Normalization of series of fundus images to monitor the geographic atrophy growth in dry age-related macular degeneration

Florence Rossant, Michel Paques

Normalization of series of fundus images to monitor the geographic atrophy growth in dry age-related macular degeneration

Corresponding author:

Florence ROSSANT, Professor,
Institut Supérieur d'Electronique de Paris (ISEP)
10, rue de Vanves, 92130 Issy-les-Moulineaux, France
Florence.rossant@isep.fr
33 1 49 54 52 62

Co-author :

Michel PAQUES, Professor
Clinical Investigation Center 1423, Quinze-Vingts Hospital, 28 rue de Charenton, Paris, France
mpaques@15-20.fr

Normalization of series of fundus images to monitor the geographic atrophy growth in dry age-related macular degeneration

Florence Rossant [a,] and Michel Paques [b]

[a]*ISEP, Institut Supérieur d'Electronique de Paris, 10 rue de Vanves, 92130 Issy-les-Moulineaux, France, florence.rossant@isep.fr*
[b] *Clinical Investigation Center 1423, Quinze-Vingts Hospital, 28 rue de Charenton, 75012 Paris, France*

A B S T R A C T

Background and Objective: Age-related macular degeneration (ARMD) is a degenerative disease that affects the retina, and the leading cause of visual loss. In its dry form, the pathology is characterized by the progressive, centrifugal expansion of retinal lesions, called geographic atrophy (GA). In infrared eye fundus images, the GA appears as localized bright areas and its growth can be observed in series of images acquired at regular time intervals. However, illumination distortions between the images make impossible the direct comparison of intensities in order to study the GA progress. Here, we propose a new method to compensate for illumination distortion between images.

Methods: We process all images of the series so that any two images have comparable gray levels. Our approach relies on an illumination/reflectance model. We first estimate the pixel-wise illumination ratio between any two images of the series, in a recursive way; then we correct each image against all the others, based on those estimates. The algorithm is applied on a sliding temporal window to cope with large changes in reflectance. We also propose morphological processing to suppress illumination artefacts.

Results: The corrected illumination function is homogeneous in the series, enabling the direct comparison of grey-levels intensities in each pixel, and so the detection of the GA growth between any two images. We present numerous experiments performed on a dataset of 18 series (328 images), manually segmented by an ophthalmologist. We demonstrate that the GA growth areas are better contrast after the proposed processing and we propose first segmentation methods to analyse the GA growth, with quantitative evaluation. Colored maps representing the GA evolution are derived from the segmentations.

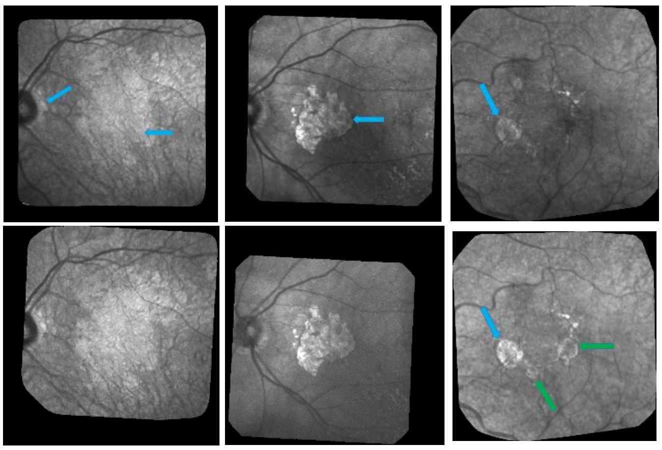
Conclusion: To our knowledge, the proposed method is the first one which corrects automatically and jointly the illumination inhomogeneity in a series of fundus images, regardless of the number of images, the size, shape and progression of lesion areas. This algorithm greatly facilitates the visual interpretation by the medical expert. It opens up the possibility of treating automatically each series as a whole (not just in pairs of images) to model the GA growth.

Keywords – eye fundus images, normalization of series of images, dry age-related macular degeneration (ARMD), geographic atrophy (GA), GA progression detection, GA progression representation.

1. Introduction

Age-related macular degeneration (ARMD) is the leading cause of visual loss in our countries. It affects the retina, the photosensitive layer located in the back of the eye. While the so-called "wet" form of ARMD can be managed by a specific treatment, there is currently no therapeutic options for its "dry" form. The latter is characterized by the progressive, centrifugal expansion of retinal degeneration, giving rise to what has been compared to a geographic-like lesion, hence the name geographic atrophy (GA). The intimate mechanisms of the growth of the lesions remain elusive.

The diagnosis of ARMD is done on fundus photographs, of which many technical modalities are available. The disease is visible under the form of a localized change in the coloration of the fundus. Indeed, the affected layer is called the retinal pigment epithelium, which contains a pigment, melanin. Hence the disappearance of this layer causes a change in the coloration of the fundus, which can be observed most specifically in fundus imaging using long wavelengths such as red or infrared (IR) light (Fig. 1). This is why infrared imaging of the fundus through confocal scanning laser ophthalmoscopy (cSLO) is being widely used. This modality is comfortable for the patient, more robust and less invasive than fundus autofluorescence (FAF) imaging; it has higher resolution and higher contrast between normal and diseased areas than color imaging, an older technology. In this work, we process cSLO images in IR.



(a) (b) (c)

Fig. 1. Examples of cSLO fundus images, acquired in infrared from three patients with dry ARMD. Each column shows two consecutive images extracted from the series. The arrows point to the GA (green arrows to point to new GA zones, blue arrows otherwise).

Fig. 1 shows three pairs of consecutive images, taken at an interval of 6 months. The GA appears as brighter regions in the macula and around the optical disk. Automatic processing to follow up these areas is obviously very challenging given the quality of the images: uneven illumination inside images (a,b,c), even saturation to black or white, illumination distortion between images (a,b,c), GA poorly contrasted (a) with retinal structures interfering (vessel, optical disk), blur, etc. The difficulty relies also in the high variability of the lesions in shape, size and number. The lesion boundary is quite smooth in some cases (b) and very irregular in others (a). At any time, new spots can appear (c) while older lesions can merge. All these features make the interpretation, in terms of GA delineation and growth, very difficult. On one hand, the segmentation task of the GA in every image is very challenging, especially long and tedious to perform manually, and even experts cannot always be sure of their manual delineation. On the other hand, differential analysis between images cannot be applied without first compensating for illumination distortion.

So, our objective is to automatically process series of eye fundus images acquired in infrared at regular time intervals, in order to detect the appearance of new atrophic areas and quantify the GA growth. The ultimate goal is to better understand the lesion evolution mechanisms and to propose predictive models of the disease progress. This article deals with the first step of this project, image preprocessing: the normalization of intensities between the images, which is a prerequisite to perform differential analysis, and the attenuation of artefacts thanks to morphological processing (Section 4). Experiments (Section 5) on a set of 18 series (328 images manually segmented by a medical expert, Section 3) enable us to evaluate the benefits of the proposed processing, in terms of homogeneity between images and in terms of contrast between healthy and unhealthy areas. Preliminary segmentation results of the GA growth are also presented, demonstrating the interest of the proposed approach. All these results are discussed in Section 6.

2. State of the art

Automatic analysis of fundus images with dry ARMD has been an important research field for two decades, for diagnosis [11] or follow up [6] purposes. Ophthalmologists can observe pathologic features such as Drusen that occur in early stages of the ARMD, and GA progression at different stages of degeneration. Imaging modalities are most often color eye fundus images, fundus autofluorescence (FAF) images, confocal scanning laser ophthalmoscopy (cSLO) in infrared (IR), but also optical coherence tomography (OCT). This review includes all modalities of eye fundus images (color, IR, FAF).

In the literature, most works deal with segmentation of single images. Standard processing methods are applied, such as region growing [6], region oriented variational methods with level set implementation [3][8], texture analysis [7]. To cope with GA variability, Köse proposed to first segment all healthy regions to get the GA as the remaining areas [7]. This approach requires segmenting separately the blood vessels, which is known to be also a difficult task. The method involves many steps and parameters that can be monitored by an user. Besides, several researchers acknowledged the difficulty to design fully automatic tools reaching the required level of performance and they developed user-guided segmentation frameworks [1][9]. Machine learning has also been experimented, with random forest [2] or k-nearest neighbor classifiers [4]. Feature vectors include image intensity, local energy, texture descriptors, values derived from multi-scale analysis and distance to the image center. Nevertheless, these algorithms are supervised, they require training the classifier from annotated data, which brings us back to the difficulty of manually segmenting GA areas. On the contrary, Ramsey's algorithm is unsupervised, based on fuzzy c-means clustering [12]. However, if the method reaches good performances for FAF images, it performs less well for color fundus photographs.

The difference between the segmentations obtained in consecutive images gives the GA growth. However, the previous study shows the high difficulty in segmenting accurately the GA without training or user intervention. That is why a few works propose another approach, based on differential analysis, to catch the GA growth between two images. Troglio proposed an automatic algorithm to detect changes between two color fundus images, based on Kittler and Illingworth (K&I) method [13]. This method is unsupervised and allows to label pixels in two classes, "change" and "no change" in a Bayesian framework. The non-uniform illumination is corrected in each image using a homomorphic filtering technic, where the image intensity is modeled as a product of a luminance component and a reflectance component. The images are not corrected with respect to each other. Residual illumination variation are taken into account by applying the K&I algorithm in sub-windows and adopting a multiple classifier voting approach. The lack of joint processing induces extra complexity and computational cost. In contrast, Marrugo proposed to correct the images by pairs, by multiplying the second image by a polynomial surface of degree 4, whose parameters are estimated in the least-squares sense [10]. In this way, illumination distortion is compensated and the image difference enhances the areas of changes, which are segmented via a statistical test applied locally. This gives a mask of the change-free areas, which are used to estimate the point spread function (PSF) in each image and apply deconvolution. The illumination compensation method is limited to pairs of images and cannot process all of the images in a series jointly. On the contrary, the blind deconvolution can be extended and is very interesting to compensate for blur present in some eye fundus images.

This review shows that the issues mentioned in the introduction are not yet solved. To the best of our knowledge, the differential approach has not yet been fully explored. Especially, images have been processed in pairs and the proposed algorithms do not deal with large series. However, the differential analysis seems more promising than the segmentation approach, since it may enable us to better handle the high shape variability of the GA areas and the interference with retinal structures, as well as the lack of contrast. This requires designing an efficient illumination normalization algorithm and/or a differential algorithm robust to non-structural changes.

This article focuses on the joint processing of the whole series, to obtain comparable intensities in the whole stack of images and attenuate illumination artefacts (Section 4). This result is in itself interesting since it facilitates the visual interpretation by the medical expert. We also present very preliminary results regarding the detection of the GA growth and many experiments assessing the benefit of the proposed image processing (Sections 5, 6). But before, we describe our dataset and the manually segmented subset used for the evaluation (Section 3).

3. Dataset

Our images were all acquired at the Quinze-Vingts National Ophthalmology Hospital in Paris, in cSLO with IR illumination. The study was carried out according to the tenets of the Declaration of Helsinki and followed international ethical requirements. Informed consent was obtained from all patients. These patients have been followed-up during a few years, hence we have series of retinal fundus images, often for both eyes, showing the progression of the GA. The images were acquired at a time interval of about six months. The average number of images in one series is 13 with a standard deviation of 12. All pictures are in grayscale and vary greatly in size, but the most common size is 650 x 650 pixels. As mentioned previously, we notice many imperfections such as blur, artifacts and, above all, non-uniform illumination inside the images and between them (Fig. 1). All images were spatially aligned with i2k software. In every image, the area of useful data does not cover the entire image and is surrounded by black borders. The automatic detection of these black zones in each image gives a mask of the useful data, and the intersection of all masks the common retinal region where changes can be searched for.

An ophthalmologist segmented manually a sub-database of eighteen series, for the quantitative evaluation of the proposed processing methods (Table 1). These series feature different characteristics in terms of disease progress, lesion shape, size, image quality. For 3 patients, we have series for both left and right eyes. The number of images per series ranges from 5 to 52 (18.2±15.9 in average), with a total of 328 segmented images. We have images at the very beginning of the disease for 5 series. In this case, there is no apparent GA in the first image(s).

P. ID	Eye	N	P. ID	Eye	N	P. ID	Eye	N
001	R	24	008	L	47	089	R	10
001	L	15	010	R	6	109	R	16
003	R	5	016	L	31	110	L	11
005	R	9	018	R	9	112	R	8
005	L	9	019	R	11	115	R	9
008	R	50	020	L	52	117	L	6

Table 1.Sub-database used in the quantitative evaluation Patient ID, processed eye, left (L) or right (R), number of images in the series (N). Background in light gray when the acquisitions start with no GA (5 series).

We developed several user-guided segmentation tools to make the ground truth, based on classical segmentation algorithms: thresholding applied locally on a rectangle defined by the user, parametric active contour model initialized by the user, simple linear interpolation between points entered by the user. The user chooses the most appropriate tool to locally delimit the lesion border, and thus progresses step by step. Automatic thresholding or active contour algorithm initialized by the user leads to segmentations that depend less on the user than the only use of interpolation, and the expert was encouraged to apply these tools as often as possible. However, the segmentation remains mostly manual, user-dependent and tedious. An ophthalmologist realized all segmentations used in our experiments. This has been a very hard and long task: as previously mentioned, GA boundaries may be very indented; moreover, there is very often a strong ambiguity on the true position of the GA border and on the appearance or not of new injured areas. Consequently, the expert had often to go through the images to make or review a decision. On average, it took 10 minutes to process one single image.

4. Methods

Let us denote by $X_n, n \in [1, N]$ the series of acquired images, to be processed. The grayscale pixels are coded by floating point numbers in the interval $[0,1]$. Our method is made of two main steps: an image normalization to compensate for illumination variation between images, and morphological processing to enhance the consistent changes between images and attenuate artefacts.

4.1. Normalization step

We consider the model where each acquired image $X_n, n \in [1, N]$, is the pixel-wise product of an image of reflectance R_n and an image of illumination I_n:

$$X_n = R_n I_n, n \in [1, N] \quad (1)$$

Our goal is to compensate for illumination variation (i.e. the fact that $I_n \neq I_m$ for $n \neq m$) to get homogeneous illumination between all processed images. In other words, we aim at calculating from X_n a new series of images, $Y_n = R_n J_n$, where J_n is the corrected illumination of the n^{th} image, with $J_n = J_m$ for all pairs. Note that the illumination can remain uneven over the image domain, but this inhomogeneity must be the same in all images of the series to allow differential analysis.

The illumination variation is generally smooth, and we can assume that the pixel-wise ratios I_{n+1}/I_n only contain low frequencies. As there are only few and only small structural changes between two consecutive images, we can also consider that the lowpass filtered version of the ratio X_{n+1}/X_n is a good estimate of I_{n+1}/I_n. So we calculate

$$U_n = \widehat{\left(\frac{I_{n+1}}{I_n}\right)} = \frac{X_{n+1}}{X_n} * f_\sigma, n \in [1, N-1] \qquad (2)$$

where f_σ is a Gaussian kernel with standard deviation σ and * the convolution operator. σ is a scale parameter, chosen to remove structural change areas between consecutive images. Then we estimate the ratio of illumination for any pair of images in a recursive way:

$$\begin{gathered} \forall (n,m) \in [1,N]^2, m \geq n, \\ V_{m,n} = \widehat{\left(\frac{I_m}{I_n}\right)} = U_{m-1} V_{m-1,n} \\ V_{n,m} = \frac{1}{V_{m,n}} \end{gathered} \qquad (3)$$

So, the lowpass filter f_σ is only applied on ratios of consecutive images, so with little change from one to the other, and not on any other pairs where the reflectance is likely to be very different, with large areas of change. Our goal is to apply an algorithm that corrects the images, so that the new illumination ratio becomes close to 1 at each pixel, and that for any pair (m,n). To achieve that, our preliminary idea was to propose an iterative algorithm that modifies each image against the $(N-1)$ others:

$$\forall (m,n), m > n, \begin{cases} Y_n^{(i+1)} = Y_n^{(i)} V_{m,n}^{(i)\frac{\alpha}{2}} \\ Y_m^{(i+1)} = Y_m^{(i)} V_{m,n}^{(i)-\frac{\alpha}{2}} \end{cases} \qquad (4)$$

where i denotes the iteration, $Y_n^{(0)} = X_n$, $V_{m,n}^{(i)}$ is the updated estimate of the illumination ratio.

Let us assume that the illumination at a given pixel (x,y) is higher in $Y_m^{(i)}$ than in $Y_n^{(i)}$. Then its illumination in the n^{th} image will be multiply by a factor $V_{m,n}^{(i)\frac{\alpha}{2}}(x,y) > 1$ while its illumination in the m^{th} image will be divided by the same factor, resulting that the new illumination values in $Y_n^{(i+1)}$ and $Y_m^{(i+1)}$ become closer. Processing all pairs in this way enables us to correct every image against all the others. The parameter α controls the amount of correction brought by each image of the series to process another one. Thus, we can assume that the illumination distortion between images progressively decreases so that, after L iterations, we get a new series $Y_n = Y_n^{(L)}$ with same illumination in all images: $I_m^{(L)} = I_n^{(L)}$ for all pairs (n,m). Note that the algorithm in equation (4) is equivalent to

$$\forall n, Y_n^{(i+1)} = Y_n^{(i)} \prod_{k=1}^{N} V_{k,n}^{(i)\frac{\alpha}{2}} \qquad (5)$$

Let us now go farther in the interpretation by studying the convergence of the proposed algorithm. For that, we consider the particular case where all images have the same reflectance but were acquired under different illuminations. For any pair, we have $V_{k,n}^{(i)} = I_k^{(i)}/I_n^{(i)} = Y_k^{(i)}/Y_n^{(i)}$ and the correction algorithm becomes

$$\forall n, Y_n^{(i+1)} = Y_n^{(i)} \prod_{k=1}^{N} \left(\frac{Y_k^{(i)}}{Y_n^{(i)}}\right)^{\frac{\alpha}{2}}$$

We denote by $Z_n^{(i)}$ the logarithm of $Y_n^{(i)}$. We have

$$\forall n, Z_n^{(i+1)} = Z_n^{(i)} + \frac{\alpha}{2} \sum_{k=1}^{N} \left(Z_k^{(i)} - Z_n^{(i)}\right) \qquad (6)$$

At iteration i, we define the difference between any two images:

$$\Delta_{n,m}^{(i)} = Z_n^{(i)} - Z_m^{(i)}$$

At the next iteration, we have

$$\Delta_{n,m}^{(i+1)} = Z_n^{(i+1)} - Z_m^{(i+1)} = \left(1 - \frac{\alpha N}{2}\right) \Delta_{n,m}^{(i)}$$

For $0 < \alpha < \frac{2}{N}$ we have $\lim_{i \to \infty} \Delta_{n,m}^{(i)} = 0$ for any pair of images, meaning that the algorithm converges to a common image. From equation (6) we notice that the mean value of all images is constant and equal to $\frac{1}{N} \sum_{n=1}^{N} \log X_n$. Consequently the final corrected image, after convergence, is the geometric average of all input images:

$$\forall n, \lim_{i \to \infty} Y_n^{(i)} = \left(\prod_{k=1}^{N} X_k\right)^{\frac{1}{N}}$$

Thus, our final corrected images, all equal to the geometric average of the input images, are given by

$$Y_n = X_n F_n \quad \text{with } F_n = \frac{\left(\prod_{k=1}^{N} X_k\right)^{\frac{1}{N}}}{X_n} = \prod_{k=1}^{N} \left(\frac{I_k}{I_n}\right)^{\frac{1}{N}} \qquad (7)$$

As we are not in this ideal case where all images have the same reflectance, our strategy will be to estimate all illumination ratios as described by equations (2) and (3), and to calculate in just one pass the corrected images:

$$Y_n = X_n F_n \text{ with } F_n = \prod_{k=1}^{N} V_{k,n}^{\frac{1}{N}} \qquad (8)$$

Finally, our correction algorithm is as follows:

Input : series of images $X_n, n \epsilon [1, N]$

- Calculate U_n for $n \epsilon [1, N-1]$ (2)
- Calculate $V_{k,n}$ for all pairs (k,n) (3)
- Calculate new images Y_n (8)

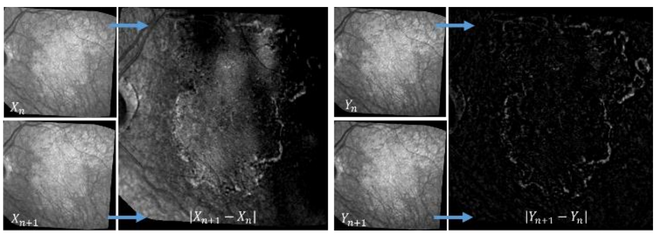


Fig. 2. Joint normalization applied on series 005-L. On the left, two consecutive source images X_n.and X_{n+1} and their difference $|X_{n+1} - X_n|$. On the right, the same after normalization.

The low-pass filtering in equations (2) and (3) gives only an approximation of the illumination ratio between two images. Therefore, considering all pairs (m,n) of images, even when the two medical exams were done at a large time interval, with important reflectance changes, is probably not appropriate, even with our recursive approach. For that reason, we propose to implement the algorithm with a sliding temporal window: each image is corrected considering $2K + 1$ images, K images acquired before and K images acquired after the processed image. The overlap ensures that all images of the series will converge to the same illumination. Experiments showed that $2 \leq K \leq 7$ (5 to 15 images in the window) are good values for our dataset (Section 5.2). Fig. 2 shows an example of two consecutive images extracted from a series of 9 images. We observe that the initial absolute difference does not provide meaningful information, while it reveals the GA growth after normalization.

The proposed algorithm corrects each image with respect to all the others. However, the new illumination generally remains spatially uneven. It is certainly possible to compensate for this distortion when the series begins without any GA, thanks to classical background subtraction technics. This has not been explored yet.

4.2. Enhancement step

The second step of our algorithm exploits prior knowledge about the studied disease: we know that the GA, which is brighter than the surrounding structures, growths continuously, without any chance of recovery. It means that the GA in image Y_n is included in the GA in $Y_{m>n}$. Therefore, we apply morphological reconstructions to all triplets $Y_{n-1}, Y_{n,}, Y_{n+1}$ as follows:

$$Z_n \; = \left(Y_n \;\; R_D^{Y_n}\left(Y_{n-1}\right) R_E^{Y_n}\left(Y_{n+1}\right)\right)^{\frac{1}{3}} \tag{9}$$

where $R_D^M(S)$ (resp. $R_E^M(S)$) denotes the reconstruction by dilation (resp. erosion) of the marker S in the mask M. The first reconstruction ensures that we keep only bright areas that were already present in the previous image. The second one suppresses bright areas that are not present in the next image. The pixel-wise product to the power 1/3 merges all these conditions and enhances the meaningful bright areas in image n. The normalizing algorithm is applied again on the images $Z_{n,}$, to ensure that the new series is properly normalized and to further reduce the influence of artefacts.

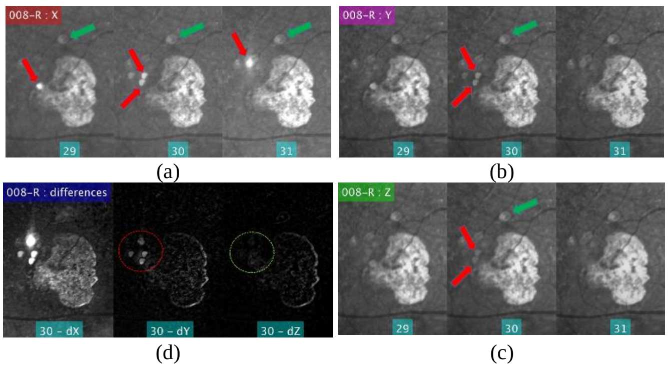


Fig. 3. Normalization and image enhancement. (a) Three consecutive images $X_{n-1}, X_n, X_{n+1}, n = 30$; the corresponding normalized images before (b) and after (c) morphological processing. The green arrows point to a part of the actual GA and the red arrows to bright artefacts; (d) corresponding difference images, emphasized by a factor 3, before normalization (left), after normalization (middle), after normalization and morphological processing (right). The normalization reveals the GA growth but light artefacts are also enhanced (inside the red circle); the morphological processing reduces the intensity of artefacts (green circle).

Fig. 3 shows an example of the whole processing on 3 consecutive images. The difference image is calculated by $max\{|X_{n+1} - X_n|, |X_n - X_{n-1}|\}$. The first step (normalization) enables us to enhance the differences between the images, while the second step (morphological processing) attenuates the bright artefacts that are not new lesions.

5. Results

We rely on the manual segmentations performed by our expert (18 series, 328 images in total) to evaluate the proposed methods, in two different ways: though the estimation of the contrast between the GA and the background (i.e. surrounding tissues and anatomical structures) and through the evaluation of preliminary segmentation results based on automatic thresholding.

5.1. Ground truth

Let us denote by GT_n the binary image of the GA segmentation made by the medical expert in the n^{th} image of a series. The inside of the GA delineated by the expert is set to 1 and the outside to 0. To deal with possible errors and inconsistencies (such as $GT_n \not\subset GT_{m>n}$), we estimate the inside $L_n^{(in)}$ of the GA in image n and the outside $L_n^{(out)}$ by:

$$\begin{aligned} L_n^{(in)} &= \cup_{k=1}^{n} GT_k \\ L_n^{(out)} &= \overline{L_n^{(in)}} \end{aligned} \tag{10}$$

where $\bar{E}$ denotes the complement of the set E. This choice ensures that $L_n^{(in)} \subseteq L_{n+1}^{(in)}$.

To deal with series where a GA is already present in the first image (13 series over 18), we define the total growth of the GA from the first image to the n^{th} by:

$$GTh_n = L_n^{(in)} \cap \overline{\cap_{k=1}^{N} GT_k} \tag{11}$$

Finally, the growth between two images is given by

$$gth_{n:m>n} = GTh_m \cap \overline{GTh_n} \tag{12}$$

These three equations gives us the ground truth for the GA segmentation (10), the growth from the first image to the n^{th} (11) and the growth between two images (12). We rely on these definitions to evaluate our methods.

We also estimate the contrast between two regions, denoted by R_1 and R_2, in a given image X, by calculating the mean signal power $P(X,R)$ in these two regions, and then the difference $\Delta P(X, R_1, R_2)$ in dB:

$$P(X,R) = 10 \log_{10}\left[\frac{1}{|R|}\sum_{(x,y)\in \mathrm{R}}\left(X(x,y)\right)^2\right] \tag{13}$$

where $|R|$ denotes the number of pixels in the region R.

$$\Delta P(X, R_1, R_2) = P(X, R_1) - P(X, R_2) \tag{14}$$

The higher $|\Delta P(X, R_1, R_2)|$, the higher the mean contrast in X between the two regions R_1 and R_2 and the easier the two regions can be segmented. This measure will help us to evaluate the benefit of our algorithms by comparing the contrast between healthy and unhealthy areas in images derived from X_n (source images), Y_n (after normalization) and Z_n (after morphological processing).

5.2. Evaluation from variance images calculated along the temporal axis

Our algorithm aims at correcting illumination variations between images. As a result, the illumination may be still uneven in every image, but it should be the same in all images of the series. So, the gray level of a pixel belonging to an healthy region (outside the GA in all images) must be stable in the series (low variance), contrary to a pixel belonging to a region of GA growth whose gray level increases significantly at a given time. Consequently, we can evaluate the benefit of our normalization algorithm by estimating the contrast between these two regions in a variance image.

We consider a sliding temporal window of $M = 2m + 1$ consecutive images (Fig. 4). For each image $n \in [m + 1, N - m]$ we take the images X_{n-m} to X_{n+m} to calculate the variance of the gray levels at each pixel. We denote by $X_n^{(VAR)}$ the resulting image. We consider the heathy region $R_n^{(healthy)} = L_{n+m}^{(out)}$ and the region of growth $R_n^{(growth)} = gth_{n-m:n+m}$, and we evaluate the contrast between both: $\Delta P\left(X_n^{(VAR)}, R_n^{(growth)}, R_n^{(healthy)}\right)$ (13)(14). We average the results in the series and over all series to get mean values, denoted by $\Delta P(X^{VAR})$.

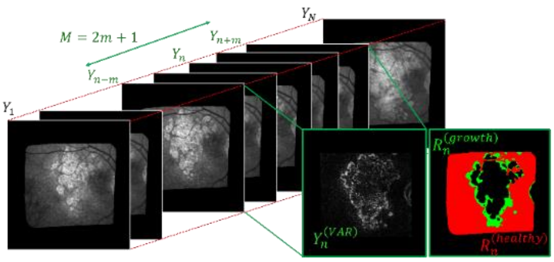
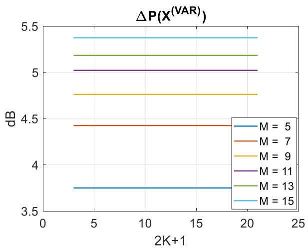
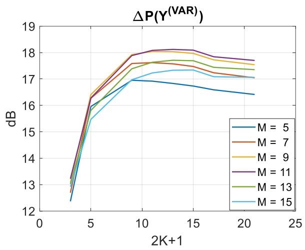


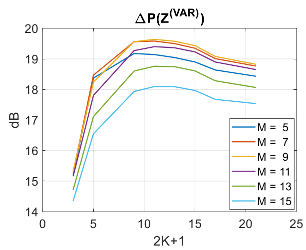
Fig. 4. Variance images calculated over $M = 2m + 1$ consecutive images. Mean power values are interpreted by considering the GA growth over this temporal window ($R_n^{(growth)}$) and the still healthy area ($R_n^{(healthy)}$).



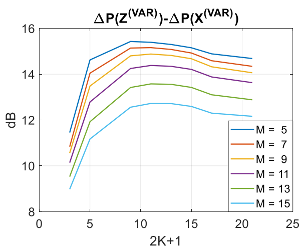
(a) Source images



(b) After normalization



(c) After the whole processing



(d) Gain (whole processing)

Fig. 5. Contrast enhancement between the GA growth area and the healthy area for several parameters K (temporal window in the normalization algorithm) and M (temporal window to compute the variance).

Fig. 5 displays the results for different sizes of sliding window in the normalization algorithm ($2K + 1$, Section 4.1), different sizes of sliding window to compute the variance images ($M = 2m + 1 \in \{5,9,11,13,15\}$), and at different steps of the algorithm: on the source images (X_n), after normalization (Y_n) and after morphological processing (Z_n). We consider only the 6 series having more than 15 images (maximum value for M) with GA, for the sake of comparison. These results demonstrate the benefits of our algorithm with great improvements of the contrast whatever the parameter K and the length of the analysis (M): the gain ranges from 11 to 15dB (Fig. 5d) after the whole processing. Most of the gain results from the normalization algorithm but the morphological processing contributes also to enhance the GA growth area (about +1.5dB). Setting $2K + 1 = 11$ leads to the highest contrasts (b,c) whatever M. But this experiment shows also that K is not a sensitive parameter since the performances remain similar for $(2K + 1) \in [5,21]$.

5.3. Evaluation from an enhanced image of the GA total growth

Let us define the image $G_n^{(X)}$ enhancing the area of growth between the first and the n^{th} image:

$$G_n^{(X)} = \max_{k \in [2,n], l < k} (X_k - X_l), n \in [2, N] \tag{15}$$

Note that maximizing over all differences, $(X_k - X_l)$ with $l < k$, makes sense only because all images have been jointly normalized. We compare the signal power in the total growth area GTh_N (11) to the power outside the GA, $L_N^{(out)}$ (10), by calculating $\Delta P\left(G_N^{(X)}, GTh_N, L_N^{(out)}\right)$ (13)(14) at the different steps of the algorithm. Table 2 summarizes the results averaged on the 18 series of the experimentation sub-database. The main gain obviously results from the normalization step, but the morphological treatment improves generally slightly the contrast, (+0.5dB). The results are again stable over K, with slightly better contrasts for $2K + 1 = 9$.

$\Delta P(G_N^{(X)})$	K	$\Delta P(G_N^{(Y)})$	Gain	$\Delta P(G_N^{(Z)})$	Gain
2.76 ± 3.47	2	8.44 ± 2.42	5.68 ± 2.44	9.52 ± 2.10	6.76 ± 3.03
	4	$\mathbf{9.25 \pm 2.13}$	$\mathbf{6.49 \pm 3.67}$	$\mathbf{9.70 \pm 1.76}$	$\mathbf{6.93 \pm 3.62}$
	5	9.06 ± 2.06	$6.30 \pm 3; 79$	9.52 ± 1.81	6.75 ± 3.50
	6	8.95 ± 2.08	$6.19 \pm 3; 70$	9.49 ± 1.91	6.73 ± 3.36

Table 2. Contrast between the area of GA growth GTh_N and the outside $L_N^{(out)}$ in the enhanced image (15) calculated from X_n, Y_n or Z_n. The gains are calculated with respect to the initial contrast $\Delta P\left(G_N^{(X)}\right)$. Values are in dB

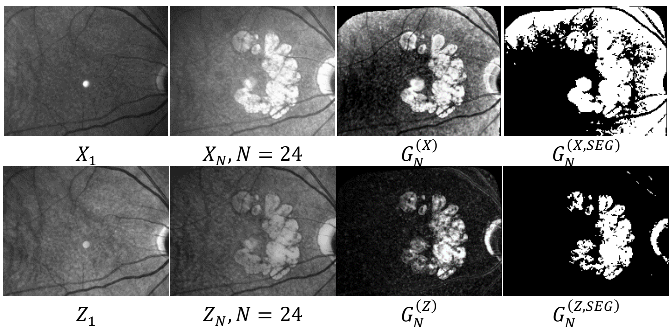


Fig. 6. GA growth enhancement and segmentation, applied on a series of 24 images with no lesion in the first image (Patient 001-R). The first row shows the images before normalization (X_n) and the second row after processing (Z_n). The enhanced GA growth is shown in the third column and the corresponding segmentation in the fourth column (F1 score: 0.77 for $G_N^{(Z,SEG)}$).

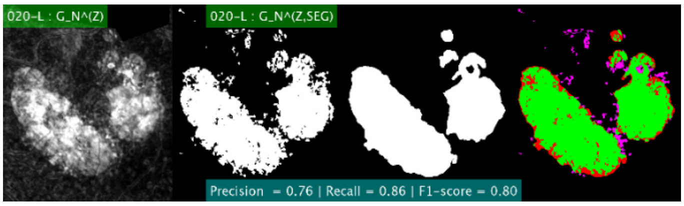
Fig. 6 shows the enhanced GA growth (15) before and after processing. In this example, the set of images starts from the very early stages of the disease to advanced ones and the whole GA is enhanced with a very dark background ($G_N^{(Z)}$) compared to the result without processing ($G_N^{(X)}$). The growth of diseased tissues around the optical disk is also clearly visible.

A simple thresholding with Otsu's method enables us to segment the total GA growth, thanks to the proposed processing, while it was impossible with the source images (Fig. 6). We denote by $G_N^{(X,SEG)}$ the binarized $G_N^{(X)}$ image. The segmentation is evaluated in Table 3 via classical metrics: precision, recall and F1-

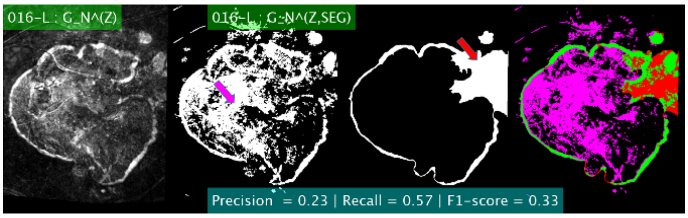
score, which is the harmonic mean of precision and recall, given the ground truth Gth_N (11).

		Lesion (5 series)			**Growth** (13 series)		
	T_{Otsu}	Prec.	Rec.	F1-sc.	Prec.	Rec.	F1-sc.
$G_N^{(X,SEG)}$	0.44 ±0.04	0,29 ±0,10	0,86 ±0,15	0,42 ±0,13	0,27 ±0,23	0,63 ±0,16	0,33 ±0,17
$G_N^{(Y,SEG)}$	0.33 ±0.02	0,82 ±0,15	0,75 ±0,15	0,77 ±0,10	0,50 ±0,27	**0,68 ±0,12**	0,53 ±0,20
$G_N^{(Z,SEG)}$	0,37 ±0,02	**0,86 ±0,09**	**0,76 ±0,14**	**0,80 ±0,09**	**0,52 ±0,26**	0,65 ±0,12	**0,53 ±0,18**

Table 3. Evaluation of the GA growth segmentation by Otsu's method at the different steps of the algorithm.



(a) A series starting from a very early stage where no GA is observable.



(b) A series starting from an advanced stage. Texture changes inside the lesion are classified as lesion growth areas (↘), resulting in low precision. The recall is low since the expert segmented the right part as lesion in the last image (↘), contrary to the algorithm.

Fig. 7. Segmentation of the GA growth. Left to right: enhanced image $G_N^{(Z)}$, proposed segmentation $G_N^{(Z,SEG)}$, ground truth Gth_N, and comparison (true positives in green, false negatives in red and false positives in magenta).

The best scores are achieved with the whole processing, normalization and morphological treatment. We notice that Otsu's threshold (T_{Otsu}) is remarkably stable over the series (0.37 ± 0.02). The results for the detection of the whole growth are especially good with a F1-score equal to 0.80±0.09 (Table 3, Fig. 6a). For the other series, the F1-score is much lower, equal to 0.53±0.18. The main problem is that the texture changes inside the GA over time, resulting in a lot of false positives and lower precision (Fig. 6b).

5.4. GA progress between consecutive images

Finally, still with the aim of evaluating our image processing algorithms, we tried to segment the GA growth between consecutive images. This process is a prerequisite to elaborate predictive models of the disease. For that, we consider the pixels belonging to the GA growth area, so satisfying $G_N^{(Z,SEG)}(x,y) = 1$ (Section 5.3), and we study their gray levels in the whole series. Let us denote by $V^{(Z,(x,y))}(n), n \in [1,N]$, the gray levels of such a pixel (x,y) in the processed series Z_n. We assume that this pixel is not part of the GA in the first image and that it will be included in the GA at a given time t. This means that the gray levels in $V^{(Z,(x,y))}$ may be low for $1 \leq n < t$ and much higher for $t \leq n \leq N$ and roughly stable in the two states. For each index $k \in [2,N]$, we classify the elements of $V^{(Z,(x,y))}$ in 2 classes: $V^{(Z,(x,y))}(n) \in C_0(k)$ if $n < k$, $V^{(Z,(x,y))}(n) \in C_1(k)$ otherwise. We calculate the inter-class variance according to Otsu's criterion for each index k:

$$\sigma_{Otsu}^{2\,(Z,(x,y))}(k) = \frac{k-1}{N}\left(\overline{V_{1:k-1}^{(Z,(x,y))}} - \overline{V_{1:N}^{(Z,(x,y))}}\right)^2 + \frac{N-k+1}{N}\left(\overline{V_{k:N}^{(Z,(x,y))}} - \overline{V_{1:N}^{(Z,(x,y))}}\right)^2 \quad (16)$$

where $\overline{V_{k:l}^{(Z,(x,y))}}$ denotes the mean value of the elements k to l of vector $V^{(Z,(x,y))}$. We calculate a segmentation map by maximizing the interclass variance $\sigma_{Otsu}^{2\,(Z,(x,y))}$:

$$MAP^{(Z)}(x,y) = \arg\max_k \sigma_{Otsu}^{2\,(Z,(x,y))}(k) \quad (17)$$

This method is not relevant for pixels that do not switch from an heathy to an unhealthy state. Consequently the method is only applied in the GA growth area, so for pixels satisfying $G_N^{(Z,SEG)}(x,y) = 1$ (Section 5.3). So we calculate the series $SEG_n^{(Z)}, n \in [1,N]$, of binary images of the GA from image 1 to image N by

$$SEG_n^{(Z)}(x,y) = \begin{cases} 1 \text{ if } n \geq MAP^{(Z)}(x,y) \text{ and } G_N^{(Z,SEG)}(x,y) = 1 \\ 0 \text{ otherwise} \end{cases} \quad (18)$$

Finally we calculate the growth between two consecutive images by

$$SEG_{n-1:n}^{(Z)} = SEG_n^{(Z)} \,\&\, \overline{SEG_{n-1}^{(Z)}} \quad (19)$$

which reveals the progression of the GA between the $(n-1)^{\text{th}}$ and the n^{th} images.

Table 4 shows the quantitative evaluation of this segmentation method, applied on the processed series Y_n and Z_n. The method is obviously not applicable before normalization. The automatic segmentation $SEG_{n-1:n}^{(Y,Z)}$ is compared to the manual segmentation $gth_{n-1:n}$ (12) using the classical metrics, which are averaged over the images n and then over the series. Fig. 8 shows samples of growth segmentation between consecutive images ($SEG_{n-1:n}^{(Z)}$, (19)) in one series, and Fig. 9 colored representations of the GA growth ($MAP^{(Z)}$, (17)) for two different series, along with the corresponding ground truth.

	Lesion (5 series)			**Growth** (13 series)		
	Prec.	Rec.	F1-sc.	Prec.	Rec.	F1-sc.
$SEG_{n-1:n}^{(Y)}$ $2K+1=9$	0.21 ±0.12	0.23 ±0.13	0.19 ±0.11	0.20 ±0.12	0.28 ±0.12	0.21 ±0.11
$SEG_{n-1:n}^{(Z)}$ $2K+1=5$	**0.25 ±0.12**	**0.24 ±0.12**	**0.19 ±0.10**	**0.23 ±0.13**	**0.31 ±0.14**	**0.24 ±0.12**

Table 4. Evaluation of the segmentation of the GA progress between consecutive images. The results are given for the series Y_n and Z_n with the optimal value of K in both cases.

Overall, the segmentation images $SEG_{n-1:n}^{(Z)}$ (Fig. 8) look very consistent, and the colored maps (Fig. 9) look similar to their equivalent deduced from the ground truth. However the average metrics are quite low (Table 4). There are several explanations for this. First, it should be remembered that the manual segmentation is very difficult to achieve and is not accurate either. Second, the GA progress between two consecutive images is structurally very thin, which explains that a small difference between the segmented area and the ground truth results immediately in low metrics.

Especially, small shifts in the change detection maps lead to an increase of false positives and false negatives, therefore to weaker metrics, although the automatic GA growth detection is not bad (Fig. 8).

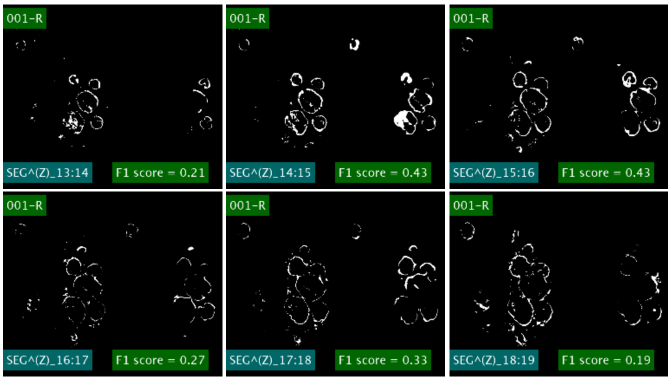


Fig. 8. Examples of very consistent change detections between consecutive images, with F1 scores. Every time, the automatic segmentation $SEG_{n-1:n}^{(Z)}$ (19) is on the left and the ground truth $gth_{n-1:n}$ (12) on the right.

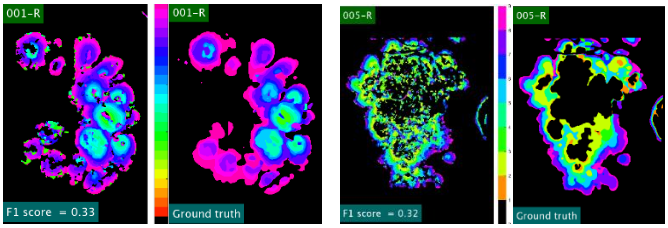
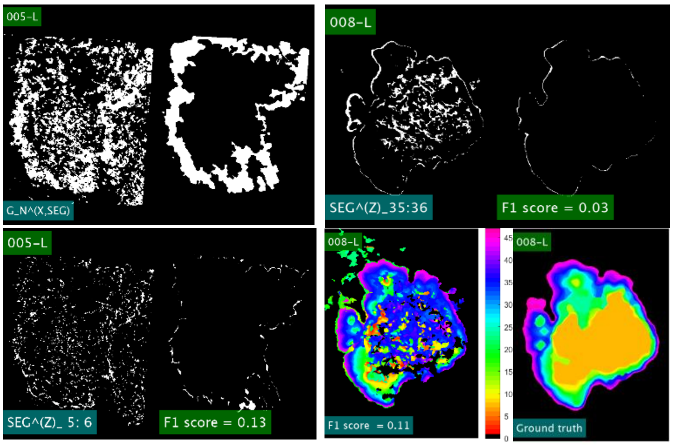


Fig. 9. Colored map of the GA progress in the series (17). The n^{th} color in the colormap corresponds to a GA progress between image n-1 and image n. The automatic result is on the left and the ground truth on the right. The F1 score is the average of all F1 scores obtained for each pair of consecutive images (19).



(a) (b)

Fig. 10. Limits of the segmentation method. In each illustration, the automatic result is on the left and the ground truth on the right. (a) Case of a series of 9 images corresponding to late stages of ARMD. The area of total GA growth $G_N^{(Z,SEG)}$ is over segmented (first row) leading to many false positives in the detection of changes between consecutive images, inside and outside the GA (second row). (b) Case of a series starting from the beginning of the disease. Texture changes over time result in false positives inside the GA, and the F1 score is low despite the good detection of the border growth (top). This results also in differences in the colored maps $MAP^{(Z)}$ (bottom).

Finally, the main limits of the proposed segmentation method come from the texture changes inside the lesion during the growth, disturbing the decision criterion (high variance in the class C_1). In addition, an over-segmentation of the total growth ($G_N^{(Z,SEG)}$) allows false positives inside and outside the GA (18), and this has a strong impact on the precision and on the F1 score, especially for series starting with advanced GA for which we have a low precision in the $G_N^{(Z,SEG)}$ segmentations (0.52 ± 0.26, Table 3). Fig. 10 illustrates these problems. All this explains the rather low metrics presented in Table 4, despite consistent detections of changes. The best results are obtained on average with the normalized images processed by mathematical morphology (Z_n), for $K = 5$, with a F1 score around 0.21 ± 0.12. For the normalized series Y_n, the best scores are for $K = 9$, and are slightly lower than those for the images Z_n. The morphological treatment improves the results by introducing a better consistency along the temporal axis.

However, these first results are already interesting, given that the segmentation method is basic, without any spatial regularization. Despite all previously mentioned problems, the normalization algorithm and the morphological processing make it possible to produce interpretable colored maps, and better results are likely to be obtained with more advanced segmentation algorithms.

6. Discussion

We have presented a set of experiments evaluating the advantages of our normalization algorithm for the direct comparison of gray levels in a series of fundus images, and, ultimately, the detection of the GA growth by differential analysis. We first calculated the variance at each pixel, which must be low in the areas that remain healthy and, if applicable, within the GA present in the first image, and high in the region of GA growth (Section 5.2). We found that our processing increases the contrast between the growth area and the rest of the image by 11 to 15dB (Fig. 5), which is very important. It means that the visual inspection of the series, by scrolling through the images, is greatly facilitated, since illumination distortions no longer disturb the interpretation. The GA expansion becomes much more visible. The main gain results from the illumination normalization, but the morphological treatment attenuates local artefacts that could be mistaken for new GA zones. This experiment demonstrates also the advantage of normalizing on a sliding window and helps us to set its optimal size (around 11 consecutive images).

In addition, the proposed preprocessing enables us to perform a differential analysis, in order to detect the GA growth between any two examinations. Thus, in Section 5.3, we calculated an enhanced image of the total GA growth and we demonstrated that the contrast between the GA growth area and the outside (still heathy region) is around 10dB while it was only around 3dB without treatment (Table 2). The gain enables a segmentation of the GA growth area with a F1-score equal to 0.80 ± 0.09 for series starting from the earliest stages of the disease, while it was only equal to 0.42 ± 0.13 without treatment (Table 3). The segmentation methods is basic, based on Otsu's thresholding, but the results are already good despite this. The segmentation of the GA growth from an already advanced stage is less good, with an average F1-score equal to 0.53 ± 0.18 (0.33 ± 0.17 without treatment). This is due to texture changes inside the lesion during time, resulting in higher values in the difference images, and therefore lower precision. Finally, we explored the possibility of calculating colored maps of the GA progress, in order to obtain a visualization of the growth and speed of growth in a single image (Section 5.4). The proposed method is based on the idea that the grey-levels are stable in each state, healthy (low values) or unhealthy (high value). We faced again the problem of texture variations inside the GA that disturb the detection of the time at which a pixel has switched from a healthy state to an unhealthy one. However, we got nice maps in some cases (Fig. 9), close to the ground truth, and also very good sets of binary images showing the GA growth between two consecutive images. The F1-scores evaluating the segmentation of the GA growth between consecutive images are not high (around 0.21 ± 0.12 on the whole database), but these poor results should be put into perspective, as the structures to be detected are very thin and the manual segmentation prone to inaccuracies. We measured weak metrics in many cases despite the binary images of change looked very consistent (Fig. 8).

All these experiments demonstrate that the proposed processing enables the detection of changes between any two images of a series of fundus images, regardless the number of images, the size and shape of the GA. This is very new compared to the literature, which has focused much more on the segmentation of the GA in single images (Section 2), and much less on change detection in series [10][13]. Contrary to [10], we normalize all images jointly and not just in pairs, and the differential analysis proposed in [13], where the images are preprocessed separately by a homomorphic filter, could benefit from our global normalization.

7. Conclusion and perspectives

Very large databases of patients available through routine imaging systems offer an opportunity to study a vast amount of data on dry ARMD. Especially, deriving models of GA progression from fundus images is of high medical interest, in order to better understand the intimate mechanisms of lesion growth or to assess the benefit of treatments. However, there is a real lack of automatic algorithms to process all these images, and especially to detect and quantify the GA progression between two medical examinations. Many research works deal with the GA segmentation in single images, but there is not yet an automatic and reliable solution. A promising alternative is to carry out a differential analysis between consecutive images, but the first obstacle is the great illumination inhomogeneity inside the images and especially between the images. This article deals with the correction of illumination inhomogeneity between all the images in a series in order to allow the differential analysis between any two images. For that, we have proposed a method which recursively estimates the illumination ratio between any two images in the series and correct the illumination of every image to the geometric mean of all illuminations. Then, we have presented a morphological processing that enhances the consistent changes between consecutive images and attenuate the others, based on the fact that the GA increases continuously. Finally, we have presented many experiments realized on a database of 18 series (328 images) acquired from 15 patients and segmented manually by a medical expert. The experiments show the benefit of the normalization algorithm and of the morphological processing, in terms of contrast between the healthy area and the GA growth area, and in terms of ability to detect cumulative changes in the series or between consecutive images. The experimental results demonstrate that the normalization algorithm makes possible the differential analysis, and that the morphological processing brings some improvements by attenuating light artefacts. Our first attempt to detect the GA growth in consecutive images shows that it is possible to process all the images of the series jointly, since the gray levels are now comparable at each pixel. We got very nice colored maps of GA progression, offering an intuitive understanding for clinicians, in terms of growth and speed of growth. However, a more robust segmentation algorithm, integrating spatial and temporal regularization, is required to obtain more reliable and more accurate maps of the GA growth.

8. References

[1] Deckert, A., Schmitz-Valckenberg, S., Jorzik, J., Bindewald, A., Holz, F. G., & Mansmann, U. (2005). Automated analysis of digital fundus autofluorescence images of geographic atrophy in advanced age-related macular degeneration using confocal scanning laser ophthalmoscopy (cSLO). BMC ophthalmology, 5(1), 8.

[2] Feeny, A. K., Tadarati, M., Freund, D. E., Bressler, N. M., & Burlina, P. (2015). Automated segmentation of geographic atrophy of the retinal epithelium via random forests in AREDS color fundus images. Computers in biology and medicine, 65, 124-136.

[3] Hu, Z., Medioni, G. G., Hernandez, M., Hariri, A., Wu, X., & Sadda, S. R. (2013). Segmentation of the geographic atrophy in spectral-domain optical coherence tomography and fundus autofluorescence images. Investigative ophthalmology & visual science, 54(13), 8375-8383.

[4] Hu, Z., Medioni, G. G., Hernandez, M., & Sadda, S. R. (2015). Automated segmentation of geographic atrophy in fundus autofluorescence images using supervised pixel classification. Journal of Medical Imaging, 2(1), 014501.

[5] I2k Retina Technology: https://www.dualalign.com/retinal/image-registration-montage-software-overview.php

[6] Köse, C., Şevik, U., & Gençalioğlu, O. (2008). Automatic segmentation of age-related macular degeneration in retinal fundus images. Computers in Biology and Medicine, 38(5), 611-619.

[7] Köse, C., Şevik, U., Gençalioğlu, O., İkibaş, C., & Kayıkıçıoğlu, T. (2010). A statistical segmentation method for measuring age-related macular degeneration in retinal fundus images. Journal of medical systems, 34(1), 1-13.

[8] Lee, N., Laine, A. F., & Smith, R. T. (2007, August). A hybrid segmentation approach for geographic atrophy in fundus auto-fluorescence images for diagnosis of age-related macular degeneration. In 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 4965-4968). IEEE.

[9] Lee, N., Smith, R. T., & Laine, A. F. (2008, October). Interactive segmentation for geographic atrophy in retinal fundus images. In 2008 42nd Asilomar Conference on Signals, Systems and Computers (pp. 655-658). IEEE.

[10] Marrugo, A. G., Millan, M. S., Sorel, M., & Sroubek, F. (2011). Retinal image restoration by means of blind deconvolution. Journal of Biomedical Optics, 16(11), 116016.

[11] Priya, R., & Aruna, P. (2011, April). Automated diagnosis of Age-related macular degeneration from color retinal fundus images. In 2011 3rd International Conference on Electronics Computer Technology (Vol. 2, pp. 227-230). IEEE.

[12] Ramsey, D. J., Sunness, J. S., Malviya, P., Applegate, C., Hager, G. D., & Handa, J. T. (2014). Automated image alignment and segmentation to follow progression of geographic atrophy in age-related macular degeneration. Retina, 34(7), 1296-1307.

[13] Troglio, G., Alberti, M., Benediksson, J. A., Moser, G., Serpico, S. B., & Stefánsson, E. (2010, April). Unsupervised change-detection in retinal images by a multiple-classifier approach. In International Workshop on Multiple Classifier Systems (pp. 94-103). Springer, Berlin, Heidelberg.

Conflict of interest statement

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript