Trabajo de Fin de Máster

Automatic wide-field registration and mosaicking of noisy OCTA images using template matching and differential evolution

Junio 2020

Autor: Alejandro Moya García

Director: Enrique J. Carmona Suárez

Máster Universitario en Investigación en Inteligencia Artificial

Escuela Técnica Superior de Ingeniería Informática

Universidad Nacional de Educación a Distancia (UNED)

Table of contents

Abstract

Optical Coherence Tomography Angiography (OCTA) is a novel non-invasive ophthalmological technique used to perform a detailed examination of the eye fundus vascularity. However, each of the images obtained by this technique only cover a small retinal area. Thus, ophthalmologists have to take complementary images of the eye fundus from different angles in order to obtain a complete visualization of patients' eye fundus. In particular, each set of images must be manually registered by a clinician, being a tedious and time-consuming process. In this work, we propose an approach based on template matching and differential evolution to automatically register a set of OCTA images characterized by containing noise and artifacts. The proposed method is divided into three main steps. First, a preprocessing step used to extract the main vascular network is applied on every image. Then, an algorithm based on differential evolution is run on every 2-combination of OCTA images in order to find the best overlap between them. Finally, a greedy algorithm iteratively selects the best pairs of images (according to their fitness) to create the complete mosaic. The proposed method was evaluated via the registration of several sets of OCTA images with the purpose of building their associated mosaics. Results show that our approach is robust and able to achieve a good approximation to the optimal mosaic.

Keywords: OCTA, retinal imaging, template matching, differential evolution, mosaicking.

1 Introduction

Optical Coherence Tomography Angiography (OCTA) is a non-invasive imaging technique used to check the retinal vascularity. There are several works showing that this technique can be useful to obtain information about pathologies such as diabetic retinopathy (Falavarjani et al., 2018), glaucoma (Holló, 2018) and choroidal neovascularization (Chalam and Sambhav, 2016). Another OCTA's useful feature is the extraction of biomarkers, which have been shown useful to identify, diagnose and monitor pathologies such as vision loss (Tang et al., 2019) and diabetic macular edema (Hsieh et al., 2019), among others.

In contrast to other techniques, like Fluorescein Angiography (FA) and Indocyanine Green Angiography (ICGA), OCTA is a fast and non-invasive technique based on movement contrast rather than intravenous dye. Furthermore, OCTA allows for the observation of the retinal vascularity and the blood vessel structure at the same time (de Carlo et al., 2015). However, OCTA devices only cover small retinal areas and ophthalmologists need to take multiple overlapping images from different retinal zones to then build a wide-field mosaic of the retina. When the mosaic is manually performed, is tedious, time-consuming and subject to inaccuracies, especially when the OCTA images contain noise and artifacts. Therefore, a method to automate this process is desired.

There are only two works in the current literature focused on this problem. The first of them, (Wang et al., 2018), uses the Speeded Up Robust Features (SURF) feature detector and descriptor to find and establish keypoint matches between pairs of images. Then, uses Random Sample Consensus (RANSAC) algorithm to extract the homography matrix needed to perform an affine transformation in order to register the pairs of images. After obtaining the best keypoint matches between every 2-combination of images, the mosaic is iteratively constructed based on the quality of keypoints matches. In each iteration, the image with closer keypoints to the images registered in the current mosaic is added to the mosaic. This process is repeated until all the images are registered. Finally, a

post-processing consisting of flow signal compensation and seamless blending is applied in order to hide the seams and achieve a better continuity of the vasculature. The second one, (Díaz et al., 2019), is inspired by the former. The main difference with the previous work is the addition of a pre-processing stage that segments the vessels in the image. In this way, the keypoints analysis is restricted to keypoints belonging to the vascular network.

Unlike the OCTA images used in (Wang et al., 2018), the images used in this work present significant levels of noise and artifacts due to the capture process. As a consequence, the process of finding an optimal affine transformation to iteratively rebuild the complete mosaic, adding one image at each step, it is difficult, has a high computational cost and the final mosaic obtained is far from optimal. Note that every time a new image is registered into the current mosaic, error is accumulated due to the deformation introduced by the affine transformation, becoming increasingly hard to register the next image.

Hence, in this work, we propose a new approach to obtain the wide-field retinal mosaic from a set of OCTA images, assuming that: (i) the considered OCTA images contain noise and artifacts; and (ii) the use of a transformation based on translation is enough to obtain a good approximation to the mosaic. Basically, we apply a first preprocessing stage to obtain the vascular tree of each OCTA image. Then, we calculate the best registration for every 2-combination of images using an algorithm based on differential evolution and template matching. Finally, the final mosaic is obtained by a greedy algorithm that iteratively registers the images one by one using the fitness and coordinates obtained by the evolutionary algorithm for each 2-combination of images.

The rest of this manuscript is organized as follows: in Section 2, we provide a introduction to the main techniques used in this work; then, in Section 3, we describe the dataset of OCTA images used in the experiments; the proposed approach is presented in Section 4; next, we detail the performed experiments and the obtained results in Section 5; then, we analyse the results and compare them with previous works in Section 6; and, finally, we present the conclusions and future work in Section 7.

2 Background

The core of this work is based on two well-known techniques: template matching (TM) and differential evolution (DE). In this section we review these techniques and their main features.

2.1 Template matching

A first approach to register two images is to find their common parts, something that can be achieved with techniques such as TM (Rosenfeld, 1969), (Brunelli, 2009). In fact, TM has already proven to produce good results not only in retinal images (Gong et al., 2019), (Mora et al., 2013), but also in other fields of medical imaging (Shahine et al., 2013), (Gurunath Bharathi et al., 2018) where normalized cross-correlation (NCC) has been used as the similarity function.

TM is a technique used for finding those parts of an image (the frame image) that match with another image (the template). A basic implementation of TM finds the optimal matching coordinates of the template by moving its center over each point in the frame, from left to right and from top to bottom. In each point, the similarity between the area spanned by the template over the frame is measured. After all the possible positions of the template with respect to the frame are considered, we can conclude which position is the one with the highest score, i.e., the position where the matching is optimal. There are several cost functions to measure the similarity such as the normalized cross-correlation, the mean absolute error and the mean squared error.

2.2 Differential evolution

Differential evolution (Storn and Price, 1997) is an evolutionary algorithm which can optimize non-differentiable, nonlinear and multimodal cost functions by iteratively improving a set of n-dimensional parameter vectors (solutions) which are used as a population for each generation. Its structure follows the general scheme of an evolutionary algorithm: population initialization, parent selection, recombination, mutation and survivor selection. The variation mechanism depends on the selected mutation scheme, crossover scheme and number of vectors involved. In this work, the most widely used DE variant, known as "rand/1/bin", is selected. The notation "DE/x/y/z" is interpreted as follows: the label "x" refers to how the individual playing the role of "a" in the mutation is selected (in our case, randomly, i.e., "x = rand"); the label "y" denotes how many couples of individuals are involved in the computation of the subtraction shown in line 7 (in our case, "y = 1"); and the label "z" refers to the type of recombination used (in our case, binomial, i.e., "z = bin", such as is shown in lines 5 - 11 of Algorithm 1). The pseudocode of this variant is shown in Algorithm 1 (assuming a maximization problem).

Algorithm 1
Pseudocode of DE variant "rand/1/bin".

```
   Input: P /* Population*/
   CR ∈ [0, 1] /* Crossover rate*/
   F ∈ [0, 2) /* Differential weight */
   Output: BS /* Best solution*/
1: while termination criteria is not met do
2:    for x in P do
        /* Randomly pick 3 vectors from P all different from x */
3:      a, b, c := sample_rand_different(P, x)
        /* Mutation and Recombination */
4:      j := rand_uni() * length(x)
5:      for i := 0;i < length(x); i++ do
6:         if random_uniform() < CR || i == j then
        /* Vector new_x is created as a combination of a, b and c */
7:            new_x[i] := a[i] + F * (b[i] - c[i])
8:         else
9:            new_x[i] := x[i]
10:        end if
11:     end for
        /* Evaluation and selection */
12:     fitness := eval(x)
13:     new_fitness := eval(new_x)
14:     if new_fitness > fitness then
15:        replace x with new_x
16:     end if
17:   end for
18: end while
19: BS = argmax_{x∈P} eval(x)
```

3 Material

This study was conducted in accordance with the principles of the Helsinki Declaration[1] and was carried out retrospectively on existing images that have previously been anonymized. The used OCTA image dataset is the same as in (Díaz et al., 2019). It contains sets of OCTA images from 10 different patients. Each OCTA image was taken using the capture device DRI OCT Triton Topcon Corp. There are between three and five images per eye, making a total of 54 images with resolution 320x320 pixels, which corresponds to a real size of 6mm^2.

[1] https://web.archive.org/web/20091015082020/http://www.wma.net/en/30publications/10policies/b3/index.html

The experiments have been carried out in Python on a computer with Intel© Core™ i5-4210H @ 2.9GHz CPU and 8GB RAM. Our implementation is based on the DE module from SciPy's library, and the source code is available at GitHub[2].

4 Methodology

The proposed methodology consists of three main stages. First, OCTA images undergo a pre-processing stage consisting of a gaussian filter and binarization in order to extract the main vessels. Then, every 2-combination of the preprocessed images is registered using an evolutionary algorithm. Finally, a greedy algorithm iteratively selects the best pairs of images to incrementally create the final mosaic. Fig. 1 summarizes the proposed methodology applied to the case of three OCTA images. In the following sections, we describe with more details each one of these stages. Note that the greedy algorithm needs to select the two best 2-registrations to build the mosaic of three images.

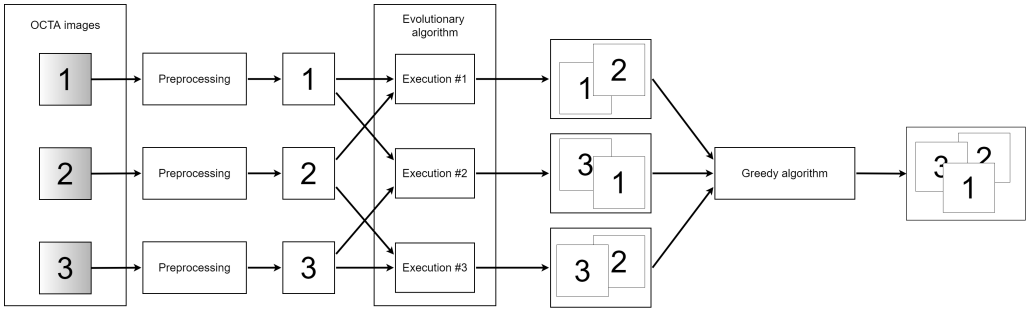


Figure 1. Stages of the proposed methodology for a set of three OCTA images.

4.1 Preprocessing

In order to avoid the evolutionary algorithm getting stuck in local optima, we need to extract the relevant information associated with OCTA images. We explored and evaluated the performance of multiple filters to reduce noise and extract the retinal vessels. We concluded that the best preprocessing pipeline for this particular scenario consists of a gaussian filter followed by a binarization filter. Fig. 2 shows an example of the stages of preprocessing applied on one image.

(a) (b) (c)

Figure 2. Example of preprocessing stages. (a) Original OCTA image. (b) Resulting image after applying gaussian filter. (c) Resulting image after applying binarization.

[2]https://github.com/amoyag00/OCTA-Image-Registrar

4.2 Search based on evolutionary algorithm

Although there are optimized TM libraries, they do not take into account that the overlapping area between two OCTA images is not known a priori: the overlapping area must be calculated in every iteration of the TM algorithm in order to extract the template and frame images. Furthermore, these libraries are only optimized for evaluating similarities based on simple cost functions. However, when the cost function used to evaluate the matching is complex, the task of computing TM between two OCTA images at original resolution is computationally expensive due to brute-force search has to be used. Thus, we have decided to use a more efficient search algorithm than brute-force search (see Appendix A). In particular, our approach is based on an evolutionary algorithm.

4.2.1 Using differential evolution

We propose a DE-based algorithm as an evolutionary algorithm. There are different works in which evolutionary algorithms, and DE in particular, have produced good results in TM (López-Franco et al., 2016), (Furukawa et al., 2019). Furthermore, there are other works which show that DE outperforms other evolutionary algorithms such as bee colonies and particle swarm optimization in numerical benchmark problems (Vesterstrom and Thomsen, 2004), (Li et al., 2010).

In order to evaluate the similarity between different subareas of images with DE, we needed to build a global coordinate system where images can move. We performed experiments in which all the images belonging to an eye can be moved in the coordinate system. The results showed that the search space and the number of centers of the set of images, were too big and, therefore, the computational cost associated with the evolutionary process was too high. Thus, we decided to register the images in pairs. In particular, we registered every 2-combination of images. This result will be used then by a greedy algorithm to obtain the complete mosaic from those 2-combination image registrations (see Section 4.3). In order to register a pair of images, we only need a 2D vector to represent each possible solution, since one of images plays the role of frame and remains at the center of the coordinate system while the other one plays the role of template and can move over the frame.

In order to always ensure an area between the template and frame images, we have established bounds that force a minimum overlap of 1 pixel. The coordinate system size is $(2 \cdot W_t + W_f - 2 \times 2 \cdot H_t + H_f - 2)$ px, where W_t and H_t are the width and height of the template image and W_f and H_f are the width and height of the frame image (in our set of OCTA images $W_t = W_f = H_t = H_f$). Fig. 3 shows a scheme of the coordinate system.

4.2.2 Population initialization

DE is able to obtain good image registrations, but it needs between 300 and 400 generations to converge. Thus, we studied how to further reduce the maximum number of generations by using a intelligent mechanism to initialize the population. For it, we decided to include in the initial population the solutions obtained by TM based on brute force, but using downscaled images to speed-up the process. We have found that the downscales which offers the best commitment between effectiveness and performance are $\frac{1}{2^2}$ and $\frac{1}{2^3}$ (see Appendix A).

Since DE implements an elitism policy by definition, we do not need to include multiple copies of the solutions obtained by brute force. On the other way, there is no guarantee that the two solutions obtained by brute force are the same or that the optimal solution to our problem is included in the set formed by both solutions. That is why our evolutionary algorithm is used to refine and guarantee the search for the optimal solution. The rest of

the population is initialized by using the latin hypercube sampling method (McKay et al., 1979).

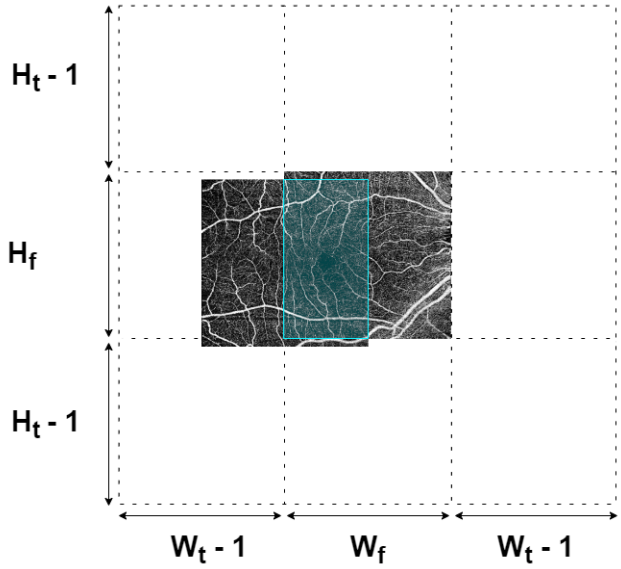


Figure 3. Scheme of the used coordinate system, where $W_t \times H_t$ and $W_f \times H_f$ correspond to the sizes of the template and frame images, respectively. The frame image remains anchored at the center while the template image is moving around. An optimal registration of a pair of images is also shown with the overlapping area colored in blue (see online version for colors).

4.2.3 Fitness function

Since our objective is to find the common area between two OCTA images, we need a fitness function that measures the similarity between the two overlapped image areas. Inspired by the use of NCC in previous works (Shahine et al., 2013), (Gurunath Bharathi et al., 2018), we have designed a NCC based fitness function specifically adapted to this problem. We have created this function based on two criteria: (i) maximize the NCC of the overlapping areas; and (ii) maximize the NCC of the overlapping subareas' perimeter. In particular, perimeter correlation rewards those registrations with good vessel continuity and also helps to avoid the algorithm getting stuck in local optima. In the same line, to avoid local optima associated with overlapping areas with few pixels, we have weighted the area-based NCC and perimeter-based NCC by the area and perimeter coverage, respectively. This way, we use the NCC as a similarity measure but we also reward those registrations in which the common area and perimeter are bigger.

The complete fitness function is presented in Eq. 1, where S_i represents a subarea of the image I_i, with $i = 1, 2$; $W_{area}(S)$ is given by Eq. 2 and calculates the area coverage of S with regard to the image total area, where $|x|$ stands for the cardinality of x; $W_{per}(S)$ is given by Eq. 3, where $per(S)$ is a function that extracts the perimeter of S with a width of 3 pixels; $NCC(S_1, S_2)$ is given by Eq. 4 and calculates the zero normalized cross-correlation between the overlapping areas S_1 and S_2, where μ_i, σ_i, and n are the average of the pixel intensity, the standard deviation of pixel intensity and the number of pixels, respectively, in S_i; and, finally, w_a and w_p are weighting factors used to control the influence of terms based on area and perimeter, respectively.

$$Fitness(S_1, S_2) = w_a \cdot NCC(S_1, S_2) \cdot W_{area}(S_1) + w_p \cdot NCC(per(S_1), per(S_2)) \cdot W_{per}(S_1) \quad (1)$$

$$W_{area}(S) = \frac{|S|}{min(|I_1|, |I_2|)} \quad (2)$$

$$W_{per}(S) = \frac{|per(S)|}{min(|per(I_1)|, |per(I_2)|)} \quad (3)$$

$$NCC(S_1, S_2) = \frac{1}{n} \sum_{i=1}^{n} \frac{(S_1(i) - \mu_1) \cdot (S_2(i) - \mu_2)}{\sigma_1 \sigma_2} \quad (4)$$

4.2.4 Early stopping and local search

In addition, to further to optimize the algorithm's performance on this particular problem, we have added two new strategies: early stopping and local search.

Our early stopping implementation consists of prematurely ending the execution of DE when there is no improvement after a number of generations. As for local search, our implementation consists of exploring the closest neighbours of the N best vectors at the end of every generation. If any of the neighbours is better than the reference individual, we replace it by that neighbour. Here, the idea is to accelerate the exploitation phase of the best solution found by the canonical DE algorithm. The neighbours are selected based on an Euclidean distance criteria. Fig. 4 shows an example of neighbourhood considering the 12 closest neighbours to the black pixel.

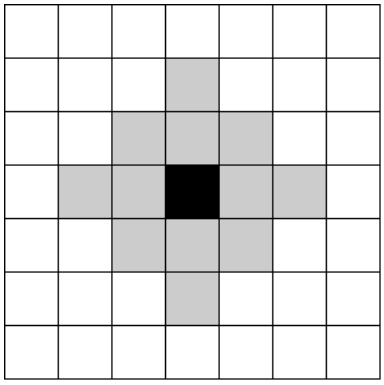


Figure 4. Example of neighbourhood: grey pixels represent the 12 closest neighbours of the black pixel.

4.3 Greedy algorithm

The last step of the proposed methodology consists of a greedy algorithm that iteratively builds the mosaic based on the fitness associated with each pair of images registered by DE. The pseudocode is presented in Algorithm 2 and it works as follows. Every registration is made up of the following four elements: a frame image; a template image; the coordinates of the template image; and the registration's fitness obtained by the evolutionary stage. First, the fittest registration (and therefore, its two images) are added to the fittest list (`selected_regs`) and removed of the registration list (`registrations`) (lines 1 - 5). Then, in every iteration, the best registration with only one of its images present in the fittest

list is added to this list (lines 6 - 10). This process is repeated until the fittest list stores the minimum number of images needed to build the mosaic. Finally, the mosaic is built based on the coordinates present in the fittest list. This list is read in reverse order so that fitter registrations are presented above in the mosaic (lines 11 - 14). Fig. 5 shows a mosaic example of three images.

Algorithm 2
Pseudocode of the greedy algorithm.

Input: `registrations` /* Set of registrations obtained by the evolutionary algorithm */
`num_images` /* Number of images of the eye */
Output: `mosaic`
1: `selected_regs := ∅`
2: `mosaic := ∅`
3: `best_reg =` $\underset{reg \in registrations}{argmax}$ `eval(reg)`
4: `selected_regs := selected_regs ∪ {best_reg}`
5: `registrations := registrations \ {best_reg}`
6: **while** `|selected_regs| < num_images - 1` **do**
7: `best_reg :=` $\underset{reg \in registrations}{argmax}$ eval(reg) `such that` $|\{images \in reg\} \cap \{images \in selected\_regs\}| = 1$
8: `selected_regs := selected_regs ∪ {best_reg}`
9: `registrations := registrations \ {best_reg}`
10: **end while**
11: `selected_regs := reverse(selected_regs)`
12: **for** `registration in selected_regs` **do**
13: `mosaic := mosaic ∪ {registration}`
14: **end for**

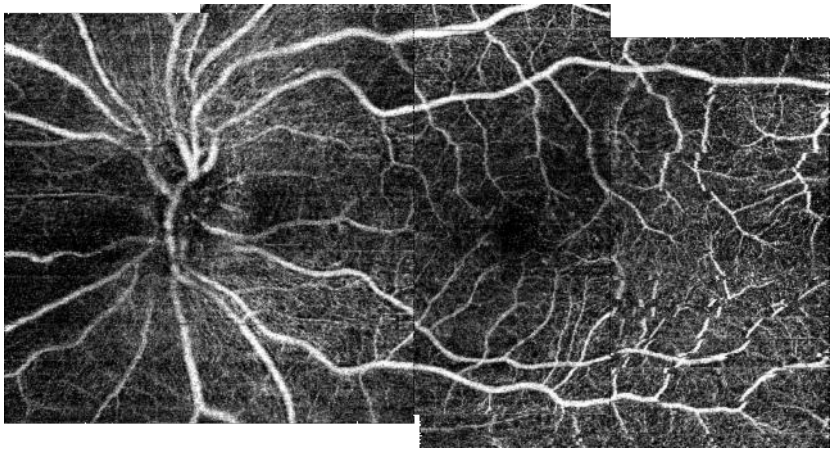


Figure 5. Example of mosaic of 3 OCTA images.

5 Experimental results

In this section we provide information about the algorithm configuration and evaluate the quality of the obtained solutions.

5.1 Algorithm configuration

We evaluated multiple combinations of gaussian kernel size and binarization threshold, including the OTSU's method (Otsu, 1979). The results showed that a kernel size of 3×3 and a binarization threshold of 150 produce the best results. The gaussian filter reduces the high frequency noise in the image, allowing for a cleaner vessel extraction. On the other way, the used binarization threshold value is high enough to eliminate the

background noise and low enough to preserve the main vessels. In any case, the value of 150 is not critical and any other close value could have been selected. Extracting only the most important vessels has proved to be helpful for the evolutionary algorithm exploration (see Section 4.2), decreasing the number of times it gets stuck in local optima.

In regard to DE, after considering and evaluating variants such as rand/1/bin, best/1/bin, best/1/exp, rand/1/exp, rand/2/bin, rand/2/exp, we concluded that rand/1/bin is the scheme which provides better results in terms of solution's quality and convergence speed in this particular problem. Regarding the control parameters of the variation operators, there are only two values to tune: F, the differential weight, a parameter in range [0, 2) used in the mutation step; and CR, the crossover rate. To tune these hyperparameters we carried out a grid search hyperparameter optimization. The results showed that a differential weight equal to 1.9 and a crossover rate equal to 0.75 were the best values.

In addition to variation operators, convergence and population diversity are another factors that must be analysed. Thus, we also explored different values of population size and maximum number of generations concluding that 400 generations and a population size of 30 produce the best and more efficient results for this particular problem. We also evaluated different values for weighting factors of the fitness function, w_a and w_p, and concluded that values of 0.8 and 0.2 respectively are the best choice. Regarding the custom code optimizations (early stopping and local search), we have set an early stopping of 50 generations, a neighbourhood size of 12 and we have only considered the best vector to perform the local search in each generation. Table 1 summarizes the parameter configuration of the proposed approach.

Table 1. Parameter configuration of the proposed approach.

Parameter	**Value**
Gaussian kernel size	3×3
Binarization threshold	150
Differential evolution variant	rand/1/bin
Crossover rate (CR)	0.75
Differential Weight (F)	1.9
Maximum number of generations	400
Population size	30
w_a	0.8
w_p	0.2
Early stopping generations	50
Local search neighbourhood size	12
Number of best local search vectors (N)	1

5.2 Results

In order to measure the quality of the registered images, we calculated the mean Euclidean distance between pairs of matching keypoints used as the ground truth. In order to obtain this ground truth, we selected four pairs of matching keypoints from the set of pairs of matching keypoints obtained by the SURF method (Bay et al., 2006). These pairs of keypoints were manually selected for each pair of images with overlapping in the mosaic manually built. Note that no keypoints have been obtained in those cases where overlapping areas are so extremely small that SURF algorithm do not produce results. The

mean Euclidean distance function is calculated according to Eq. 5, where N is the number of pairs of matching keypoints; (x_i, y_i) are the coordinates of the keypoint i belonging to the template image; and (x'_i, y'_i) are the coordinates of the keypoint i belonging to the frame image.

$$\overline{D} = \frac{\sum_{i=1}^{N} \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2}}{N} \tag{5}$$

On the one hand, we evaluate the quality of the set of registrations selected by the greedy algorithm. On the other hand, we evaluate the quality of the mosaic itself. The first case allows us to evaluate the quality of the solutions produced by the evolutionary algorithm and selected by the greedy algorithm, while the second one allows us to evaluate the quality of the solutions obtained by the whole method (including the extra overlaps that appear in the mosaic but do not belong to the selected registrations to build it). Note that the distances obtained in the first case are usually less than or equal to the distances obtained in the second case, since in the second one all the overlapping areas are considered, that is, not only the ones selected by the greedy algorithm. The results concerning the first and second case are presented in the columns $\overline{D}_{pairs}$ (proposed method) and $\overline{D}_{mosaic}$ of Table 2, respectively. Since our approach is stochastic, the values shown in the mentioned columns are the mean values obtained after 20 executions. We have also established a success threshold of 10 pixels to decide the success of an execution. Note that 10 pixels is approximately 3% of W, with $W \times W$ being the size of our OCTA images. The success rate of both cases is presented in columns SR_{pairs} and SR_{mosaic}, respectively.

We have also evaluated the mean weighted NCC of the overlapping areas in both cases by using Eq. 6, where C_i and A_i are the correlation and area of the ith overlapping area, respectively. Again, we have calculated the mean value after 20 executions. The results are presented in columns $\overline{NCC}_{pairs}$ (proposed method) and $\overline{NCC}_{mosaic}$.

$$\overline{NCC} = \frac{\sum_{i=1}^{n} C_i \cdot A_i}{\sum_{i=1}^{n} A_i} \tag{6}$$

In addition, we have evaluated an affine transformation (AT) based approach and inspired by (Díaz et al., 2019) in order to evaluate if it is possible an improvement of the results obtained by our method. In the new approach, we have also measured the mean Euclidean distance and the mean weighted NCC of the pairs of images selected by the greedy algorithm. These results are presented in columns $\overline{D}_{pairs}$ (AT-based method) and $\overline{NCC}_{pairs}$ (AT-based method) of Table 2.

In regards the computational cost, the proposed method needs a mean execution time of 10.22 seconds to register a pair of images and 68.12 seconds to register the complete mosaic, while the AT-based method needs 11.28 seconds to register a pair of images. Finally, in Appendix B, we show the mosaics obtained by the proposed method in every set of OCTA images.

Table 2. Mean Euclidean distance, mean weighted NCC and SR metrics of the proposed method and AT-based method.

Patient ID	Eye	$\overline{\mathbf{D}}_{\mathbf{pairs}}$		$\mathbf{SR}_{\mathbf{pairs}}$	$\overline{\mathbf{D}}_{\mathbf{mosaic}}$	$\mathbf{SR}_{\mathbf{mosaic}}$	$\overline{\mathbf{NCC}}_{\mathbf{pairs}}$		$\overline{\mathbf{NCC}}_{\mathbf{mosaic}}$
		Proposed method	**AT-based method**	**Proposed method**	**Proposed method**	**Proposed method**	**Proposed method**	**AT-based method**	**Proposed method**
1	Right	2.83	1.26	100%	2.83	100%	0.47	0.52	0.47
	Left	5.00	11.90	100%	5.00	100%	0.48	0.19	0.48
2	Right	2.14	1.53	100%	2.41	100%	0.47	0.57	0.46
	Left	2.42	0.85	100%	2.42	100%	0.39	0.49	0.39
3	Right	1.45	1.33	100%	1.64	100%	0.52	0.56	0.49
	Left	2.42	21.13	100%	3.78	100%	0.44	0.25	0.43
4	Right	2.10	1.98	100%	2.17	100%	0.51	0.60	0.49
	Left	1.79	1.42	100%	2.05	100%	0.42	0.57	0.39
5	Right	2.58	1.27	100%	2.65	100%	0.52	0.62	0.49
	Left	2.08	1.26	100%	2.08	100%	0.61	0.68	0.61
6	Right	2.11	0.61	100%	2.17	100%	0.42	0.62	0.41
7	Right	1.80	3.90	100%	1.81	100%	0.53	0.60	0.48
	Left	2.00	2.85	100%	1.86	100%	0.52	0.59	0.50
8	Right	1.78	114.39	100%	1.78	100%	0.57	0.33	0.57
	Left	Undetermined	142.08	0%	195.90	0%	0.21	0.21	0.21
9	Right	3.74	3.41	100%	3.48	100%	0.38	0.40	0.37
	Left	5.15	44.02	100%	6.17	100%	0.29	0.36	0.25
10	Right	1.91	0.86	100%	2.46	100%	0.42	0.68	0.39
	Left	2.45	1.16	100%	4.92	100%	0.37	0.69	0.27

6 Discussion

Column $\overline{D}_{pairs}$ (proposed method) of Table 2 shows that the evolutionary algorithm is able to achieve a mean Euclidean distance lower than 5.15 pixels in every eye except for the left eye of patient 8. The reason behind this is the absence of a central image that encompasses the area between the optic disc and fovea, a situation that only occurs in the eyes of patient 8, and makes some overlapping areas of the left eye of patient 8 to be very small (see Figs. 36 and 37 of Appendix B). Since our fitness function (Eq. 1) rewards registrations with higher correlation and bigger overlapping areas, it becomes harder to achieve a good fitness in some of the pairs of images of this eye. This problem leads the greedy algorithm to a situation where it selects a registration with no real overlapping due to the registration with real overlapping presenting a worse fitness. The result is that the method fails in this case (see Figs. 38 and 39 of Appendix B). Thus, the mean Euclidean distance of the left eye of patient 8 is labelled *undetermined*, that is, it cannot be calculated because it includes a pair of images without real overlap. Since the lack of a central image only occurs in patient 8, we have marked this eye as an outlier. Leaving out this case, the proposed method achieves a mean Euclidean distance of 2.54 pixels in

the whole dataset. Note that, as the SR_{pairs} column indicates, the proposed method is able to reach SR equal to 100% in every eye except for the outlier.

Regarding the column $\overline{D}_{mosaic}$, the mean Euclidean distances of the mosaics are a little worse than the ones obtained in column $\overline{D}_{pairs}$ (proposed method). This result is consistent, since the mosaic is build based on all the overlapping areas, that is, not only on the overlapping areas selected by the greedy algorithm. Leaving out the left eye of patient 8, the proposed method achieves a mean Euclidean distance of 2.88 pixels in all the mosaics. As the SR_{mosaic} column shows, the proposed approach also achieves a SR equal to 100% in every eye without considering the outlier case.

In regards to the $\overline{NCC}$ metric, we observe in columns $\overline{NCC}_{pairs}$ (proposed method) and $\overline{NCC}_{mosaic}$ that smaller Euclidean distances result in higher correlations. As with columns $\overline{D}_{pairs}$ and $\overline{D}_{mosaic}$, values of column $\overline{NCC}_{mosaic}$ are a little worse than the values of column $\overline{NCC}_{pairs}$. These results are also consistent, taking into account that a better registration (shorter Euclidean distance) will produce a better correlation.

The comparison of the $\overline{D}_{pairs}$ results obtained by the AT-based method and the proposed method shows that the former generally slightly improves the proposed method, but there are some cases where the results obtained are substantially worse than those associated with the latter (see left eye for patients 1, 3 and 9 and right eye for patient 8). In these cases, the pairs of keypoints produced by the SURF algorithm (which is used by the AT-based method) are wrong and, therefore, the affine transformation calculated from them is not satisfactory. In addition, affine transformations produce a distortion of the transformed images, hindering the iterative process of adding a new image to the mosaic in each iteration. Note that this deterioration in quality will accumulate after each iteration: the more images the mosaic contains, the more difficult it will be to build. In regards the comparison of the $\overline{NCC}_{pairs}$ columns for both methods, the results are consistent again: shorter Euclidean distances produce a better correlation and vice versa.

We have observed that with our approach, based on translation-only transformations, we cannot achieve a perfect vessel continuity. Due to this, we have evaluated an evolutionary affine transformation (EAT) algorithm which performs affine transformations over one of the two registered images by our evolutionary translation-only (ETO) algorithm.

The affine transformation is performed in three main steps described in Fig. 6. First, starting from the registration obtained by the ETO algorithm, we calculate the new coordinates of the overlapping area corners (Fig. 6a). The new coordinates can be located in a 15×15 px neighbourhood area. This range is depicted by a red square. Then, we calculate the matrix M which transforms the original corners' coordinates to the new ones (Fig. 6b). Finally, we calculate the new overlapping area (Fig. 6c).

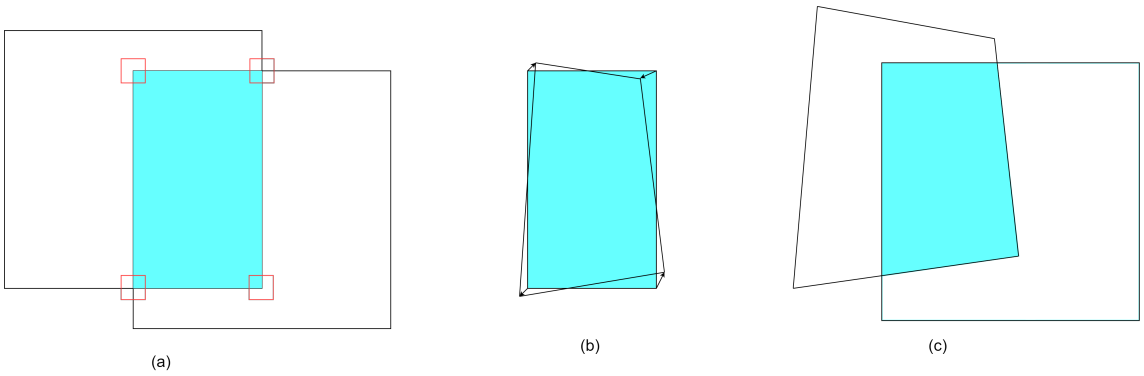


Figure 6. Steps of affine transformation. (a) Bounds of corners' new coordinates (red squares). (b) Affine transformation of the overlapping area corners. (c) New overlapping area.

In order to achieve the affine transformations with the EAT algorithm, we have to increase the dimensionality of our population individuals. Since we need two 2-dimensional coordinates for each one of the corners, this EAT algorithm takes 8-dimensional vectors as population individuals. We have kept the same value of F and CR as well as the variant used in the ETO algorithm (rand/1/bin). The EAT algorithm has been run during 200 generations with a population size of 30 individuals. We have not included any optimization mechanism such as local search or early stopping. In regards to the fitness function, we have observed that the mean absolute error of the overlapping area's perimeter with a width of 25 px offers better results than the previous fitness function.

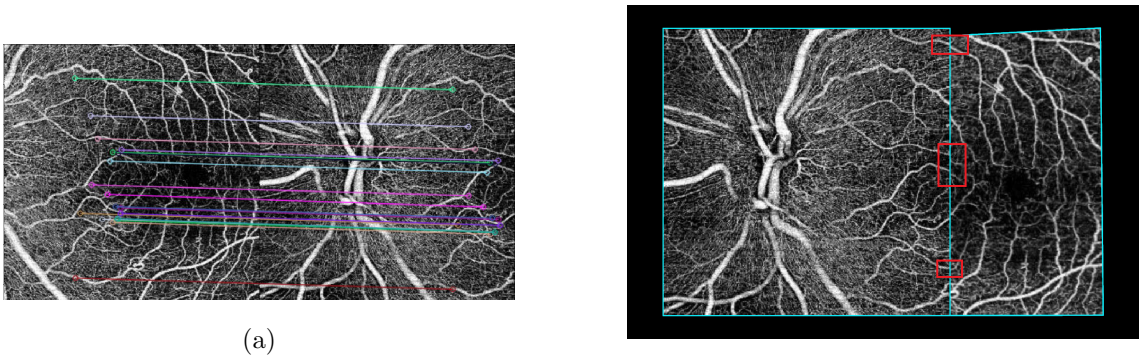
We tested this new approach on five different eyes and the results showed that the mean Euclidean distance between matched keypoints is about four times worse than the one obtained with ETO transformations. Results are presented in Table 3 with more detail. Furthermore, the mean execution time spent by the new approach to register a pair of images is 5 minutes, which in comparison with the translation-only approach, is too computationally expensive.

Table 3. Comparison of mean Euclidean distance between the evolutionary translation-only (ETO) algorithm and the evolutionary affine transformation (EAT) algorithm.

PatientID	**Eye**	$\overline{\mathbf{D}_{\mathbf{pairs}}}$	
		ETO	EAT
1	Right	2.82	5.34
2	Left	2.42	12.69
4	Right	2.10	8.33
5	Left	2.08	14.80
7	Left	2.00	6.83
Total		2.24	9.26

Furthermore, in order to prove that not even an affine transformation based on a manual approach can guarantee a perfect vessel continuity on a pair of OCTA images, we performed another experiment in which we manually filtered SURF keypoints matches to guarantee that they are homologous points. Then, we calculated a homography based on the selected keypoints an applied the transformation in one of the two images. The selected keypoints are presented in Fig. 7 (a) and the obtained registration is shown in Fig. 7 (b). As can be seen in red squares, the vessel continuity is further from perfect (Fig. 7).

The information presented in Table 3 and Fig. 7 provide evidence that the type of noise associated with the images of the used dataset is of such a nature that an affine transformation is not sufficient to obtain optimal overlaps. Therefore, we believe that the use of this type of transformation to overlap each pair of images is not justified. Additionally, the deformation produced by each transformation can difficult the mosaic composition when this is made by means of an iterative process as is made here or that in (Wang et al., 2018) or (Díaz et al., 2019): each time a new image is added to the current mosaic, it has to match the current mosaic formed by images that used a non-optimal affine transformation. This process of matching is made increasingly difficult due to the accumulation of errors produced by the previous transformations.



(a) (b)

Figure 7. Example of registration based on an affine transformation: (a) a subset of 18 SURF keypoints matches are manually selected to guarantee the quality of the transformation. (b) the composition obtained after applying the affine transformation. Note how the transformed image is deformed and, additionally, the little red squares reveal a lack of vessel continuity at the border.

7 Conclusions

We have proposed a new approach to register a set of OCTA images to obtain a final mosaic. Given the noisy nature of the input images, our approach cannot guarantee a perfect vessel continuity in the borders of the different overlapping images, but it is able to obtain a robust approximation in a reasonable amount of time. This supports our initial hypothesis that the use of a translation-based transformation is sufficient to obtain a good approximation to the mosaic. In addition, since not every ophthalmologist possesses state-of-art capture devices, which obtain less noisy and higher resolution OCTA images than those used in our study, it is also convenient to design registration methods for OCTA images obtained by less cutting-edge devices. We have also provided evidence that the use of affine transformations in this type of noisy images causes deformations that difficult the iterative process of adding the n-th image to the mosaic as the value of n increases.

In relation to the computational cost of the proposed method, it could be significantly reduced if the registration process of each pair of images were parallelized and implemented in a computer cluster. This is not a difficult task since the control flow of an evolutionary algorithm is easily parallelizable.

Acknowledgements

This work is supported by the Ministerio de Ciencia, Innovación y Universidades, Government of Spain, through the RTI2018-095894-B-I00 research project. We would also like to express our gratitude to the Varpa Research Group[3] for allowing us the access to the OCTA image database that was used here to evaluate our method.

References

H. Bay, T. Tuytelaars, and L. V. Gool. SURF: Speeded up robust features. In *Computer Vision – ECCV 2006*, pages 404–417. Springer Berlin Heidelberg, 2006.

R. Brunelli. *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley Publishing, 2009.

[3] http://www.varpa.es/

K. Chalam and K. Sambhav. Optical coherence tomography angiography in retinal diseases. *Journal of Ophthalmic and Vision Research*, 11(1):84, 2016.

T. E. de Carlo, A. Romano, N. K. Waheed, and J. S. Duker. A review of optical coherence tomography angiography (OCTA). *International Journal of Retina and Vitreous*, 1(1), 2015.

M. Díaz, J. de Moura, J. Novo, and M. Ortega. Automatic wide field registration and mosaicking of OCTA images using vascularity information. *Procedia Computer Science*, 159:505–513, 2019.

K. Falavarjani, J. Khadamy, and K. Aghdam. An update on optical coherence tomography angiography in diabetic retinopathy. *Journal of Ophthalmic and Vision Research*, 13 (4):487, 2018.

H. Furukawa, J. Sato, T. Yamada, K. Ito, and S. Ito. Grasping position detection using template matching and differential evolution for bulk bolts. In *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, volume 1, pages 5622–5627, 2019.

C. Gong, N. B. Erichson, J. P. Kelly, L. Trutoiu, B. T. Schowengerdt, S. L. Brunton, and E. J. Seibel. Retinamatch: Efficient template matching of retina images for teleophthalmology. *IEEE Transactions on Medical Imaging*, 38(8):1993–2004, 2019.

P. Gurunath Bharathi, A. Agrawal, S. Pareek, and A. A. Prince. Brain abnormality detection using template matching. *Bio-Algorithms and Med-Systems*, 14, 2018.

G. Holló. Optical coherence tomography angiography in glaucoma. *Turkish Journal of Ophthalmology*, 48(4):196–201, 2018.

Y.-T. Hsieh, M. N. Alam, D. Le, C.-C. Hsiao, C.-H. Yang, D. L. Chao, and X. Yao. OCT angiography biomarkers for predicting visual outcomes after ranibizumab treatment for diabetic macular edema. *Ophthalmology Retina*, 3(10):826–834, 2019.

H. Li, K. Liu, and X. Li. A comparative study of artificial bee colony, bees algorithms and differential evolution on numerical benchmark problems. In Z. Cai, H. Tong, Z. Kang, and Y. Liu, editors, *Computational Intelligence and Intelligent Systems*, pages 198–207, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

C. López-Franco, J. Hernández-Barragán, M. Lopez-Franco, M. Reynoso, E. Nuño, and A. López-Franco. Real-time image template matching algorithm based on differential evolution. In *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pages 573–578, 2016.

M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239, 1979.

A. D. Mora, J. Soares, and J. M. Fonseca. A template matching technique for artifacts detection in retinal images. *2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pages 717–722, 2013.

N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

A. Rosenfeld. Picture processing by computer. *ACM Computing Surveys*, 1(3):147–176, 1969.

O. Shahine, H. Kelash, G. Attiya, and O. Faragallah. Breast cancer detection based on dynamic template matching. *Wulfenia Journal*, 20(12):193–205, 2013.

R. Storn and K. Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

P. H. Tang, R. Jauregui, S. H. Tsang, A. G. Bassuk, and V. B. Mahajan. Optical coherence tomography angiography of RPGR-associated retinitis pigmentosa suggests foveal avascular zone is a biomarker for vision loss. *Ophthalmic Surgery, Lasers and Imaging Retina*, 50(2):e44–e48, 2019.

J. Vesterstrom and R. Thomsen. A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. In *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)*, volume 2, pages 1980–1987 Vol.2, 2004.

J. Wang, A. Camino, X. Hua, L. Liu, D. Huang, T. S. Hwang, and Y. Jia. Invariant features-based automated registration and montage for wide-field OCT angiography. *Biomedical Optics Express*, 10(1):120, 2018.

Appendix A

In order to prove that our evolutionary algorithm is less computationally expensive than template matching based on brute force, we consider the number of evaluations needed to register two images in both approaches.

Brute force

The size of our coordinate system is $(2 \cdot W_t + W_f - 1, 2 \cdot H_t + H_f - 1)$, where W_t is the width of the template image; W_f is the width of the frame image; H_t is the height of the template image; and H_f is the height of the frame image. However, center coordinates only take values in a coordinate system with size:

$$(2 \cdot W_t + W_f - 2 \cdot \frac{W_t}{2} - 1, 2 \cdot H_t + H_f - 2 \cdot \frac{W_f}{2} - 1) = (W_t + W_f - 1, H_t + H_f - 1).$$

This means that the brute force approach would need a number of fitness evaluations equal to $(W_m + W_f - 1) \cdot (H_m + H_f - 1)$. In particular, our dataset is comprised of square images, that is, images where $W_m = H_m = W_f = H_f$, so we can approximate the previous expression to $4N^2$, where N is the size of the image's side. Since our images are 320×320 px, we would need 409600 fitness evaluations.

Evolutionary algorithm

Our evolutionary algorithm initializes the population by calculating brute force solutions on downscaled images. We consider downscale factors of $\frac{1}{2^2}$ and $\frac{1}{2^3}$, meaning that the population initialization is:

$$(\frac{W_t + W_f}{4} \cdot \frac{H_t + H_f}{4} - 1) + (\frac{W_t + W_f}{8} \cdot \frac{H_t + H_f}{8} - 1) \quad \text{evaluations}$$

Again, considering that $W_t = H_t = W_f = H_f$ we can approximate the previous expression to $\frac{5N^2}{16}$. Since our images are 320×320 px, we would need 32000 fitness evaluations in order to initialize the population.

In the worst case scenario, our evolutionary algorithm will run the maximum number of generations, NG. If the population size is P, then it will be necessary $P \cdot NG$ evaluations. In our case, $30 \cdot 400 = 12000$ evaluations. We also have to consider the local search mechanism, which adds E_{local} additional evaluations per generation, that is, $(P + E_{local}) \cdot NG$ evaluations. In our case $(30 + 12) \cdot 400 = 16800$ evaluations.

In total, in the worst case scenario, our evolutionary algorithm will perform 48800 evaluations. That is about eight times less evaluations than the brute force search approach. As the size of images increases, this proportion also increases in favour of the evolutionary algorithm.

Note that our fitness function is more complex than the traditional cross correlation function. Because of this, we cannot use optimized template matching implementations such as OpenCV's. Furthermore, we do not know the template nor its size a priori. We have to extract the overlapping area in every iteration and then perform the fitness evaluation, making harder the task of developing an optimized template matching approach specifically adapted to our scenario.

Appendix B

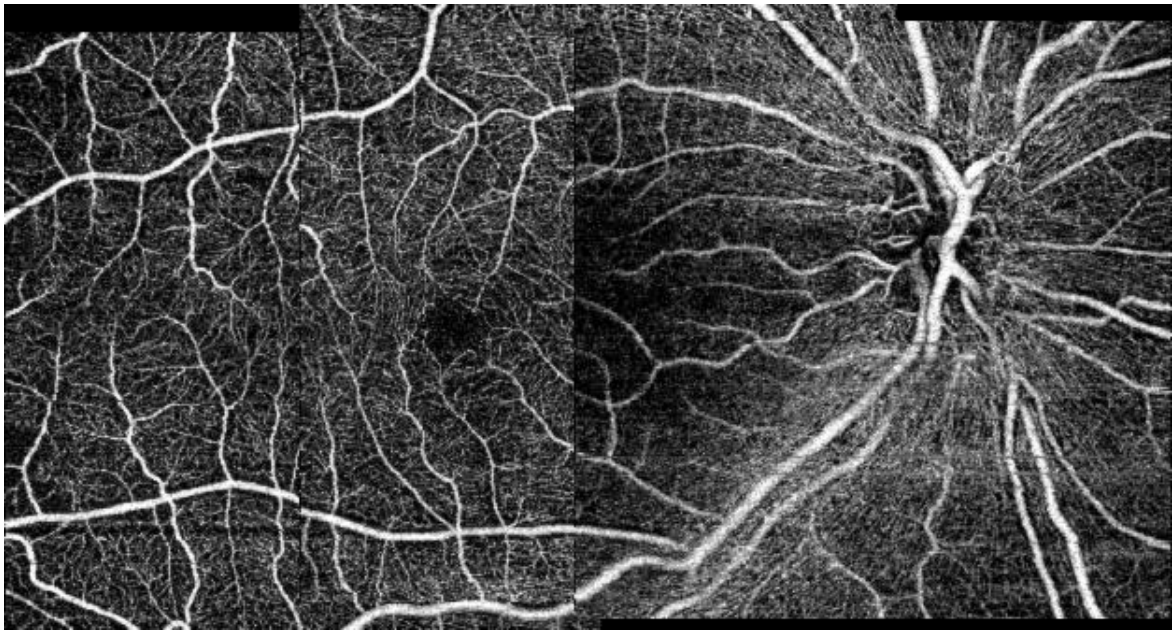


Figure 8. Mosaic for the right eye of patient 1 composed by the proposed method.

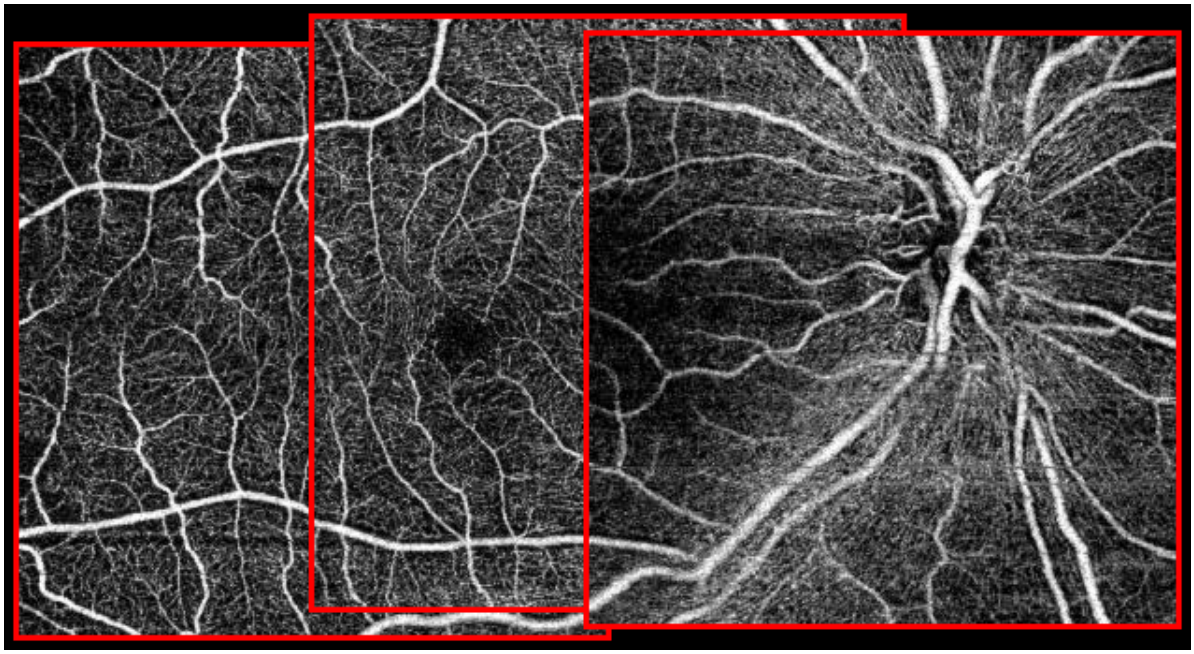


Figure 9. Mosaic for the right eye of patient 1 composed by the proposed method (the borders of each image are also highlighted).

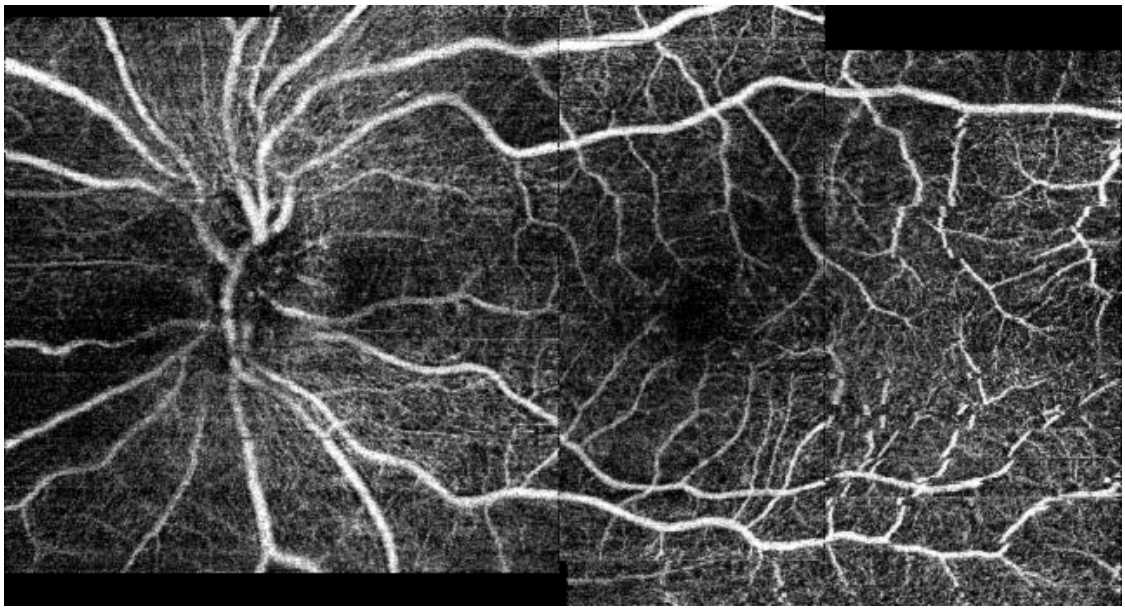


Figure 10. Mosaic for the left eye of patient 1 composed by the proposed method.

Figure 11. Mosaic for the left eye of patient 1 composed by the proposed method (the borders of each image are also highlighted).

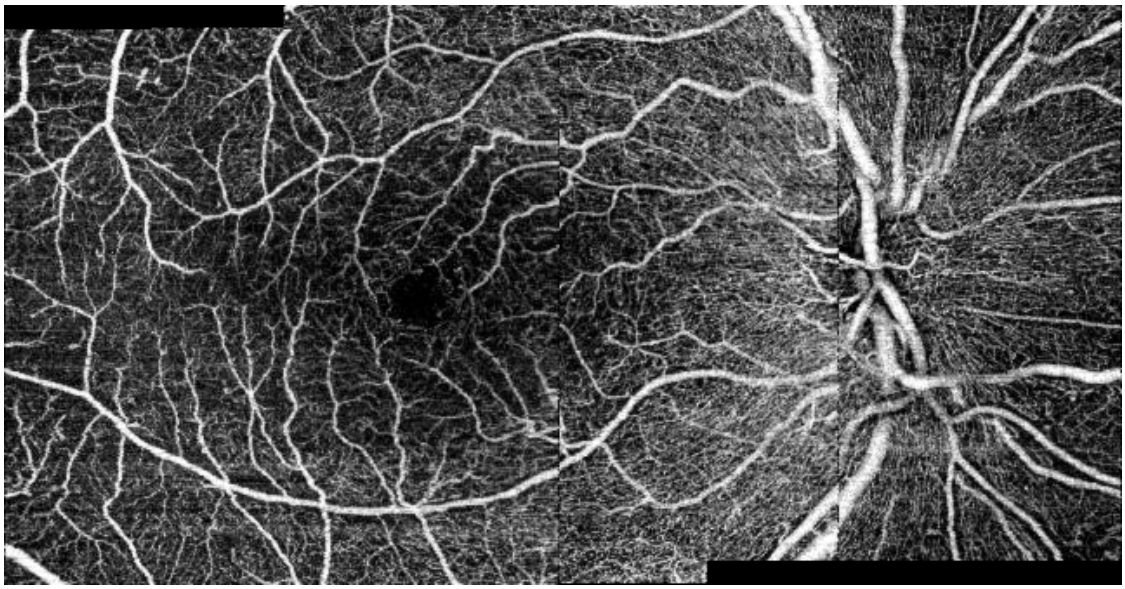


Figure 12. Mosaic for the right eye of patient 2 composed by the proposed method.

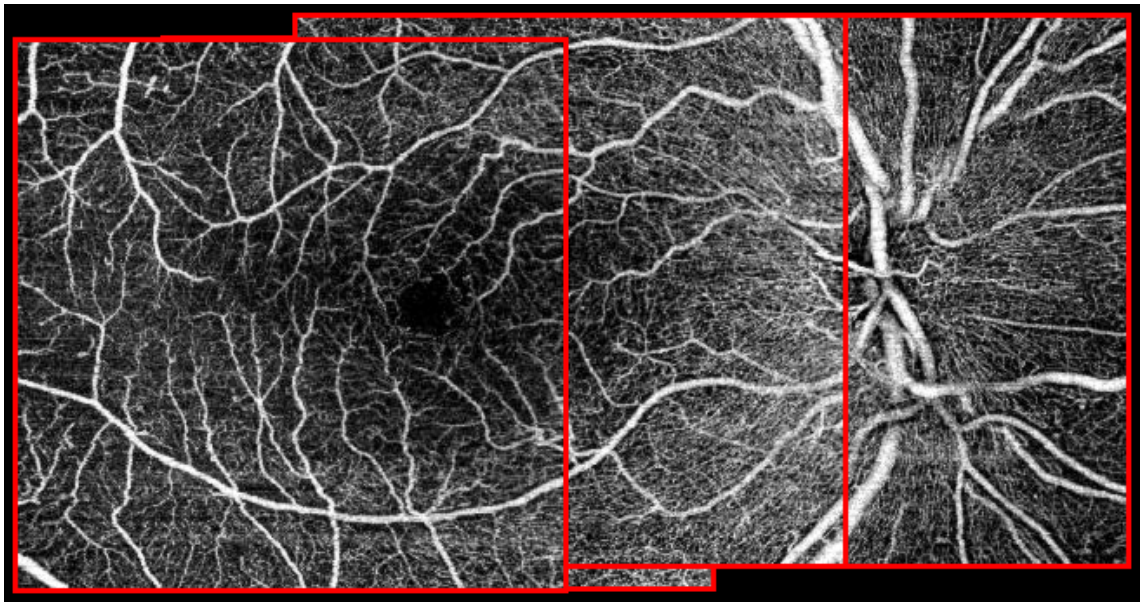


Figure 13. Mosaic for the right eye of patient 2 composed by the proposed method (the borders of each image are also highlighted).

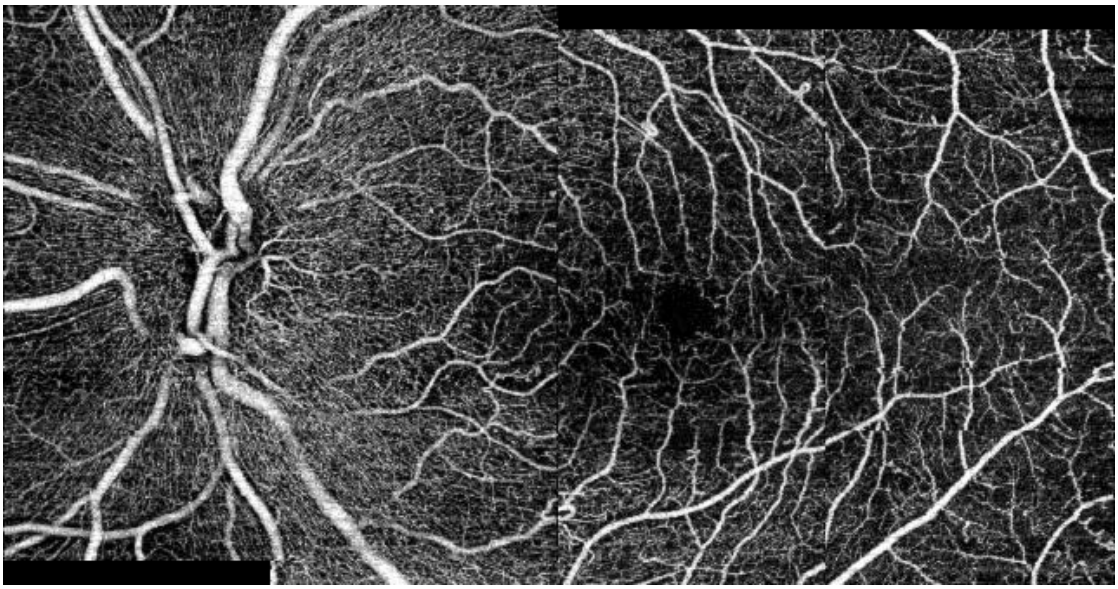


Figure 14. Mosaic for the left eye of patient 2 composed by the proposed method.

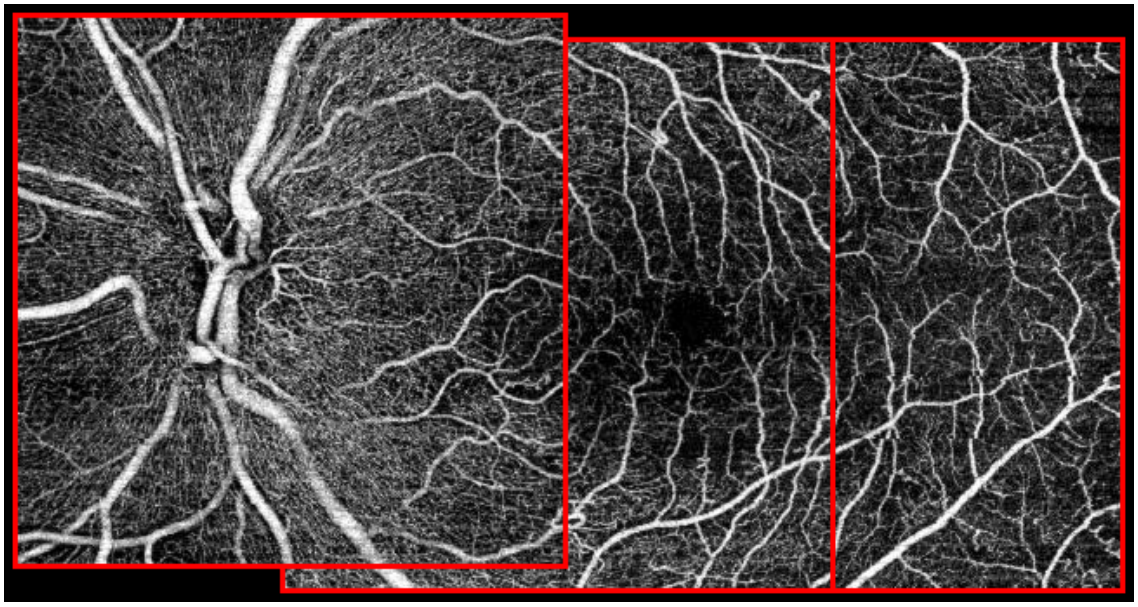


Figure 15. Mosaic for the left eye of patient 2 composed by the proposed method (the borders of each image are also highlighted).

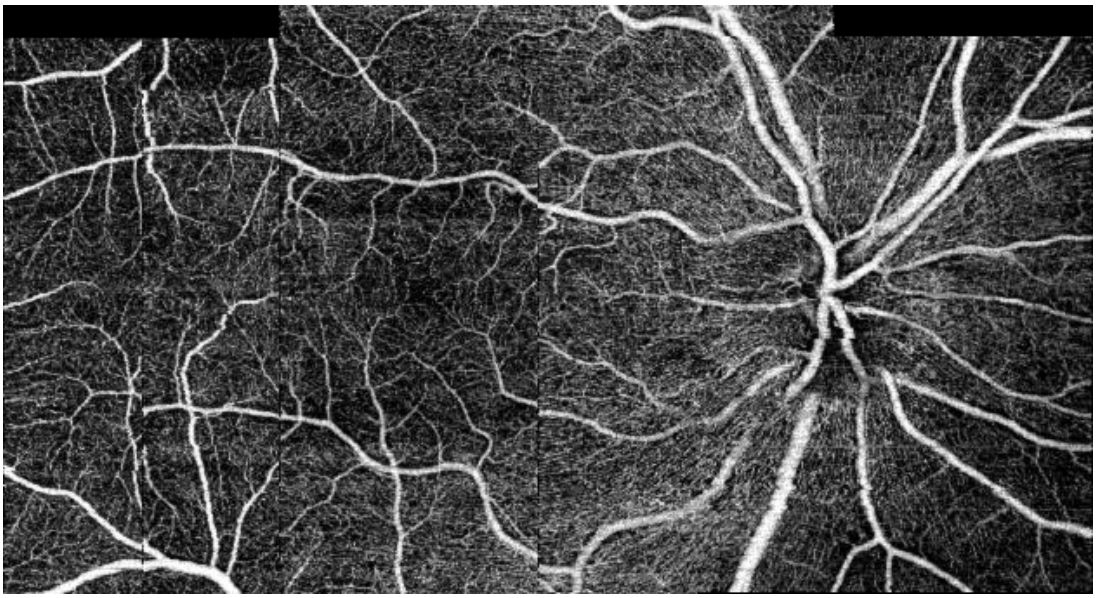


Figure 16. Mosaic for the right eye of patient 3 composed by the proposed method.

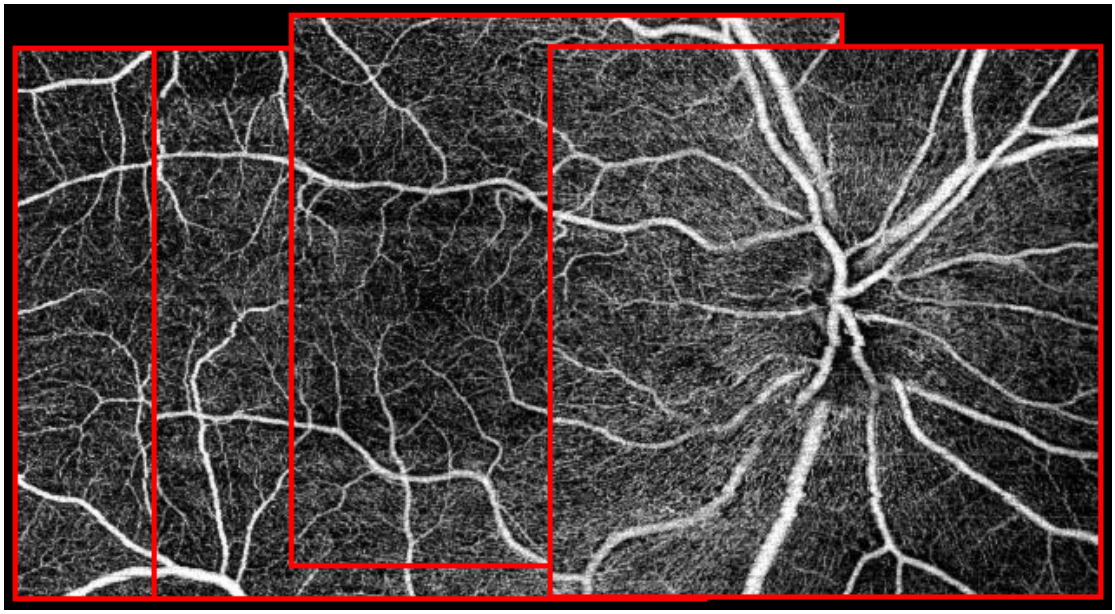


Figure 17. Mosaic for the right eye of patient 3 composed by the proposed method (the borders of each image are also highlighted).

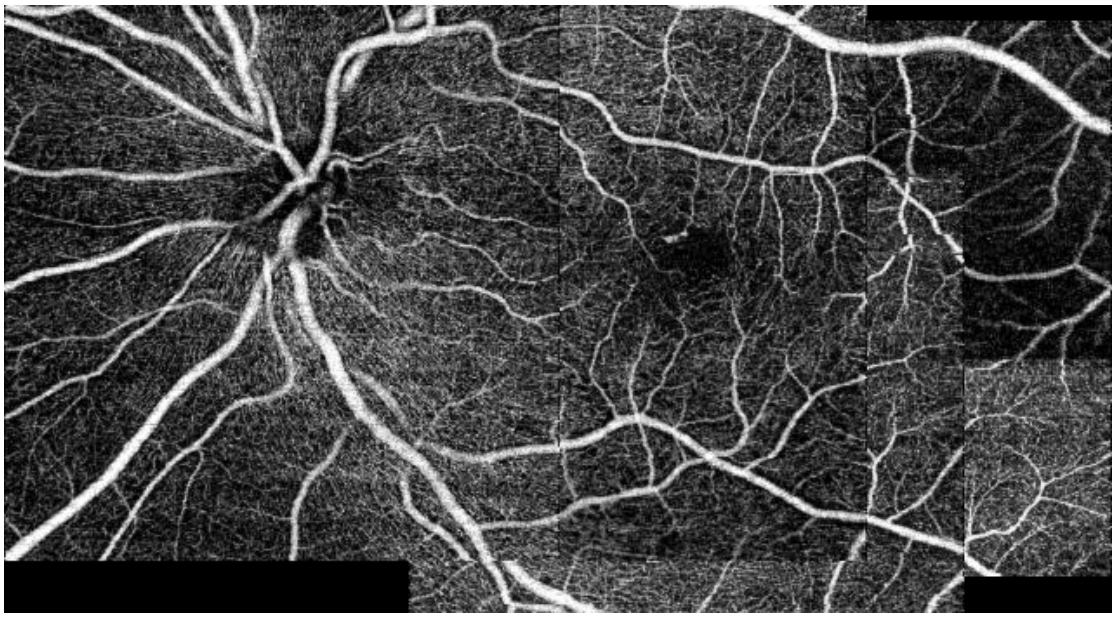


Figure 18. Mosaic for the left eye of patient 3 composed by the proposed method.

Figure 19. Mosaic for the left eye of patient 3 composed by the proposed method (the borders of each image are also highlighted).

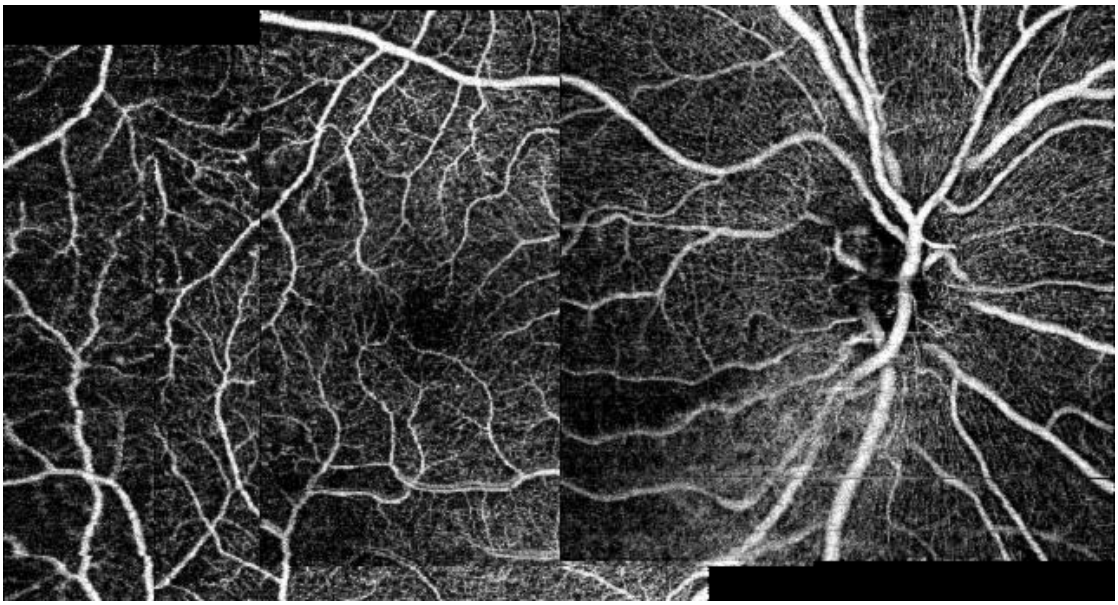


Figure 20. Mosaic for the right eye of patient 4 composed by the proposed method.

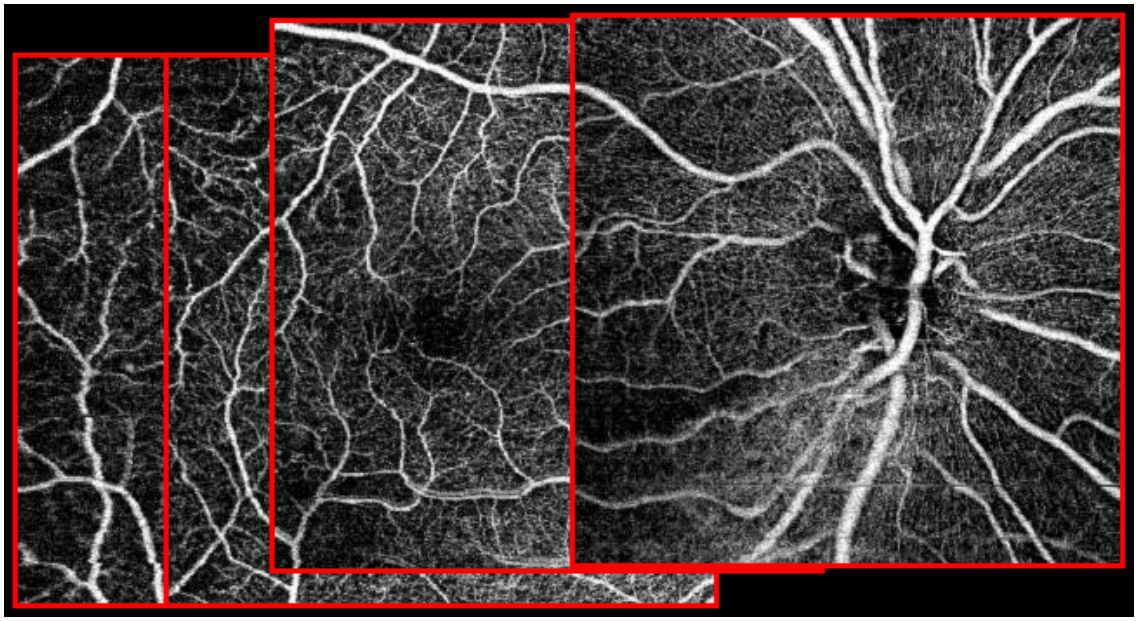


Figure 21. Mosaic for the right eye of patient 4 composed by the proposed method (the borders of each image are also highlighted).

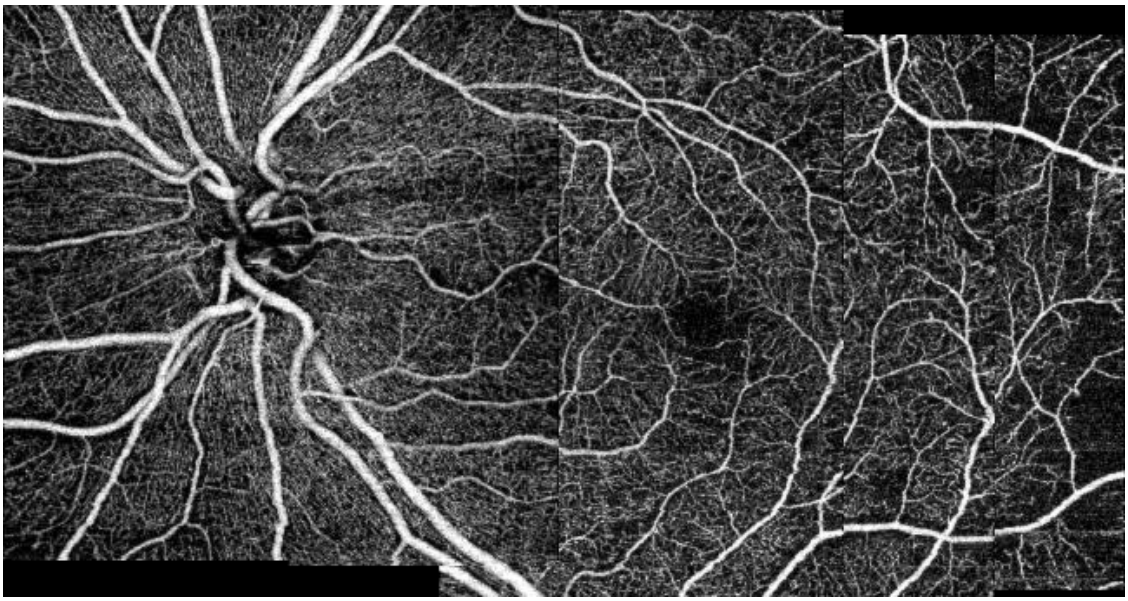


Figure 22. Mosaic for the left eye of patient 4 composed by the proposed method.

Figure 23. Mosaic for the left eye of patient 4 composed by the proposed method (the borders of each image are also highlighted).

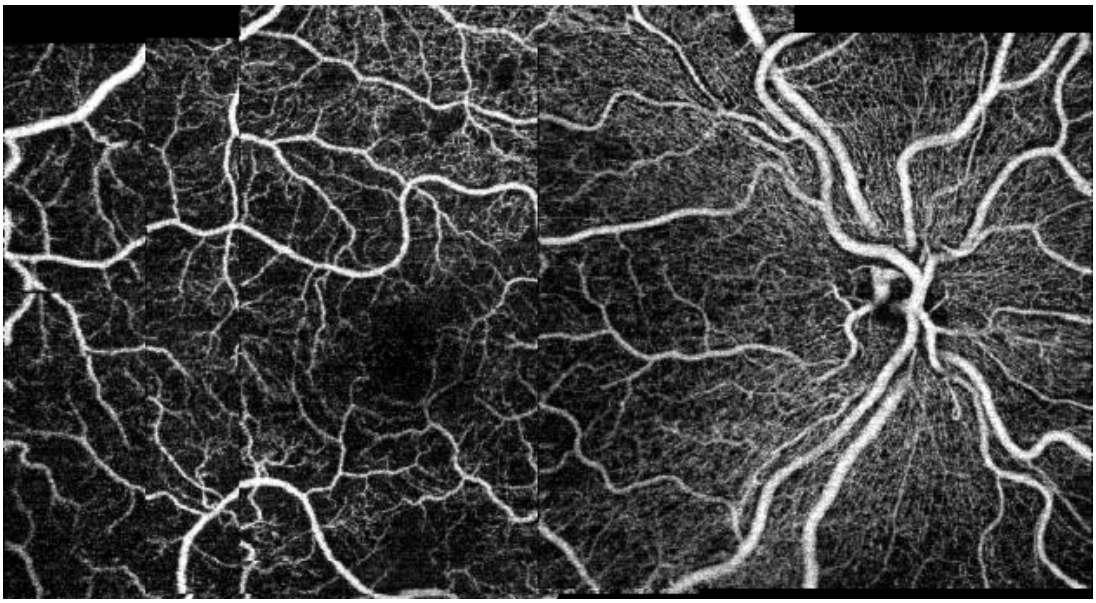


Figure 24. Mosaic for the right eye of patient 5 composed by the proposed method.

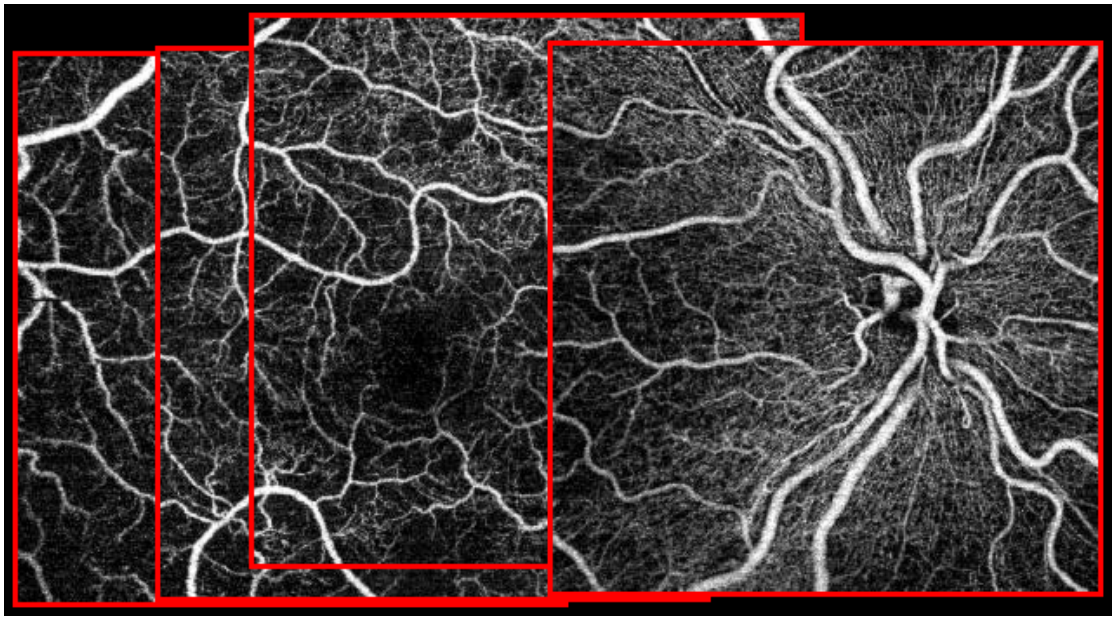


Figure 25. Mosaic for the right eye of patient 5 composed by the proposed method (the borders of each image are also highlighted).

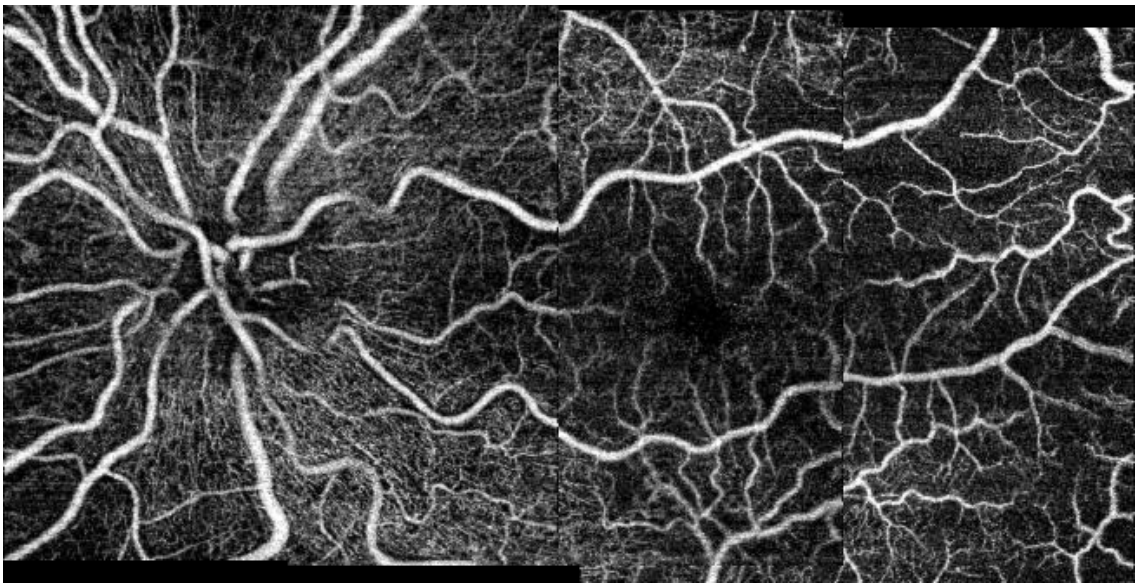


Figure 26. Mosaic for the left eye of patient 5 composed by the proposed method.

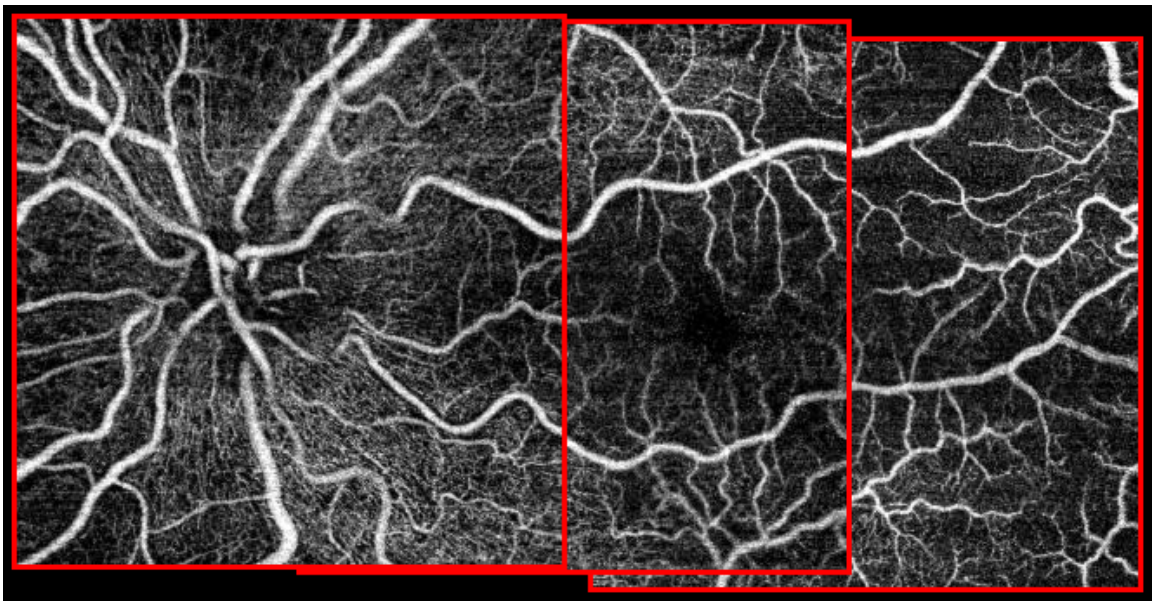


Figure 27. Mosaic for the left eye of patient 5 composed by the proposed method (the borders of each image are also highlighted).

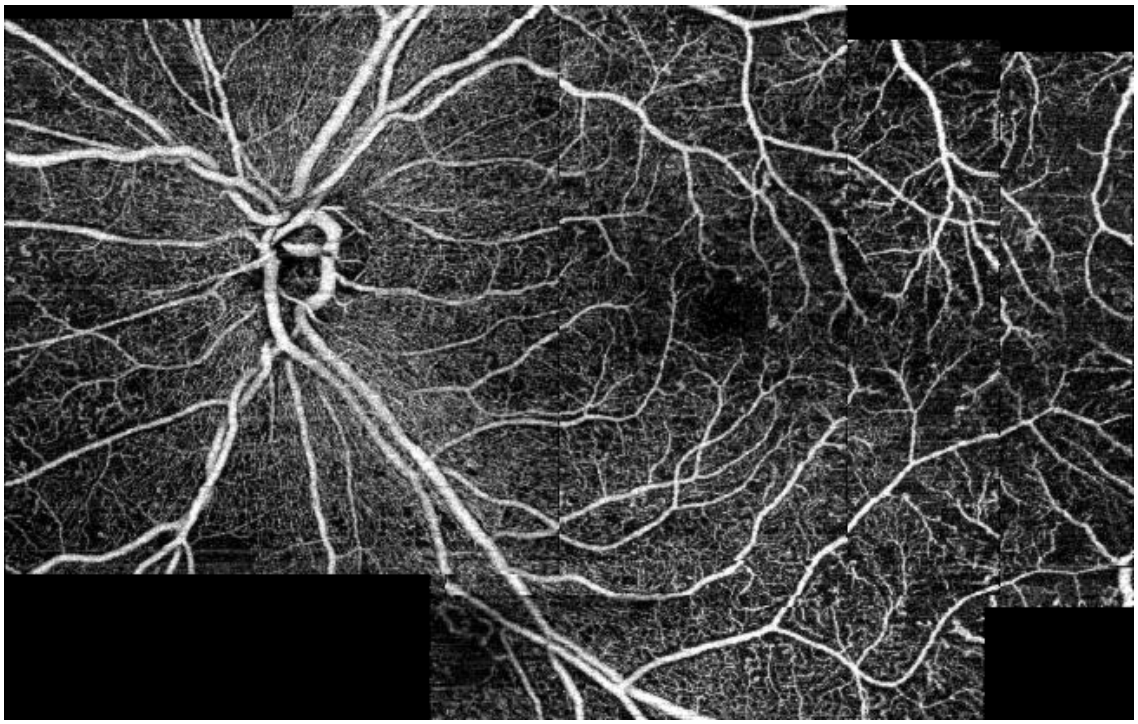


Figure 28. Mosaic for the right eye of patient 6 composed by the proposed method.

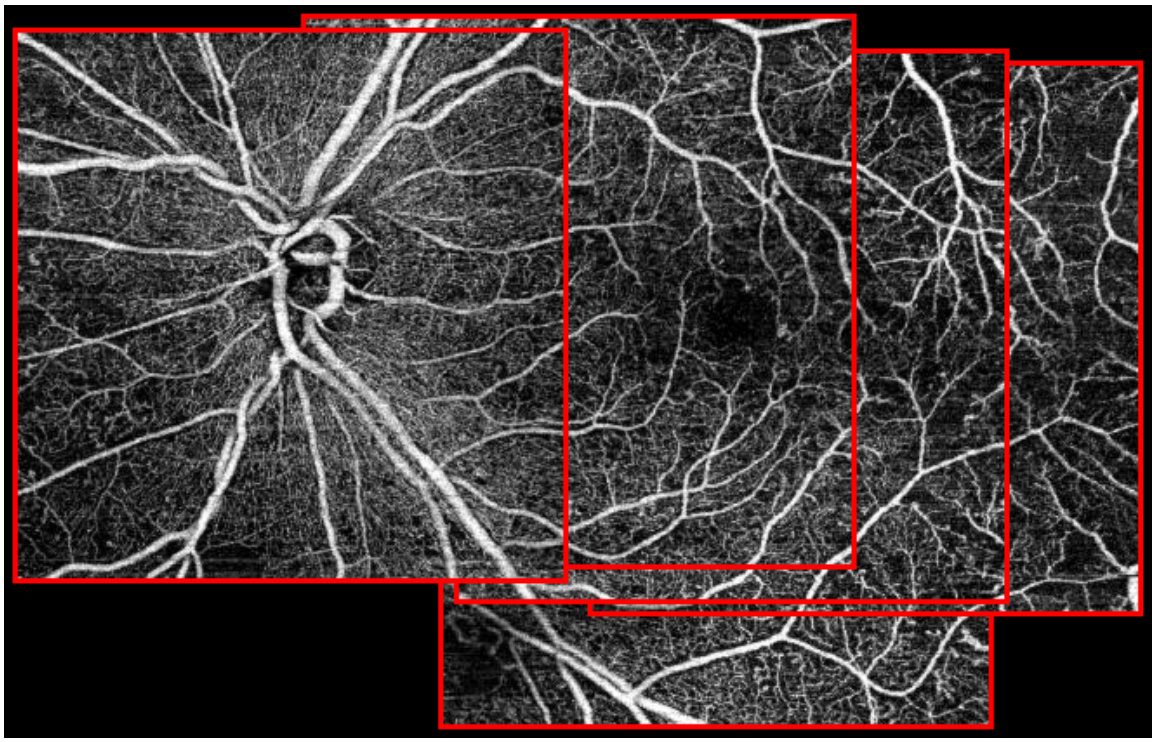


Figure 29. Mosaic for the right eye of patient 6 composed by the proposed method (the borders of each image are also highlighted).

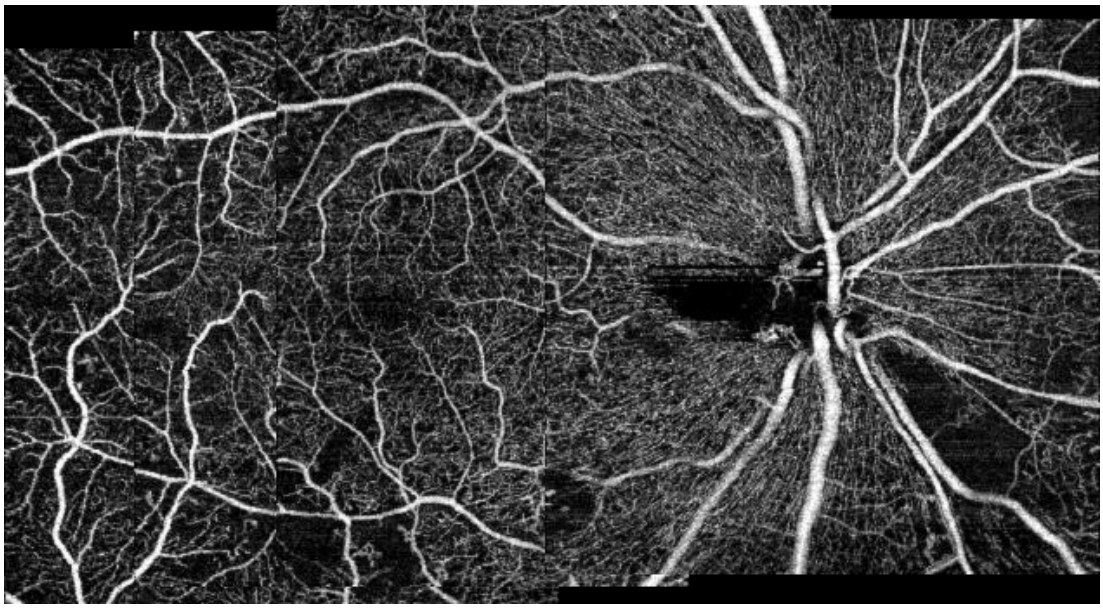


Figure 30. Mosaic for the right eye of patient 7 composed by the proposed method.

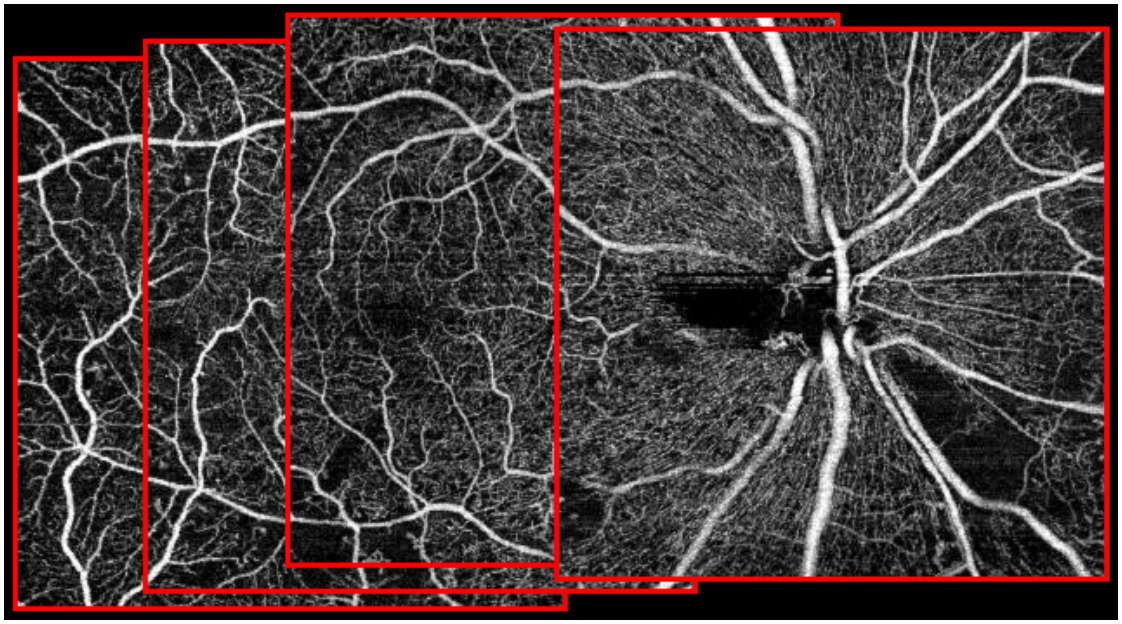


Figure 31. Mosaic for the right eye of patient 7 composed by the proposed method (the borders of each image are also highlighted).

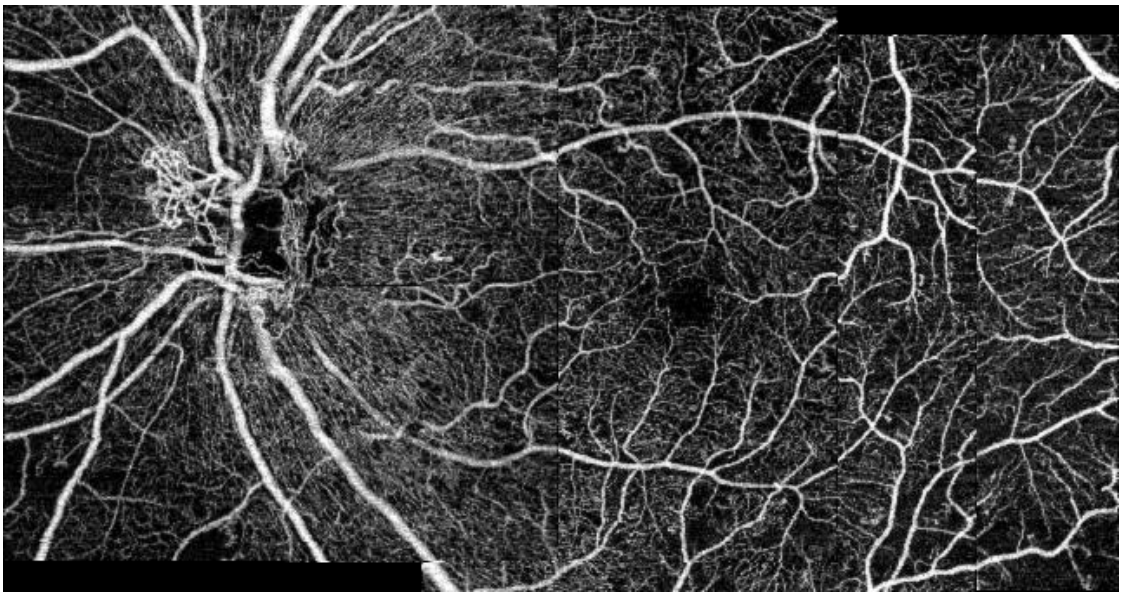


Figure 32. Mosaic for the left eye of patient 7 composed by the proposed method.

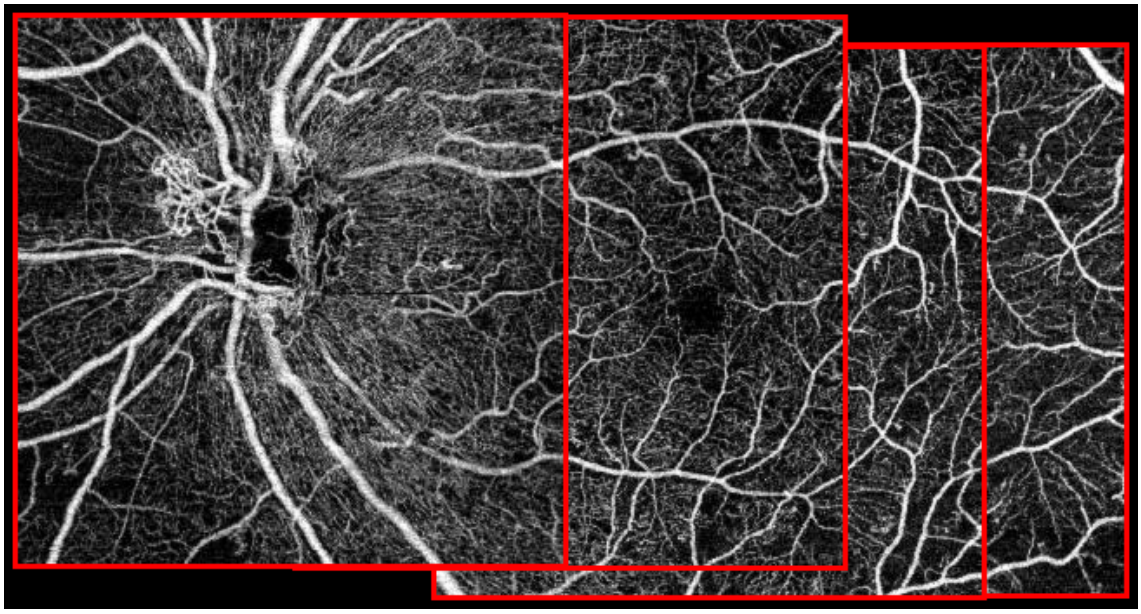


Figure 33. Mosaic for the left eye of patient 7 composed by the proposed method (the borders of each image are also highlighted).

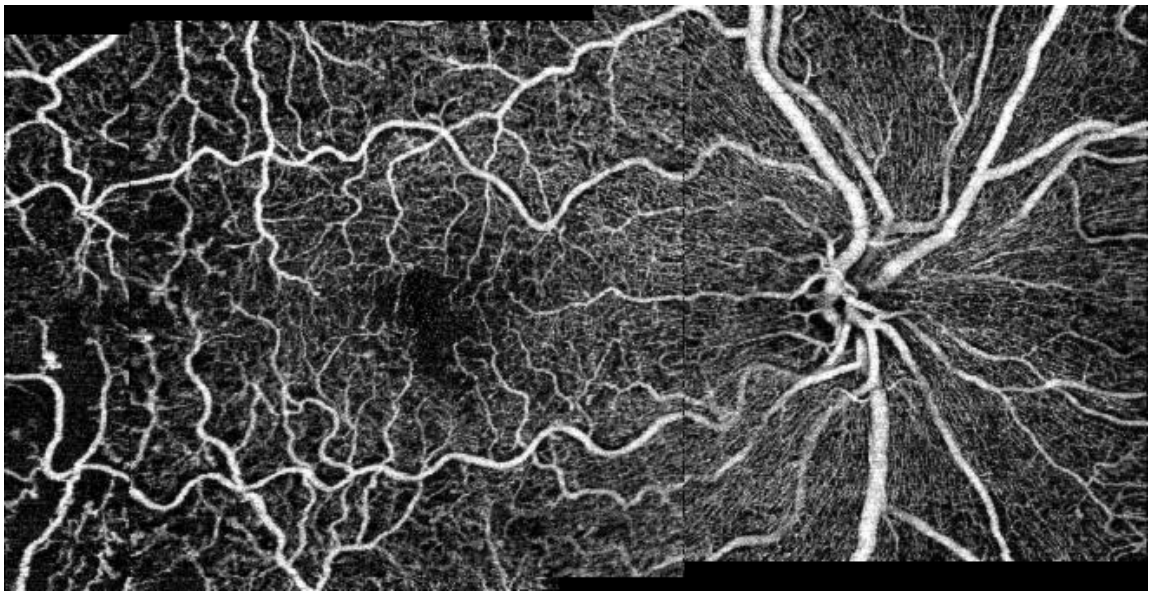


Figure 34. Mosaic for the right eye of patient 8 composed by the proposed method.

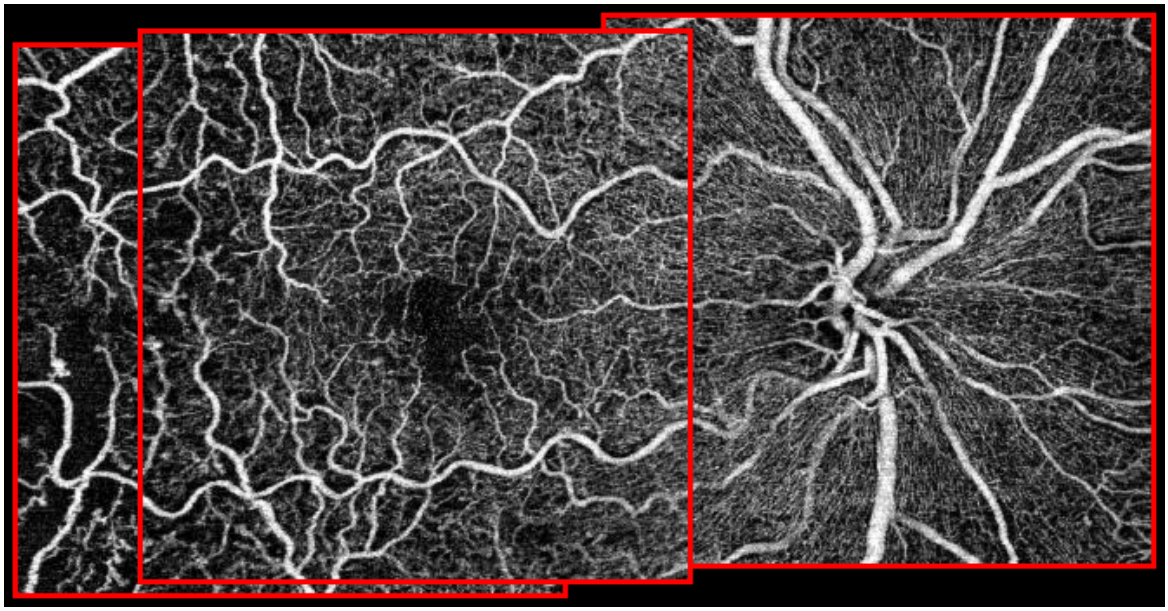


Figure 35. Mosaic for the right eye of patient 8 composed by the proposed method (the borders of each image are also highlighted).

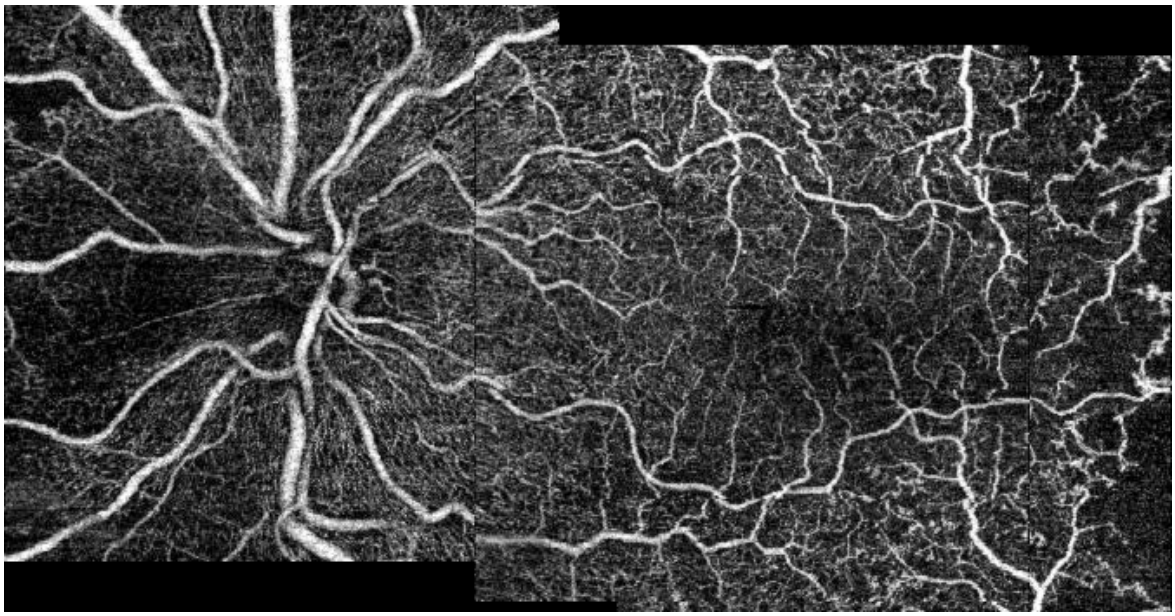


Figure 36. Mosaic for the left eye of patient 8 composed manually.

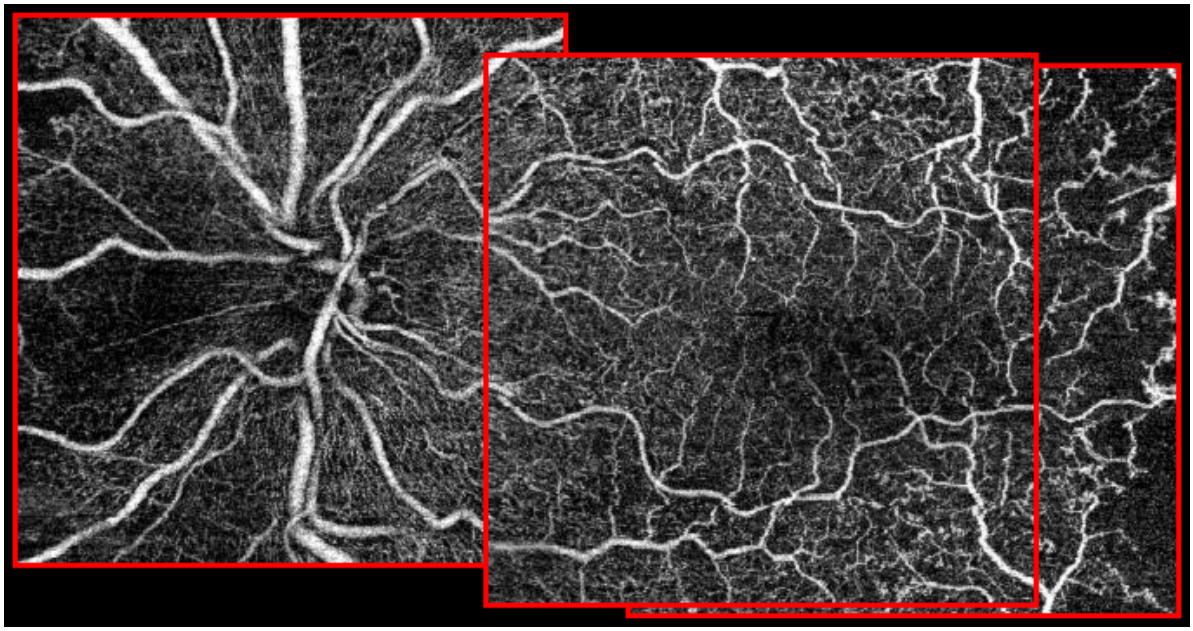


Figure 37. Mosaic for the left eye of patient 8 composed manually (the borders of each image are also highlighted).

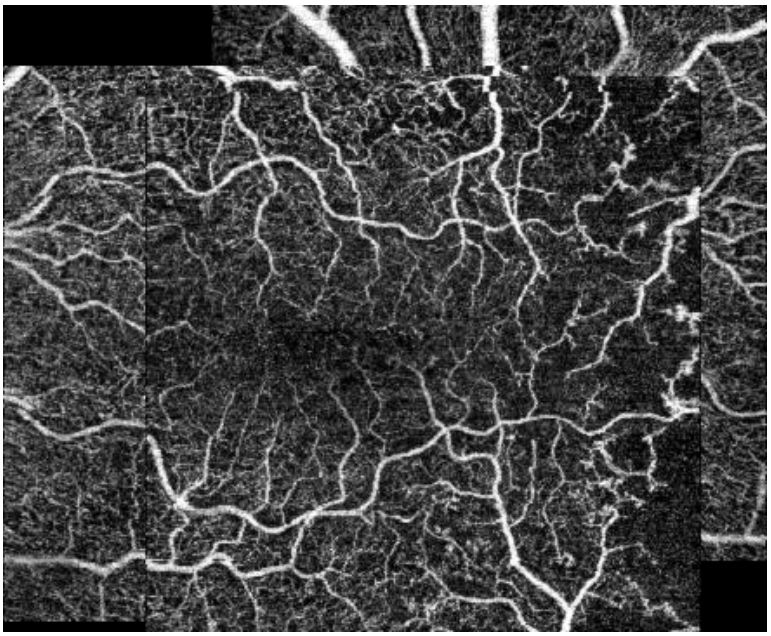


Figure 38. Mosaic for the left eye of patient 8 composed by the proposed method. The method fails in this case.

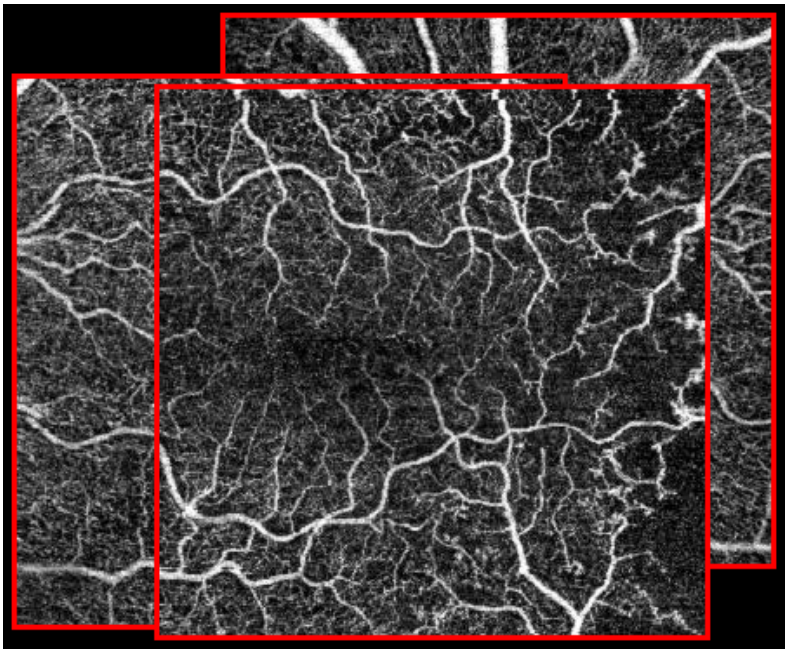


Figure 39. Mosaic for the left eye of patient 8 composed by the proposed method. The method fails in this case (the borders of each image are also highlighted).

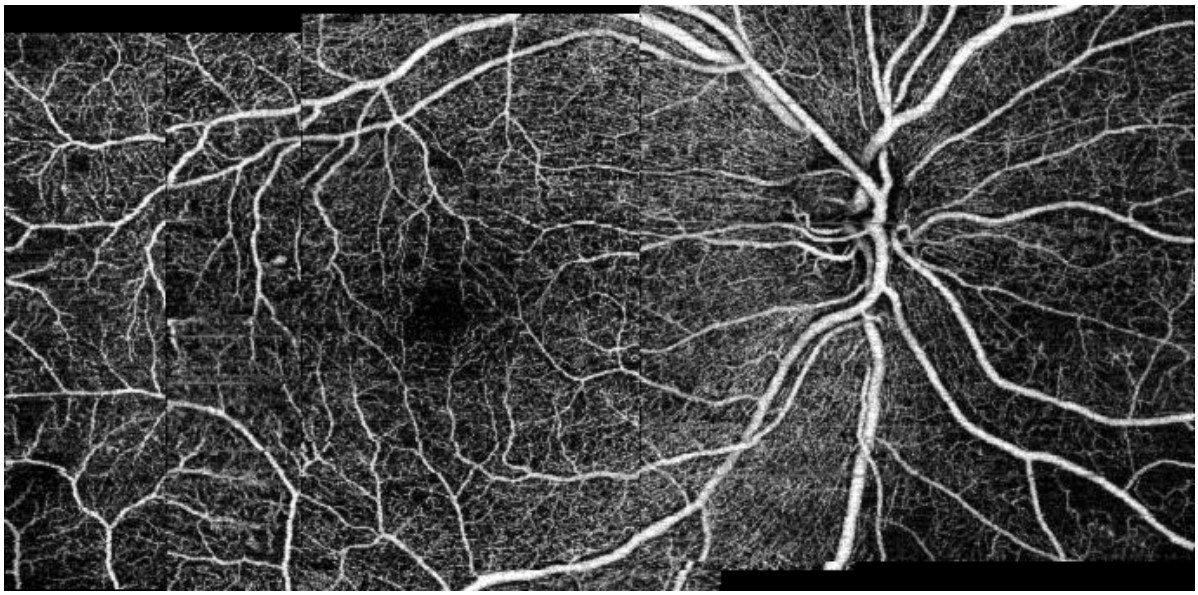


Figure 40. Mosaic for the right eye of patient 9 composed by the proposed method.

Figure 41. Mosaic for the right eye of patient 9 composed by the proposed method (the borders of each image are also highlighted).

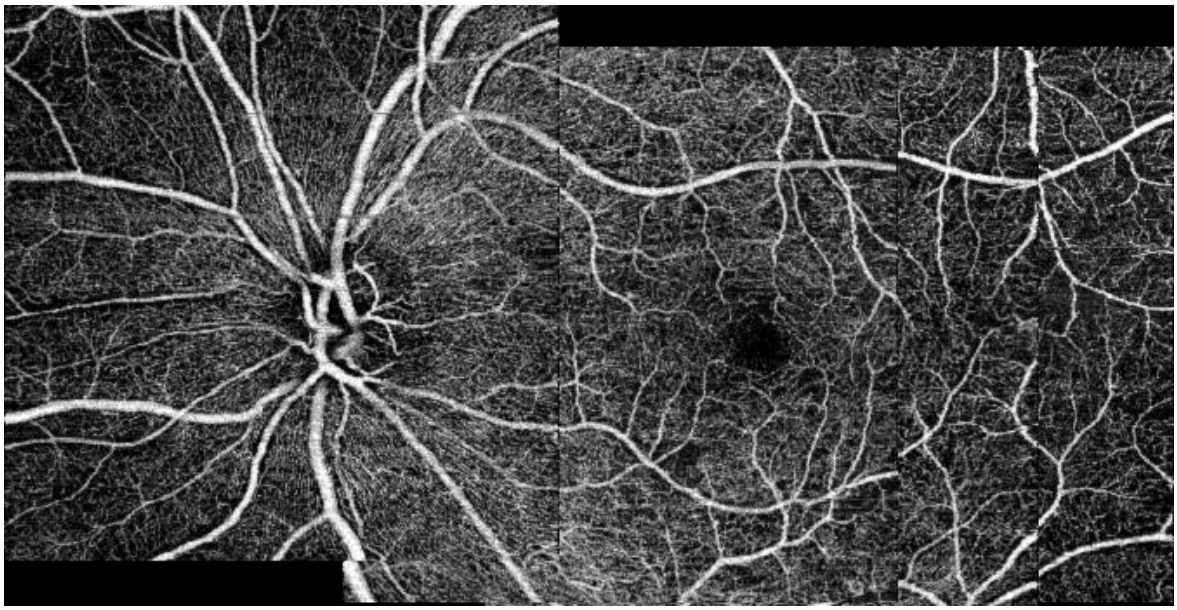


Figure 42. Mosaic for the left eye of patient 9 composed by the proposed method.

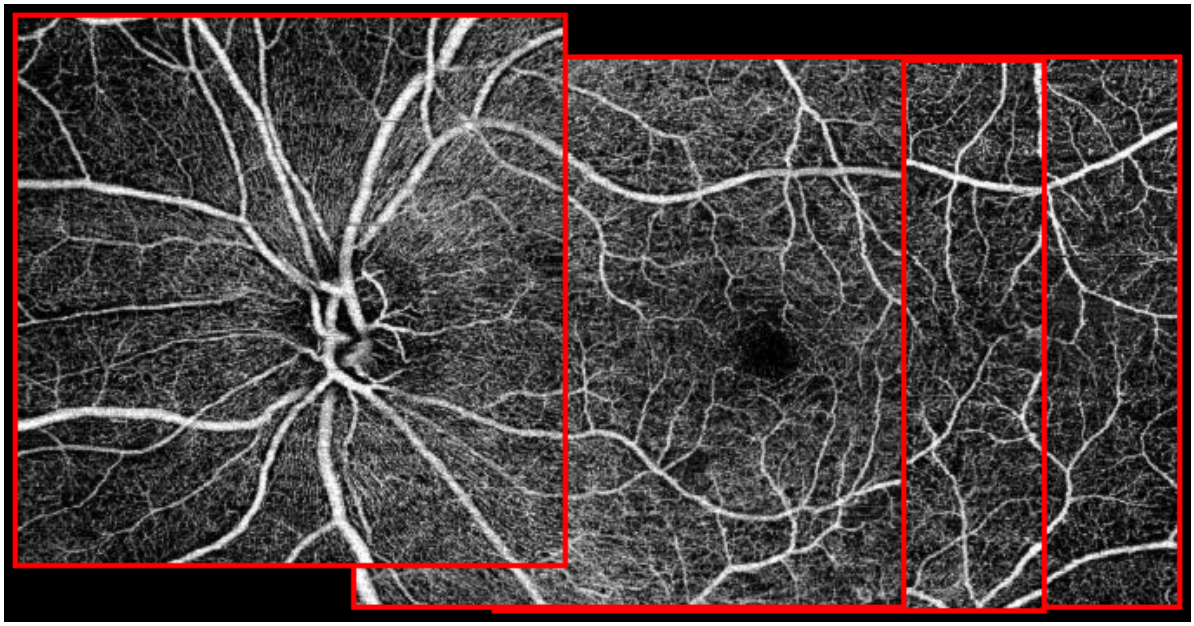


Figure 43. Mosaic for the left eye of patient 9 composed by the proposed method (the borders of each image are also highlighted).

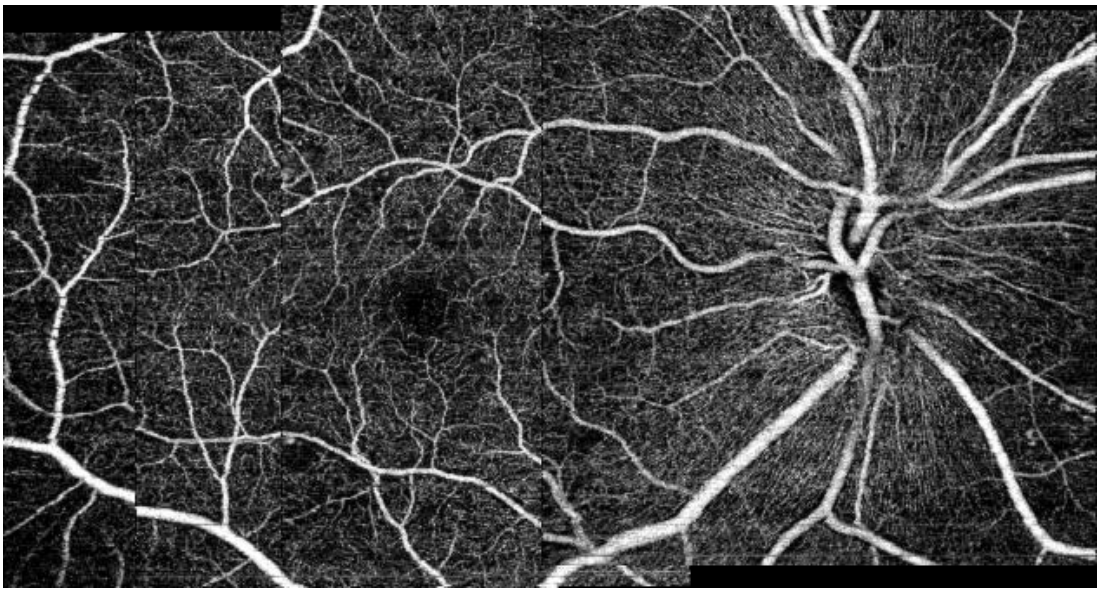


Figure 44. Mosaic for the right eye of patient 10 composed by the proposed method.

Figure 45. Mosaic for the right eye of patient 10 composed by the proposed method (the borders of each image are also highlighted).

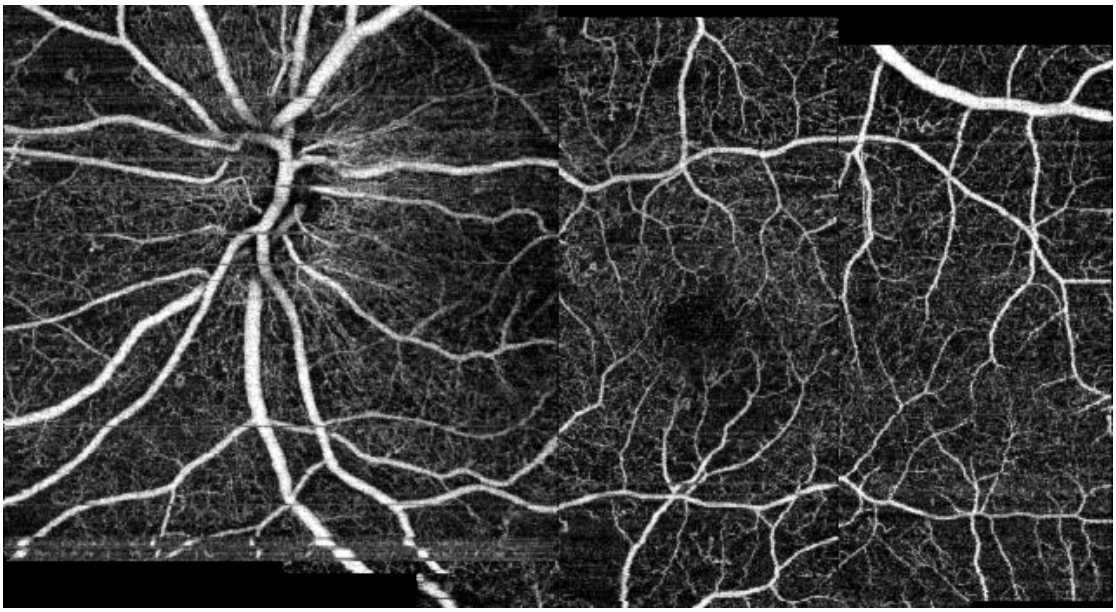


Figure 46. Mosaic for the left eye of patient 10 composed by the proposed method.

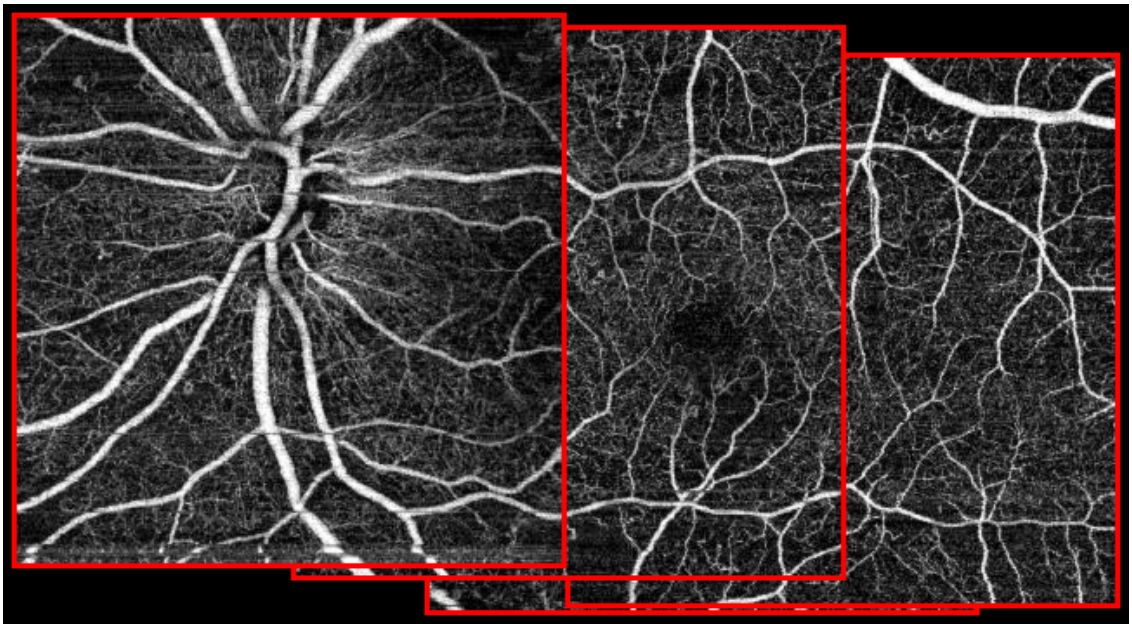


Figure 47. Mosaic for the left eye of patient 10 composed by the proposed method (the borders of each image are also highlighted).