Automatic labeling of vascular structures with topological constraints via HMM

Xingce Wang[1], Yue Liu[1], Zhongke Wu[1], Xiao Mou[1], Mingquan Zhou[1], Miguel A. González Ballester[2,3], and Chong Zhang[2]

[1] Beijing Normal University, China
[2] SimBioSys, DTIC, Universitat Pompeu Fabra, Spain
[3] ICREA, Spain

Abstract. Identification of anatomical vessel branches is a prerequisite task for diagnosis, treatment and inter-subject comparison. We propose a novel graph labeling approach to anatomically label vascular structures of interest. Our method first extracts bifurcations of interest from the centerlines of vessels, where a set of geometric features are also calculated from. Then the probability distribution of every bifurcation is learned using a XGBoost classifier. Finally a Hidden Markov Model with a restricted transition strategy is constructed in order to find the most likely labeling configuration of the whole structure, while also enforcing topological consistency. In this paper, the proposed approach has been evaluated through leave-one-out cross validation on 50 subjects of centerlines obtained from MRA images of healthy volunteers' Circle of Willis. Results demonstrate that our method can achieve higher accuracy and specificity, while obtaining similar precision and recall, when comparing to the best performing state-of-the-art methods. Our algorithm can handle different topologies, like circle, chain and tree. By using coordinate independent geometrical features, it does not require prior global alignment. Source code and data are available under http://doi.org/10.5281/zenodo.809931.

1 Introduction

Automatic anatomical labeling approaches for tubular-like structures has been investigated for a couple of decades. Methods exist from 2D atlas registration [6], 3D branch matching [8], to the maximum likelihood estimation [7], *etc.* These methods are applicable to tree-like structures with short and straight branches, such as the airways and abdominal aorta. However, they are not feasible for structures with large variations in geometry or/and topology. For example, the Circle of Willis (CoW) is genus 1 in terms of topology, with total of 83 variations [4]. Majority of human have one or more missing arteries in their CoW. Additionally, vessels can be twisted and intertwined, which create complex geometries with large range in branch length, curvature, torsion and radius. These rise difficulties for atlas based labeling approaches [1]. Even for the same person, his/her vessel structure and morphology change over time. All these make the automatic CoW labeling a challenging problem.

In [2], a maximum a posteriori (MAP) classification is firstly used to identify five branches. It then combines the MAP with a graph matching method to label the CoW in the form of three separate trees. The MAP inferences in [9, 10] are formulated as a quadratic binary programming problem. This formulation can handle non tree-like vasculature with high efficiency. It simultaneously segments and labels CoW with an integer programming. However, their methods do not differentiate symmetrically located bifurcations.

This paper describes a novel and generic approach to anatomically labeling vascular structures. Our method first extracts bifurcations of interest from the centerlines of vessel structures, where a set of geometric features are calculated. Then the probability distribution of these bifurcations is generated using an XGBoost classifier. Finally a Hidden Markov Model (HMM) with a restricted transition strategy is constructed to find the most likely labeling configuration of the whole structure, while constraining all branches or bifurcations with a feasible topology. Our main contribution is threefold: (1) Coordinate independent geometrical features are used to describe the bifurcations, which alleviate the global alignment that is required in other approaches. (2) XGBoost method is used to learn the probability distribution of bifurcations of interest as well as those not of interest. (3) The topology identification is considered as a matching problem which is solved by an HMM.

2 Method

Problem formulation. Each vascular structure is modeled as a centerline graph together with a set of radii corresponding to all centerline points. Based on this initial setting, although many features like centerline length, curvature, etc, could be calculated, these features alone are not sufficient to classify each anatomical structure. We propose to classify bifurcation points based on extracted geometrical features associated with them. That is, for each bifurcation point, we calculate features from the three branches, which are uniquely identified as the Head, Left and Right, connected to the bifurcation as well as a few combined features derived from them. Each vascular structure is modeled as an undirected weighted graph $G = (V, E, A)$, where a sets of vertex $V = \{v_1, \ldots, v_n\}$ denote the bifurcation points of the structure, $E \subset V \times V$ is a set of edges representing vessel connections with these bifurcations, and $A : V \mapsto \mathbb{R}^f$ are features associated with V. The bifurcation $c = \{V_c, E_{c\_head}, E_{c\_left}, E_{c\_right}, A_c\}$, $c = 1, \ldots, f$, is generated by G. The vessel labeling problem can be considered as a bifurcation mapping problem, based on an available graph set $\{\hat{G}\}$ as well as known bifurcation set $\{\hat{C}\}$. All these form a knowledge base $\mathcal{K}$. For a given target graph G^t, the target labeling is defined as $L : C^t \mapsto \{\hat{C}\} \cup \{\varnothing\}$, where $C^t = \{c_i^t : i = 1, \ldots, m\}$represents bifurcation sets on the G^t, and the label $\varnothing$ represents certain bifurcation point that is not of interest. Obviously, a surjective mapping to $\varnothing$ is allowed. As for the remaining subset of C^t, the inclusion map $C^t \mapsto \{\hat{C}\}$ is injective but not necessarily surjective. It is possible that multiple solutions that fulfill these conditions exist. Let the joint probability distribution

of every possible solution be $P(C^t, L|\mathcal{K})$. We aim at finding the label L^* with the maximum a posteriori (MAP) probability:

$$L^* = \arg\max_{L\in\mathcal{L}} P(C^t, L|\mathcal{K}), \tag{1}$$

where $\mathcal{L}$ denotes the labeling solution space.

The proposed model. By observing the vascular anatomy, it seems that anatomical bifurcations can only appear in a certain order, starting from a reference point, e.g. internal carotid artery and vertebral basilar artery in the CoW. However, there always exists anatomical variabilities. For example, CoW exhibits large variability in its topology [2], including the absence of bifurcation points of interests. This also means that the absence of one bifurcation can imply that those further downstream cannot be present either. Despite of those variations, it is still helpful to impose topology constraints during the inference.

We propose to formulate the labeling process as an HMM with a restricted transition strategy. In our model, a bifurcation's label is assumed to depend only on those of its immediate neighbors while independent on the rest. Therefore, the bifurcation points with feature sets can be considered as a sequence of observations. Their labels are considered as a state sequence with state transition probabilities as topology constraints from the *prior* training population. Feature set A^t generated by G^t gives information about the sequence of states. Let $S=\{s_i\}, s_i\in\{l_1,\ldots,l_k\}\cup\{\varnothing\}$ be the state space, containing labels of different anatomical bifurcation points and a null label $\varnothing$ for those not of interest. Let $O=\{o_i\}, o_i\in\{c_1,\ldots,c_m\}$ be the observation space of all anatomically possible bifurcations. For a given observed sequence of length L, $Y=\{y_1,\ldots,y_L\}$, $y_i\in O$, we can extract its corresponding state sequence $X=\{x_1,\ldots,x_L\}, x_i\in S$ from $\{\hat{G}\}$. Then, we can define the state transition probability matrix Q and the output probability matrix B with dimensions equal to that of the state space:

$$\begin{aligned} Q&=\left[q_{ij}\right]_{k*k}, q_{ij}=P(x_{t+1}=s_j|x_t=s_i)\\ B&=\left[b_{ij}\right]_{k*m}, b_{ij}=P(y_t=o_j|x_t=s_i) \end{aligned} \tag{2}$$

The initial probabilities on the state space is defined as $\Pi=\{\pi_i\}_k, \pi_i=P(x_1=s_i)$. Following the Markov property, we have:

$$\begin{aligned} &P(x_t|x_1,x_2,\ldots,x_{t-1}) = P(x_t|x_{t-1}), \text{ and}\\ &P(y_t|y_1,y_2,\ldots,y_{t-1},x_1,x_2,\ldots,x_t) = P(y_t|x_t). \end{aligned} \tag{3}$$

The problem in Eq 1 can be formulated as: $X^* = \arg\max_X P(Y,X)$, which can be solved by the Viterbi algorithm [5]. Let $V_{i,k}$ be the maximum probability of a sequence with length i and end state s_k. With dynamic programming, we have:

$$\begin{aligned} V_{1,k}&=P(y_1|s_k)\cdot\pi_k\\ V_{i,k}&=P(y_i|s_k)\cdot\max_{x\in[1,L]}(q_{x,k}\cdot V_{i-1,x}). \end{aligned} \tag{4}$$

Table 1: Geometrical features defined on vessel centerline models.

Type	Parameter	#
basis	lengths (*len*), average radii (*rad*) , average curvatures (*cur*), average torsion(*tor*), angles between bifurcation's three tangential vectors (*tan*), angles between bifurcation and its three branches' end points (*seg*)	6×3
combined	$\int len \times cur$ (*alc*), rad/len (per), ratio of lengths between its left & right branches from any branch (*bal*)	3×3

The solution can be found: $x_L = \arg\max_{x\in[1,L]} V_{L,x}$. The critical task is to convert the graph G^t into a chain, and to find the optimal implicit state sequence Y by the observed sequence X and model. It remains to calculate Q and B(Eq 2).

Note that the topology of graph G^t is always different, due to the fact that, some branches may be missing in certain cases, and some branches may not be what we are interested in. Such heterogeneity could be caused by either the patient anatomy or the process of segmentation and skeletonization. Bifurcations are"sorted" by the angles between each vector from E (Fig 1(a)) to a bifurcation and a specific reference vector. For example, assuming that the CoW is complete, the order can normally be determined by traversing. For seemingly "multiple-tree" structures that are split by the gap of some missing blood vessels, and if a bifurcation is not connected with the others, the next one in the clock-wise order can be considered as the next bifurcation. With this rule, all bifurcations can be virtually connected even with missing branches. Additionally in this way, even if symmetric bifurcations have similar geometric features, they have different position in the chain. For any given sequence Y, except the $\varnothing$, the probability of each elements s_i in state space appear in a particular location is different. The vascular structure of interest can be regarded as a Markov chain, including the CoW where multiple trees maybe found instead of a "circle".

Then for the prediction step, we construct the chain model of the output/observed sequence, $P(x_t{=}s_i)$ can be calculated from the occurrence rate of s_i in the training set, and $P(x_t{=}s_i|y_t{=}o_j)$ can be obtained from a bifurcation label classifier trained on $\hat{C}$.

Bifurcation label classifier. In order to estimate the likelihood of a bifurcation o_j with label s_i, $P(s_i|o_j)$, we propose to use the XGBoost algorithm with probability estimation [3]. It is an improved gradient tree boosting based on classification and regression tree. This method is particularly suitable for our vascular datasets which spans in large space but also with larger sparsity. The connected three branches are uniquely identified, such that f features of each branch could be combined to $3f$ features for the corresponding bifurcation. Since the CoW ring lies approximated on a plane, the relative positions of the branches around any bifurcation are the same. Then we just consider a branch, whose one

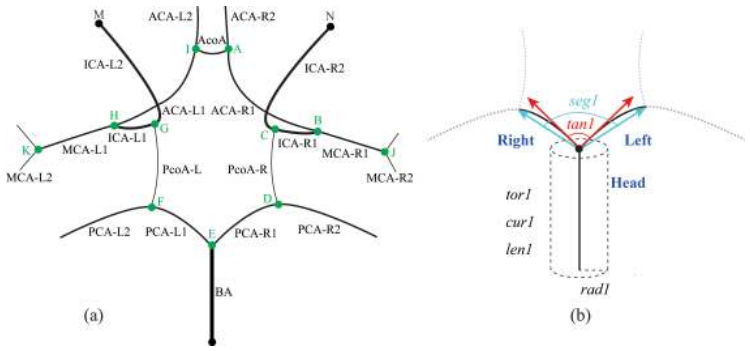

Fig. 1: (a) A sketch of CoW, together with the 11 labels (A-K) of the bifurcations of interest. (b) A sketch of the basis features (as in Table 1) of one bifurcation. Only the corresponding ones for one of the three branches are indicated, those of the other two can be similarly defined.

end is on the circle structure and the other not, as the Head one. With the bifurcation as the origin, from the Head branch anti-clockwise in the ring plane we can identify the children branches (Fig 1(b)). Specifically, 27 coordinate invariant features are calculated such that the global alignment can be avoided (Table 1).

3 Experiments and Results

Data and technical details. Our proposed approach has been evaluated on the public dataset distributed by the MIDAS Data Server at Kitware Inc.. It contains 50 MRA images of the cerebral vasculature from healthy volunteers together with theirs egmentations and centerlines [2], as well the ground truth manually indicted by an expert. We first prune the centerline model to a region around the CoW. This results in about 18 bifurcations in each case. In this step, we use features of Euclidean distance to three base points (i.e. where blood enters the whole CoW) and the smallest average radius of the bifurcation. Due to the heterogeneity of the different cases, according to the expert's experience, we are mainly interested in 11 bifurcation points, which are labeled as A-K in Fig 1. In the 27 features reported in Table 1, *tor* is removed by a Least Significant Difference metric, which results in 24 features. In the approximate plane of the CoW, the orthogonal vectors of tangent from the bifurcation point of the BA branch (E) is considered as the reference vector. Branches are sorted by the angles between each vector from E to a bifurcation and the reference vector.

Evaluation. A leave-one-out cross-validation was performed to asses the performance of the proposed method. Exemplar labeled segmentation of a normal CoW (Fig 2(a)) and a varied CoW (Fig 2(b)) are shown, where branches of interest, i.e. those determined by the 11 bifurcations of interest, are color-coded, with grey-colored ones being not of interest. We could see that our approach is able to handle topological disturbances around the CoW (e.g. grey-colored small branches in the enlarged views).

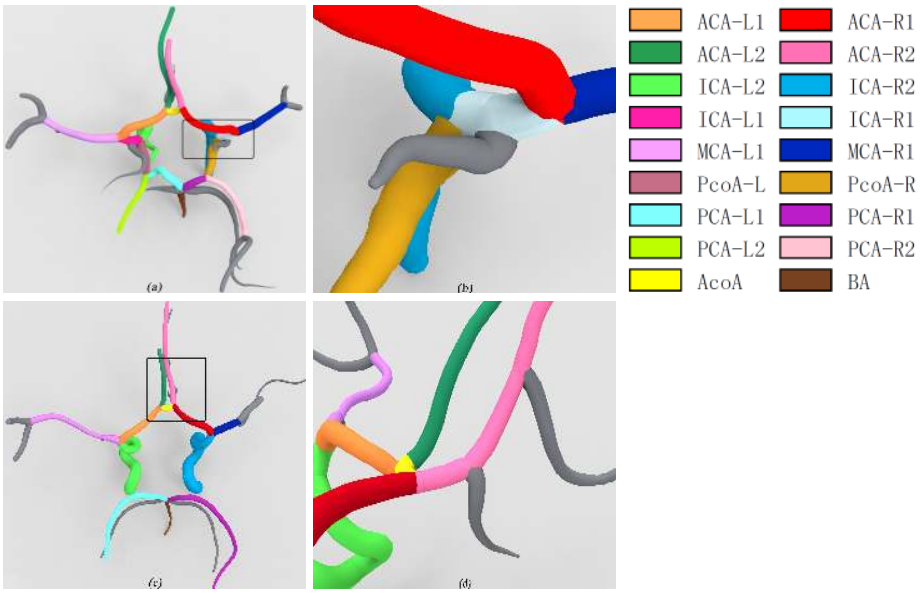

Fig. 2: Color-coded example labeling results of: (a)(c) normal and (b)(d) varied CoW models (with grey-colored ones being not of interest).

Table 2: Evaluation comparison of the 11 bifurcations of interest (Fig. 1) with Accuracy (A), precision (P), recall (R) and Specificity (S) are reported for the proposed method with (*Ours*) and without (*Ours w/o*) the topological constraints, as well as for the best performing state-of-the-art methods [2, 10].

L	Ours				Ours w/o				Bogunović et al. [2]				Robben et al. [10]			
	A	P	R	S	A	P	R	S	A	P	R	S	A	P	R	S
A	99.3	95.9	94.0	99.7	**99.5**	**97.9**	94.0	**99.9**	96	95	**100**	88	*96*	*100*	*95*	*n.a.*
I	99.3	**100**	87.8	**100**	98.8	97.6	81.6	99.9	**100**	**100**	**100**	**100**				
B	99.2	93.9	92.0	**99.6**	98.7	88.5	92.0	99.2	**100**	**100**	**100**	n.a.	*100*	*100*	*100*	*n.a.*
H	**99.8**	96.2	**100**	99.7	99.6	94.3	100	99.6	98	**100**	98	**100**				
C	**99.4**	**100**	90.0	**100**	98.7	93.5	86.0	99.6	98	**100**	**97**	**100**	*98*	*100*	*97*	*n.a.*
G	**99.5**	96.0	96 .0	99.7	99.5	**100**	92.0	**100**	98	97	**100**	93				
D	**98.0**	92.3	72.0	**99.6**	97.5	79.6	78.0	98.7	96	**95**	**100**	87	*97*	*100*	*95*	*n.a.*
F	97.6	77.8	84.0	98.5	98.1	85.4	82.0	99.1	**100**	**100**	**100**	**100**				
E	**99.5**	92.6	**100**	**99.5**	97.7	79.2	84.0	98.6	96	**96**	**100**	n.a.	92	98	93	n.a.
J	**98.2**	**90.7**	78.0	**99.5**	94.6	55.6	50.0	97.4	80	80	**100**	n.a.	*87*	*87*	*100*	*n.a.*
K	**97.4**	79.6	78.0	**99.5**	94.2	51.8	58.0	96.5	84	**84**	**100**	n.a.				

Table 2 shows the accuracy, precision, recall and specificity for each bifurcation of interest separately, using our method with and without the topological

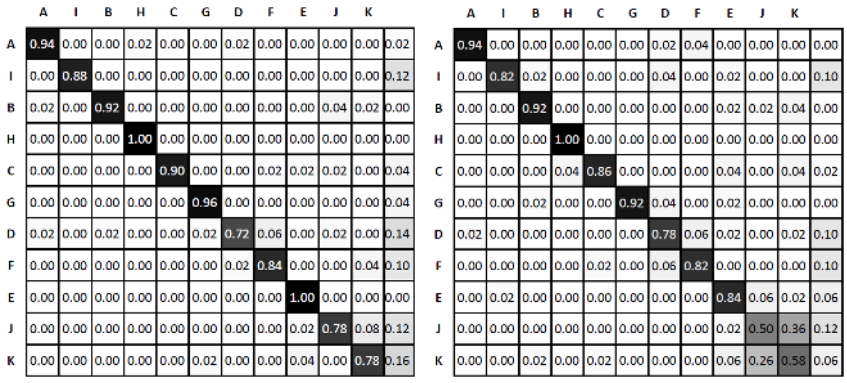

Fig. 3: The confusion matrices with (*left*), without (*right*) topological constraints.

constraints. They are compared with two best performing methods [2,10]. In general, our method offers higher accuracy and specificity, while those by [2] offers higher recall. These suggests that our method also tries to label both bifurcations of and not of interest correctly. This is probably more suitable when the data is more generic and not trimmed only for those of interests. But for datasets that contain only relevant branches, the method by [2] could perform better. With these said, for all of the bifurcations, the differences between the methods are not significant, as have also been reported in [10]. However, note that the method by [10] does not distinguish symmetrical bifurcation points, which are one of the main error sources in the CoW classification problem. Thus, we do not intend to compare and interpret their numbers directly. The error source could be seen in the plotted confusion matrices shown in Fig 3, where the "confused" symmetrically located bifurcation points could be seen, e.g. J and K, D and F. From the confusion matrices, we could see that the bifurcation points J and K are the most wrongly classified. In addition to confusing each other, it is due to the difference in the length of the MCA-1 vessel and the presence of other bifurcations. The errors of the bifurcation A, I, D and F are mainly caused by missing vessels, e.g. AcoA and PCoA. More interestingly, due to the obvious features of ICA, the bifurcation C and G are hardly affected by the reason above.

4 Conclusions

We have proposed a supervised learning method for anatomical labeling of vascular branches. It learns the probability distribution of every anatomical branch bifurcation with a XGBoost classifier, which is then used as inputs to a Hidden Markov Model so as to enforcing the topology consistency. Our method is able to solve a labeling problems of trees, loops and chains. Furthermore, it has demonstrated that it can handle a large anatomical variability such as those present in the topology of the CoW, and is able to map labels to cases containing different subsets of bifurcation points of interest. The leave-one-out cross validation performed on 50 cases has shown higher accuracy that the best performing

state-of-the-art methods. Additionally, our method uses coordinate independent features, reducing the need for a global rigid registration step before the labeling process. Pending issues such as validations on larger population, on structures with different topologies, the parallelization of the bifurcation classifier and the optimization of topological constrained model fall into our future work.

Acknowledgments

This research was partially supported by the Chinese High-Technical Research Development Foundation (863) Program (No.2015AA020506), Beijing Natural Science Foundation of China(No.4172033), the Spanish Ministry of Economy and Competitiveness, through the Maria de Maeztu Programme for Centres/Units of Excellence in R&D (MDM-2015-0502), and the Spanish Ministry of Economy and Competitiveness (DEFENSE project, TIN2013-47913-C3-1-R). We thank the authors of [2] for sharing their centerline delineations.

References

1. H. Bogunović, J.M. Pozo, R. Cárdenes, and A.F. Frangi. Anatomical labeling of the anterior circulation of the Circle of Willis using maximum a posteriori classification. In *MICCAI*, pages 330–337, 2011.
2. H. Bogunović, J.M. Pozo, R. Cárednes, L. San Román, and A.F. Frangi. Anatomical labeling of the Circle of Willis using maximum a posteriori probability estimation. *IEEE T Med Imaging*, 32(9):1587–1598, 2013.
3. T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *ACM SIGKDD*, pages 785–794. ACM, 2016.
4. O.H. Del Brutto, R.M. Mera, M. Zambrano, and J. Lama. Incompleteness of the circle of willis correlates poorly with imaging evidence of small vessel disease. a population-based study in rural ecuador (the Atahualpa project). *J Stroke Cerebrovasc Dis*, 24(1):73–77, 2015.
5. G.D. Forney. The viterbi algorithm. *Proc IEEE*, 61(3):268–278, 1973.
6. K. Haris, S.N. Efstratiadis, N. Maglaveras, C. Pappas, J. Gourassas, and G. Louridas. Model-based morphological segmentation and labeling of coronary angiograms. *IEEE T Med Imaging*, 18(10):1003–1015, 1999.
7. T. Matsuzaki, M. Oda, T. Kitasaka, Y. Hayashi, K. Misawa, and K. Mori. Automated anatomical labeling of abdominal arteries and hepatic portal system extracted from abdominal CT volumes. *Med Image Anal*, 20(1):152–161, 2015.
8. K. Mori, M. Oda, T. Egusa, Z. Jiang, T. Kitasaka, M. Fujiwara, and K. Misawa. Automated nomenclature of upper abdominal arteries for displaying anatomical names on virtual laparoscopic images. In *International Workshop on Medical Imaging and Virtual Reality*, pages 353–362. Springer, 2010.
9. D. Robben, E. Türetken, S. Sunaert, V. Thijs, G. Wilms, P. Fua, F. Maes, and P. Suetens. Simultaneous segmentation and anatomical labeling of the cerebral vasculature. In *MICCAI*, pages 307–314, 2014.
10. D. Robben, E. Türetken, S. Sunaert, V. Thijs, G. Wilms, P. Fua, F. Maes, and P. Suetens. Simultaneous segmentation and anatomical labeling of the cerebral vasculature. *Med Image Anal*, 32:201–215, 2016.