Reconstructing Loopy Curvilinear Structures Using Integer Programming Appendix

Engin Türetken[1] Fethallah Benmansour[1] Bjoern Andres[2] Hanspeter Pfister[2] Pascal Fua[1]

[1]Computer Vision Laboratory (EPFL), CH-1015 Lausanne, Switzerland

[2]Visual Computing Group (Harvard University), MA-02138 Cambridge, US

{engin.turetken,fethallah.benmansour,pascal.fua}@epfl.ch {bandres,pfister}@seas.harvard.edu

We prove here that the negative log likelihoods of Eq. 2 and 3 can be written as linear and quadratic functions of the x_{ijk} indicator variables defined in Section 3.1.

1. Likelihood Term Derivation

In Eq. 2, we wrote

$$-\log(P(I,G|\mathbf{X}=\mathbf{x})) = \sum_{e_{ijk}\in F} w_{ijk}x_{ijk}, \tag{A.1}$$

where w_{ijk} is a cost term that accounts for the quality of the geodesic paths associated with the edge pair e_{ijk}. To prove this, we introduce two sets of auxiliary random variables into our formulation: one denoting presence of edges in the final solution and the other accounting for compatibility of consecutive edge pairs. Let $\mathbf{Y} = \{Y_{jk}\}$ be the vector of binary random variables denoting whether edges $\{e_{jk}\}$ truly belong to the underlying curvilinear structure, and $\mathbf{y} = \{y_{jk}\}$ the corresponding set of indicator variables.

Since we do not allow edges to have more than one active incoming edge pair, we have $y_{jk} = \sum_{e_{ij}\in E} x_{ijk} \leq 1$. As a result, for each edge e_{jk} in the solution, there can be at most one parent edge e_{ij} such that $x_{ijk} = 1$. Let Z_{jk} be the random variable standing for the parent of e_{jk} and $\mathbf{Z}$ be the vector of all such variables. That is, Z_{jk} can take values from the set $\{e_{ij} \mid e_{ij} \in E \setminus \{e_{kj}\}\}$. There is a one to one deterministic mapping between $\mathbf{X}$ and $(\mathbf{Y},\mathbf{Z})$. More specifically, we have

$$X_{ijk} = Y_{jk} \ \ \mathbb{1}(Z_{jk} = e_{ij}), \qquad \forall e_{ijk} \in F \tag{A.2}$$

where $\mathbb{1}(.)$ is an indicator function. We express the likelihood term of Eq. 1 in terms of $\mathbf{Y}$ and $\mathbf{Z}$ and drive the unary objective of Eq. 2 as follows:

$$\begin{aligned}
P(I,G|\mathbf{X}=\mathbf{x}) &= P(I,G|\mathbf{Y}=\mathbf{y},\mathbf{Z}=\mathbf{z}) && \text{(A.3)}\\
&= \prod_{e_{jk}\in E} P(I_{jk},E_{jk}|Y_{jk}=y_{jk},Z_{jk}=z_{jk}) && \text{(A.4)}\\
&= \prod_{e_{jk}\in E} \frac{P(Z_{jk}=z_{jk}|Y_{jk}=y_{jk},I_{jk},E_{jk})P(Y_{jk}=y_{jk}|I_{jk},E_{jk})P(I_{jk},E_{jk})}{P(Y_{jk}=y_{jk},Z_{jk}=z_{jk})} && \text{(A.5)}\\
&\propto \prod_{e_{jk}\in E} P(Z_{jk}=z_{jk}|Y_{jk}=y_{jk},I_{jk},E_{jk})P(Y_{jk}=y_{jk}|I_{jk},E_{jk}) && \text{(A.6)}\\
&\propto \prod_{e_{jk}\in E} [P(Z_{jk}=z_{jk}|Y_{jk}=1,I_{jk},E_{jk})P(Y_{jk}=1|I_{jk},E_{jk})]^{y_{jk}} \quad \times \\
&\qquad [P(Z_{jk}=z_{jk}|Y_{jk}=0,I_{jk},E_{jk})P(Y_{jk}=0|I_{jk},E_{jk})]^{1-y_{jk}} && \text{(A.7)}\\
&\propto \prod_{e_{jk}\in E} \left[\left[\prod_{e_{ij}\in E} P(Z_{jk}=e_{ij}|Y_{jk}=1,I_{jk},E_{jk})^{x_{ijk}}\right] P(Y_{jk}=1|I_{jk},E_{jk})\right]^{y_{jk}} \quad \times \\
&\qquad \left[\frac{1}{deg^*(v_j)} \quad P(Y_{jk}=0|I_{jk},E_{jk})\right]^{1-y_{jk}} && \text{(A.8)}\\
&\propto \prod_{e_{jk}\in E} \left[\left[\prod_{e_{ij}\in E} P(Z_{jk}=e_{ij}|Y_{jk}=1,I_{jk},E_{jk})^{x_{ijk}}\right] \left[\frac{P(Y_{jk}=1|I_{jk},E_{jk})deg^*(v_j,v_k)}{P(Y_{jk}=0|I_{jk},E_{jk})}\right]^{\sum_{e_{ij}\in E} x_{ijk}}\right] && \text{(A.9)}
\end{aligned}$$

where E_{jk} denotes the set of edges containing e_{jk} and its incoming edges in the graph, and I_{jk} denotes the image evidence around these edges. The term $deg^*(v_j, v_k)$ is the number of in-edges of vertex v_j excluding the edge e_{kj}, if it exists.

Eq. A.4 is obtained using the assumption that the image evidence around edge pairs are conditionally independent given that we know whether they belong to the curvilinear structures or not. In Eq. A.5 and A.6, we first use Bayes' rule and then remove the constant terms $P(I_{jk}, E_{jk})$ and $P(Y_{jk} = y_{jk}, Z_{jk} = z_{jk})$, assuming a uniform prior for both. We derive Eq. A.7 and A.8 by using the fact that $y_{jk}, x_{ijk} \in \{0, 1\}$ and substituting $P(Z_{jk} = z_{jk} | Y_{jk} = 0, I_{jk}, E_{jk})$ with $\frac{1}{deg^*(v_j)}$. Finally, in Eq. A.9, we drop the constant terms and express y_{jk} in terms of x_{ijk}'s. Taking the negative logarithm of Eq. A.9 and substituting $p^q_{jk} = P(Y_{jk} = 1 | I_{jk}, E_{jk})$ and $p^c_{ijk} = P(Z_{jk} = e_{ij} | Y_{jk} = 1, I_{jk}, E_{jk})$, we obtain

$$\sum_{f_{ijk} \in F} -log\left(\frac{p^c_{ijk} p^q_{jk} deg^*(v_j, v_k)}{(1 - p^q_{jk})}\right) x_{ijk} = \sum_{f_{ijk} \in F} w_{ijk}\, x_{ijk}, \tag{A.10}$$

which is what we wanted to prove. The probability p^q_{jk} denotes the likelihood that edge e_{jk} belongs to the curvilinear structure given the associated geodesic path and corresponding image evidence. This is an image-based term that accounts for the quality of the paths associated with the edges. In practice, instead of relying only on the image evidence around the edge e_{jk}, we evaluate the path classifier on a larger neighbourhood including its in edges $\{e_{ij}\}$ and use the corresponding probabilities in the above summation. Therefore, for each p^q_{jk} term in the above summation, we use the path probability corresponding to the edge pair e_{ijk}, which we obtain using the path classification approach of [2].

The term p^c_{ijk} denotes the probability that the edge pair e_{ijk} belongs to the structures given that its target edge e_{jk} belongs to them. In our experiments, this probability is expressed as a sigmoid function of a distinctive feature, which helps reconstructing the right connectivity at crossovers such as the one of Fig. 2. More specifically, for the brightfield micrographs shown in the third row of Fig. 4, this feature is taken as the tortuosity of the path (z axis is discarded) associated to e_{ijk}, because most of the fibers appear as linear filaments in the x-y plane. For the brainbow stacks, it is taken as the sum of the squared color distances in the CIELAB color space of pairs of vertices (v_i, v_j), (v_j, v_k) and (v_i, v_k). In both cases, the sigmoid function parameters are learned from the same training samples used for training the path classifier [1].

For inherently loopy structures such as blood vessels and road networks, we assume a uniform probability for p^c_{ijk} since for these structures, we don't need to disambiguate crossovers to obtain a loop-free solution, and hence, we are not interested in the true states of Z_{jk} variables. Therefore, we substitute $\frac{1}{deg^*(v_j, v_k)}$ for p^c_{ijk} in Eq A.10, which then simplifies to the negative log-likelihood ratio of the probabilities p^q_{jk}.

2. Prior Term Derivation

In Eq. 3, we wrote

$$-\log(P(\mathbf{X} = \mathbf{x})) = -\sum_{e_{ij} \in E}\left[\sum_{e_{mi} \in E} \log(p^t) x_{mij} + \sum_{e_{jn} \in E} \log\left(\frac{p^c}{p^t}\right) x_{ijn} \; + \sum_{e_{jn} \in E} \sum_{\substack{e_{jk} \in E \\ k<n}} \log\left(\frac{p^b p^t}{(p^c)^2}\right) x_{ijn} x_{ijk}\right],$$

where p^t, p^c and p^b are probabilities introduced in the main text. They are defined in terms of $M_{ij} = \sum_{e_{mij} \in F} X_{mij}$ and $O_{ij} = \sum_{e_{ijn} \in F} X_{ijn}$, two latent variables that denote the true number of incoming and outgoing edge pairs into and out of edge e_{ij}, respectively. Note that the M_{ij} are binary variables since we limit the number of active incoming edge pairs into an edge to one. Without loss of generality, assuming that the O_{ij} variables take values from the set $\{0, 1, 2\}$ and using a Bayesian network to model the dependencies between the variables M_{ij} and O_{ij}, we get

$$P(\mathbf{X} = \mathbf{x}) = \prod_{e_{ij} \in E} P(\mathbf{X}_{ij} = \mathbf{x}_{ij} | O_{ij} = o_{ij}) P(O_{ij} = o_{ij} | M_{ij} = m_{ij}) \tag{B.1}$$

$$\propto \prod_{e_{ij} \in E} P(O_{ij} = o_{ij} | M_{ij} = m_{ij}) \tag{B.2}$$

$$\propto \prod_{e_{ij} \in E} P(O_{ij} = o_{ij} | M_{ij} = 1)^{m_{ij}} \;\; P(O_{ij} = o_{ij} | M_{ij} = 0)^{(1-m_{ij})} \tag{B.3}$$

$$\propto \prod_{e_{ij} \in E} P(O_{ij} = o_{ij} | M_{ij} = 1)^{m_{ij}} \tag{B.4}$$

$$\propto \prod_{e_{ij} \in E} \left[P(O_{ij} = 0 | M_{ij} = 1)^{\mathbb{1}(o_{ij}=0)} \;\; P(O_{ij} = 1 | M_{ij} = 1)^{\mathbb{1}(o_{ij}=1)} \;\; P(O_{ij} = 0 | M_{ij} = 2)^{\mathbb{1}(o_{ij}=2)}\right]^{m_{ij}} \tag{B.5}$$

where $\mathbf{X}_{ij}$ denotes the vector of random variables $X_{ijn}, \ \ \forall e_{ijn} \in F$. In this work, we assume that all configurations $\mathbf{X}_{ij}$ are equally likely for an edge e_{ij} given that we know the total number of outgoing edge pairs out of it. Under this assumption, we obtain Eq. B.2, which we then decompose into two terms in Eq. B.3 using the fact that $m_{ij} \in \{0, 1\}$. In Eq. B.4, we remove the second term $P(O_{ij} = o_{ij}|M_{ij} = 0)^{(1-m_{ij})}$ in the product because we have $o_{ij} = 0$ when $m_{ij} = 0$ due to the connectedness constraints we impose, and hence, the term is always equal to 1. Finally, we drive Eq. B.5 by expressing the probability $P(O_{ij} = o_{ij}|M_{ij} = 1)$ as a product of three admissible event probabilities, namely termination, continuation and bifurcation, only one of which contribute to the product for each edge e_{ij}. The indicator functions are defined as follows:

$$\mathbb{1}(o_{ij} = 2) = \sum_{e_{jn} \in E} \sum_{\substack{e_{jk} \in E \\ k<n}} x_{ijn} \ \ x_{ijk} \tag{B.6}$$

$$\mathbb{1}(o_{ij} = 1) = \sum_{e_{jn} \in E} x_{ijn} - 2 \sum_{e_{jn} \in E} \sum_{\substack{e_{jk} \in E \\ k<n}} x_{ijn} \ \ x_{ijk} \tag{B.7}$$

$$\mathbb{1}(o_{ij} = 0) = \sum_{e_{mi} \in E} x_{mij} - \sum_{e_{jn} \in E} x_{ijn} + \sum_{e_{jn} \in E} \sum_{\substack{e_{jk} \in E \\ k<n}} x_{ijn} \ \ x_{ijk} \tag{B.8}$$

Note that multiplying these functions with $m_{ij} = \sum_{e_{mi} \in E} x_{mij}$ results in themselves since they are all equal to zero when $m_{ij} = 0$. Substituting them in Eq. B.5 and taking the negative logarithm, we obtain the desired result

$$-\sum_{e_{ij} \in E} \left[\sum_{e_{mi} \in E} \log(p^t) x_{mij} + \sum_{e_{jn} \in E} \log\left(\frac{p^c}{p^t}\right) x_{ijn} \ + \sum_{e_{jn} \in E} \sum_{\substack{e_{jk} \in E \\ k<n}} \log\left(\frac{p^b p^t}{(p^c)^2}\right) x_{ijn} x_{ijk} \right], \tag{B.9}$$

where $p^t = P(O_{ij} = 0|M_{ij} = 1)$, $p^c = P(O_{ij} = 1|M_{ij} = 1)$ and $p^b = P(O_{ij} = 2|M_{ij} = 1)$. We estimate these probabilities from the training data by first counting the total number of graph edges that intersect with the ground truth tracings and then finding the ratio of the number of bifurcating, continuing and terminating edges to this number.

Note that Eq. B.9 always results in positive values, which act as a regularizer. When the task is to reconstruct a tree structure, this helps penalize spurious bifurcations and early terminations at branch crossings. However, for loopy networks, it also penalizes legitimate bifurcations that are part of the loops. We therefore don't use this term for the blood vessel and the road network datasets.

References

[1] J. Platt. Advances in Large Margin Classifiers, chapter Probabilistic Outputs for SVMs and Comparisons to Regularized Likelihood Methods. MIT Press, 2000. 2

[2] E. Turetken, F. Benmansour, and P. Fua. Automated Reconstruction of Tree Structures Using Path Classifiers and Mixed Integer Programming. In CVPR, June 2012. 2