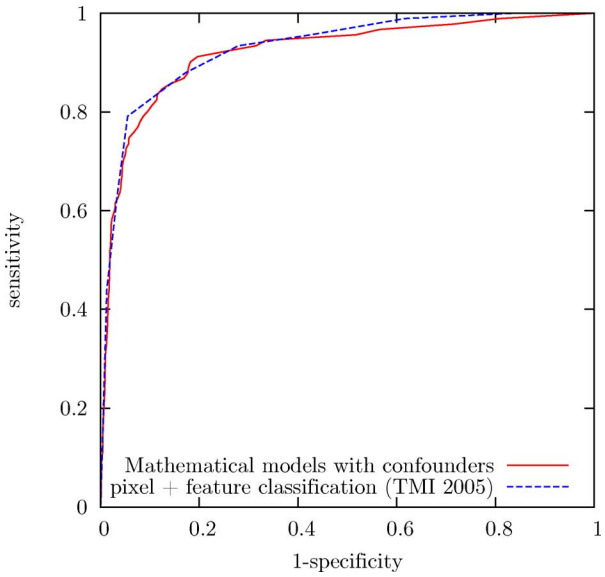

Fig. 7. ROC of microaneurysm detection at an image level on a set of 2739 patient images, including 67 with referable DR.

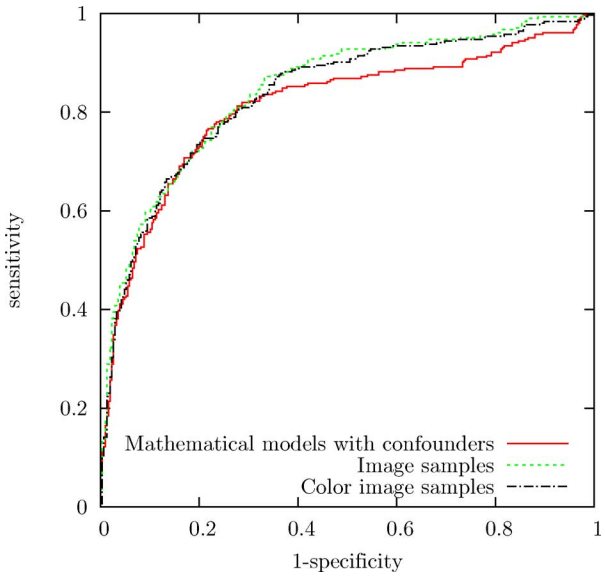

Fig. 8. ROC of drusen versus flecks classification.

timal filter set, image samples representative of the three types of lesions are collected and the filters are defined as the principal components of these samples. Special attention is given in the proposed approach to ensure the representativeness of the reference image samples. Two completely different approaches for obtaining these *maximally representative samples* are proposed and compared: an expert- and a data-driven approach. The optimal filter set is used as a reference to classify new image samples. Because the feature space spanned by these filters is compact, the classification procedure, based on an approximate nearest neighbor search, is fast. Although we used a k-NN classifier in this study, because it is fast, the proposed optimal filter framework can be used with any other instance-based classifier, including the support-vector machines. We do not claim that the proposed filter framework, in combination with any specific classifier, is in any way optimal. We do claim, and establish empirically, that the filter framework itself is optimal, irrespective of the classifier. The nice property of the couple (proposed filter framework; k-NN classifier) is that it allows state-of-the-art classification performance at a very low computational cost. This property is particularly important in medical applications such as the instantaneous assessment of retinal diseases.

The optimal filter framework has been tested on the detection of two important lesions in eye disease: microaneurysms and drusen. Our results show that the detection performance achieved by the new optimal filter framework is comparable to that of previous algorithms presented by our group [15] (see Figs. 5 and 7), however the new method is much faster: each retinal image is processed in less than a second, as opposed to several minutes with the previous approach. This implies that instantaneous assessment of retinal images for DR is achievable, while maintaining the performance achieved by previously published algorithms. The results show that the optimal filter framework also has good performance in the differentiation of drusen from Stargardt's disease flecks, illustrating its generality.

The optimal filter sets obtained varied little with the method through which they were obtained (see Fig. 3). The distribution of confounders types (e.g., round microaneurysms, elongated microaneurysms, vessel branching, etc.) in the expert-driven set of reference samples may differ from their distribution in the data-driven set. This difference of distribution can explain that the ordering of the principal components is changed. This being said, it mattered little whether the representative samples (from which the optimal filters are derived) are obtained from a mathematical model designed by a *nonclinician* domain expert, i.e., expert driven, or by directly sampling a representative set of images at locations annotated by one or more expert clinicians, i.e., data-driven, following the sampling procedure described in this paper. Thus, an optimal filter framework, with the set of domain optimal filters obtained directly from image samples, and only requiring expert clinician input for annotating, performs close (for DR detection) or equivalent (for drusen versus flecks differentiation) to a system in which the filters require the design of a model by a nonclinician expert. The data-driven approach is expected to generalize well to other retinal image analysis problems, and possibly, to other medical image analysis problems, because only limited clinician expert annotation effort is required, and no further domain knowledge is required on the designer of the framework. The results show that the use of color does not decrease performance despite widespread knowledge that the green channel gives best discrimination [43]; it even significantly increases per-lesion sensitivity for some numbers of false positives per image ($p = 0.014$ for 0.1 false positives per image). This is primarily important for the data-driven approach, because it shows that no expert intervention for the use or dis-use of specific color channels is required. The expert-driven approach was superior to the data-driven approach when the designer of the mathematical models (GQ) had sound nonclinician expert knowledge (microaneurysm detection). It shows that, if expert knowledge is available, it is possible to push performance further through

mathematical modeling. Moreover, should the method be applied to a new dataset, we would just have to change the value of a few parameters, while in the data-driven approach additional annotation might have been required. In conclusion, although the mathematical modeling can lead to higher classification rates, it needs to be adapted to each lesion type or each particular type of confounders a dataset may contain; if one does not want to spend time designing new models, the data-driven approach should be used.

Finally, the results indicate that using a set of filters obtained from factorial coding of the data—in our case, PCA—may be superior, and at least equal, to the set of filters obtained through feature selection on a Gaussian derivative or Gabor wavelet or filterbank (see Fig. 5) [44], [15], [37], [29], [45]. We suspect that *on average*, across all object detection problems, from all of the potential filter sets in the filter, or feature, space, a small set Gabor or Gaussian derivative filters may be optimal, but that for *specific* object detection problems, a locally optimal, better performing filter set exists, and can be found in this space through the framework described in this paper. Such a tool for finding good image features automatically is advantageous for researchers: without it, they may spend hours looking through the existing image feature literature and never find satisfactory features.

This study has some limitations. First, Stargardt patients as a group tend to be younger than AMD patients, and so may have a more reflective inner limiting membrane (ILM) than AMD patients. Although we primarily use shape, not color or intensity, as a feature, the difference in ILM reflectivity might influence the classification score reported in Fig. 8. Second, the reference standards used in all experiments were obtained by a single expert, which certainly affects performance. Finally, the proposed approach also has a limitation: if the scale range for the target lesion is very important, then several classifiers may have to be trained using a different size of samples each time.

Since a fast lesion detection algorithm is available, it will be possible in future works to analyze several frames of an eye fundus movie and combine lesion detections from consecutive analyzed frames, in order to improve detection performance. Several frames may be analyzed while the operator is shooting the movie, in order to give him/her some feedback: the algorithm could indicate parts of the retina where there might be lesions (based on automated analysis of previous frames) but where a finer analysis is needed to confirm.

In summary, a novel, general, optimal filter framework for the detection of lesions in retinal images, derived through factor analysis of the lesions and their confounders, has been presented in this paper and was successfully applied to two important retinal image analysis problems. Potentially, this optimal filter framework can be applied to other lesion detection problems in retinal images, where rapid system development and instantaneous performance are required.

REFERENCES

[1] D. C. Klonoff and D. M. Schwartz, "An economic analysis of interventions for diabetes," *Diabetes Care*, vol. 23, no. 3, pp. 390–404, Mar. 2000.

[2] J. Kinyoun, F. Barton, M. Fisher, L. Hubbard, L. Aiello, and F. Ferris, "Detection of diabetic macular edema: Ophthalmoscopy versus photography—Early treatment diabetic retinopathy study (ETDRS) report 5," *Ophthalmology*, vol. 96, no. 6, pp. 746–750, June 1989.

[3] ETDRS Research Group, "Early photocoagulation for diabetic retinopathy—ETDRS report 9," *Ophthalmology*, vol. 98, no. 5, pp. 766–785, May 1991.

[4] G. H. Bresnick, D. B. Mukamel, J. C. Dickinson, and D. R. Cole, "A screening approach to the surveillance of patients with diabetes for the presence of vision-threatening retinopathy," *Ophthalmology*, vol. 107, no. 1, pp. 19–24, Jan. 2000.

[5] M. D. Abràmoff and M. S. A. Suttorp-Schulten, "Web-based screening for diabetic retinopathy in a primary care population: The eyecheck project," *Telemed. J. E. Health*, vol. 11, no. 6, pp. 668–674, Dec. 2005.

[6] T. Teng, M. Lefley, and D. Claremont, "Progress towards automated diabetic ocular screening: A review of image analysis and intelligent systems for diabetic retinopathy," *Med. Biol. Eng. Comput.*, vol. 40, no. 1, pp. 2–13, Jan. 2002.

[7] R. J. Winder, P. J. Morrow, I. N. McRitchie, J. R. Bailie, and P. M. Hart, "Algorithms for digital image processing in diabetic retinopathy," *Comput. Med. Imaging Graph.*, vol. 33, no. 8, pp. 608–622, Dec. 2009.

[8] M. D. Abràmoff, M. Niemeijer, M. S. A. Suttorp-Schulten, M. A. Viergever, S. R. Russell, and B. van Ginneken, "Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes," *Diabetes Care*, vol. 31, no. 2, pp. 193–198, Feb. 2008.

[9] T. Spencer, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus," *Computers and Biomedical Research*, vol. 29, no. 4, pp. 284–302, Aug. 1996.

[10] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "A fully automated comparative microaneurysm digital detection system," *Eye*, vol. 11, no. 5, pp. 622–628, 1997.

[11] J. H. Hipwell, F. Strachan, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "Automated detection of microaneurysms in digital red-free photographs: A diabetic retinopathy screening tool," *Diabetic Medicine*, vol. 17, no. 8, pp. 588–594, Aug. 2000.

[12] L. Gang, O. Chutatape, H. Lui, and S. M. Krishnan, "Abnormality detection in automated mass screening system of diabetic retinopathy," in *Proceedings of 14th IEEE Symposium on Computer-Based Medical Systems*, 2001, pp. 132–137.

[13] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Medicine*, vol. 19, no. 2, pp. 105–112, Feb. 2002.

[14] M. Larsen, J. Godt, N. Larsen, H. Lund-Andersen, A. K. Sjølie, E. Agardh, H. Kalm, M. Grunkin, and D. R. Owens, "Automated detection of fundus photographic red lesions in diabetic retinopathy," *Invest. Ophthalmol. Vis. Sci.*, vol. 44, no. 2, pp. 761–766, Feb. 2003.

[15] M. Niemeijer, B. van Ginneken, J. Staal, M. S. A. Suttorp-Schulten, and M. D. Abràmoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Trans. Med. Imaging*, vol. 24, no. 5, pp. 584–592, May 2005.

[16] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Trans. Med. Imaging*, vol. 25, no. 9, pp. 1223–1232, Sept. 2006.

[17] T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, and J.-C. Klein, "Automatic detection of microaneurysms in color fundus images," *Med. Image Anal.*, vol. 11, no. 6, pp. 555–566, Dec. 2007.

[18] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE Trans. Med. Imaging*, vol. 27, no. 9, pp. 1230–1241, Sept. 2008.

[19] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, "Hierarchical detection of red lesions in retinal images by multiscale correlation filtering," in *Proceedings of SPIE Medical Imaging 2009: Computer Aided Diagnosis*, Feb. 2009, vol. 7260, pp. 72 601L-1–12.

[20] C. I. Sánchez, R. Hornero, A. Mayo, and M. García, "Mixture model-based clustering and logistic regression for automated detection of microaneurysms in retinal images," in *Proceedings of SPIE Medical Imaging 2009: Computer Aided Diagnosis*, Feb. 2009, vol. 7260, pp. 72 601M-1–8.

[21] A. Mizutani, C. Muramatsu, Y. Hatanaka, S. Suemori, T. Hara, and H. Fujita, "Automated microaneurysm detection method based on double-ring filter in retinal fundus images," in *Proceedings of SPIE Medical Imaging 2009: Computer Aided Diagnosis*, Feb. 2009, vol. 7260, pp. 72 601N-1–8.

[22] M. Niemeijer, M. D. Abràmoff, and B. van Ginneken, "Information fusion for diabetic retinopathy CAD in digital color fundus photographs," *IEEE Transactions on Medical Imaging*, vol. 28, no. 5, pp. 775–785, 2009.

[23] E. Chaum, T. P. Karnowski, V. P. Govindasamy, M. Abdelrahman, and K. W. Tobin, "Automated diagnosis of retinopathy by content-based image retrieval," *Retina*, vol. 28, no. 10, pp. 1463–1477, 2008.

[24] M. Niemeijer *et al.*, "Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs," *IEEE Trans. Med. Imaging*, vol. 29, no. 1, pp. 185–195, Jan. 2010.

[25] R. van Leeuwen, C. C. W. Klaver, J. Vingerling, A. Hofman, and P. T. V. M. deJong, "The risk and natural course of age-related maculopathy: Follow-up at 6 1/2 years in the rotterdam study," *Arch. Ophthalmol.*, vol. 121, pp. 519–526, 2003.

[26] A. Barthes, J. Conrath, M. Rasigni, M. Adel, and J. P. Petrakian, "Mathematical morphology in computerized analysis of angiograms in age-related macular degeneration," *Med. Phys.*, vol. 28, pp. 2410–2419, 2001.

[27] K. Rapantzikos, M. Zervakis, and K. Balas, "Detection and segmentation of drusen deposits on human retina: Potential in the diagnosis of age-related macular degeneration," *Med. Image Anal.*, vol. 7, pp. 95–108, 2003.

[28] R. T. Smith, J. K. Chan, T. Nagasaki, U. F. Ahmad, I. Barbazetto, J. Sparrow, M. Figueroa, and J. Merriam, "Automated detection of macular drusen using geometric background leveling and threshold selection," *Arch. Ophthalmol.*, vol. 123, pp. 200–206, 2005.

[29] M. Niemeijer, B. van Ginneken, S. R. Russel, M. S. A. Suttorp-Schulten, and M. D. Abràmoff, "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for early diagnosis of diabetic retinopathy," *Inv. Ophthalm. Vis. Sci.*, vol. 48, pp. 2260–2267, 2007.

[30] A. D. Fleming, S. Philip, K. A. Goatman, G. J. Williams, J. A. Olson, and P. F. Sharp, "Automated detection of exudates for diabetic retinopathy screening," *Phys. Med. Biol.*, vol. 52, pp. 7385–7396, 2007.

[31] S. Walia and G. A. Fishman, "Natural history of phenotypic changes in stargardt macular dystrophy," *Ophthalmic Genet.*, vol. 30, pp. 63–68, 2009.

[32] P. McDonnell, "Retinator: Revenge of the machines," *Ophthalmology Times*, vol. 35, no. 13, p. 4, 2010.

[33] M. D. Abràmoff, J. M. Reinhardt, S. R. Russell, J. C. Folk, V. B. Mahajan, M. Niemeijer, and G. Quellec, "Automated early detection of diabetic retinopathy," *Ophthalmology*, vol. 117, no. 6, pp. 1147–1154, 2010.

[34] M. D. Abràmoff, M. Garvin, and M. Sonka, "Retinal imaging and image analysis," *IEEE Reviews in Biomedical Engineering*, 2010, (in press).

[35] A. J. Bell and T. J. Sejnowski, "The "independent components" of natural scenes are edge filters," *Vision Res.*, vol. 37, no. 23, pp. 3327–3338, 1997.

[36] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *J. Opt. Soc. Am. A*, vol. 14, pp. 1393–1411, 1997.

[37] M. D. Abràmoff, W. L. M. Alward, E. C. Greenlee, L. Shuba, C. Y. Kim, J. H. Fingert, and Y. H. Kwon, "Automated segmentation of the optic nerve head from stereo color photographs using physiologically plausible feature detectors," *Invest Ophthalmol. Vis. Sci.*, vol. 48, no. 4, pp. 1665–1673, Apr. 2007.

[38] I. T. Jolliffe, "Principal Component Analysis," in *Statistics*, ser. Springer Series, 2nd ed. , NY: Springer, 2002.

[39] M. Bethge, "Factorial coding of natural images: How effective are linear models in removing higher-order dependencies," *J. Opt. Soc. Am. A Opt. Image Sci. Vis.*, vol. 23, no. 6, pp. 1253–1268, 2006.

[40] S. Arya and D. M. Mount, "Approximate nearest neighbor queries in fixed dimensions," in *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms*, 1993, pp. 271–280.

[41] P. Bunch, J. Hamilton, G. Sanderson, and A. Simmons, "A free-response approach to the measurement and characterization of radiographic-observer performance," *J. Appl. Photographic Eng.*, vol. 4, pp. 166–171, 1978.

[42] C. E. Metz, B. A. Herman, and J.-H. Shen, "Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data," *Statistics in Medicine*, vol. 17, no. 9, pp. 1033–1053, Dec. 1998.

[43] M. Foracchia, E. Grisana, and A. Ruggeri, "Luminosity and contrast normalization in retinal images," *Medical Image Analysis*, vol. 9, no. 3, pp. 179–190, June 2005.

[44] B. M. ter Haar Romeny, *Front-End Vision and Multi-Scale Image Analysis*. : Kluwer Academic Publisher, 2003.

[45] I. C. Sluimer, M. Prokop, I. Hartmann, and B. van Ginneken, "Automated classification of hyperlucency, fibrosis, ground glass, solid and focal lesions in high resolution CT of the lung," *Medical Physics*, vol. 33, pp. 2610–2620, 2006.