RETINAL VASCULATURE SEGMENTATION USING PRINCIPAL SPANNING FORESTS

Erhan Bas

Janelia Farm Research Center
Howard Hughes Medical Institute
Ashburn, VA. 20147 USA

Esra Ataer-Cansizoglu,
Deniz Erdogmus

Cognitive Systems Laboratory
Northeastern University
Boston, MA. 02115 USA

Jayashree Kalpathy-Cramer

Athinoula A. Martinos Center
Massachusetts General Hospital
Boston, MA. 02129 USA

ABSTRACT

We propose an automated morphology reconstruction method for curvilinear network analysis. The proposed approach first projects samples to the ridge of the intensity image of the curvilinear system. Then, a manifold deviation measure is utilized to approximate the ridge with piecewise linear segments between the projected samples. A nonparametric system workflow based on the kernel interpolation and density estimation is provided for the derivations without any user defined meta-parameter, i.e. hard threshold for segmentation. Lastly, a rigorous sampling strategy using the manifold deviation measure that can be used for robust sparse tree reconstruction is provided. The proposed approaches have been tested on a small set of representative retinal scans. Preliminary qualitative results indicate the effectiveness of the method.

***Index Terms*—** Principal graphs, resampling on manifolds

1. INTRODUCTION

Curvilinear network analysis in images has been an active research area for the last few decades. The aim of curvilinear analysis is to provide a simple, yet biologically meaningful summary of a complex network, which can be achieved through morphology (geometry and topology extraction). Hence, various approaches have been proposed to extract the skeleton/ridge/centerline/principal curve of the data and represent it in an ordered fashion[1]. Airways, vascular and neuronal networks are such examples that have been studied extensively in the literature [1, 2, 3]. Although it is impossible to cover all approaches here, it is reasonable to categorize previous work into two categories: *i*) local and *ii*) global approaches.

Local approaches are explanatory methods that rely on local decisions to traverse/trace the data space. Since decisions are purely based on the local evidence, they are computationally feasible. However, a common problem in local tracing methods is suboptimal traces. Problems related to imaging, such as noise, gaps or crossovers are frequently present in the data and any decision limited to an immediate local neighborhood is in general sensitive to such variations in the data. On the other hand, global approaches are more robust to the such variations. Moreover, any a priori constraint on the morphology can be incorporated to ridge estimation such as in active contours. However, they are computationally demanding given the size of the data. In general, global methods are applied to pre-segmented data or initialized to some initial morphology to reduce computational complexity. In fact, any error in the pre-segmentation or initialization stage propagates throughout the estimation process. Moreover, given the variations in the appearance and shape of curvilinear structures within a dataset, it is practically infeasible to come up with a global approach or a trained classifier that can handle all scenarios. Fig. 1 shows such a case where maximum intensity projections (MIP) of two confocal microscopy images [4] from the central region of Drosophila, where intensity and shape variations are highly noticeable, albeit being acquired using the same modality and preparation and belonging to same region. Moreover, each curvilinear system has it is own dynamics and characteristics. Blood vessels in the eye might form highly sparse tree-like vascular networks [5] as seen in Fig. 2(a), whereas neuronal arbors such as neuromuscular junction fibers [6] as in Fig. 2(b) form dense bundled networks.

In this work, we discuss an automated global spanning algorithm that uses local statistics to adapt variations in the data that aims to fill the gap between local and global approaches. The proposed approach is based on ridge analysis [7, 8] and manifold connectivity concepts that we introduced in our previous work on neuronal morphology extraction [9]. Briefly, given a dataset, we first map sam-

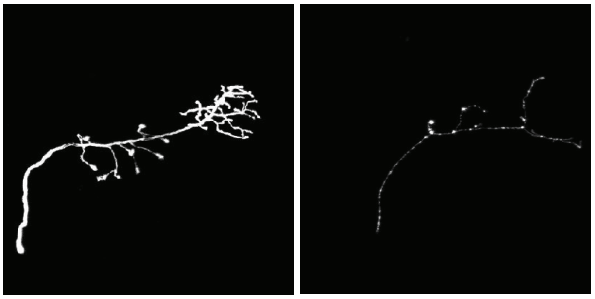
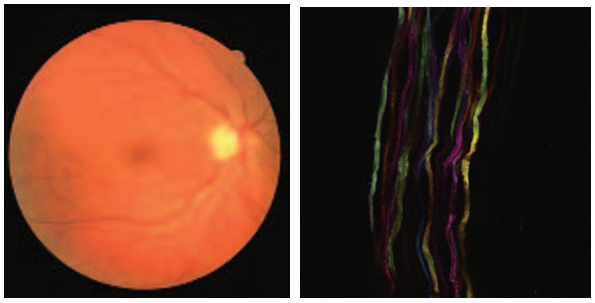


Fig. 1. Olfactory datasets obtained from confocal microscopy.



(a) (b)

Fig. 2. Different curvilinear systems. a) Retinal scan. b) Neuronuscular junction axonal projection

[1]Here, order is used as in hierarchical representation.

ples to their underlying ridge. Then, in order to extract robust morphological reconstruction and to prevent aforementioned data dependencies, a nonparametric twice differentiable ridge function $s(\mathbf{x})$ is constructed from these projections directly. Finally, an affinity based spanning algorithm based on this ridge function is utilized to cover the whole dataset.

2. RIDGE ANALYSIS

The centerline of a multivariate function whose high-valued regions tightly follow a curvilinear structure (manifold) is defined as the curve that passes through the middle of this high-value region, or in the case of data clouds from the middle of the high density region [7, 10, 8]. In our approach, we use the local critical set/manifold definitions [7, 8] to decide if a point is on the ridge of a function. The following sections are adopted from our previous work [9, 11].

Let $f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ be an at least twice continuously differentiable function with $\mathbf{g}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$ be its gradient and Hessian. Let $\mathbf{q}_i$ and λ_i be the i^{th} eigenvector/eigenvalue pair of the Hessian matrix of $f(\mathbf{x})$ such that $\lambda_1 \leq \cdots \leq \lambda_n$. For the sake of argument, we assume that appearance model of $f(\mathbf{x})$ is concave, i.e. has local maxima along its ridge. Since the underlying morphology is curvilinear and the function is concave, a point is on the ridge of a function if the gradient is collinear with one of the eigenvectors and the eigenvalues corresponding to the remaining eigenvectors are all negative.

Note that mentioned local conditions are generalizations of the local maxima/minima conditions; a point on the ridge is the local maximum of the function in the subspace spanned by the $n-1$ eigenvectors, $S_\perp = \mathrm{span}(\mathbf{q}_2, \ldots, \mathbf{q}_n)$, assuming without loss of generality that the gradient is aligned with the first eigenvector.. Let $\mathbf{C}(\mathbf{x}) = \mathbf{H}(\mathbf{x})^{-1}$ be the inverse of the local Hessian matrix and $\mathbf{C}_\perp(\mathbf{x}) = \sum_2^n (\lambda_i^{-1} \mathbf{q}_i \mathbf{q}_i^T)$ be the orthogonal component of $\mathbf{C}(\mathbf{x})$; a measure for being on the ridge can be formulated as

$$\gamma(\mathbf{x}) = \frac{\mathbf{g}(\mathbf{x})^T \mathbf{C}_\perp(\mathbf{x}) \mathbf{g}(\mathbf{x})}{\|\mathbf{C}(\mathbf{x})\mathbf{g}(\mathbf{x})\| \|\mathbf{g}(\mathbf{x})\|} \tag{1}$$

where the measure $\gamma(\mathbf{x})$ becomes 0 on the ridge. In fact, any measure that quantifies whether the gradient vector lies in the null space of $S_\perp$ can be used instead of Eqn.1. However, the proposed measure has only one local minimum at 0 (on the ridge), and it is monotonically increasing elsewhere except at the umbilic locations[2], hence captures the deviation from the underlying ridge in the extended vicinity of the ridge [9] and can be used to construct an affinity measure between samples as discussed in Section 3.

Given $f(\mathbf{x})$, all samples are mapped to their underlying principal curve by constructing a second order gauge coordinate system and projecting them by solving an ordinary differential equation[3]. In order to prevent redundant calculations in the projection step, we only keep the pixels where the underlying ridge passes through. Consequently, only a small fraction of the total pixel count (voxel count in 3D images) is processed with the principal curve projection iterations.

In order to obtain a twice continuously differentiable function $f(\mathbf{x})$, kernel regression technique is utilized. Let $I(\mathbf{x})$ be the appearance, i.e. intensity value, at position $\mathbf{x}$, $f(\mathbf{x})$ can be written as

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^{N} I(\mathbf{x}_i) G_{\sigma_\mathbf{c}}(I(\mathbf{x}) - I(\mathbf{x}_i)) G_{\mathbf{\Sigma}_i}(\mathbf{x} - \mathbf{x}_i) \quad (2) \\ &= \sum_{i=1}^{N} w(\mathbf{x}_i) G_{\mathbf{\Sigma}_i}(\mathbf{x} - \mathbf{x}_i) \end{aligned}$$

where $w(\mathbf{x}_i) = I(\mathbf{x}_i) G_{\sigma_\mathbf{c}}(I(\mathbf{x}) - I(\mathbf{x}_i))$ is the weight for the i^{th} data sample $\mathbf{x}_i$, σ_c is the scale of the intensity values $G_{\sigma_c}(\mathbf{x}_i) = C_{\sigma_\mathbf{c}} e^{-\frac{1}{2}(I(\mathbf{x}) - I(\mathbf{x}_i))^2 / \sigma_c^2}$ with normalization term $C_{\sigma_\mathbf{c}}$ and $\mathbf{\Sigma}_i$ is the variable kernel[4] covariance of the Gaussian kernel $G_{\mathbf{\Sigma}_i}(\mathbf{x}_i) = C_{\mathbf{\Sigma}_i} e^{-\frac{1}{2}\mathbf{x}^T {\mathbf{\Sigma}_i}^{-1} \mathbf{x}}$.

The gradient and the Hessian of $f(\mathbf{x})$ are:

$$\mathbf{g}(\mathbf{x}) = -\sum_{i=1}^{N} w(\mathbf{x}_i) G_{\mathbf{\Sigma}_i}(\mathbf{x} - \mathbf{x}_i) {\mathbf{\Sigma}_i}^{-1} (\mathbf{x} - \mathbf{x}_i) \tag{3}$$

$$\mathbf{H}(\mathbf{x}) = \sum_{i=1}^{N} w(\mathbf{x}_i) G_{\mathbf{\Sigma}_i}(\mathbf{x} - \mathbf{x}_i) (\mathbf{u}_i \mathbf{u}_i^T - {\mathbf{\Sigma}_i}^{-1}) \tag{4}$$

Here $\mathbf{u}_i = {\mathbf{\Sigma}_i}^{-1} (\mathbf{x} - \mathbf{x}_i)$.

Once we map the pixels onto the ridge of $f(\mathbf{x})$, a ridge function $s(\mathbf{x})$ is constructed similar to Eqn. 2 using variable-width kernel density estimate (KDE) based on the projected sample locations. We use uniform weighting $w(\mathbf{x}_i) = 1$ for the density estimation. Note that, it is also possible to use $f(\mathbf{x})$ to construct an affinity, but any outlier present in the data will manifest itself in the regression due to selected weighting scheme. For that reason, ridge function $s(\mathbf{x})$ is evaluated only using the projected samples, hence invariant to appearance and shape variations in the data. The gradient, Hessian and the resultant manifold deviation measure, $\kappa(\mathbf{x})$, of $s(\mathbf{x})$ are calculated similar to Eqns. (3,4 and 1) respectively.

In the literature, there are various techniques to estimate the kernel size from given data [13]. In this paper, we used leave-one out cross validation to estimate the width of anisotropic kernels. Although density estimation using variable kernel size or anisotropic kernels is more robust to outliers and is capable of fitting the underlying density variations better, the estimation procedure requires *k*-nearest neighbor (*k*-nn) information where selection of *k* is crucial (we typically selected $k = \sqrt{N}$).

3. SIMILARITY ON MANIFOLDS

In order to extract morphology, one needs to order samples with respect to a reference or a root vertex. A common approach for curvilinear analysis is to represent the data as a graph with vertices as the locations on the ridge and edges between consecutive vertex pairs. We construct a graph representation of the data by first detecting all possible connections/edges between samples, then represent them in a hierarchical formation that governs the topology. In order to simplify the morphological reconstruction, we also provide a sampling strategy based on the manifold deviation concept that we will cover in this section.

In order to define affinity between samples, we first define a dissimilarity measure as a function of how close the edge between them is to the underlying ridge [9]. We use the manifold deviation value of $s(\mathbf{x})$ similar to Eqn. 1, and define the dissimilarity as the integral of this value over the edge.

Let $\wp(\mathbf{a}, \mathbf{b})$ be the pairwise dissimilarity score of points $\mathbf{a}$ and $\mathbf{b}$, given as the line integral of scalar valued function, $\kappa(.)$, from $\mathbf{a}$ to $\mathbf{b}$ evaluated on the curve $l(t)$

[2] Umbilics are the regions where eigenvalues of local Hessian matrix change sign.

[3] We skip details due to space constraints. Technical details can be found in [9, 12].

[4] Assuming Gaussian kernels here for simplicity. Any suitable interpolating kernel could be used depending on smoothness assumptions.

Given confidence/compression threshold, *thr* and a root sample $\mathbf{x}_0$, obtain $\mathbf{M}(\mathbf{a},\mathbf{b}) = \Gamma\wp(\mathbf{a},\mathbf{b}) > thr$ and initialize the set of extracted branches $B_0 = \emptyset$. Then, repeat below until all samples are visited. At the i^{th} iteration

1. Find the furthest sample $\mathbf{x}_k$ to $\mathbf{x}_0$ constrained to the connectivity matrix M.
2. Run a shortest path algorithm to obtain the set of samples on a branch $b = \mathbf{x}_k, \ldots, \mathbf{x}_0$ from $\mathbf{x}_k$ to $\mathbf{x}_0$.
3. Expand the branch with unvisited nodes $B_i = B_{i-1} \bigcup b$
4. Delete any nodes along b.

Table 1. Principal Spanning Forest Algorithm

$$\wp(a,b) = \int_0^1 \kappa(\mathbf{l}(t))[\dot{l}^T(t)\dot{\mathbf{l}}(t)]^{\frac{1}{2}}dt \tag{5}$$

here we parameterized $\mathbf{l}(t) = \mathbf{a} + t(\mathbf{b}-\mathbf{a})$ as a line with $\mathbf{l}(0) = \mathbf{a}$, $\mathbf{l}(1) = \mathbf{b}$ and $\dot{\mathbf{l}}(t) = (\mathbf{b}-\mathbf{a})$.

Note that each sample has its own dedicated second order gauge coordinate system, so the local principal curve rankings (sorted eigenvalue/eigenvector pairs) might not coincide with the ones in the neighborhood manifolds. Consequently, we used the ranking at location $\mathbf{a}$ as reference and select the ranking of eigenvalues at an intermediate step based on the pairwise inner product of reference eigenvectors with the ones at the intermediate step. Let $q_j^{\mathbf{l}(t)}$ be the j^{th} eigenvector at $\mathbf{l}(t)$ and $q_{\perp,k}^{\mathbf{a}}$ be the k^{th} eigenvector that spans $S_\perp(\mathbf{a})$ at reference $\mathbf{a}$, where $j = 1,\ldots,n$ and $k = 2,\ldots,n$. Ranking of the k^{th}principal curve can be obtained as

$$\Theta_k = \arg\min_{i=1,\ldots,n} (q_i^{\mathbf{l}(t)})^T q_{\perp,k}^{\mathbf{a}} \tag{6}$$

If the line integral passes through an umbilical region or a region where local convexity is violated, $\kappa(\mathbf{x})$ will attain negative values. Since we initially assumed that the data lies very close to the manifold, we only calculate the score between pairs where the connecting path also lies inside the locally defined convex region such that

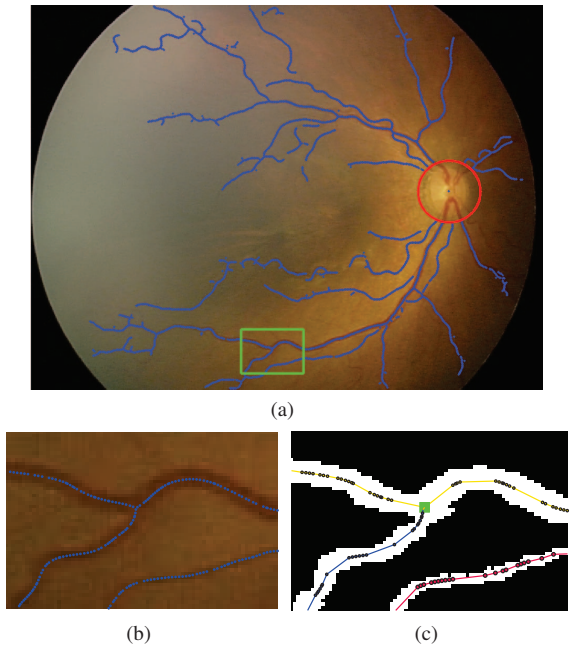
$$\bar{\wp}(\mathbf{a},\mathbf{b}) = \begin{cases} \wp(\mathbf{a},\mathbf{b}) & \text{if } \forall t \in [0,1]\lambda_{2,\ldots,n}(\mathbf{l}(t)) > 0 \\ \infty & \text{otherwise} \end{cases} \tag{7}$$

Note that the defined score is not symmetric. This is implied by the fact that in the presence of multiple local manifolds, the significance ranking of manifold dimensions vary as one travels through the space. Although there are two overlapping paths (actually same path defined by the local principal curve) that connects data pairs, they have different rankings.

4. SAMPLING ON PRINCIPAL MANIFOLDS

We propose the strategy outlined in Tab. 1 to approximate any smooth principal curve with piecewise linear lines.

The constructed dissimilarity matrix summarizes all possible edges in the data, where each edge weight is a measure of having two samples on the same ridge/principal curve. In fact, by thresholding this measure, then applying a connected component analysis one can cluster samples with a compression. Given a compression factor, we define the deviation from the original curve invariant of path length as $\bar{\wp}^*(\mathbf{a},\mathbf{b}) = \max(\bar{\wp}(\mathbf{a},\mathbf{b}),\bar{\wp}(\mathbf{b},\mathbf{a}))/L$, where $L = \int_0^1 [\dot{\mathbf{l}}^T(t)\dot{\mathbf{l}}(t)]^{\frac{1}{2}}dt$ is the arc length of the path. Let $\Gamma\wp(\mathbf{a},\mathbf{b}) = e^{-\bar{\wp}(\mathbf{a},\mathbf{b})/L}$ be the affinity matrix obtained from this path normalized pairwise score matrix and $\mathbf{M}(\mathbf{a},\mathbf{b})$ be the neighborhood mask obtained from $\Gamma\wp(\mathbf{a},\mathbf{b})$ by thresholding with *thr*.



(a)

(b) (c)

Fig. 3. A sample retinal scan. (a) Blue dots represent the projected samples and red circle denotes the optical disk region. (b) Zoomed version of the green rectangle in (a). (c) Reconstruction result overlayed on the Frangi fiter output. Black dots represent the selected nodes in the spanning algorithm. Green square depicts the bifurcation location.

5. RESULTS

We tested the proposed spanning approach on retinal scans [14] obtained and provided to us by researchers at the Casey Eye Institute at OHSU (Portland, OR)[5]. In our calculations, we only used the green channel, where the curvilinear networks are observable. An initial Frangi filter [15] is utilized to enhance the curvilinear structures followed by a connected component analysis to eliminate small isolated regions in the data. In order to guarantee a single connected component, a region growing algorithm that grows from the root location is employed in a recursive fashion to merge any disconnected regions. Optical disk regions (red circle) are manually selected and excluded from the calculations show in Fig. 3(a) and its center is selected as the root sample. Mapped samples after the projection step are depicted with blue dots and overlayed on the original image, as well as on the zoomed region as shown in Fig. 3(a and b) respectively. Fig. 3(c) shows the reconstruction result with resampled samples on the ridge (shown as black dots). Note that reconstruction is highly sparse compared to Fig. 3(b). Each branch is labeled with a distinct color and the bifurcation location is depicted with a green square.

Fig. 4 shows the result of the spanning method on 6 different retinal images. Left column depicts the original images, whereas middle column displays the Frangi filter output. Morphological reconstruction result is displayed in the right column, where each color denotes a branch. In general the proposed technique extracts the morphology information accurately. Common mistakes in the reconstruction are missing branches. Since we used the filtering result as the input for the projection step, any misses in the filtering step would result in missing branches. In addition to this, we eliminated

[5]We thank Dr. Michael Chiang and colleagues for providing a set of images with manual segmentation of vessels. In this work we will not present quantitative measures of overlap between traces and ground truth vessel areas. Future work based on the results presented here will include extensive assessment along these lines.

clusters having small number of samples which might potentially eliminate small isolated vessels.

6. DISCUSSION

In this paper, we presented an automated curvilinear morphology reconstruction method. The proposed approach aims to fill the gap between global and local approaches that have been studied in the curvilinear network analysis literature. We start with a projection step that sparsely maps the samples to their underlying principal curve using local statistics. A nonparametric ridge function is constructed from this projection to highlight the overall topology. Morphology of the networks is obtained by piecewise linearly connecting samples based on the ridge function, such that each connection remains in the concave region of the function and does not deviate from the ridge more than an amount determined by a user selected threshold. Overall, the proposed approach enables us: (*i*) to decide if two samples belong to the same curvilinear network and (*ii*) to sparsify highly dense curvilinear samples with a user defined parameter and satisfying the constraint that approximations remain within the concave region around the ridge. Although we did not mention the usage of the second property in detail due to space limitations, sparse representation of such networks is essential for various clinical applications, such as retrieval for disease classification. Current and future work will focus on such research for clinical applications.

Acknowledgement: This work is partially supported by NSF.

7. REFERENCES

[1] D. Aykac, E.A. Hoffman, G. McLennan, and J.M. Reinhardt, "Segmentation and analysis of the human airway tree from three-dimensional x-ray ct images," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 8, pp. 940–950, Aug. 2003.

[2] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys*, vol. 36, no. 2, pp. 81–121, 2004.

[3] E. Meijering, "Neuron tracing in perspective," *Cytometry Part A*, vol. 77, no. 7, pp. 693–704, 2010.

[4] G.S.X.E. Jefferis, C.J. Potter, A.M. Chan, E.C. Marin, T. Rohlfing, C.R. Maurer, et al., "Comprehensive maps of Drosophila higher olfactory centers: spatially segregated fruit and pheromone representation," *Cell*, vol. 128, no. 6, pp. 1187–1203, 2007.

[5] J. Staal, M.D. Abràmoff, M. Niemeijer, M.A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 4, pp. 501–509, 2004.

[6] J. Livet, T.A. Weissman, H. Kang, R.W. Draft, J. Lu, R.A. Bennis, J.R. Sanes, and J.W. Lichtman, "Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system," *Nature*, vol. 450, no. 7166, pp. 56–62, 2007.

[7] D. Eberly, *Ridges in image and data analysis*, Kluwer Academic Pub, 1996.

[8] U. Ozertem and D. Erdogmus, "Locally defined principal curves and surfaces," *Journal of Machine Learning Research*, vol. 12, pp. 1249–1286, 2011.

[9] E. Bas and D. Erdogmus, "A ridge score with application to piecewise linear neural reconstruction," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*. IEEE, 2011, pp. 1341–1344.

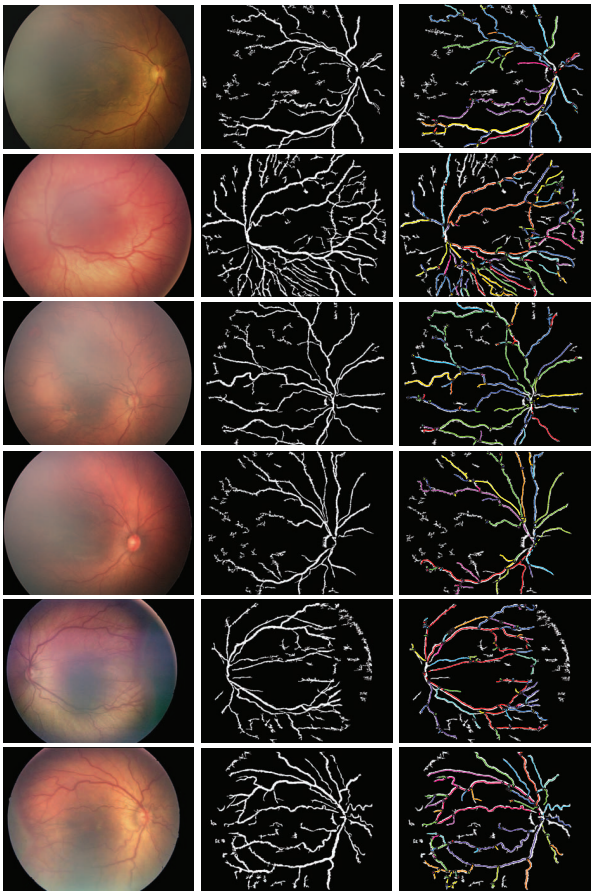


Fig. 4. Principal spanning forest result. Left column: Retinal scans. Middle column: Frangi filter output before connected component analysis. Right column: Reconstruction result after connected component analysis followed by a region growing algorithm. Clusters having small number of samples or outliers around the prephiler region (fifth row) are eliminated before the reconstruction.

[10] T. Hastie and W. Stuetzle, "Principal curves," *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 502–516, 1989.

[11] E. Bas and D. Erdogmus, "Sampling on locally defined principal manifolds," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2276–2279.

[12] E. Bas, *Extracting structural information on manifolds from high dimensional data and connectivity analysis of curvilinear structures in 3D biomedical images*, Ph.D. thesis, Northeastern University, 2012.

[13] B.W. Silverman, *Density estimation for statistics and data analysis*, Chapman & Hall/CRC, 1998.

[14] M.F. Chiang, L. Jiang, R. Gelman, Y.E. Du, and J.T. Flynn, "Interexpert agreement of plus disease diagnosis in retinopathy of prematurity," *Archives of ophthalmology*, vol. 125, no. 7, pp. 875, 2007.

[15] A.F. Frangi, W.J. Niessen, K.L. Vincken, and M.A. Viergever, "Multiscale vessel enhancement filtering," *Medical Image Computing and Computer-Assisted Intervention, MICCAI 1998*, p. 130, 1998.