Combining Efficient Hand-Crafted Features with Learned Filters for Fast and Accurate Corneal Nerve Fibre Centreline Detection

Roberto Annunziata*[1], Ahmad Kheirkhah[2], Pedram Hamrah[2] and Emanuele Trucco[1]

***Abstract*— We propose a new approach to corneal nerve fibre centreline detection for *in vivo* confocal microscopy images. Relying on a combination of efficient hand-crafted features and learned filters, our method offers an excellent compromise between accuracy and running time. Unlike previous solutions using sparse coding to learn small filter banks, we employ K-means to efficiently learn the high amount of filters needed to cope with the multiple challenges involved, e.g., low contrast and resolution, non-uniform illumination, tortuosity and confounding non-target structures. The use of K-means for dictionary learning allows us to learn banks of 100 filters in less than 30 seconds compared to several days needed when using sparse coding. Experimental results using a dataset including 100 images show that our approach outperforms significantly state-of-the-art methods in terms of precision-recall curves.**

I. INTRODUCTION

Recent clinical studies have reported significant correlations between corneal nerve fibre morphometric features and disease conditions including diabetic retinopathy [1], dry eye disease [2] and herpes simplex virus keratitis [3], [4]. Corneal nerves, particularly the subbasal nerve plexus, are often captured through *in vivo* confocal microscopy (IVCM), a rapid and non-invasive imaging technique. In these images, several morphometric features can be assessed, for instance fibre tortuosity [5] or branch and fibre density [6]. However, measuring these parameters manually is time-consuming, thus limiting the volume of data examinable within fixed-time projects. Besides, in order to achieve a sufficient level of repeatability, experienced specialists have to perform such analysis as reported in [7], which contributes to increasing healthcare costs and opthalmologists' workload. Several solutions have been proposed to cope with the multiple challenges of curvilinear structure segmentation, including low signal to noise ratio at small scales, confounding non-target structures, non-uniform illumination, low resolution limiting the accuracy, complex configurations (e.g. bifurcations and crossings) [8]–[13]. Most approaches are based on a local *tubularity* measure (called *vesselness* in [9], [11]–[13]) estimated via hand-crafted features (henceforth, HCFs) [8], [13], or learned from training data [14]. HCFs may rely on assumptions which are violated in some cases (e.g. at bifurcations and crossing points); learned features address this problem by learning key configurations on training data, but can be demanding computationally since they typically require more filters than efficient HCFs such as [11], [13]. Recently, combining HCFs with a few learned filters has proven successful as it exploits the efficiency of a fast HCF approach to reduce the amount of learned filters [10], desirable when dealing with large volumes of image data. Although this method outperforms state-of-the-art approaches such as [11], [13] on retinal blood vessels and neuronal trees, it provides modest results on IVCM images, mainly due to lower contrast and resolution, in addition to tortuosity and fibre fragmentation (Figure 1). The main reason is that a few amount of learned filters (in [10] 9 filters

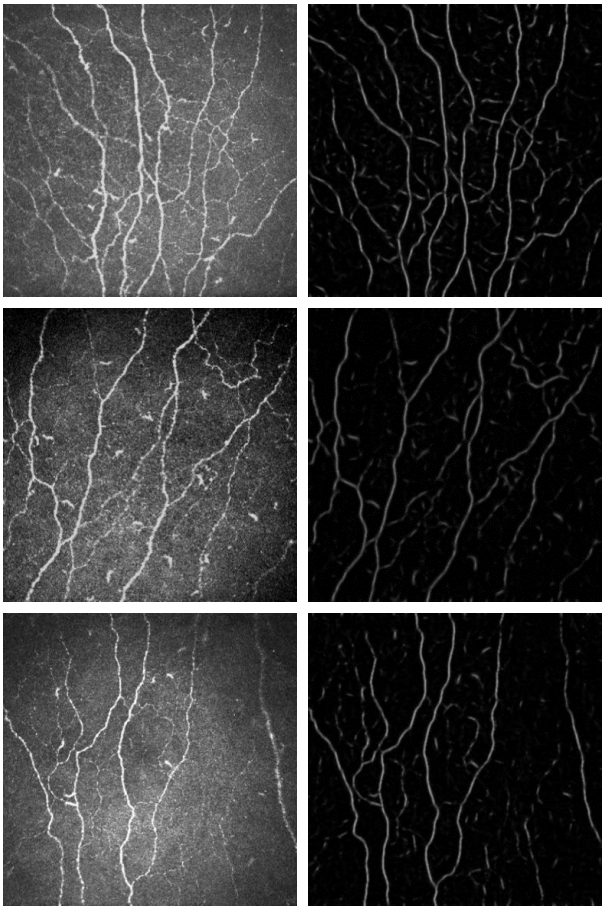

Fig. 1. IVCM images of corneal subbasal nerve fibres (first column) and their tubularity maps obtained using the proposed approach (second column). Strong variation of contrast, illumination, fibre fragmentation and tortuosity seen in these images characterizes the whole dataset.

*Corresponding author (r.annunziata@dundee.ac.uk). This research was supported by the EU Marie Curie ITN REVAMMAD, project number 316990.

[1]Roberto Annunziata and Emanuele Trucco are with the Computer Vision and Image Processing group, School of Computing, University of Dundee, Dundee, UK.

[2] Ahmad Kheirkhah and Pedram Hamrah are with the Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, USA.

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

were used) are not able to capture all these complexities and the use of sparse coding for dictionary learning would require several days to learn filter banks including tens of filters as reported in [10].

We improve in two ways on the state-of-the-art of curvilinear structure segmentation to achieve reliable detection of the centreline of subbasal corneal nerve fibres. First, we use K-means clustering for filter learning as it offers higher flexibility than sparse coding in selecting the dictionary size, requiring less than 1 minute to learn filter banks of hundreds of filters. Second, we use Canny-like non-maxima suppression to refine centreline detection as opposed to mathematical morphology techniques such as thinning and skeletonization.

II. MATERIALS

100 images of the subbasal nerve plexus of the central cornea were acquired using laser scanning IVCM (Heidelberg Retina Tomograph 3 with the Rostock Cornea Module, Heidelberg Engineering GmbH, Dossenheim, Germany) by the clinical authors. The diode laser source of this microscope has a 670-nm red wavelength and the microscope is equipped with a $63\times$ objective lens with a numerical aperture of 0.9 (Olympus, Tokyo, Japan). The images obtained by this confocal microscope represent a coronal section of the cornea of 400×400 μm which can be of any corneal layer. In this study, images from the corneal subbasal layer were used. For imaging, a disposable sterile polymethylmethacrylate cap (Tomo-Cap; Heidelberg) was filled with a layer of hypromellose 0.3% gel (GenTeal; Alcon, Fort Worth, TX) and mounted on the Cornea Module. After topical anesthesia with 0.5% proparacaine hydrochloride (Alcon, Fort Worth, TX), a drop of hypromellose 0.3% gel was applied to both eyes. One drop of this gel was also put on the outside tip of the cap to improve optical coupling. Then, the Cornea Module of the confocal microscope was advanced until the gel contacted the central surface of the cornea. Digital images were then recorded at a rate of 3 frames per second with the sequence mode, including 100 images per sequence, from all layers of the cornea. Images from the subbasal nerves, typically at a depth of 50 to 80 μm, were chosen for this study. Nerves were manually traced by clinical authors using NeuronJ, an add-on plug-in for the ImageJ software. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board of the clinical authors.

III. METHODS

A. Unsupervised Filter Learning

Several algorithms can be used to learn filters from unlabeled data. Some of them are non-trivial to train due to the high amount of parameters to be tuned. Recently, using K-means clustering as the unsupervised learning module to learn filters from unlabeled data has been shown to yield excellent performance, often competing with state-of-the-art systems [15]. The main advantage, at a parity of classification performance [15], is that K-means is fast and easily adaptable

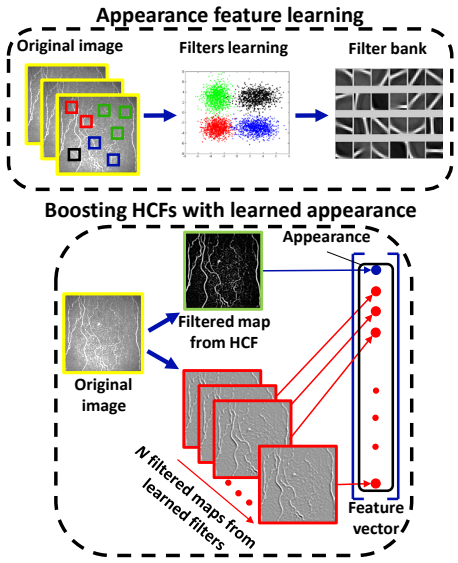

Fig. 2. Leveraging HCFs with learned appearance filters using K-means clustering.

to different datasets as it requires a few intuitive parameters to tune. Moreover, although it is not designed to learn sparse representations such as sparse coding, it has been shown empirically that K-means tends to discover sparse projections of the data under the right conditions [15], i.e., sufficiently large number of patches given the patch size, patch-level whitening to remove correlations between nearby pixels. Because of this property and given the large amount of training data available to us (Section IV-B), we can employ K-means-based filter learning algorithm instead of the ones explicitly designed to obtain sparse representations.

Our goal is to learn a dictionary $\mathbf{D} \in \mathbb{R}^{n\times k}$ of k vectors so that a data vector $\mathbf{x}^{(i)} \in \mathbb{R}^n, i = 1,\ldots,m$ can be mapped to a code vector that minimizes the reconstruction error. Following [15], before running the learning algorithm we first normalize the brightness and contrast of each input data point (i.e. patches) $\mathbf{x}^{(i)}$. Then, we apply patch-level whitening through the ZCA transform so that $\mathbf{x}^{(i)}_{ZCA} = \mathbf{V}(\mathbf{\Sigma} + \epsilon_{ZCA}\mathbf{I})^{-1/2}\mathbf{V}^T\mathbf{x}^{(i)}$, where $\mathbf{V}$ and $\mathbf{\Sigma}$ are computed from the eigenvalue decomposition of the data points covariance $\mathbf{V\Sigma V}^T = \text{cov}(\mathbf{X})$, and ϵ_{ZCA} is a small constant controlling the trade-off between whitening and noise amplification. After pre-processing the patches, we optimize the following objective function:

$$\begin{aligned} &\underset{\mathbf{D},\mathbf{c}}{\text{argmin}} \sum_i \left\| \mathbf{D}\mathbf{c}^{(i)} - \mathbf{x}^{(i)}_{ZCA} \right\|_2^2 \\ &\text{subject to} \quad ||\mathbf{c}^{(i)}||_0 \leq 1, \forall i = 1,\ldots,m \\ &\text{and} \quad ||\mathbf{d}^{(j)}||_2 = 1, \forall j = 1,\ldots,k \end{aligned} \tag{1}$$

where $\mathbf{c}^{(i)}$ is the code vector related to the pre-processed input $\mathbf{x}^{(i)}_{ZCA}$, and $\mathbf{d}^{(j)}$ is the $j-th$ column of the dictionary $\mathbf{D}$. The second line in Eq. (1) forces the code to have at most one non-zero entry, while the third prevents learned filters from becoming arbitrarily large or small. The optimal

code $\mathbf{c}^{(i)}$ is found by computing a few iterations (about 10 [15]) of:

$$\mathbf{c}_j^{(i)} = \begin{cases} {\mathbf{d}^{(j)}}^T \mathbf{x}_{ZCA}^{(i)} & \text{if } j = \underset{l}{\text{argmax}} |{\mathbf{d}^{(l)}}^T \mathbf{x}_{ZCA}^{(i)}| \quad \forall j, i \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

$$\begin{aligned} \mathbf{D}_{new} &= \mathbf{X}_{ZCA}\mathbf{C}^T + \mathbf{D}_{old} \\ \mathbf{d}^{(j)} &= \mathbf{d}_{new}^{(j)} / ||\mathbf{d}_{new}^{(j)}||_2 \quad \forall j. \end{aligned} \tag{3}$$

Note that the first and second line in Eq. (3) are used to prevent small clusters from being moved too much in a single iteration ("damped" updates).

B. Description Vector and Supervised Classification

Once a filter bank is learned, learned filters are used to compute multiple filtered maps $\mathbf{L}^{(j)}$:

$$\mathbf{L}^{(j)} = \left\| \mathbf{D}^{(j)} - \mathbf{X}_{ZCA}^{(i)} \right\|_2^2, \quad i = 1, \ldots, m_I \tag{4}$$

where $\mathbf{D}^{(j)}$ is the $j-th$ learned filter, $\mathbf{X}_{ZCA}^{(i)}$ is the $i-th$ normalised patch (arranged as matrix here) in the training or test image and m_I is the total number of patches within an image. We select the OOF [11] as HCF, since it tends to outperform other efficient HCFs on our dataset (Section IV). Thus, for each image location (u, v), we construct the following description vector combining HCFs and learned filters:

$$[\underbrace{\mathbf{OOF}(u,v)}_{\textbf{HCF}}, \underbrace{\mathbf{L}^{(1)}(u,v), \ldots, \mathbf{L}^{(N)}(u,v)}_{\textbf{learned}}]^T. \tag{5}$$

We then apply a Random Decision Forest [10], [16] to classify each pixel as lying on a corneal nerve fibre or background. The resulting probability/tubularity map (examples shown in Figure 1) leads indeed to a more accurate and less noisy segmentation than other methods (precision-recall curves in Section IV). Finally, centreline detection is obtained using Canny-like non-maxima suppression on the tubularity map. Local orientation is estimated using OOF.

IV. Results and Discussion

A. Performance Evaluation

As usually done to evaluate methods extracting one-pixel-wide curves [14], we introduce a tolerance factor ρ: a predicted centreline point is considered a true positive if it is at most ρ pixels distant from a ground truth centreline point. In our experiments we set $\rho = 2$ pixels. Following the usual benchmarking procedure [10], we average performance measures over 5 random sub-sampling cross-validation runs, using 50 images for training and the rest for testing in each run. The resulting precision-recall curves are reported in Figure 3 (mean and standard deviation of the results from individual runs). We compare the proposed method against 3 widely used methods for curvilinear structure detection, namely Frangi [13], Gabor [12] and the recent Optimally Oriented Flux (OOF) [11].

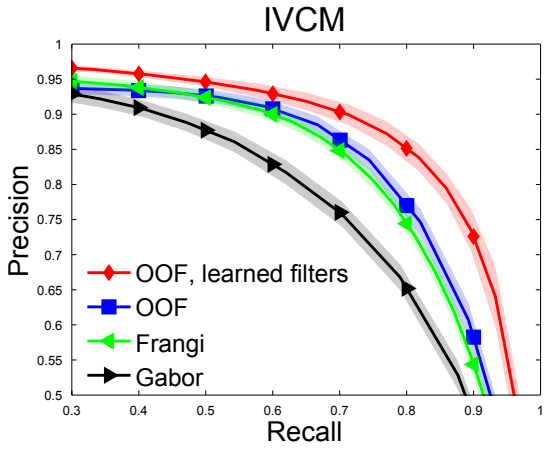

Fig. 3. Precision-recall curves for pixel-level classification. Shaded color bands represent 1 standard deviation of the results from individual runs.

B. Experimental Setup

We evaluated the performance of 3 state-of-the-art HCFs to decide which HCF had to be included in our framework: OOF [11], Frangi [13] and Gabor filtes [12] widely acknowledged as excellent tubular structure detectors at a limited computational cost. We applied our method to OOF since we found experimentally that it outperforms Gabor and tends to show better performance than Frangi on the IVCM dataset adopted here for validation (Figure 3). In terms of pre-processing, we first normalize images to zero mean and unit standard deviation. Then, we apply OOF and learned filters. When HCFs (Frangi, Gabor and OOF) are used as baselines, all parameters are tuned separately to achieve best performance. For the OOF, we set $\sigma = \{2, 3, 4\}$ (Eq. (8) [11]) and $R = \{2, 3, 4\}$ (Eq. (5) [11]). ϵ_{ZCA} was set to 0.001, considering the trade-off between noise amplification and filters sharpness [15]. Filters size was set to 15×15 pixels, a good compromise between speed and pixel context information. Notice that the chosen patch size allows us to collect a sufficient number of patches to learn dictionaries, in agreement with the guidelines in [15][1]. We learned filter banks including 100 filters (i.e., $N = 100$) for our method to exploit at most the capability of the learning part to handle complexities detailed above, without increasing too much the running time. We used Random Decision Forests consisting of 100 random trees to make the predictions fast. We trained classifiers using the same number of positive and negative samples and empirically estimated priors. Experiments were run on a 8-core 64-bit architecture using MATLAB implementations.

C. Experimental Results and Discussion

Figure 3 shows precision-recall curves regarding the corneal fibre centreline detection task (pixel-level classification) for the proposed method and the baselines on our dataset. The proposed method outperforms significantly

[1] We selected about $300,000$ patches, although in [15] $100,000$ 16×16 patches are considered sufficient.

TABLE I

EFFECT OF PATCH AND DICTIONARY SIZE ON THE AREA UNDER PRECISION-RECALL CURVES (MEAN/STANDARD DEVIATION).

IVCM	**Patch size (pixels)**			**Number of learned filters (N)**		
	11×11	15×15	21×21	10	100	200
OOF, learned filters	0.8665/0.0114	0.8878/0.0059	0.8870/0.0102	0.8524/0.0057	0.8878/0.0059	0.8944/0.0033
	100 learned filters			**Patch size:** 15×15 **pixels**		

others as it better deals with low signal-to-noise ratio due to low resolution, tortuous and fragmented structures. In fact, while at low recall values the gain in precision is limited with respect to OOF and Frangi, the use of learned filters proves decisive for medium to high recall values. In the latter region, thresholds on the tubularity maps are such that many false positives could be produced due to noise (i.e. clutter) and non-target structures present on the background. Normalizations at patch level proved helpful in dealing with variations in contrast and illumination, and quantitative results suggest a better robustness against those problems.

We evaluated the effect of the patch and dictionary size on the performance, and report in Table I experiments for different combinations of parameters. As discussed above, combining HCFs with 100 learned filters leads to significantly better performance than using 9 filters as experimented by the authors in [10]. We notice that increasing the amount of learned filters would contribute to improve performance further, confirming the difficulty of detecting fibre centrelines using a small subset of learned filters for IVCM images. For this reason we employ K-means instead of sparse coding as done in [10]. In fact, less than 30 seconds are typically required to learn 100 filters using our approach which compares favourably with several days reportedly needed to learn 121 filters using sparse coding [10]. Using 100 learned filters seems to be a good compromise between accuracy and running time. In fact, since the size of our IVCM is relatively small (384×384), about 1 second is needed to apply our filter bank to an image and less than 9 seconds are typically required to obtain a tubularity map (classification) using unoptimized MATLAB code. Notice that the proposed framework is easily parallelizable and the running time could be reduced further.

V. CONCLUSIONS

We presented an algorithm for efficient and accurate corneal nerve fibre centreline detection combining fast HCFs with learned filters. Experimental results obtained using a dataset of 100 IVCM images show that the proposed approach outperforms state-of-the-art approaches such as OOF [11], Gabor [12] and Frangi [13]. Unlike a previous solution [10] using sparse coding to learn a small filter bank, we employ K-mean algorithm for dictionary learning. We showed experimentally that increasing the amount of learned filters improves performance considerably when dealing with corneal nerve fibres captured through *in vivo* confocal microscopy. The use of K-means for dictionary learning allows us to learn filter banks in less than 30 seconds compared to several days needed when using sparse coding. Quantitative results in terms of accuracy and running time suggest that a corneal nerve fibre centreline detection pipeline for a different dataset can be set up and get state-of-the-art performance within a few minutes. Our future work will investigate the combination of learned filters with our recently proposed HCF, SCIRD [17], and the design of new ones to exploit optimally the information obtained by the HCF used.

REFERENCES

[1] S. De Cillà, S. Ranno, E. Carini, P. Fogagnolo, G. Ceresara, N. Orzalesi, and L. M. Rossetti, "Corneal subbasal nerves changes in patients with diabetic retinopathy: an in vivo confocal study," *IOVS*, 2009.

[2] E. Villani, D. Galimberti, F. Viola, C. Mapelli, and R. Ratiglia, "The cornea in sjögrens syndrome: an in vivo confocal study," *IOVS*, 2007.

[3] P. Hamrah, A. Cruzat, M. H. Dastjerdi, H. Prüss, L. Zheng, B. M. Shahatit, H. A. Bayhan, R. Dana, and D. Pavan-Langston, "Unilateral herpes zoster ophthalmicus results in bilateral corneal nerve alteration: an in vivo confocal microscopy study," *Ophthalmology*, 2013.

[4] P. Hamrah, A. Cruzat, M. H. Dastjerdi, L. Zheng, B. M. Shahatit, H. A. Bayhan, R. Dana, and D. Pavan-Langston, "Corneal sensation and subbasal nerve alterations in patients with herpes simplex keratitis: an in vivo confocal microscopy study," *Ophthalmology*, 2010.

[5] R. Annunziata, A. Kheirkhah, S. Aggarwal, B. M. Cavalcanti, P. Hamrah, and E. Trucco, "Tortuosity classification of corneal nerves images using a multiple-scale-multiple-window approach," in *OMIA Workshop, MICCAI*, 2014.

[6] I. N. Petropoulos, U. Alam, H. Fadavi, A. Marshall, O. Asghar, M. A. Dabbah, X. Chen, J. Graham, G. Ponirakis, A. J. Boulton *et al.*, "Rapid automated diagnosis of diabetic peripheral neuropathy with in vivo corneal confocal microscopy," *IOVS*, 2014.

[7] I. N. Petropoulos, T. Manzoor, P. Morgan, H. Fadavi, O. Asghar, U. Alam, G. Ponirakis, M. A. Dabbah, X. Chen, J. Graham *et al.*, "Repeatability of in vivo corneal confocal microscopy to quantify corneal nerve morphology," *Cornea*, 2013.

[8] R. Annunziata, A. Garzelli, L. Ballerini, A. Mecocci, and E. Trucco, "Leveraging multiscale hessian-based enhancement with a novel exudate inpainting technique for retinal vessel segmentation," *IEEE Journal of Health and Biomedical Informatics*, 2015.

[9] J. Hannink, R. Duits, and E. Bekkers, "Crossing-preserving multi-scale vesselness," in *MICCAI*, 2014.

[10] R. Rigamonti and V. Lepetit, "Accurate and efficient linear structure segmentation by leveraging ad hoc features with learned filters," in *MICCAI*, 2012.

[11] M. W. Law and A. C. Chung, "Three dimensional curvilinear structure detection using optimally oriented flux," in *ECCV*, 2008.

[12] J. V. Soares, J. J. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification," *IEEE TMI*, 2006.

[13] A. Frangi, W. Niessen, K. Vincken, and M. Viergever, "Multiscale vessel enhancement filtering," in *MICCAI*, 1998.

[14] A. Sironi, V. Lepetit, and P. Fua, "Multiscale centerline detection by learning a scale-space distance transform," in *CVPR*, 2014.

[15] A. Coates and A. Y. Ng, "Learning feature representations with k-means," in *Neural Networks: Tricks of the Trade*, 2012.

[16] A. Criminisi and J. Shotton, *Decision forests for computer vision and medical image analysis*. Springer, 2013.

[17] R. Annunziata, A. Kheirkhah, P. Hamrah, and E. Trucco, "Scale and curvature invariant ridge detector for tortuous and fragmented structures," in *MICCAI*, 2015.