

# Retina Mosaicing Using Local Features

Philippe C. Cattin, Herbert Bay, Luc Van Gool, and Gábor Székely

Computer Vision Laboratory, ETH Zurich, 8092 Zurich, Switzerland  
{cattin, bay, vangool, szekely}@vision.ee.ethz.ch

**Abstract.** Laser photocoagulation is a proven procedure to treat various pathologies of the retina. Challenges such as motion compensation, correct energy dosage, and avoiding incidental damage are responsible for the still low success rate. They can be overcome with improved instrumentation, such as a fully automatic laser photocoagulation system.

In this paper, we present a core image processing element of such a system, namely a novel approach for retina mosaicing. Our method relies on recent developments in region detection and feature description to automatically fuse retina images. In contrast to the state-of-the-art the proposed approach works even for retina images with no discernable vascularity. Moreover, an efficient scheme to determine the blending masks of arbitrarily overlapping images for multi-band blending is presented.

## 1 Introduction

With the dramatic demographic shift towards an older population, age-related eye diseases, such as Glaucoma and Macular Degeneration, are about to become increasingly prevalent and to develop into a serious public health issue. However, age-related eye diseases are not the only emerging causes of visual loss. Diabetic retinopathy that is directly linked to the increase of diabetes in the developed world has already surpassed age-related diseases as the main cause of blindness.

All the aforementioned potential causes of blindness can be treated with laser photocoagulation. The small bursts from the laser cause controlled damage that can be used to either seal leaky blood vessels, destroy abnormal blood vessels, reattach retinal tears or destroy abnormal tumorous tissue on the retina. Photocoagulation laser surgery is an outpatient procedure. The treatment is performed while the patient is seated in a chair and eye drops will be given to dilate the pupil, immobilise and anaesthetise the eye. The laser treatment itself is currently performed manually and lasts from a few minutes up to half an hour, depending on the type of treatment and number of laser spots to apply. Although photocoagulation is the most widely used treatment for the aforementioned conditions, the success rate is still below 50% for both the first treatment and the possible re-treatments [1]. Many of these failures can be attributed to the manual nature of the procedure, in particular to the difficulty to quickly respond to rapid eye and head movements of the patient that can not be completely stopped, and the problems to quantitatively measure and control the applied laser energy. As it takes several weeks before knowing if laser surgery has been successful or re-treatment is required to prevent further vision loss, the success rate of the

procedure has to be maximised. Additionally, the visual recovery declines with each re-treatment.

A fully automatic surgery under controllable and reproducible conditions would therefore be desirable. Computer vision methods for mosaicing retina images would be a useful tool to achieve these goals. Such mosaics could be used on the one hand for diagnosis and interventional planning, and on the other hand as a spatial map for the laser guidance system during treatment.

Many retinal image registration methods have been proposed in the literature [2,3,4]. The automatic stitching method proposed by Can *et al.* marked an important milestone in retinal image mosaicing. They used bifurcations of the vascular tree on the retinal surface to establish image correspondences. These bifurcations are easy to detect but sometimes lack distinctiveness, and their localisation is often quite inaccurate. To improve the landmark extraction Tsai proposed in [5] a model-based approach that greatly increased the accuracy and repeatability of estimating the locations where vascular structures branch or cross over. Stewart proposed in [6] a different approach that uses one or more initial correspondences defining the mapping only in a small area around these bootstrap regions. In each of these regions the transformations are iteratively refined using only information from this same area. The region is then expanded and tested to see if a higher order transformation model can be used. The expansion stops when the entire overlapping region of the images is covered. Finally, in [7] Tsai evaluated the performance of the two methods proposed by Can and Stewart in a clinically-oriented framework.

All the previously mentioned methods are limited to cases with clearly visible vascular structures. Quite often, however, bleedings or tumorous tissue limits the number of detectable bifurcations or even prevents their proper segmentation, see Fig. 3(a) for an example.

The work described below is part of a project aiming towards a fully automatic laser photocoagulation system. As will be argued, a key component of such a system is the automatic mosaicing method capable to stitch and blend the captured retina images. The method should be robust against all morphologies seen in the various pathologies subjected to laser photocoagulation.

## 2 Methods

Recent interest point detection/description schemes, such as SIFT [8] have brought about important progress in image mosaicing [9,10]. Yet, in our experiments, these state-of-the-art methods failed to identify sufficient number of reliable, distinct landmarks allowing to build retinal mosaics. The use of our very recently developed framework [11] proved more successful.

The method for obtaining a mosaic image from multiple images of the retina, using region detectors/descriptors, can be summarised with the following steps. (1) Interest points are detected at different scales and their feature descriptors are calculated in each image individually. (2) The region descriptors are robustly matched for each image pair to find the overlapping images using the second-nearest-neighbour ratio matching strategy [12]. (3) A graph theoretical algorithm

is used to find the *anchor* image that is connected to all the other images through the shortest path. (4) The mappings from all images into the coordinate system of the anchor image are linearly estimated and then globally optimised. (5) All images are warped into the coordinate frame of the anchor image and multi-band blending is used to seamlessly fuse the individual retina images. The subsequent paragraphs describe the aforementioned steps in more detail.

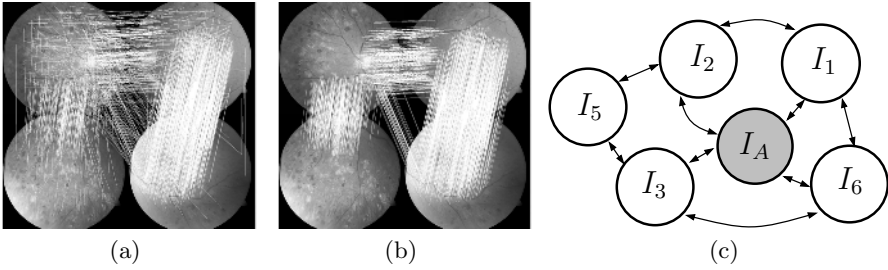
(1) *Interest point detection and feature descriptors*: The first task for image mosaicing is the identification of a possibly high number of interest points in all images of the same scene. As described in the introduction, the previously presented retina mosaicing methods used the branches and crossovers of the vascular tree as interest points. Although this approach works very well it is bound to fail for pathological retinas, where the vascular tree is seriously occluded by bleeding or tumorous tissue, see Fig. 3(a) for an example. Only our recent developments in interest point detection and characterisation [11] made the application of this technology possible for the highly self-similar regions generally seen in retina images. Our algorithm proved to be superior to the state-of-the-art methods with regard to repeatability and stability. It uses a Hessian-matrix based detector in order to identify blob-like interest points and a 128-dimensional distribution-based vector to describe the neighbourhood of every interest point.

Given a set of  $N$  input images  $I_1, \dots, I_N$  for the construction of the mosaic, let each image  $I_i$  have a number  $N_i$  of interest points. As a first step the image coordinates  $(u, v)$  of these interest points are extracted yielding  $p_{i,1}, \dots, p_{i,N_i}$ . Their respective feature descriptors are defined by  $c_{i,1}, \dots, c_{i,N_i}$ . Once the interest points have been found, the overlapping images can be identified by establishing the correspondences between all image pairs.

(2) *Matching*: Given a pair of images  $I_i, I_j$  with their respective interest points and feature descriptors, for every interest point in the first image  $I_i$ , we calculate the euclidian distance to all feature descriptors in the second image  $I_j$ . If the ratio of the nearest neighbour  $e_{1NN}$  and the second-nearest neighbour  $e_{2NN}$  is smaller than the predefined threshold of 0.8, see [8,10], a match is assumed to be correct and is therefore added to the list of putative matches, see Fig. 1(a). Note, that any given interest point  $p_{j,1..N_j}$  in  $I_j$  can (in contrast to the interest points  $p_{i,1..N_i}$  in  $I_i$ ) be matched by more than one interest point from image  $I_i$ .

Even if the human retina is not planar, the images appear flat as the ophthalmoscope captures only a small part of the whole retina. Therefore, homographies can be used to roughly approximate the transformation of the putative interest point matches for a given image pair. Let  $\mathbf{X}$  be a point on a planar surface with projections,  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , in two images taken from different viewpoints. The homography  $\mathbf{H}$  describes the transformation that connects  $\mathbf{x}_i$  and  $\mathbf{x}_j$  for any point  $\mathbf{X}$  on that surface, thus  $\mathbf{x}_j = \mathbf{H}\mathbf{x}_i$ . By robustly estimating the homography with the classical RANSAC scheme [13], the transfer error  $d(\mathbf{x}_j, \mathbf{H}\mathbf{x}_i)$  can be used to filter out possible mismatches from the list of putative matches.

As the initial matches are paired using a very conservative feature vector comparison, many pairs are not matched in the first step. The advantage of



**Fig. 1.** (a) Correspondences after matching SURF features (b) after the guided matching. (c) Example graph of the connected images after pairwise matching.

this initial conservative selection is that the ratio of inliers vs. outliers is better and thus a correct homography can be found using RANSAC. Once the initial homography is established, more correspondences can be found with a *guided* matching method. Knowing the homography, the final correspondences are established by ignoring the feature descriptors and just matching interest points if they are spatially close enough, see Fig. 1(b) for an example.

(3) *Anchor Image Selection:* Once the correspondences between all image pairs have been determined, the anchor image  $I_A$  is identified. The selection of the anchor image among the available images  $I_A \in I_1, \dots, I_N$ , plays a crucial role in image mosaicing as the anchor defines the base coordinate system towards which all other images are warped. A central image having direct correspondences with all other ones in the data set would be an ideal candidate. However, such an image usually does not exist. The next logical candidate would thus be an image that connects to all other images through the least number of connections, see Fig. 1(c). Graph theoretical methods [14] allow to calculate the minimum distances between nodes given the connectivity matrix. If more than one node is connected to the rest with the minimal number of connections, the node with the highest number of correspondences with its neighbouring images is selected.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} & \theta_{15} & \theta_{16} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} & \theta_{25} & \theta_{26} \end{pmatrix} \begin{pmatrix} x^2 \\ xy \\ y^2 \\ x \\ y \\ 1 \end{pmatrix} \quad (1)$$

(4) *Mapping Estimation:* Even if overlapping regions of image pairs can be related with homographies, a planar transformation model for the mosaic would result in a false representation of distances and angles. The curved nature of the retina can best be taken into account by using a quadric transformation model  $\Theta$  as proposed by Can *et al.* in [4]. It transforms a point  $\mathbf{x} = (x, y)^\top$  of an input image to a point  $\mathbf{x}' = (x', y')^\top$  in the anchor image  $I_A$  coordinate system.

In a first step the transformations  $\Theta_{i,A}$  for each image  $I_i \in I_1, \dots, I_N \setminus I_A$  overlapping with the anchor image  $I_A$  are estimated solving the linear equation

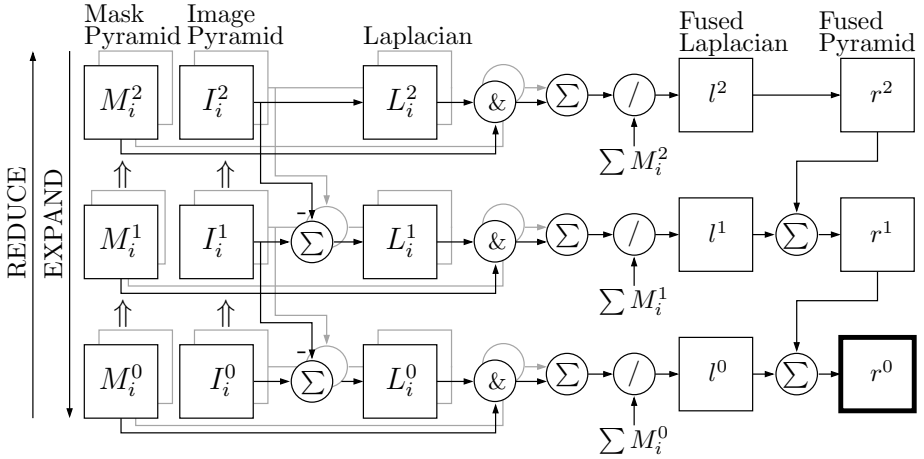
system with SVD shown in Eq. 1. For all other images having no overlap with the anchor image a different, indirect approach is followed. As an example, the mapping of image  $I_5$  to the anchor image  $I_A$  in Figure 1(c) can be estimated indirectly via the already known transformations  $\Theta_{2A}$  and  $\Theta_{3A}$ . To ensure the most accurate estimation for the indirect transformations, all possible paths<sup>1</sup> are used when building the linear equation system. These initial transformation estimates  $\Theta_{i,A}$  are then subject to a global optimisation in which the sum of squared reprojection errors is minimised (Sec. 5.2 in [15]).

(5) *Image Warping and Blending:* At this stage, the geometrical relationship between the anchor image  $I_A$  and all its surrounding images is known. In a first step, all images  $I_i$  are warped into the spatial domain of the anchor image  $I_A$  forming a new image  $I_i^0$ . As only the forward mapping  $\Theta_{iA}$  is known and inverting this transformation proved to be difficult [15], interpolation is used to calculate all pixels in  $I_i^0$ . The intensity at each pixel location  $\mathbf{x} = (x, y)^\top$  in  $I_i^0$  is the weighted average of all the intensities in  $I_i^0$  falling within a pixel radius.

With all images mapped into the same coordinate system, the images are combined to form a single visually appealing mosaic with invisible image boundaries. Ideally, corresponding image pixels should have the same intensity and colour on every image. In practice, however, this is not the case. Possible causes are inhomogeneous illumination, changes in aperture and exposure time, mis-registrations of the mosaicing procedure, etc. Thus, advanced image blending methods are required to yield visually pleasing results. In order to preserve the fine structures seen on retina images while smoothing out low frequency variations caused by irregular illumination, we decided to use the multi-band blending method proposed by Burt [16]. The idea behind this approach is to blend the high frequencies over a small spatial range and the low frequencies over a large spatial range. This can be performed for multiple frequency bands. In our implementation, we consider 6 octaves using a Laplacian pyramid.

The main difficulty with this approach is the design and implementation of suitable and computationally efficient *blending masks* for the arbitrarily overlapping images. In order to determine this blending mask for every image, we assigned a weight  $w_i(x, y)$  to each pixel in an image that varies from 1 in the center of the image to 0 at the edge. The choice of this weighting function was inspired by the observation that (a) retina images tend to be more focused in the middle and (b) the illumination gets inferior with increasing distance from the image center. The weights have to be warped into the anchor coordinate system together with the image pixels. In case pixels from multiple images are mapped to the same pixel location in the anchor domain, the pixel of image  $I_i^0$  with the strongest weight  $w_i(x, y)$  is considered. By evaluating this decision for each pixel location we obtain a binary blending mask  $M_i^0$  for each warped image  $I_i^0$ . For the blending to work properly, see [16], pixels at blending borders have to be averaged. This can be easily implemented by dilating all the masks  $M_i^n$  at each level  $n$  of the Laplacian pyramid by one pixel and then dividing the fused

<sup>1</sup> Applied to the example in Fig. 1(c)  $\Theta_{5A}$  is estimated via  $\Theta_{2A}$  as well as via  $\Theta_{3A}$ .



**Fig. 2.** A summary of the steps involved for the multi-band blending of two images (the steps for the second image are in light gray) over three octaves. First, the warped images  $I_i^0$  and their blending masks  $M_i^0$  are used to generate the Gaussian pyramid levels  $I_i^1, I_i^2, \dots$  and masks sizes  $M_i^1, M_i^2, \dots$ . The Laplacian pyramid is computed as the difference between adjacent Gaussian levels (one of them must be expanded first). The Laplacian pyramid is then masked with the dilated blending mask ( $\&$ ) and summed for every image in the pyramid. Finally, the resulting blended image  $r^0$  is generated by expanding and summing all levels of the pyramid.

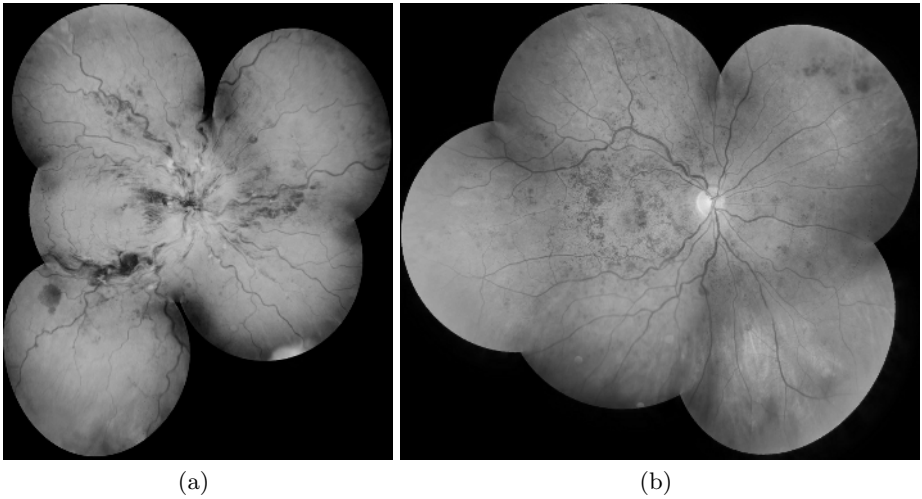
Laplacian image at each level by the number of images that contributed to a certain pixel. The data flow diagram to blend two images over three octaves is depicted in Fig. 2.

### 3 Results

We have applied the proposed technique to acquisitions from a patient database with retinal images showing various pathologies. In particular, we also investigated cases where the vascular tree is not clearly visible and thus the state-of-the-art methods would fail. The images were taken with a high resolution digital camera  $2256 \times 2032$  pixels but scaled to half their size, i.e.  $1128 \times 1016$  pixels, to assess the performance on the more common lower resolution images.

We evaluated the robustness of the matching step that automatically finds overlapping image pairs using a set of 100 pairs. The matching method failed in 6 cases. All of these cases had an overlap of less than 20.2% (average 11.4%). On the other hand it successfully managed to match 12 image pairs with an overlap of 20% or less (average 14.3%). Per image pair an average of 311 correspondences (max 1763, min 9) were found and none for the failed cases.

The sample mosaics depicted in Fig. 3 give a visual impression of the efficacy of the proposed method. It can be clearly seen, that even retinas without visible vascularisation in the overlapping region can be correctly mosaiced. The average reprojection error for these mosaics was  $1.01 \pm 0.01$  pixels.



**Fig. 3.** Two contrast enhanced sample mosaics with different pathologies composed of (a) 6, (b) 7 retina images and multi-band blending over 6 octaves

## 4 Conclusions

In this paper, we described an approach to mosaic retina images relying on our recently published results for region detection and characterisation [11]. We demonstrated that the SURF method can be efficiently used to mosaic highly self-similar retina images even for cases with no discernable vascularisation. We also proposed an efficient way to determine the blending masks required for multi-band blending of arbitrarily overlapping images.

The algorithm is currently integrated into the interventional planning system. At the same time we are working on approaches allowing to utilise the applied feature matching method for intra-operative navigation support.

## Acknowledgments

This work has been supported by the CO-ME/NCCR research network of the Swiss National Science Foundation (<http://co-me.ch>). We thank the University Hospital in Berne, Switzerland for providing the retina images.

## References

1. Zimmer-Galler, I., Bressler, N., Bressler, S.: Treatment of choroidal neovascularization: updated information from recent macular photocoagulation study group reports. *Int. ophthalmology clinics* **35** (1995) 37–57
2. Hart, W.E., Goldbaum, M.H.: Registering retinal images using automatically selected control point pairs. In: *Proc. IEEE International Conference on Image Processing (ICIP)*. Volume 3. (1994) 576–81

3. Becker, D.E., Can, A., Turner, J.N., Tanenbaum, H.L., Roysam, B.: Image processing algorithms for retinal montage synthesis, mapping, and real-time location determination. *IEEE Trans. on Biomedical Engineering* **45** (1998) 105–18
4. Can, A., Stewart, C.V., Roysam, B.: Robust hierarchical algorithm for constructing a mosaic from images of the curved human retina. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Volume 2. (1999) 286–92
5. Tsai, C.L., Stewart, C.V., Tanenbaum, H.L., Roysam, B.: Model-based method for improving the accuracy and repeatability of estimating vascular bifurcations and crossover from retinal fundus images. *IEEE Transactions on Information Technology in Biomedicine* **8** (2004) 122–30
6. Stewart, C.V., Tsai, C.L., Roysam, B.: The dual-bootstrap iterative closest point algorithm with application to retinal image registration. *IEEE Trans. Med. Imaging* **22** (2003) 1379–1394
7. Tsai, C.L., Majerovics, A., Stewart, C.V., Roysam, B.: Disease-oriented evaluation of dual-bootstrap retinal image registration. In: *MICCAI*. Volume 2878 of LNCS. (2003) 754–761
8. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* (2004)
9. Brown, M., Lowe, D.G.: Recognising panoramas. In: *10th International Conference on Computer Vision*. (2003) 1218–25
10. Brown, M., Szeliski, R., Winder, S.: Multi-image matching using multi-scale oriented patches. In: *International Conference on Computer Vision and Pattern Recognition (CVPR)*. (2005) 510–517
11. Bay, H., Tuytelaars, T., Gool, L.V.: SURF: Speeded up robust features. In: *ECCV*. Volume 3951 of LNCS. (2006) 404–17
12. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. *Transactions on Pattern Analysis and Machine Vision* (2005) Accepted to PAMI.
13. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University Press (2000)
14. Diestel, R.: *Graph Theory*. New York: Springer-Verlag (1997)
15. Can, A., Stewart, C., Roysam, B., Tanenbaum, H.: A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: Application to mosaicing the curved human retina. In: *IEEE Conf. on Computer Vision and Pattern Recognition*. Volume 2. (2000) 585–91
16. Burt, P.J., Adelson, E.H.: A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics* **2** (1983) 217–36