Enhancement of Optical Coherence Tomography Images of the Retina by Normalization and Fusion

Florence ROSSANT, Frederic AMIEL, Thomas EA , C. MARAVER MARTINEZ [1], *Michel PÂQUES* [2]

[1] Institut Supérieur d'Electronique de Paris, ISEP, 21 rue d'Assas 75006 Paris, France
[2] Centre d'Investigation Clinique - Centre Hospitalier National d'Ophtalmologie des Quinze-Vingts, 28 rue de Charenton 75012 Paris, France

florence.rossant| frederic.amiel|thomas.ea @isep.fr, michel.paques@gmail.com

Abstract. *This paper describes an image processing method applied to Optical Coherence Tomography (OCT) images of the retina. The aim is to achieve improved OCT images from the fusion of sequential OCT scans obtained at identical retinal locations. The method is based on the normalization of the acquired images and their fusion. As a result, a noise reduction and an image enhancement are reached. Thanks to the resulting improvement in retinal imaging, clinical specialists are able to evaluate more efficiently eyes pathologies and anomalies. This paper presents the proposed method and gives some evaluation results.*

Keywords

Optical coherence tomography, denoising, filtering, correlation, fusion.

1. Introduction

Optical coherence tomography (OCT) is a cross-sectional imaging technique allowing micrometric-scale resolution of retinal structures. It enables reliable demonstration of changes in overall retinal thickness, detection of fluid in and behind the neurosensory retina, and identification of the retinal nerve fiber and photoreceptor layers. It is routinely used for diagnosis of retinal diseases.

The OCT principle is to obtain a histology image of the retina by measuring the reflection intensity of a low coherence infra-red light beam on the retina. The fovea of the eye, region where the analysis is done on, has a length between 0.5 and 1mm. The cut takes place on a dimension of 3 to 10 mm and on 512 points. The best resolution is, therefore, obtained for a length of 3 mm. The depth of 20 μm is measured by the tomograph on 1024 points [1]. The obtained reflection coefficients are displayed as an image (512x1024 pixels) in false colors.

Although these images give an excellent idea about the retina, the excessive presence of noise makes the intervening layers of the neurosensory retina be only vaguely discernible. A primary goal of our study was to obtain a better characterization of the outer retina, that is, the interface between the retinal pigment epithelium and the photoreceptors. With ultra high resolution instruments, based on titanium sapphire lasers [3], further improvements would be possible. However, this instrument is not commercialized yet. This is the reason why the research of a method that increases the signal to noise ratio and enhance the image has become an issue to study [4].

Image fusion generally results in a better quality, since it leads to an increased signal-to-noise ratio. In OCT imaging of the retina, a series of scans ($I_1 \ldots I_N$) are acquired from identical fundus locations. A first pre-processing step is applied to each one, in order to normalize the images. This step is called "alignment" in what follows. It allows a better readability of each image taken separately, and is necessary to achieve the fusion. The fusion step itself is based on an average, but other functions could be considered. The block diagram of the global system is shown in Fig. 1.

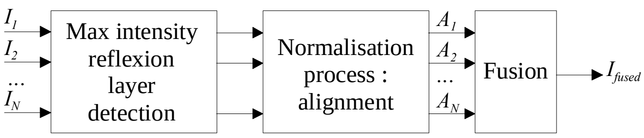


Fig. 1. Global system.

We have explored this idea using data from the conventional StratusOCT instrument [1], [2]. We have worked directly with the reflection coefficients instead of the processed false color image. Indeed, the reflection coefficients provide much more information (about 1500 different values), whereas the given output images are quantized on about 245 different values. So, the information obtained by the StratusOCT tomography, to be processed by our method, is a matrix of reflection coefficients (1024 rows per 512 columns). The developed algorithm treats this matrix as an image, where each value corresponds to a pixel.

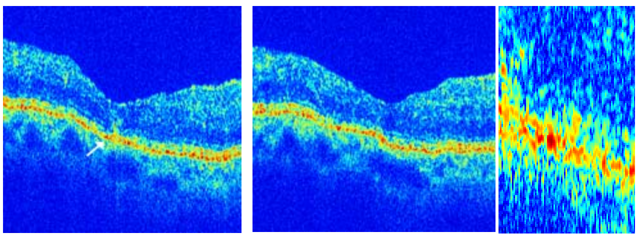
The paper is organized as follows. Sections 2 and 3 present the normalization method, which is based on the detection of the maximal intensity reflection layer (Section

2), followed by an image transformation, that results in the alignment of this layer (Section 3). Section 4 describes the fusion method. Finally, some evaluation results are analyzed in Section 5. The last section presents a conclusion and introduces future work.

2. Maximum Intensity Reflection Layer Detection

2.1 Region of Interest

Fig. 2 shows two scans of retinal OCT images, I_1 and I_2, captured consecutively from the same patient. The images are displayed in false colors. They have been also cropped and resized for a better readability of the figure. Our coordinate system is defined by the origin at the upper left corner of the image, the vertical x-axis and the horizontal y-axis. It is obviously not possible to fuse directly these images: some distortions can be observed between both, since they do not exactly represent the same retina cut (because of ocular movements). So, an essential preprocessing step consists in detecting a common region of interest, the maximal intensity reflection layer (Fig. 2(a)), and aligning all images, so that the pixels belonging to this layer are located at the same vertical position (Fig. 5(b)). Thus, the retinal structure will be at the same position in all images, and the fusion can be achieved. Another important goal of this transformation is that it allows a better clinical interpretation of the retina OCT images.



(a) **(b)** **(c)**

Fig. 2. (a)(b) Two scans taken from the same patient (I_1, I_2), (c) a zoom of figure (a). The maximal intensity reflection layer (in red) is the common region of interest used as reference to achieve the normalization.

The major difficulty for this detection is the high noise presence (Fig. 2(c)). The dominating noise source in OCT images is usually speckle noise arising from interference between coherent waves backscattered from nearby scatters in the measuring volume of the retina. Some of the classic methods used for denoising are averaging each pixel with its neighbors (spatial low-pass filtering), applying a median filtering or a low pass filtering in the frequency domain [5]. These three techniques are used in different steps all along the method that is described below.

2.2 Detection Method

The aim is to find an internal line on the retinal layer. In the OCT image, this layer corresponds to a roughly horizontal set of pixels taking the highest values (displayed in red). Because of the low signal-to-noise ratio, such simple methods like looking for the maximal pixel value column by column, do not work. In addition, known algorithms about boundary detection, as for example active contours [6], do not work properly either, as the noisy pixels block the evolution of the contour.

Consequently, more complex treatments are proposed. Firstly, a pixel belonging to the region of interest, called "internal point", is found with certainty. This pixel is then used to initialize an iterative algorithm that allows to deduce the median line of the maximal intensity layer. This algorithm uses correlation information between adjacent pixel columns, in order to reach a more robust and representative result. It is also applied on images that have been beforehand low-pass filtered, in order to reduce the noise. The block diagram of the detection method, applied to each source image I_n ($1 \le n \le N$), is shown in Fig. 3.

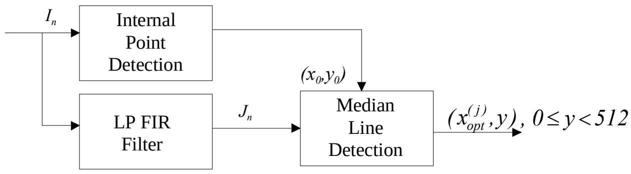


Fig. 3. Maximal reflection intensity layer detection method.

Searching for an internal point: The image I_n is first smoothed, using a large convolution mask (10*40 coefficients equal to 1/400). The pixel reaching the highest value, at the (x_0,y_0) coordinates, is certain to belong to the maximal intensity reflection layer, and consequently, it is taken as the internal point. In our experimentations, it is proved that a rectangular mask leads to better results, since it is more representative of the globally horizontal structure we want to detect.

Low-pass filtering: The initial image I_n is filtered in the frequency domain, using a FIR filter of order 16, with a normalized cut-off frequency equal to 0.05. As a result, the signal to noise ratio is increased while the boundaries of the region of interest are not significantly delocalized. The output image, denoted by J_n, is passed to the median line detection algorithm.

Median line detection: Starting from the detected internal point (x_0,y_0), the algorithm deduces iteratively, column by column, the median line of the maximal intensity reflection layer. Let consider the next column at the horizontal coordinate y=y_0+1. The pixels of this column are low-pass filtered, using a convolution mask of K = 21 coefficients, all equal to 1/K:

$$S(x,y) = \frac{1}{K} \sum_{k=-\lfloor K/2 \rfloor}^{\lfloor K/2 \rfloor} J_n(x+k,y). \tag{1}$$

The highest outputs of this filter are located inside the maximal intensity reflection layer. In order to avoid false detections and smooth the searched median line, a recursive low-pass filter is also applied to the $S(x,y)$ coefficients, along the y coordinate. Let denote by $C(x,y)$ the output of this second filter (Eq. 2). The decision at column y is taken by retaining the $x^{(y)}_{\mathrm{opt}}$ vertical coordinate corresponding to the maximum output value (Eq. 3):

$$C(x,y) = (1-\alpha)S(x,y) + \alpha C(x,y-1)\,, \tag{2}$$

$$\max_x \{C(x,y)\} = C(x^{(y)}_{opt}, y)\,. \tag{3}$$

In this method, the $S(x,y)$ coefficients are continuously integrated to provide the decision at the column y. The parameter α of this recursive filter expresses the relative importance between the current local results, and the previous filter outputs. In our experiments, we use α=0.8. This value results from a compromise: with a greater value, the median line is not accurately followed; with a smaller value, the algorithm is too sensitive to bright noise pixels.

The same algorithm is applied in the other direction, for decreasing y coordinates, starting also from the first detected internal point (x_0,y_0). The result obtained after applying the method to the image I_1 (Fig. 2(a)) is presented below, superimposed on the filtered image J_1. The detected median line (in dark blue) follows accurately the maximal intensity reflection layer.

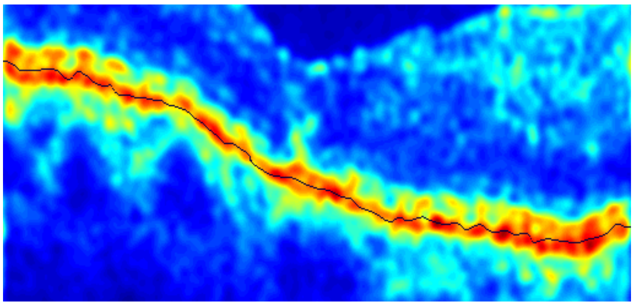
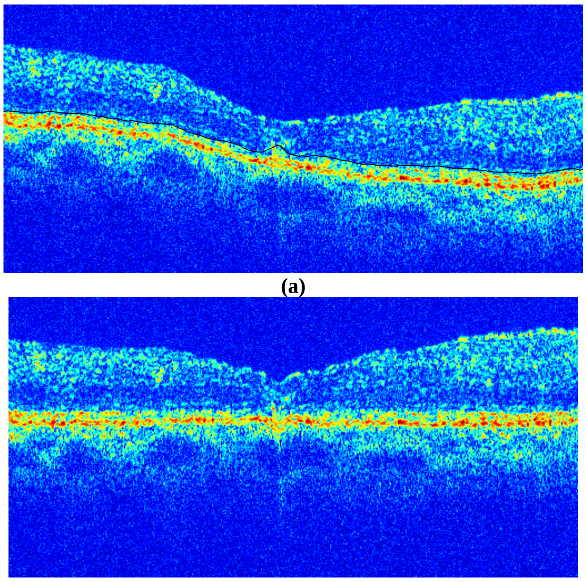


Fig. 4. Maximal intensity reflection layer detection. The dark line represents the median position of this layer.

3. Alignment

The aim of the alignment is to normalize the different source images, so that they can be fused, and also to improve their readability for the clinical diagnostic. The alignment is not carried out against the median internal line found by the method previously described, but against the superior boundary of the maximum intensity reflection layer. This boundary can be deduced through a region growing algorithm, whose seed is the internal median line. In order to suppress the noise, a median filter is applied beforehand column per column. Its size is equal to the typical width of the maximum intensity reflection layer. The region growing stops when a pixel value is below a threshold T that is dynamically set to the half of the mean pixel value M of the region of interest ($T = 0.5M$). M is estimated by averaging the pixels on a small area around the median line. Fig. 5(a) shows the superior boundary found with this method.



(a)

(b)

Fig. 5. Detection of the superior boundary of the maximum reflection intensity layer (a), and alignment against this line (b).

Finally, the alignment is carried out. It consists on a column by column simple shift (Fig. 5(b)). The resulting images are denoted by A_n in what follows.

4. Fusion

The aim of this section is to obtain an enhanced image in order to satisfy the medical interest. The idea is to generate from several aligned images a new one, with a higher signal-to-noise ratio.

Before fusing, the correct superposition between the images has to be found. Translation, rotation and homothety transformation could be considered. But only translation is studied, because a rotation will give a different cut and thus a different image. Homothety transformation is not significant because the patient does not move his head during snapshots.

The images are correlated two by two, in order to find the optimal superposition between all couples. The higher is the correlation score, the higher is the similarity between both images. So, the maximum correlation value corresponds to the optimal translation shift. The maximum correlation scores obtained over all the aligned images allow also to choose a reference image, from which all the translations will be achieved, and to reject images that are not enough similar to it. The reference image A_R is chosen as follows: it is the one that leads, in average, to the two highest correlation scores with the other images. Images that get a correlation score below S=0.5 (threshold experimentally found) are considered to be erroneous ones, and will not be included in the fusion.

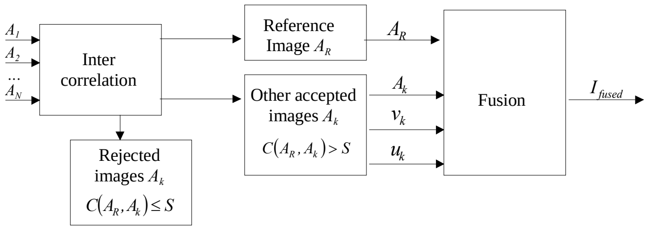


Fig. 6. Fusion process.

Different fusion functions can be envisaged. In our first experiments, we use a simple average function that applies spatially on each retained image A_k and also between all these images:

$$A_f(x,y) = \frac{1}{9K}\sum_{k=1}^{K}\sum_{i=-1}^{1}\sum_{j=-1}^{1} A_k(x+u_k+i, y+v_k+j) \quad (4)$$

where K ($K \leq N$), the number of accepted images, indexed by k, including the reference image A_R, (u_k,v_k) the horizontal and vertical shifts between the image A_k and the reference image A_R.

The results can be still improved, by excluding from the average the pixels that are likely to be noise pixels. $A_f(x,y)$ represents the mean pixel value, calculated over the K retained images, on a small 3x3 neighborhood around (x,y). In the same way we compute an estimation of the standard deviation $\sigma_f(x,y)$. The pixels of the neighborhood that differ too much from the mean value $A_f(x,y)$ are excluded from the averaging (5). The resulting image I_{fused} is consequently improved, compared to the mean image A_f.

$$\left|A_f(x,y) - A_k(x+u_k+i, y+v_k+j)\right| > 0.5\sigma_f(x,y)$$
$$\Rightarrow A_k(x+u_k+i, y+v_k+j) \text{ excluded from the fusion.} \quad (5)$$

An example of fusion is shown in Fig. 8. Four images A_k were fused. This process results in a higher signal to noise ratio and an enhanced contrast (see next section).

5. Results

Our study has been done over 10 patients, healthy or presenting different pathologies. Three to fifteen images were captured per patient. The proposed method has been evaluated and quantified based on the following procedure. We have considered three homogenous areas where the grey levels should be almost constant without the presence of noise: the fovea (FOVEA), the ganglion cell layer and the inner plexiform layer (GCL+IPL), and the outer nuclear layer (ONL) (Fig. 7). We have estimated the noise power by calculating the image variance on these regions (manually delimitated). Let us denote by $\sigma_k^{(i)}$ the standard deviation of the region i in the image k, and by $\sigma_f^{(i)}$ the corresponding measure in the fused image. Then, the ratio between the mean variance calculated on the original images and the variance calculated on the fused image (6) provides an estimation of the signal to noise ratio improvement. This study has been conducted only on images of healthy retinas, otherwise the three regions cannot be all correctly defined. Tab. 1 indicates the number of images used for the fusion and the gains (in dB) obtained for each region.

$$G^{(i)} = -10\log\left(\frac{\sigma_f^{(i)2}}{\sigma_m^{(i)2}}\right), \sigma_m^{(i)} = \frac{1}{K}\sum_{k=1}^{K}\sigma_k^{(i)} \quad (6)$$

Number of images (K)	$G^{(1)}$ (dB) FOVEA	$G^{(2)}$ (dB) GCL+IPL	$G^{(3)}$ (dB) ONL
3	8.5	3.5	6.3
3	9.4	2.7	5.2
4	8.4	3.3	5.4
6	7.5	2.9	6.1
9	11.9	8.0	8.4
10	14	7.0	8.4
15	10.3	6.2	7.5

Tab. 1. The signal to noise ratio gain measured on three layers, for 7 cases of healthy eye.

The signal to noise ratio is increased in all cases, with a gain between 3 and 12 dB. The greater is the number of fused images, the greater is the SNR gain. The Fovea and the ONL regions are clearly denoised (gain over 6 dB) while the GCL+IPL layer has been significantly smoothed.

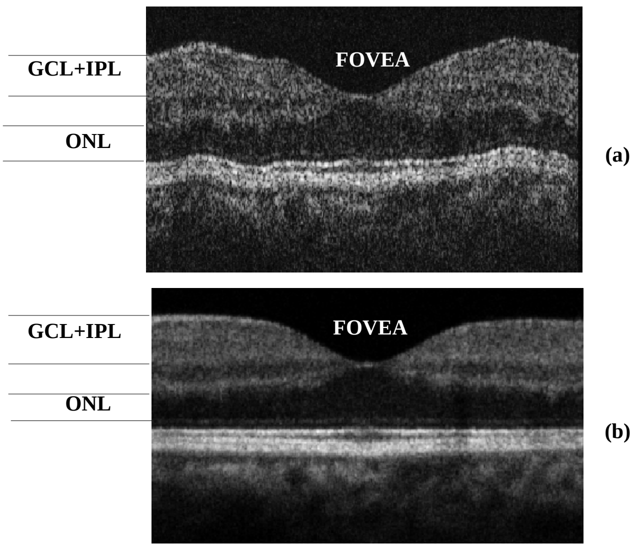


Fig. 7. Fusion of 10 images: (a) one of the original images, (b) the result image. Images are represented in grey levels and normalized between 0 and 1.

The results were also submitted to clinical specialists from the Quinze-Vingts hospital (Paris). Doctors assessed in all cases that the normalization and fusion method leads to an enhanced image that makes the clinical interpretation easier and more accurate. It is clearly apparent from Fig. 7 and 8 that the contrast between the retinal layers is improved. In particular, the reflective layer attributed to the external limiting membrane, that is, the frontier between the inner and outer segment of the photoreceptors, becomes clear, while it was slightly apparent in non-processed images.

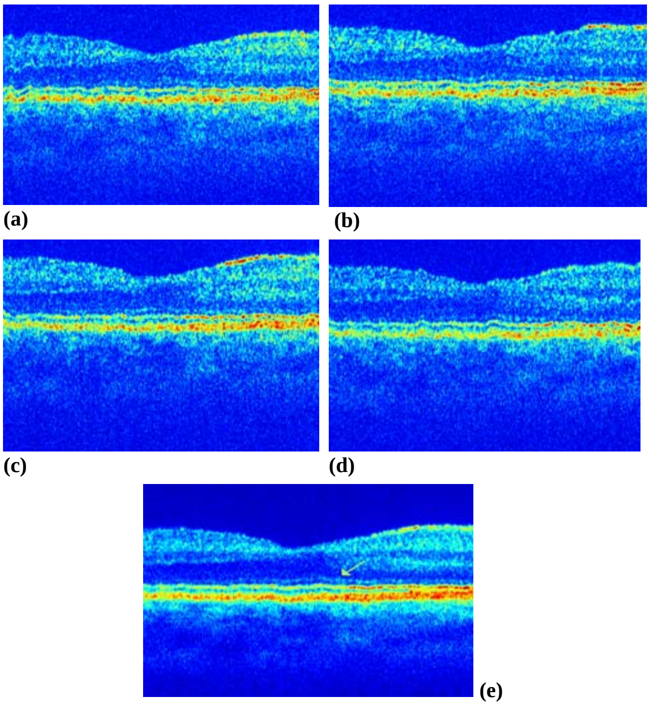


Fig. 8. Fusion of 4 aligned images (a)(b)(c)(d) to provide the enhanced output image (e).

In the example shown in Fig. 9, the interruption of the photoreceptor line is evident. Alterations of the intraretinal structure are also observed, consisting of collection of oedema within the retina.

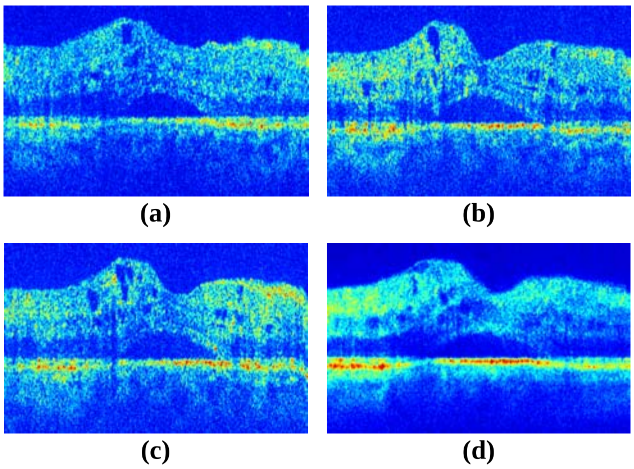


Fig. 9. Fusion of 6 aligned images (three of them are represented in figures (a)(b)(c)) to provide the enhanced output image (d).

6. Conclusion

This paper describes a method of optical coherence tomography imaging enhancement of the retinal area of the eye. The methods and algorithms proposed in this article, i.e. reflection layer detection, image alignment and fusion, have proved to increase considerably the quality and the signal to noise ratio. In fact, fusion of a collection of image OCT scans from the same retinal area enhances enough the quality of imaging to reveal new details of the retina.

To date, only total retinal thickness was considered in most studies of OCT; alterations of retinal layering, which is an indicator of retinal diseases, could not be reliably defined. Averaging OCT scans thus provides additional information about the intraretinal structures. This is in agreement with the results presented by [4]. The higher definition of the retinal layering offers the opportunity to better define the alterations of retinal structures. Such improved precision contributes to a better diagnosis of retinal alterations and hence to the cause of visual impairment.

Future work will include investigation of new fusion techniques, in order to improve the processing speed and the quality of the result. The automatic detection of the different physiological layers of the retina, in order to calculate some measures useful for the clinical diagnostic, is also a subject of further study.

References

[1] StratusOCT tomography, Zeiss documentation, http://www.zeiss.com/C125679E0051C774/Contents-Frame/2BF7095D5578B41D882572430063F9AD.

[2] Carl Zeiss Meditec AG, Goeschwitzer Str 51-52, 07745 Jena, Germany, http://www.zeiss.com/.

[3] KENT, C. Taking OCT technology to the next level. *Review of Ophthalmology*, 2006, vol. 13:06, Issue 6/7/2006.

[4] SANDER, B., LARSEN, M., THRANE, L., HOUGAARD, J. L., JØRGENSEN, T. M. Enhanced optical coherence tomography imaging by multiple scan averaging. *Br. J. Ophthalmol*. 2005; vol. 89; p. 207-212 doi:10.1136/bjo.2004.045989, December 2005.

[5] GONZALES, R. C., WOODS, R. E. *Digital Image Processing.* New Jersey: Prentice Hall, 2002.

[6] SETHIAN, J. A. *Level Set Method and Fast Matching Methods.* Cambridge University Press, 1999.