Local Binary Pattern Matching for Fast Retina Map Relocalization Using the Slit-Lamp

Rodrigo T.F. Linhares, Rogério Richa, Ricardo Moraes,
Eros Comunello, and Aldo von Wangenheim

Brazilian National Institute for Digital Convergence - INCoD
`rodrigo.linhares@posgrad.ufsc.br`

Abstract. Computer assistance has the potential for increasing safety and accuracy during retinal laser treatment using the slit-lamp. In this context, intra-operative retinal mapping is a fundamental requirement to overlay relevant pre-operative information for surgeons. Retinal mapping using the slit-lamp is a challenging task, due to disturbances such as lens distortions, occlusions and glare. Such disturbances have a negative impact on the duration of the mapping procedure, consequently affecting its acceptance in clinical practice. To cope with these visual tracking interruptions, we propose a fast retina map relocalization strategy based on template-matching, using local binary patterns, which are suitable for the retina's texture. We perform extensive experiments to show the superior accuracy and computational efficiency of the proposed approach in comparison with feature-based methods.

1 Introduction

The slit-lamp biomicroscope is the most commonly used device for laser treatment of sight-threatening diseases, such as diabetic retinopathy and retinal vein occlusions. These pathologies generate large areas of retinal ischemia, triggering the proliferation of abnormal neovascularization, consequently leading to bleeding and retinal detachment. Laser treatment can prevent blindness by destroying ischemic areas. In this context, computer assistance can aid surgeons by overlaying relevant information intra-operatively, such as pre-operative plans and images (i.e. fundus or angiographic images), thus improving surgery safety and efficiency. Toward this purpose, the creation of an accurate intra-operative retina map is required. In the literature, similar methods for intra-operative view expansion and mapping have been proposed in the domains of minimally invasive surgery [14], confocal endomicroscopy [2,12], vitreo-retinal surgery [11] and fundus imaging [4]. In the domain of slit-lamp imaging, the pioneer work reported in [1] has proven the feasibility of an intra-operative retina mapping system.

Intra-operative retinal mapping using the slit-lamp is a challenging task. Disturbances such as occlusions, glare and patient motion often affect visual tracking, slowing the mapping process and limiting the size of the retina map. Thus, techniques for recovering from tracking loss are essential for reducing the duration of the mapping procedure. In all previously cited methods, feature detection
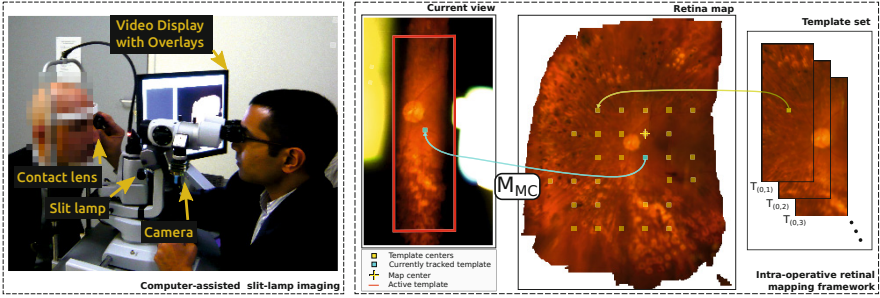
Fig. 1. (Left) The prototype slit-lamp device with information augmentation. (Right) The retinal mapping method - Using direct visual tracking, a retina map is built on-the-fly by registering adjacent templates. The retina mosaic is the result of the blending of multiple overlapping templates, whose centers are indicated by yellow squares.

and matching is used for relocalization purposes. However, retinal images acquired by the slit-lamp are very rich with texture, with less salient features than alternative imaging modalities (e.g. the fundus camera). Consequently, feature-based relocalization yields poor results, even when computationally expensive bootstrapping techniques [11] are used.

Focusing on this limitation, we have developed a fast relocalization method for intra-operative retinal mapping based on local binary patterns (LBP). Inspired in direct visual SLAM techniques [8] and recently proposed approaches for online site retargeting for optical biopsy [15], we perform relocalization by matching the current retina view against the retina map using an image similarity function based on the LBP. The LBP has the advantage of translating local texture information into a very compact code with relative illumination invariance, which allows us to perform template matching at much faster rates than similarity measures such as the normalized cross correlation or mutual information. An extensive experimental analysis reveals that the proposed template-matching approach is significantly more accurate and computationally efficient than feature-based approaches.

This paper is organized as follows. In the next section, we briefly describe the framework for retinal mapping and introduce the relocalization method based on the LBP. In Section 3, we present the extensive experiments conducted on our dataset for evaluating the practical value of the proposed strategy. Our conclusion and future work can be found in Section 4.

2 Methods

Our framework for visual tracking and mosaicking of the retina is inspired in [11]. A schematic overview is given in Figure 1. During the exam, only a small section of the retina is visible. A threshold is applied to every image acquired by the camera coupled to the biomicroscope to extract the visible part of the retina. As the human retina absorbs most of the blue spector of light, a pixel
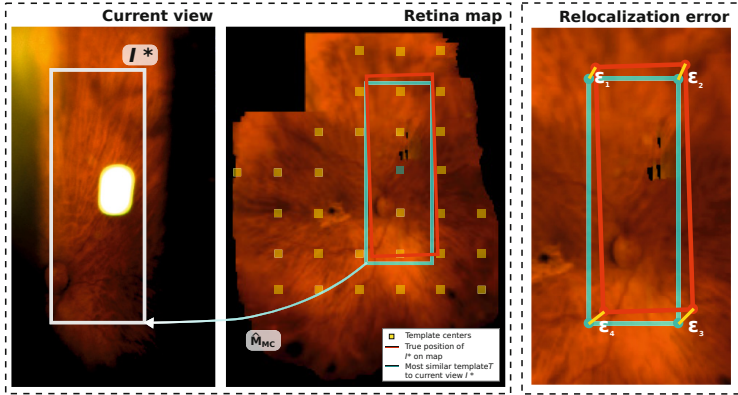
Fig. 2. The proposed template-based relocalization method. The map to camera transformation M_{mc} is approximated as the position of the template on the retina map which produces the highest score $\mathcal{L}(I^*, T)$. The relocalization error evaluated in the experimental section is shown in detail on the right.

is considered valid if $r - 0.7g > 0$, where r and g are the its red and green intensities. To initialize the mapping method, a reference image of the retina is selected. The center of this reference represents the origin of the map.

The retina map is composed of several overlapping templates T. They are evenly spaced on a grid, with significant overlap between adjacent templates. They are illustrated in Figure 1 as $T_0, T_1, ...$. During the mapping procedure, the template closest to the current view of the retina is tracked using the direct tracking method described in [13]. Due to the large overlap area between templates, the current focus of the tracker can be easily switched to a more suitable template, closer to the current position in the mosaic. The transformation M_{mc}, which maps pixel positions on the retina map onto the current retina view, is updated at every frame. For tracking the retina, we define M_{mc} as a rotation and translation model. As the operator explores the retina, additional templates are incorporated as the distance between the current visible part of the retina and the map origin increases. More details on the mapping framework can be found in [11] or in the supplementary material.

Due to frequent occlusions and degradations, tracking confidence can drop below a minimum threshold and tracking is deemed lost. To re-initialize tracking and continue the map expansion, we propose the fast LBP-based relocalization method described next.

2.1   Map Relocalization Using an LBP-Based Similarity Score

The relocalization problem consists in estimating the transformation M_{mc} to re-initialize tracking and continue the map expansion. Since patients are placed on a head rest, this transformation can be reduced to a simple translation for relocalization purposes. The most common approach to re-initialize M_{mc} is through
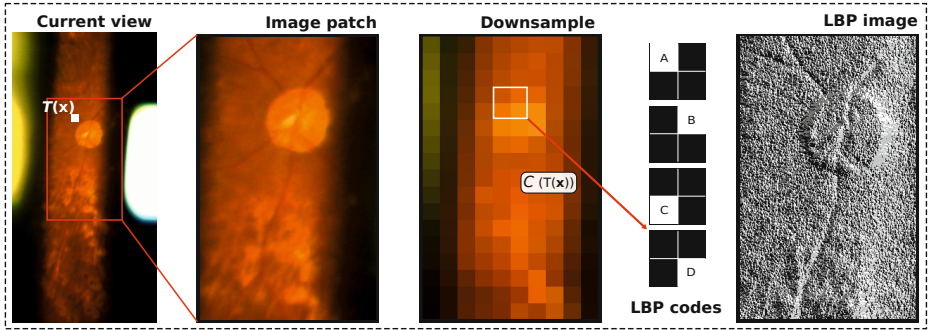
Fig. 3. Labeling process for the creation of an LBP image. Each input pixel is assigned a code $\mathcal{C}(T(\mathbf{x})) \mapsto \{A, B, C, D\}$. The code indicates the position of the brightest pixel among the input pixel and its 3 immediate lower right neighbors. The resulting labeled image is shown in the far right.

feature detection and matching. In [11], M_{mc} is estimated by SURF feature matching, using a combination of FLANN and RANSAC. Feature-based methods work well when a sufficient number of features is available. However, stable features can only be found in areas close to the optic nerve bundle, which makes relocalization impossible in the retinal periphery.

Inspired by template-based SLAM relocalization techniques [8], we solve the relocalization problem by matching the current image with the templates in the retina map. Taking advantage of the large overlap between adjacent templates, the transformation M_{mc} is approximated as the position of the template on the retina map which produces the highest similarity score in comparison with the current retina view. The proposed matching procedure is illustrated in Figure 2. First, a template-sized patch I^* centered on the visible part of the current frame is extracted. The matching process consists in evaluating the similarity between every pair I^* and T_i, where $i \in [0, N]$ and N represents the number of templates in the retina map at a given time.

Several image similarity measures can be used for template matching. Since illumination conditions can vary significantly, robust functions such as the normalized cross correlation (NCC) [8] and mutual information [6] are preferred. However, these similarity functions are computationally expensive and their calculation for a large number of templates is too slow for our purposes. To circumvent this issue, we have developed a fast approach for template matching using Local Binary Patterns.

LBP is essentially a compact texture descriptor, which encodes the relative intensity difference between neighboring pixels. Several LBP codes have been proposed in the literature [10]. In our work, we adopt the compact four-bit code used in [7,15]. The idea is illustrated in Figure 3. For a given image T, a code $\mathcal{C}(T(\mathbf{x})) \mapsto \{A, B, C, D\}$ is assigned for every pixel $\mathbf{x}$. Each code indicates the position of the brightest pixel among $\mathbf{x}$ and its 3 immediate lower right neighbors.
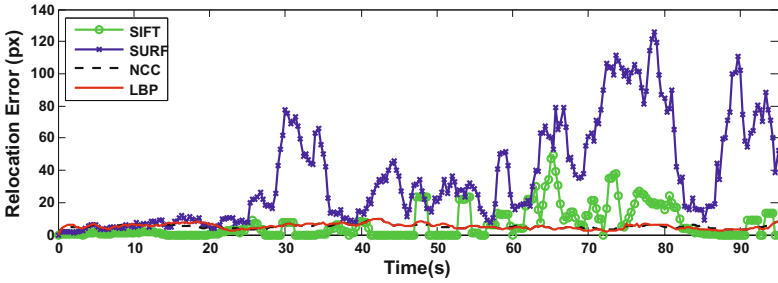
Fig. 4. Relocalization error of all four methods in a given retina video sequence. One can clearly notice the superior performance of template-based over feature detection and matching methods.

The LBP-based similarity $\mathcal{L}$ between a pair of images I and T is measured as the number of pixels with matching labels:

$$\mathcal{L}(I,T) = \sum_{\mathbf{x}} \Big[\mathcal{C}(I(\mathbf{x})) \wedge \mathcal{C}(T(\mathbf{x})) \Big] \tag{1}$$

where $\wedge$ represents a label comparison, i. e., $\mathcal{C}(I(\mathbf{x})) \wedge \mathcal{C}(T(\mathbf{x})) = 1$ if $I(\mathbf{x})$ has the same label as $T(\mathbf{x})$, and 0 otherwise.

Another aspect of the LBP-based similarity $\mathcal{L}$ is the scale of the input images. We chose to downsample images by a factor $\alpha = 0.5$ to 'lock' on a specific pattern frequency. However, experiments have shown that the discriminability of $\mathcal{L}$ is not affected for values of α within a reasonable range ($\alpha \in [0.3, 0.9]$).

3   Experiments

The proposed relocalization method was tested on a large database of videos recorded from several human patients using the device illustrated in Figure 1. The objective of the experiments is to show the superior performance of the template-based relocalization compared to similar feature-based relocalization methods used in the literature. We also highlight the advantages of the proposed LBP-based similarity score in terms of computational efficiency compared to similarity measures such as the NCC. The resulting mosaic size was fixed at 560x620 so that it could contain the whole area explored by the surgeon in each video. All methods were implemented using OpenCV and parallelized using OpenMP. The system ran on an Intel i5 3.2GHz computer with 8GBytes of RAM. All tracking parameters and system settings remained constant throughout the experiments. Some resulting videos can be found through this link.

3.1   Comparative Tests

To show the superior accuracy of the proposed LBP-based matching method over feature-based approaches, we perform a comparative study between several
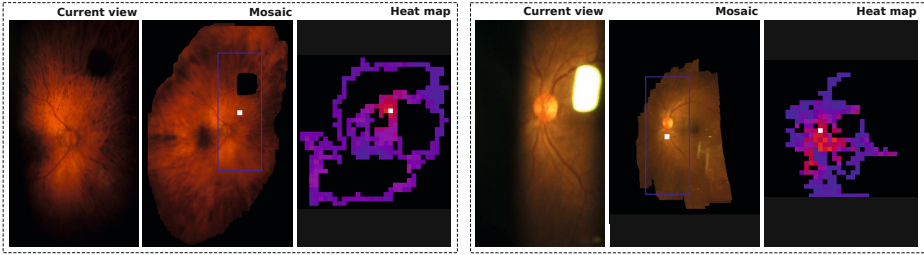
Fig. 5. A 'heat map' shows the template for which the score $\mathcal{L}$ is the highest. Colors with highest temperature (closest to violet) indicate the highest similarity scores. Notice the high accuracy of the current view position on the mosaic (blue rectangle) estimated using the proposed template matching approach.

template-based and feature-based strategies. For a performance comparison, we execute the relocalization step at every frame, in parallel with retinal mapping. This experiment is conducted on sequences where tracking is never lost, so we can focus solely on the relocalization performance evaluation.

Since ground-truth is not available in our dataset, we use the tracked transformation M_{mc} as reference for measuring the relocalization error ϵ in every frame. The error ϵ is computed as the mean of the absolute Euclidean distance between the four corners of the estimated and true positions of the current frame on the retina map (i.e. $\frac{|\epsilon_1|+|\epsilon_2|+|\epsilon_3|+|\epsilon_4|}{4}$). An illustration of the relocalization error can be found in Figure 2. To compare the proposed template-based approach to previously used feature-based methods, we tested SIFT [9] and SURF [3] feature detectors and descriptors. Due to the nature of slit-lamp retinal images, more specialized feature descriptors such as [5] did not provide sufficiently good results. Alternatively, we compared NCC and LBP-based similarity measures for template-based relocalization using the matching framework described in Section 2. Due to the excessive computational complexity, we left out similarity functions such as mutual information.

The relocalization error results for a given video sequence are plotted in Figure 4. Videos are provided as supplementary material to illustrate this experiment. A visual analysis clearly indicates the superior accuracy of template-based relocalization strategies. More specifically, the average and standard deviation relocalization error using SIFT and SURF are 8.65±22.14 pixels and 36.33±40.66, respectively. In comparison, the relocalization error using NCC and the proposed LBP-based matching strategy are 5.21±2.11 and 5.39±3.81 pixels, 37% smaller than the best performance using a feature-based approach. From the plot, it is also clear that the only period where feature-based methods have comparable performance (from [0,25]s) is when the optical nerve is visible. This is due to the fact that the optical nerve offers more stable features (such as blood vessel crossings, etc) compared to other parts of the retina. On the other hand, template-based methods display considerably more consistent results throughout the entire sequence.
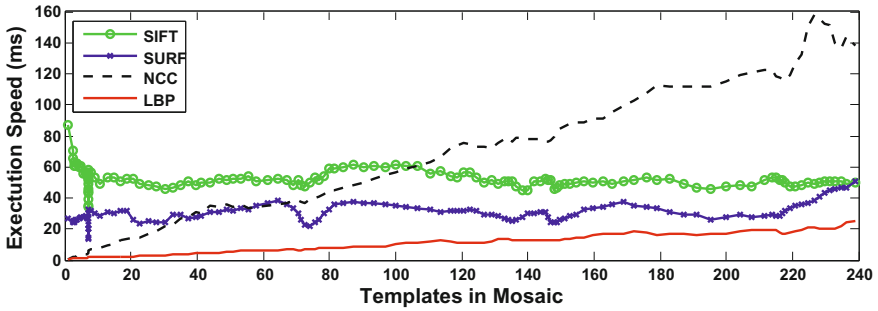
Fig. 6. A comparison between the execution speed as a function of the number of templates in the retina mosaic. The superior computational efficiency of the proposed LBP-based matching method is noticeable when the number of templates in the retina map increases.

3.2   Discriminability and Computational Effort

Another important aspect of relocalization is the template matching discriminability. In other words, the LBP-based similarity function in Equation (1) should yield the highest score for the template most similar to the current frame. To illustrate this aspect of the proposed method, we plotted in Figure 5 the LBP similarity scores during mapping for several videos in our database. The similarity scores are shown in the form of a "heat map". Next to each mosaic in Figure 5, the centers of every template on the retina map are displayed with a specific color corresponding to the LBP similarity score. Colors with higher temperatures indicate high similarity scores. As in the previous experiment, relocalization is performed at every frame. A visual inspection of Figure 5 shows a gradual decrease in the LBP similarity as we move farther away from the correct match, indicating that the LBP-based similarity function is sufficiently discriminative for relocalization purposes.

Another key aspect of our proposed method is speed. Figure 6 shows a plot of execution time in a video sequence. During the retina exploration, the number of templates in the retina map gradually increases. The plot shows that the NCC-based relocalization speed quickly decreases as the video progresses and more templates are added to the mosaic. That is due to the fact that the NCC computation is considerably more expensive than the LBP-based similarity score. Furthermore, contrary to the NCC, each template's LBP codes can be precomputed as soon as it is added to the mosaic, which allows us to greatly reduce the similarity calculation effort at run time.

4   Conclusion

In this paper, we proposed a fast relocalization method for retinal mapping using the slit-lamp device. The key element in the proposed method is the local binary pattern code used for matching retinal images. The LBP code is sufficiently discriminative to provide accurate relocalization results and faster to compute

than illumination invariant similarity measures such as the NCC. Using the proposed template-based map relocalization strategy, tracking is quickly restored in case of occlusions or degradations, allowing the creation of wide retina maps in a straightforward manner. Experiments conducted on a large database of recorded human patient videos have shown the superior accuracy and computational efficiency of the proposed method, compared to similar feature-based approaches used in the literature. Our current work focuses on improving the LBP code for increasing matching discriminability. We are also exploring unsupervised ensemble classification techniques to further increase the relocalization accuracy and computational efficiency.

References

1. Asmuth, J., et al.: Mosaicking and enhancement of slit lamp biomicroscopic fundus images. British Journal of Ophthalmology (2001)
2. Atasoy, S., Noonan, D.P., Benhimane, S., Navab, N., Yang, G.Z.: A global approach for automatic fibroscopic video mosaicing in minimally invasive diagnosis. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 850–857. Springer, Heidelberg (2008)
3. Bay, H., et al.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding (2008)
4. Broehan, A., et al.: Real-time Multimodal Retinal Image Registration for Computed-Assisted Laser Photocoagulation System. IEEE TBME (2011)
5. Can, A., et al.: A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina (PAMI) 24(3), 347–364 (2002)
6. Dame, A., Marchand, E.: Second order optimization of mutual information for real-time image registration. IEEE Trans. on Image Processing 21(9), 4190–4203 (2012)
7. Kalal, Z., et al.: P-N learning: Bootstrapping binary classifiers by structural constraints. In (CVPR), pp. 49–56 (2010)
8. Lovegrove, S.: Parametric Dense Visual Slam. Ph.D. thesis, Department of Computing, Imperiall College London (2011)
9. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision (2004)
10. Ojala, T., et al.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition (1996)
11. Richa, R., Vágvölgyi, B., Balicki, M., Hager, G., Taylor, R.H.: Hybrid tracking and mosaicking for information augmentation in retinal surgery. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 397–404. Springer, Heidelberg (2012)
12. Rosa, B., et al.: Building Large Mosaics of Confocal Endomicroscopic Images Using Visual Servoing. IEEE Transactions on Biomedical Engineering (2013)
13. Silveira, G., Malis, E.: Unified direct visual tracking of rigid and deformable surfaces under generic illumination changes in grayscale and color images. International Journal of Computer Vision (2010)
14. Totz, J., Mountney, P., Stoyanov, D., Yang, G.-Z.: Dense surface reconstruction for enhanced navigation in MIS. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 89–96. Springer, Heidelberg (2011)
15. Ye, M., Giannarou, S., Patel, N., Teare, J., Yang, G.-Z.: Pathological Site Retargeting under Tissue Deformation Using Geometrical Association and Tracking. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 67–74. Springer, Heidelberg (2013)