Optimal wavelet transform for the detection of microaneurysms in retina photographs.

Gwénolé Quellec, Mathieu Lamard, Pierre Marie Josselin, Guy Cazuguel, Béatrice Cochener, Christian Roux

To cite this version:

Gwénolé Quellec, Mathieu Lamard, Pierre Marie Josselin, Guy Cazuguel, Béatrice Cochener, et al.. Optimal wavelet transform for the detection of microaneurysms in retina photographs.. IEEE Transactions on Medical Imaging, Institute of Electrical and Electronics Engineers, 2008, 27 (9), pp.1230-41. 10.1109/TMI.2008.920619 . hal-00326135

HAL Id: hal-00326135

https://hal.archives-ouvertes.fr/hal-00326135

Submitted on 1 Oct 2008
Optimal wavelet transform for the detection of microaneurysms in retina photographs

Gwénolé Quellec$^{1,3}$, Mathieu Lamard$^{2,3}$, Pierre Marie Josselin$^{2,3,4}$, Guy Cazuguel$^{1,3}$, Member, IEEE, Béatrice Cochener$^{2,3,4}$, Christian Roux$^{1,3}$, Fellow, IEEE

Abstract—In this article, we propose an automatic method to detect microaneurysms in retina photographs. Microaneurysms are the most frequent and usually the first lesions to appear as a consequence of diabetic retinopathy. So, their detection is necessary for both screening the pathology and follow up (progression measurement). Automating this task, which is currently performed manually, would bring more objectivity and reproducibility. We propose to detect them by locally matching a lesion template in subbands of wavelet transformed images. To improve the method performance, we have searched for the best adapted wavelet within the lifting scheme framework. The optimization process is based on a genetic algorithm followed by Powell’s direction set descent. Results are evaluated on 120 retinal images analyzed by an expert and the optimal wavelet is compared to different conventional mother wavelets. These images are of three different modalities: there are color photographs, green filtered photographs and angiographs. Depending on the imaging modality, microaneurysms were detected with a sensitivity of respectively 89.62%, 90.24% and 93.74% and a positive predictive value of respectively 89.50%, 89.75% and 91.67%, which is better than previously published methods.

Index Terms—diabetic retinopathy, microaneurysms, template matching, optimal wavelet transform, genetic algorithm

I. INTRODUCTION

DIABETIC RETINOPATHY (DR) has become a major public health issue because it is one of the main sources of blindness. In particular, epidemiological studies carried out in industrialized countries classify DR amongst the four main causes of sight problems over the whole population and the first cause of blindness before 50 years old [1][2]. Economic issues are also at stake. For example, 600 million dollars could potentially be saved every year in the United States by improving early detection of DR [3].

DR is a progressive pathology: its severity is determined by the number and the types of lesions present on the retina. As a consequence, there is a need to detect those lesions either for screening DR or for measuring its progression. Currently, lesion detection is performed manually. Automating this task would primarily be interesting in terms of objectivity and reproducibility. Moreover, if screening or follow up becomes more widely accepted, automatic segmentation will be necessary to replace physicians in the time consuming process of detection.

Microaneurysms (MAs), which are small swellings appearing on the side of tiny blood vessels, are the most frequent and often the first lesions to appear as a consequence of DR. Therefore, within this study we focused on detecting this kind of lesion. Several algorithms have been proposed to detect MAs in retina photographs. They always involve a sequence of different processing tasks. The first family of processing usually involves removing vessels for instance using mathematical morphology [4][5][6], including the top-hat transform [7]. Images are then enhanced and normalized, particularly to overcome lighting variations, by median filtering [8][6], histogram equalization [8] or retina specific treatments [9]. Different features (such as shape features or brightness) are then extracted. The use of statistical classifiers has been proposed [10][7][11][9] to detect the lesions using extracted features. In [12], preprocessed images are directly classified by a neural network (without extracting features). The two previous approaches have been combined, thus [13], [7], [9] or [14] propose hierarchical classifiers that first perform a pixel-wise classification to extract candidate lesions and then a statistical classifier was used to reject.

A matched filter is used in some studies to enhance the lesion shape using Gaussian [5][9] or special-purpose filters [12][15]. Matched filters have also been widely used on retinal images to extract blood vessels in particular [16][17][18][19], usually Gaussian filters [16][19], sometimes morphological filters [20]. In [19], a binary and a Kirsch template matched filter were also used.

In this article we introduce a new template-matching based algorithm to detect MAs. A related approach has already been proposed to detect pulmonary nodules in Helical CT Images, using a Gaussian function as template [21]. The images we study present important lighting variations and are somewhat noisy. So, template-matching cannot be efficiently performed on raw images and lesion features have to be enhanced. For this reason, we propose template-matching in the wavelet domain: in this domain, without other image processing, it is possible to solve problems caused by lighting variations or high frequency noise by choosing the working subbands. Another contribution of this article is to design a wavelet family “optimally adapted” to the problem. Several researchers have put forward methods to design a wavelet to match a specified signal in the least squares sense ([22][23]). We propose a design with a higher level of criterion: finding the wavelet that is best able to discriminate lesions from lesion-free areas. The proposed detection algorithm was evaluated on a database of multimodal retinal photographs.

The article plan is as follows. Section II-A describes our image...
Fig. 1. Series of retina images from a patient file. Angiographs are obtained by injecting a contrast product and taking several snapshots to obtain a temporal sequence: (d)(e)(f).

Fig. 2. 1D generalized Gaussian function $f(r; \alpha, \beta, \gamma, \delta)$

(a) influence of $\alpha$ ($\beta=2, \gamma=0, \delta=1)$

(b) influence of $\beta$ ($\alpha=1, \gamma=0, \delta=1)$

II. MATERIAL AND METHOD

A. The Database

The database contained both 914 color photographs acquired for screening purposes and 995 multimodal photographic images acquired from 52 diabetic patients for follow up (an example is given in figure 1). These photographic modalities brought complementary information about each DR lesion type. The disease severity of the above-mentioned 52 patients, according to the disease severity scale [24], is shown in table I.

<table>
<thead>
<tr>
<th>Disease Severity</th>
<th>Number of Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (No DR)</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

These images have a definition of 1280 pixels/line for 1008 lines/image and are lossless compressed images. They were acquired by experts using a Topcon Retinal Digital Camera (TRC-50IA) connected to a computer and were then analyzed and classified by ophthalmologists. The proposed detection algorithm requires a learning step, performed with images in which lesions have been detected: an expert indicated each lesion centerpoint by clicking on the lesion and positions were recorded. Since a number of lesions were detected at random in the analyzed images, for each lesion, $\alpha$, $\gamma$, and $\delta$ were estimated to match locally the lesion, given a constant shape factor $\beta$. Image backgrounds were determined by using operators from mathematic morphology (opening and closing of size 15), and removed from images prior to parameter estimation (this preprocessing step was performed to validate the model only). For each lesion, $\alpha$ was determined from a discrete set of values $\{\alpha_1, ..., \alpha_n\}$: for each tested value $\alpha_i$, $\gamma$ and $\delta$ were estimated from the pixels found at a distance $r$ from the lesion centerpoint.

B. A Model for Microaneurysms

Despite their size and intensity variations, MAs are quite similar to each other, for a given imaging modality. Indeed, we can model them with 2-dimensional rotation-symmetric generalized Gaussian functions, defined by the following equation (1), represented in figure 2:

$$f(r; \alpha, \beta, \gamma, \delta) = \gamma + \delta \exp \left( \frac{|r|}{\alpha} \right)$$

(1)

where:

- $\alpha$ is the parameter modeling lesion size
- $\beta$ is the parameter modeling lesion sharpness: it is a shape factor (the same value is used for each lesion in images of the same modality)
- $\gamma$ is the parameter modeling the background intensity
- $\delta$ is the parameter modeling lesion height

As an illustration, some typical MA images and corresponding profiles are shown in figure 3(a) and compared to a synthetic lesion generated by the model. The generalized Gaussian function is convenient as it can model a wide range of shapes. Thus, it can model MAs with more or less sharp outlines, depending on imaging modality.

To validate the model relevance, we studied the shape deviation between real MAs and the model. To this end, 100 reference lesions were selected at random in the analyzed images. For each lesion, $\alpha$, $\gamma$, and $\delta$ were estimated to match locally the lesion, given a constant shape factor $\beta = \beta_0$. Image backgrounds were determined by using operators from mathematic morphology (opening and closing of size 15), and removed from images prior to parameter estimation (this preprocessing step was performed to validate the model only). For each lesion, $\alpha$ was determined from a discrete set of values $\{\alpha_1, ..., \alpha_n\}$: for each tested value $\alpha_i$, $\gamma$ and $\delta$ were estimated from the pixels found at a distance $r$ from the lesion centerpoint.
that most MAs are almost round and that they are never very flat.

**C. Template-matching in the Wavelet Domain**

1) **Template-matching**: usually, to perform template-matching in an image $I$, we move a window $w$ over $I$ and consider the subimages of $I$ defined by the positions of $w$, noted $I_{|w}$ (examples are given in figure 4). To classify the content of $I_{|w}$, we have to find the set of parameters (in our case: $(\alpha, \gamma, \delta)$, $\beta_0$ is given - see equation (1)) for which the parametric model best matches $I_{|w}$. If the model is close enough to $I_{|w}$, a lesion is detected at the center of $w$. The distance measure commonly used is the sum of the squared errors (SSE): it is defined as the sum of the squared differences between the value of each pixel of $I_{|w}$ and the corresponding value of the model function.

2) **Adaptation in the Wavelet Domain**: the wavelet decomposition of an image produces several subimages of coefficients (called subbands), each of them containing information at a specific scale/frequency and along a specific direction (see figure 5(a)). These subbands contain more or less relevant information to describe MAs. We propose adaptation of the previous procedure in the wavelet domain, to subbands that contain discriminant information alone. In particular, by ignoring high and low frequency subbands, we can get rid of noise and slow image variations, respectively. Moreover, the wavelet transform of $g = f - \gamma$, where $f$ is the model function and $\gamma$ is the local background intensity, differs from the wavelet transform of $f$ only in the lowest frequency subband. Thus, if we reject the lowest frequency subbnd, the offset parameter $\gamma$ can be set to 0: as a conclusion, the model can be defined in the spatial domain by only 3 parameters $(\alpha, \beta_0, \delta)$.

In the proposed method, for $n$ predetermined standard deviations $\alpha_i$, the normalized model $f(r; \alpha_i, \beta_0, \gamma = 0, \delta = 1)$ is discretized and its wavelet transform $WT M_i$ is computed. Let $J$ be the set of discriminant subbands. The SSE is now defined as the sum of the squared differences between the coefficients of $WTI_{|w}$ (the wavelet transform of $I$ restricted to the window $w$) divided by $\delta_i$ and that of $WT M_i$, in the subbands of $J$ $(\delta_i \mapsto f(r; \alpha_i, \beta_0, \gamma = 0, \delta)$ is linear and the wavelet transform is linear). The proposed procedure is summarized as follows:

- For each standard deviation $\alpha_i$, $1 \leq i \leq n$, we compute the lesion height $\delta_i$ minimizing the SSE between $WT M_i$ and its wavelet transform.
and $\frac{W T I_w (\delta_i)}{\delta_i}$ (\(\delta_i\) is computed by the least mean squared algorithm).

- If $\delta_i$ is not in a predetermined range of values, no lesions of size $\alpha_i$ are detected in $I_{|w|}$.
- Otherwise, $I_{|w|}$ is classified as a MA of size $\alpha_i$ and height $\delta_i$ if the SSE is below a corresponding threshold $t_i \in T$, $1 \leq i \leq n$.

The set of discriminant subbands $J$, the standard deviations $\alpha_i$ and the thresholds $t_i$ are learnt, as we will explain in sections II-E2 and II-E3.

3) Moving windows: it is assumed that a moving window contains at most one lesion. Problems might occur if a lesion is too close to another lesion or to a vessel (see figure 4). As a consequence, we had to use small moving windows: we used circular windows of radius $2\alpha_i$ (see section II-B). It is not necessary to move the windows to each pixel location: the moving step is chosen in inverse proportion to the typical lesion size. In our images, we evaluated $\frac{L}{4}$ window positions, where $L$ is the number of image pixels. To speed up the search, we do not evaluate areas $I_{|w|}$ where $\max(I_{|w|}) - \min(I_{|w|})$ is below the range of values for $\delta$.

D. Wavelet Adaptation to our Specific Problem

Wavelet transform is highly tunable and we have a free choice of the function basis used to decompose images. So we propose a search for a wavelet basis which may make matched filtering in the wavelet domain easier and more efficient. However, the design of matched wavelet bases is usually a difficult task, as several constraints have to be satisfied simultaneously (such as biorthogonality, regularity, vanishing moments, etc) [25]. Many constructions of wavelets have been introduced in the mathematical [26] and signal processing literature (in the context of quadrature mirror filters) [27], to satisfy those constraints. Moreover, these properties might not be sufficient to process specific signals or images. Thus, several studies have been carried out to adapt a wavelet to a signal of reference, with different constraints on the wavelet properties and on the adaptation criteria. Usually, the goal is to minimize the $L^2$ norm between the signal and an approximation derived from its wavelet decomposition (for compression purposes, for instance) [22][28]. These methods have been applied to medical classification problems, in order to match waveform signals [29][30]. However, in our case, we do not intend to match a signal of reference, which is not relevant as the lesions we study present great variations in size and amplitude. We look for the basis best adapted to our problem: template matching in the wavelet domain. The only criterion available is the system’s overall classification rate. So we have to design a wavelet empirically, by a numerical optimization procedure, in which no explicit expression of the wavelet fitness function is given.

In 1994, Sweldens introduced a convenient way to satisfy all the desired properties of wavelets by reducing the problem to a simple relation between the wavelet and scaling coefficients. This approach, named the lifting scheme [31], makes it possible to relax all the mentioned constraints and thus to reach a simple optimization process. Indeed, it is possible to generate any compactly supported biorthogonal wavelet with a compactly supported dual. A filter bank was proposed in [25] to implement the lifting scheme (see figure 6). It defines two linear filters, named predict ($P$) and update ($U$). $U$ and $P$ are two real vectors, respectively of size $N_p$ and $N_u$. The coefficients of those filters have only to satisfy two linear relations to generate biorthogonal wavelets (see appendix A). They allow a perfect reconstruction of decomposed images. Moreover, the lifting scheme makes the waveform transform faster, and is used in the Jpeg-2000 compression standard [32]. For these reasons, we decided to design lifting scheme based wavelet filters to transform our images.

In addition to the wavelet basis, the wavelet transform enabled us to choose the subbands in which images are decomposed, by selecting at each scale which subbands are further decomposed. An algorithm, called “best basis”, has been proposed by Coifman [33][34] to match such a subband selection to a given family of signals. Nevertheless, this approach is too restrictive for the proposed algorithm, since the decomposition on the best basis
preserves all the information contained in the images, whereas we want to reject irrelevant information (such as noise). So, we explored the entire space of subband subsets, in order to find the best ones, from the overall classification rate point of view.

Finally, we did not subsample subbands in the usual way in the wavelet transform. Because the typical size of MAAs is small in our images (the average diameter is about 10 pixels), we decided to use the translation invariant wavelet transform [35], in which every subband is the size of the raw image (see also appendix A). The models and an example of translation invariant decomposition are shown in figure 5(b).

E. Description of the Calibration Procedure

1) Overall Procedure (figure 7(a)): to calibrate the proposed classifier, we had several parameters to determine. These parameters were learnt from a subset of the analyzed images: the learning dataset. The remaining analyzed images (the validation dataset) were only used to evaluate the whole algorithm, when parameter selection had been accomplished. The number of images required in the learning dataset is discussed in section III. We classified the parameters into three groups:

- The model parameters (\( n \) values for \( \alpha \), a range of values for \( \delta \), a value for \( \beta \), see equation (1))
- The decomposition subband subsets used to compute the SSE: we had \( 3 \times N_1 + 1 \) binary values to set, where \( N_1 \) is the maximum number of decomposition levels considered. For our images, we set \( N_1 \) to 3, because our experiments show that there is no interest in decomposing on more than three levels: information at lower frequencies is not adapted to MA scale. Let \( S = \{1HH, 1HL, 1LH, 2HH, 2HL, 2LH, 3HH, 3HL, 3LH, 3LL\} \) be the set of available subbands (see figure 5). We had to evaluate each subset \( J_m \in 2^S \setminus \emptyset \), \( 1 \leq m \leq M \) (where \( 2^S \) is the set of subsets of \( S, M = card(2^S \setminus \emptyset) = 2^{card(S)} - 1 = 1023 \)).
- The wavelet filter support and coefficients: we evaluated both conventional and specific adapted wavelets (based on the lifting scheme)

Groups of parameters were defined in a specific order, which is discussed hereafter. Model parameters are obviously independent from the wavelet filter and the subband subset selection, therefore, they were calibrated first. A more difficult issue is whether the wavelet filters and the subband subsets should be learnt simultaneously, which would be an extremely time-consuming process. To simplify the learning procedure, the subband subsets were evaluated first with several conventional wavelets. These evaluations provided the best subband subsets, which were then used to learn the adapted wavelet filters. The learning procedure can be summarized as follows:

1) model parameter selection (see section II-E3)
2) subband subset learning (see section II-E4)
3) wavelet learning (see section II-E5)

To determine the subband subsets and the adapted wavelets, we needed to compute the \( n \) thresholds on the SSE (one for each value of \( \alpha \)) each time we evaluated a couple (wavelet, subband subset). The method is explained in the following section.

2) Computation of the Thresholds on the SSE for a Couple (wavelet filter \( W_j \), subband selection \( J_m \)): according to section II-C, we had to learn a threshold \( t_i \) for each value of \( \alpha_i, 1 \leq i \leq n \) (the model standard deviations). Elements of the threshold tuple \( T = (t_1, t_2, \ldots, t_n) \) were determined simultaneously. First, each image \( I \) in the learning dataset was decomposed with \( W_j \) and the resulting subbands in \( J_m \) were stored. Then the SSE were computed at each corresponding position in \( J_m \) of the sliding window \( w \) on these images, for each value of \( \alpha_i \) (see equation (1)). A classification score was defined (equation (3) - see explanation below). We wanted to know which \( T \) provided the best classification score, according to the manual detections provided by the expert. This was done by a simple grid search optimization: we explored the space of \( T \) with a given step and retained the best. The procedure is illustrated in figure 7(b).

To define the classification score, it is important to notice that, when optimizing \( T \), we actually have two fitness criteria: sensitivity (the percentage of lesions that are detected) and PPV (Positive Predictive Value: the percentage of relevant detections). We considered a detected object relevant if the ground truth indicated a lesion centerpoint within three pixels (as a comparison, the average lesion radius is five pixels). Sensitivity and PPV may be conflicting, then optimization involves finding the best compromise between the two objective scores. In [36], the authors proposed to combine sensitivity and PPV following equation (2).

\[
\text{score} = \sqrt{\text{sensitivity} \times \text{PPV}}
\]  

Thus, score is high if and only if both sensitivity and PPV are high. However, in a disease screening framework, the cost of
false negatives is higher than that of false positives. Indeed, it is better to send a healthy patient to an ophthalmologist for further analyses, than to miss lesions, and let an ill patient leave without treatment. If we use that fitness score, we observe that PPV is favored. Therefore we decided to penalize threshold tuples for which sensitivity is lower than PPV by using the following score definition (equation (3)).

\[
\text{score} = \zeta \left( \frac{\text{sensitivity}}{\text{PPV}} \right) \times \sqrt{\text{sensitivity} \times \text{PPV}} \\
\zeta(x) = x \text{ for } x < 1, \zeta(x) = 1 \text{ otherwise}
\]

(3)

3) Model Parameter Selection: the choice of $\beta$ is explained in section II-B. Ranges for amplitudes $\delta$ were determined on 100 reference lesions so that every sample laid within them. Finally, the standard deviations $\alpha$, the most critical parameters, were learnt from all the analyzed images. They were chosen so that the template-matching classification score was maximized. This procedure was performed taking different numbers $n$ of values for $\alpha$: it emerged from all the trials that the use of $n = 3$ values did not bring any significant improvement over $n = 2$. For instance, if the Haar wavelet was used to decompose images, sensitivity increased by 6.56% when $n$ increased from 1 to 2 and by 0.17% when it increased from 2 to 3.

4) Subband Subset Learning: we wanted to select a few subband subsets that would be used to compute the SSE, regardless of the wavelet used to decompose images. We assumed that the subband subsets that work best together with several conventional wavelets are liable to be the most efficient with any other wavelet. To select them, we evaluated each subset $J_m \in 2^N \setminus \emptyset, 1 \leq m \leq M$ with $N$ conventional wavelets. The $N = 4$ wavelets we used are given below with their support (lowpass filter support/ highpass filter support):

- the orthogonal Haar wavelet (support=2/2): $W_1$
- the Le Gall 5/3 biorthogonal wavelet, which is used in the Jpeg2000 standard part I [37] (support=5/3): $W_2$
- the Daubechies 9/7 biorthogonal wavelet, also in the Jpeg2000 standard (support=9/7): $W_3$
- the Daubechies 4-tap orthogonal wavelet [38] (support=4/4): $W_4$

For each subset $J_m$ and each conventional wavelet $W_i, 1 \leq i \leq N$, we computed the score of the couple $(W_i, J_m)$, defined as the score obtained for the best threshold tuple (see section II-E2). The global score, on all the wavelets, of each $J_m$ is given by equation (4).

\[
\text{score}(J_m) = \frac{1}{N} \sum_{i=1}^{N} \text{score}(W_i, J_m)
\]

(4)

Then we kept the $k$ best subband subsets ($J_1^{\text{opt}}, J_2^{\text{opt}} \ldots J_k^{\text{opt}}$), defined as the subsets with upper score.

5) Wavelet Learning: For each of the best subband configurations $J_i^{\text{opt}}, 1 \leq i \leq k$, we searched for the best couple (lifting scheme based designed wavelet, $J_i^{\text{opt}}$). Within the lifting scheme, predict and update filters have non-zero even filter lengths [25]. As a result, by reference to formula 13 (appendix A), the corresponding wavelet filters can have the following supports:

- $5/3$ ($N_p = 2, N_u = 2 \rightarrow 2$ parameters)
- $9/3$ ($N_p = 2, N_u = 4 \rightarrow 4$ parameters)
- $9/7$ ($N_p = 4, N_u = 2 \rightarrow 4$ parameters)
- $13/7$ ($N_p = 4, N_u = 4 \rightarrow 6$ parameters)

or larger

As we are studying small structures, we considered only the first three filter configurations.

A meta-algorithm was used to generate lifting scheme based wavelets (see appendix B). Each time a wavelet $W$ was generated, it was evaluated jointly with one of the $k$ best subband subsets on the learning dataset, using the procedure described in section II-E2.

### III. RESULTS

The algorithm parameters and wavelet filter coefficients obtained for each imaging modality are given in table III. The influence of standard deviations $\alpha$ (see equation (1)) on the classification score is illustrated in figure 8.

An important question we had to answer, was the number $p$ of images required in learning sets. We noticed that only three images per imaging modality database are necessary to learn the $n = 2$ thresholds on the SSE (one for each value of $\alpha$), given a wavelet filter, a subband selection and the model parameters (see table IV). Thus the calibration score (the score computed in
Fig. 8. Influence of the generalized Gaussian function standard deviation on the classification score. Figure (a) shows the classification score according to the standard deviation $\alpha$ in the case $n = 1$ (when template matching is performed with a single model size). Figure (b) shows the classification score according to the couple of standard deviations ($\alpha_1, \alpha_2$), $\alpha_1 < \alpha_2$, in the case $n = 2$. Classification scores are proportional to the gray level and the optimal couple is represented by a cross.

Fig. 9. Best combinations of subbands for template matching using conventional wavelets. The first 8 best combinations in decreasing order of their scores (from (a) to (h)). It emerges that the highest frequency subbands ($1HL, 1HL, 1HL$) and the lowest ($3LL$), as well as the diagonal subbands ($1HL, 2HL, 3HL$) are not used (see figure 5 for the subband names).

learning sets is approximately equal to the test score (the score computed in validation sets). However, when searching for the optimal subband selection (on conventional wavelets) or for the optimal wavelet filter, given a subband selection (see procedure in figure 7(a)), the number of parameters increases. To determine $p$, we applied the learning process of section II-E with an increasing number $p$ until the test error stopped evolving. Whatever images we chose in the learning set, the test error stopped evolving for $p = 8$ at most. The choice of $p$ had little correlation with the number of lesions in images. In any case, with too few lesions (less than 40), the objective function (equation (3)) is discontinuous and parameter estimation becomes imprecise. For each imaging modality, we used 8 images in the learning set. The best subband subsets are given in figure 9. We give the score of conventional wavelets with the first three subsets in table V(a).

Results clearly show that the efficiency of the method depends on the wavelet used, which confirms the interest of looking for an optimal wavelet. However, as stated above, optimal subband combinations depend little on the wavelet filter used.

<table>
<thead>
<tr>
<th>TABLE IV</th>
<th>LEARNING AND VALIDATION SCORES WITH AN EXAMPLE.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score values computed on the learning set (L) and the validation set (V) for the Haar wavelet on the following subband subset: ${2HL, 2LH}$.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>set</td>
</tr>
<tr>
<td>L</td>
<td>nr. of images</td>
</tr>
<tr>
<td>GFC photo.</td>
<td>3</td>
</tr>
<tr>
<td>Angiographs</td>
<td>3</td>
</tr>
<tr>
<td>Color photo.</td>
<td>3</td>
</tr>
</tbody>
</table>

| V | nr. of images | nr. of lesions | sensitivity | PPV |
| GFC photo. | 3 | 279 | 87.97% | 95.12% |
| Angiographs | 3 | 276 | 82.94% | 92.09% |
| Color photo. | 3 | 761 | 86.88% | 91.07% |

| TABLE V | CLASSIFICATION EFFICIENCY ACCORDING TO THE SUBBAND SELECTION |
| In each cell, the classification scores are given as a couple (sensitivity, PPV). Values are computed on the validation dataset. |

(a) Conventional wavelets

<table>
<thead>
<tr>
<th>subbands</th>
<th>Haar wavelet</th>
<th>5/3 biorthogonal wavelet</th>
<th>4-tap orthogonal wavelet</th>
<th>9/7 biorthogonal wavelet</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFC photo.</td>
<td>${2HL, 2LH}$</td>
<td>(87.97%,86.88%)</td>
<td>(75.48%,74.41%)</td>
<td>(72.88%,71.76%)</td>
</tr>
<tr>
<td>Angiographs</td>
<td>${2HL, 2LH}$</td>
<td>(78.88%,76.69%)</td>
<td>(71.37%,67.17%)</td>
<td>(58.52%,57.92%)</td>
</tr>
<tr>
<td>Color photo.</td>
<td>${2HL, 2LH}$</td>
<td>(88.56%,75.09%)</td>
<td>(71.05%,67.20%)</td>
<td>(58.72%,54.54%)</td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(81.89%,75.09%)</td>
<td>(58.72%,54.54%)</td>
<td>(62.97%,60.82%)</td>
<td></td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(85.82%,83.53%)</td>
<td>(84.80%,85.31%)</td>
<td>(61.80%,61.80%)</td>
<td></td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(83.91%,84.18%)</td>
<td>(84.62%,84.84%)</td>
<td>(62.97%,66.98%)</td>
<td></td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(89.33%,91.07%)</td>
<td>(83.11%,80.25%)</td>
<td>(68.04%,79.36%)</td>
<td></td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(85.82%,83.53%)</td>
<td>(84.80%,85.31%)</td>
<td>(61.80%,61.80%)</td>
<td></td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(83.91%,84.18%)</td>
<td>(84.62%,84.84%)</td>
<td>(62.97%,66.98%)</td>
<td></td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(89.33%,91.07%)</td>
<td>(83.11%,80.25%)</td>
<td>(68.04%,79.36%)</td>
<td></td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(85.82%,83.53%)</td>
<td>(84.80%,85.31%)</td>
<td>(61.80%,61.80%)</td>
<td></td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(83.91%,84.18%)</td>
<td>(84.62%,84.84%)</td>
<td>(62.97%,66.98%)</td>
<td></td>
</tr>
</tbody>
</table>

(b) lifting scheme based wavelets

<table>
<thead>
<tr>
<th>subbands</th>
<th>9/3</th>
<th>9/7</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFC photo.</td>
<td>(2HL, 2LH)</td>
<td>(90.24%,89.75%)</td>
</tr>
<tr>
<td>Angiographs</td>
<td>(2HL, 2LH)</td>
<td>(88.32%,88.06%)</td>
</tr>
<tr>
<td>Color photo.</td>
<td>(2HL, 2LH)</td>
<td>(87.41%,87.43%)</td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(87.41%,87.43%)</td>
<td>(87.35%,87.37%)</td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(88.95%,88.7%)</td>
<td>(88.94%,88.67%)</td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(87.86%,87.96%)</td>
<td>(88.11%,87.79%)</td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(87.86%,87.96%)</td>
<td>(88.11%,87.79%)</td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(87.86%,87.96%)</td>
<td>(88.11%,87.79%)</td>
</tr>
<tr>
<td>${2HL, 2LH, 3HL}$</td>
<td>(87.86%,87.96%)</td>
<td>(88.11%,87.79%)</td>
</tr>
</tbody>
</table>
Classification scores obtained for optimally designed wavelets are given in table V(b) and the optimal couples (wavelet filter, combination of subbands) found for each imaging modality database have already been given in table III. In figure 10, we gave an example of manually and automatically processed GFC image.

Finally, the proposed method (PM) was compared with those introduced by Fleming [9] and Niemeijer [14] in our database, using the same learning and validation sets. These methods were implemented following the description in the literature, where they are designed to process color photographs. We adapted them to each of the three imaging modalities to handle differences in image resolution, number of color planes, color and typical size of vessels and lesions. In particular, features extracted from the blue, red and hue image planes are ignored when processing gray level images, and vessel or lesion dependent parameters are tuned to maximize the classification score. We also evaluated our lesion detector as a candidate extractor for Niemeijer’s method (NPM), instead of mathematical morphology and pixel classification based candidate extractors. With this aim, the thresholds on the SSE were learnt so that at least 98% of the lesions in the learning set were detected (by analogy with the pixel classification based candidate extractor), with as few false positives as possible (the other parameters are set to the values of table III). The classification scores are given in table VI.

**TABLE VI**

<table>
<thead>
<tr>
<th>Method</th>
<th>GFC photo</th>
<th>Angiographs</th>
<th>Color photo</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>(90%,90%)</td>
<td>(94%,92%)</td>
<td>(90%,89%)</td>
</tr>
<tr>
<td>Fleming</td>
<td>(68%,68%)</td>
<td>(81%,77%)</td>
<td>(70%,55%)</td>
</tr>
<tr>
<td>Niemeijer</td>
<td>(69%,73%)</td>
<td>(80%,79%)</td>
<td>(57%,61%)</td>
</tr>
<tr>
<td>NPM</td>
<td>(73%,73%)</td>
<td>(79%,85%)</td>
<td>(61%,58%)</td>
</tr>
</tbody>
</table>

**IV. DISCUSSION**

In this study, we have proposed a new method for detecting MAs in the retina, based on template-matching in the wavelet domain. The wavelet basis is adapted using the lifting scheme framework. A new criteria has been proposed to evaluate a lesion detector (see equation (3)): it makes it possible to search for the wavelet basis, the relevant subbands and the template-matching parameters in an automatic fashion. The results show that individual lesions can be detected with both a high sensitivity and a high positive predictive value for different photographic modalities. Thus a sensitivity/PPV pair of (90.24\%,89.75\%) was achieved for GFC photographs, of (93.74\%,91.67\%) for angiographs and of (89.62\%,89.50\%) for color photographs. According to that criterion, this method outperforms recent publications [9][14] about MA detection. Moreover, the method is quite insensitive to quality variation between images (see figure 11 for instance). The wavelet filter used for decomposition has been designed by an optimization process within the lifting scheme framework, with the classification rate as the fitness criterion. Wavelet adaptation notably improves the results obtained by conventional wavelet filters. This is particularly true for color photographs: we observe an improvement of 6.7% in sensitivity and of 6.9% in PPV (see tables V(a) and (b)). Wavelet filters (resp. scaling filters) obtained by the calibration procedure are almost antisymmetric (resp. symmetric), like Haar wavelet/scaling filters and unlike the other wavelets filters we studied. Precisely, for Haar and the adapted wavelet, the scaling filter looks roughly like MAs. This explains why the Haar wavelet gives better results than the others (table V(a)). Indeed, if the scaling function matches MA at a given scale \( j \), information specific to MAs is concentrated in subbands of scale \( j \). Thus template-matching can be performed in those subbands alone: relevant and irrelevant information has been decorrelated. Since MAs are rotation-invariant, as discussed in section II-B, MA specific information lies equally in both horizontal (\( j_1 \)HL) and vertical (\( j_2 \)LH) subbands. This can be seen in figure 9: each of the best subband selections include both an horizontal and a vertical subband, either at the second or at the third decomposition level.

Some problems are not completely solved by the method. Firstly, a few MAs are not detected because they are too close to big vessels or clustered. A solution to these problems might be vessel elimination and iterative lesion elimination, respectively. Secondly, some punctiform hemorrhages (another lesion of DR) are detected as MAs, as well as dust stuck to camera lenses or some small structures that ophthalmologists have difficulty to differentiate from MAs.

The proposed method is fast: the learning step is processed only once; then the detection process consists in only two steps:
wavelet transform and template-matching. It does not require extracting and classifying candidates. The total time required to process a single image, of definition 1280 pixels/line for 1008 lines/image, is approximately 2 seconds on a 2 GHz AMD Athlon 64-bit processor. The method can also be used as an efficient candidate extractor and be coupled with a candidate classification method. It is also interesting in terms of reproducibility because parameter estimation is performed automatically and images do not need acquisition-specific preprocessing. Given its simplicity and reproducibility, this method may be one of the answers for disease follow up. Moreover, the method can be adapted to other structures in the retina (like vessels with directional models, or punctiform hemorrhages for instance).

In the current study, we processed images from different modalities separately. We are now processing series of multimodal registered images to fuse their complementary information, for patient follow up.

**APPENDIX A**

**THE LIFTING SCHEME**

A multiresolution approximation is a nested sequence of linear spaces $\{V_j\}_{j \in \mathbb{Z}}$ for approximating functions $f \in L_2(\mathbb{R})$ such that:

- $V_j \subset V_{j+1}$, $\forall j \in \mathbb{Z}$
- $\bigcup_{j=\infty}^{j=-\infty} V_j$ is dense in $L_2(\mathbb{R})$ and $\bigcup_{j=-\infty}^{j=\infty} V_j = \{0\}$
- $f(x) \in V_j \Rightarrow f(2x) \in V_{j+1}$, $\forall j \in \mathbb{Z}$
- $f(x) \in V_j \Rightarrow f(x-2^{-k}) \in V_j$, $\forall j, k \in \mathbb{Z}$

The approximation of a function $f \in L_2(\mathbb{R})$ at a resolution $2^j$ is defined as the projection of $f$ on $V_j$. There is a function $\Phi \in L_2(\mathbb{R})$, called scaling function, such that for any $j \in \mathbb{Z}$, $(x \mapsto \Phi_{jk}(x) = \sqrt{2^j} \Phi(2^j x - k))_{k \in \mathbb{Z}}$ is a basis of $V_j$. The scaling function $\Phi$ satisfies the refinement relation of equation (5).

$$\Phi(x) = 2 \sum_{k=-\infty}^{\infty} h_k \Phi(2x - k)$$  \hspace{1cm} (5)

The additional information available in the approximation of $f$ at the resolution $2^{j+1}$ as compared with the resolution $2^j$, is given by the projection of $f$ on the orthogonal complement of $V_j$ in $V_{j+1}$, noted $W_j$. There is a function $\Psi \in L_2(\mathbb{R})$, called wavelet function, such that for any $j \in \mathbb{Z}$, $(x \mapsto \Psi_{jk}(x) = \sqrt{2^j} \Psi(2^j x - k))_{k \in \mathbb{Z}}$ is a basis of $W_j$. The wavelet function $\Psi$ satisfies the refinement relation of equation (6).

$$\Psi(x) = 2 \sum_{k=-\infty}^{\infty} g_k \Phi(2x - k)$$  \hspace{1cm} (6)

$\{V_j\}_{j \in \mathbb{Z}}$ and $\{W_j\}_{j \in \mathbb{Z}}$ constitute what we call a multiresolution analysis framework. We define a dual multiresolution analysis framework $(\{\tilde{V}_j\}_{j \in \mathbb{Z}}, \{\tilde{W}_j\}_{j \in \mathbb{Z}})$ associated with a dual scaling function $\tilde{\Phi}$ and a dual wavelet function $\tilde{\Psi}$, that are biorthogonal to $\Phi$ and $\Psi$ (see equation (7)).

$$\begin{cases} < \Phi_{jk}, \Psi_{jk'} > = \delta_{kk'} \\ < \Psi_{jk}, \Phi_{jk'} > = \delta_{kk'} \\ < \Psi_{jk}, \Psi_{jk'} > = 0 \\ < \Phi_{jk}, \Psi_{jk'} > = 0 \end{cases}$$  \hspace{1cm} (7)

$\tilde{\Phi}$ and $\tilde{\Psi}$ satisfy refinement relations such as (5) and (6) with coefficients $\tilde{h}_k$ and $\tilde{g}_k$ respectively. A function $f \in L_2(\mathbb{R})$ can be written as follows:

$$x \mapsto f(x) = \sum_{j,k \in \mathbb{Z}} < f, \tilde{\Psi}_{j,k} > \tilde{\Psi}_{j,k}(x)$$

$$= \sum_{k \in \mathbb{Z}} < f, \tilde{\Phi}_{j,k} > \tilde{\Phi}_{j,k}(x) + \sum_{j \leq J, k \in \mathbb{Z}} < f, \tilde{\Psi}_{j,k} > \tilde{\Psi}_{j,k}(x), J \in \mathbb{Z}$$  \hspace{1cm} (8)
The lifting scheme allows the construction of compactly supported wavelets with compactly supported duals, i.e. such that the refinement filters \( h_k, g_k, \tilde{h}_k \) and \( \tilde{g}_k \) are finite filters (a finite number of coefficients are non-zero). We define the following 2π-periodic functions \( h(\omega) = \sum_{k=-\infty}^{\infty} h_k e^{-i k \omega} \) and \( g(\omega) = \sum_{k=-\infty}^{\infty} g_k e^{-i k \omega} \) and the following matrices (called modulation matrices) \( m(\omega) = \begin{pmatrix} h(\omega) & h(\omega + \pi) \\ g(\omega) & g(\omega + \pi) \end{pmatrix} \)

and \( \tilde{m}(\omega) = \begin{pmatrix} \tilde{h}(\omega) & \tilde{h}(\omega + \pi) \\ \tilde{g}(\omega) & \tilde{g}(\omega + \pi) \end{pmatrix} \). A necessary condition for biorthogonality, which ensures perfect reconstruction, is given in equation (9).

\[ \forall \omega \in \mathbb{R} : \tilde{m}(\omega)^T m(\omega) = 1 \]  

In the case of finite filters, \( \det(m(\omega)) \) is a monomial, the determinant is chosen as \( \det(m(\omega)) = -e^{-i \omega} \) which leads to:

\[ \begin{align*}
\tilde{g}(\omega) &= e^{-i \omega} h(\omega + \pi) \\
g(\omega) &= e^{-i \omega} h(\omega + \pi) \\
h(\omega)h(\omega) + h(\omega + \pi)h(\omega + \pi) &= 1
\end{align*} \]  

The lifting scheme relies on a simple relationship between all multiresolution analysis frameworks that share the same scaling function: let \( \{ \Phi^0, \tilde{\Phi}^0, \Psi^0, \tilde{\Psi}^0 \} \) be an initial set of biorthogonal scaling and wavelet functions, then a new set \( \{ \Phi, \tilde{\Phi}, \Psi, \tilde{\Psi} \} \) can be found according to equation (11) [31].

\[ \begin{align*}
\Phi(x) &= \Phi^0(x) \\
\Psi(x) &= \Psi^0(x) - \sum_k s_k \Phi^0(x-k) \\
\tilde{\Phi}(x) &= 2 \sum_k \tilde{h}_k \Phi(2x-k) + \sum_k s_{-k} \Psi(x-k) \\
\tilde{\Psi}(x) &= 2 \sum_k \tilde{g}_k \Phi(2x-k)
\end{align*} \]  

Coefficients \( s_k \) of equation (11) can be freely chosen. Hence we can start with an initial set of biorthogonal filters and generate new biorthogonal filters by the previous relation.

In [25], the authors propose a filter bank adapted from equation (11). It uses the lazy wavelet [31] as initial biorthogonal filter and two filters \( P \) and \( U \) are defined from the \( s_k \) coefficients. The lazy wavelet is the set of biorthogonal filters given in equation (12):

\[ \begin{align*}
2\tilde{h}(\omega) &= h^0(\omega) = 1 \\
2\tilde{g}(\omega) &= 2 g^0(\omega) = e^{-i \omega}
\end{align*} \]  

One step in the lazy wavelet transform consists in splitting the signal into its odd and even indexed samples. In the case of the decimated wavelet transform, one step of the wavelet decomposition is given as follows:

- split: the signal \( x[n] \) is split into its odd \( x_o[n] \) and even \( x_e[n] \) coefficients.
- predict: we generate the wavelet coefficients \( d[n] \) as the error in predicting \( x_o[n] \) from \( x_e[n] \) using predictor operator \( P: d[n] = x_o[n] - P(x_e[n]). \)
- update: combine \( x_e[n] \) and \( d[n] \) to obtain scaling coefficients representing a coarse approximation to the original signal \( x[n] \). This is done by using an update operator \( U: c[n] = x_e[n] + U(d[n]). \)

The translation invariant wavelet transform is simply two decimated wavelet transforms, intertwined at each scale, as proposed by Claypool [25]. The first transform predicts the odd coefficients from the even coefficients, as explained above, the second predicts the even coefficients from the odd coefficients. The filter bank is depicted in figure 6. Let \( N_p \) and \( N_u \) be the length of the linear filters \( P \) and \( U \). It was demonstrated [25] that, to form a biorthogonal wavelet, filters \( P \) and \( U \) only have to satisfy the following conditions:

\[ \sum_{i=1}^{N_p} P_i = 1 \quad \text{and} \quad \sum_{i=1}^{N_u} U_i = \frac{1}{2} \]

There are then \( N_p + N_u - 2 \) undetermined coefficients. Wavelets designed by the lifting scheme have a support length equal to \( s/t \) given in equation (13), where \( s \) is the support of the lowpass filter and \( t \) is the support of the highpass filter.

\[ \begin{align*}
s &= 2(N_p + N_u) - 3 \\
t &= 2N_p - 1
\end{align*} \]  

### Appendix B

**Optimization Methods**

#### A. Genetic Algorithms

When we searched for the best fitted wavelet, we had no preconceived ideas about either its coefficients (except for the two conditions given above) or the wavelet fitness function. Thus we started the search by a controlled random process. Genetic algorithms [39] are amongst the quickest and most popular methods for finding rough estimations of the best local maxima within a given region of space. The use of genetic algorithms to find a predict and an update filter has been proposed by Jones [40]. The general principle of genetic algorithms is to:

- generate a random population of solutions (a set of randomly generated couples of prediction/update filters)
- combine previous good solutions (crossover) and slightly modify some solutions (mutation) to make the population evolve
- keep the fittest solutions (according to equation (3)) at each generation

In particular, we used the steady state algorithm (at each generation, new individuals are created and added to the former population, the worst individuals of the resulting population having been removed), with the following parameters:

- population size = 50
- maximum number of generations = 30
- selection methods: tournament selector (2 individuals are selected for crossover with a probability proportional to their score, and only the fittest is actually kept)
- crossover probability (CR) = 70% (CR: probability for two selected individuals to combine their genes and form two new individuals, otherwise they are cloned)
- mutation probability (MU) = 60% (MU: probability for a gene to be swapped with another after a crossover)

This configuration was adopted for its fast convergence. Indeed, it does not impact the sensitivity/PPV score, since a descent is performed subsequently using Powell’s Direction Set Method.

#### B. Powell’s Direction Set Method

The n-dimensional Powell’s Direction Set Method is a descent algorithm based on conjugate directions [41]. The algorithm starts with an initial point \( P_0 \) and an initial direction set \( D = \{e_1, e_2, ..., e_n\} \), usually the unit vectors. It consists in minimizing
the function along each direction $e_i \in D$ iteratively. Then, the direction set is updated, and the process is reiterated until the solution stops evolving. The interest of this method is that it does not require computing the function gradient.

Once approximations of the wavelet function optimum are found by a genetic algorithm, we can use the best as initial points $P_0$ for descents to reach the local minima. The descent leads to a typical improvement of 2 or 3% in sensitivity and PPV over genetic algorithms alone.

**REFERENCES**


