CTA CORONARY LABELING THROUGH EFFICIENT GEODESICS BETWEEN TREES USING ANATOMY PRIORS

by

MEHMET AKIF GULSUN

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science in Engineering

Baltimore, Maryland

October, 2014

Abstract

We present an efficient realization of recent work on unique geodesic paths between tree shapes for the application of matching coronary arteries to a standard model of coronary anatomy in order to label coronary centerlines extracted in cardiac Computed Tomography (CT) Angiography (CTA) data. Automatically labeled coronary arteries would speed reporting coronary diseases for physicians, be used for building patient specific myocardial segment models for the correct integration of coronary anatomy with myocardial function and guide segmentation algorithms for extracting the centerline representation of coronary arteries in CTA data.

Our approach builds on Quotient Euclidean Distance metric that leverages both geometric and topological information in order to compute unique natural and continuous geodesic deformations between tree-shapes. The efficiency of our approach and the quality of the results are enhanced using the relative position of detected cardiac structures including four chambers and pericardium. We explain how to efficiently compute the geodesic paths between tree shapes using Dijkstra's algorithm and we present a methodology to account for missing side branches during matching. We address computational difficulties for labeling large and complex coronary arteries and present evaluation results on 50 expert annotated and 20 automatically detected coronary centerlines in CTA data. For nearly all labels our approach shows promise compared with recent work and we show results for 12 additional labels. The results

show the practicality and accuracy of our approach for labeling patient specific coronary centerlines extracted in CTA.

Advisors:

Dr. Russell H. Taylor from JHU

Dr. Gareth Funka-Lea from Siemens

Acknowledgement

I would like to express my gratitude to my advisors Dr. Gareth Funka-Lea from Siemens Corporation, Corporate Technology and Dr. Russell H. Taylor from The Johns Hopkins University for their valuable guidance and advice through the work of this master thesis. My deepest thanks to Dr. Gareth Funka-Lea for his tremendous support and help during the course of my master's study and for insightful conversations during the development of ideas in this thesis.

In addition, I would like to take this opportunity to thank Dr. Yefeng Zheng from Siemens Corporation, Corporate Technology who has been very supportive with sharing his tools to segment patient heart models used in this thesis work. I also would like to thank Dr. Aasa Feragen from University of Copenhagen for our valuable discussions on the QED tree-shape metric and acknowledge Siemens for providing me annotated CTA coronary data.

Finally, sincere thanks go to my family who always gave me unconditional support throughout my life.

Table of Contents

List of Tables

List of Figures

1 Introduction

1.1 Motivation / Significance

According to American Heart Association, coronary artery disease (CAD) is a leading cause of death among cardiovascular diseases in the United States [1]. Coronary arteries extracted from CT angiography (CTA) are used for advanced visualization and quantification purposes to facilitate the diagnosis and treatment workflow for radiologists and cardiologists. Automated coronary labeling can be used to facilitate this workflow for time efficient reporting, examining the correlation of coronary arteries with myocardial function and guiding coronary segmentation algorithms.

It is critical that an imaging physician report the anatomical location of pathology in a standard way to the referring physician. A principle goal of automated medical image analysis is the efficient reporting of such findings following established medical guidelines. Criteria have been established for how lesions along the coronary arteries should be reported from CT angiography (CTA) [2, 3]. The number of coronary labels varies in different standards but there is agreement on the major labels. The AHA established a standard 15 segment model in 1975 [2]. Our model follows more closely the more recent and more complete models of [3, 4]. The physician generally knows which coronary segment contains a lesion but it is time consuming to label images when more than a few labels must be applied due to the large variability in coronary anatomy.

Therefore, one of the goals of automatic coronary labeling is to speed the preparation of a report.

Coronary arteries have an important role in keeping the cardiovascular system functioning by feeding heart muscles. Correct integration of coronary anatomy with myocardial segments can be used to infer the presence, severity and location of coronary disease based on the myocardial function and perfusion which can be non-invasively evaluated using ultrasound or magnetic resonance imaging. Standard practices rely on empirical and population based models such as 17 segment myocardial model to define the correspondence between coronary arteries and myocardial segments which is not sufficient for personal medicine due to the large variability in coronary anatomy [5, 6]. Automated coronary labeling can provide an efficient way to build patient specific models to relate coronary anatomy with myocardial segments.

Detection of coronary arteries in CTA is a challenging task due to irregularity of their anatomy, diseases such as stenosis and occlusion, and imaging artifacts. Most of previous approaches are data-driven and do not exploit any high level coronary prior model. A recent work [7] uses a model-driven algorithm to extract major coronary branches to constraint the segmentation and reports very promising results. Similarly, automated labeling can be used to build coronary models from partially recovered trees to guide the segmentation by predicting the location of missing branches and leakages.

1.2 Background

In order to label the coronaries, our approach will leverage both geometric and topological information and incorporate coronary anatomy priors to define the correspondence between a labeled model and unlabeled data. We only consider the centerlines of the coronaries and not the coronary lumen. Most prior work labeling vascular or airway trees extracts an abstract graph that captures the topology of the tree but uses limited or no geometric information. A graph matching algorithm is then run to define the best correspondence between nodes in the model and in the unlabeled graph [8, 9]. This is a natural approach when attempting to label coronaries in 2D X-ray angiograms where the 3D geometry has been lost as in [8, 9].

The best known method for computing geodesic deformation between tree shapes is the tree edit distance (TED) [11]. TED defines the geodesic distance as the minimum cost for deforming one tree to the second by simple tree operations such as adding, removing or deforming a branch. This metric was previously used for matching cerebral vessels [12] although it was not used for labeling. [12] uses a bidirectional deformation between two cerebral trees in which both trees are edited concurrently by only removing their branches until they represent the same topology where the pruned trees are then deformed to match each other. With the TED metric there are frequently infinitely many geodesic paths between two tree-shapes and therefore a TED based geodesic does not satisfy the local uniqueness property which is required for assigning unique labels or performing a sound statistical analysis.

Recent work by [10] labels 3D CTA in 2 steps based on an anatomical model derived from [8]. First, a reduced coronary model is rigidly aligned using a point-set registration method and labels for the major branches are assigned based on proximity. Second, the side branches are labeled based on iteratively reducing a cost function which penalizes unlabeled centerline points. Our approach can handle more complete coronary trees. In the results section, we compare our results with theirs.

Our solution builds on the recent work defining unique geodesic paths between trees [13], specifically, the proposed Quotient Euclidean Distance (QED) metric to compute *unique* natural and continuous geodesic paths between tree-shapes by concatenating local tree deformations [13]. The paths are defined in a high dimensional space that captures the topology and geometry of the tree. The QED geodesic was previously applied to the computation of an average airway tree [14] and to matching airway trees [15]. In [15], a novel method for the automated labeling of an unseen airway tree was proposed by first finding its correspondence to labeled airway trees in a training set and then assigning a label to each branch of the unseen tree based on a majority vote scheme. It establishes the branch correspondence between two airway trees by QED deformation. Similarly, our approach uses QED deformation for propagating labels from an atlas coronary model to unseen coronary tree.

1.3 Summary of Contributions

The main contributions reported in this thesis are 1) practical definition of the 3D QED geodesic space for coronary trees based on automatically detected cardiac anatomy, 2) an efficient approximation of the QED geodesic using Dijkstra's algorithm, 3) explicit explanation of our heuristic in the QED framework to handle missing side branches, 4) the use of a coronary territory prior to augment the shape and topology information and to label large and complex coronary trees.

2 Method

2.1 Overview

The focus of this thesis is to label coronary centerlines that are either annotated by experts or automatically detected in cardiac CTA data which were acquired from patients known or suspected to have coronary artery disease. These patient CTA data were acquired with contrast agent for the purpose of diagnosing coronary artery disease in order to rule out the coronary artery catheterization.

In our method, we label a target coronary tree by finding a minimum cost deformation from a coronary atlas model. We define the deformation cost between two coronary trees as the weighted sum of a labeling cost term and a regularization term using a tree-shape metric. Specifically, the labeling cost term is based on the likelihood of the labels assigned to the target coronary branches and this likelihood is computed from the spatial distribution of coronaries over the heart surface. One labeling approach would be to only use the spatial distribution of coronaries and assign labels to coronary branches based on maximum likelihood. However, this approach can easily over fit the training data and may produce unrealistic labeling results that do not take the coronary topology and branch geometry into account. Therefore, we introduce a regularization term in our deformation cost using the QED metric between the coronary atlas model and the target coronary tree. We will refer to the minimum cost deformation between two coronary trees as the *coronary geodesic* in this thesis.

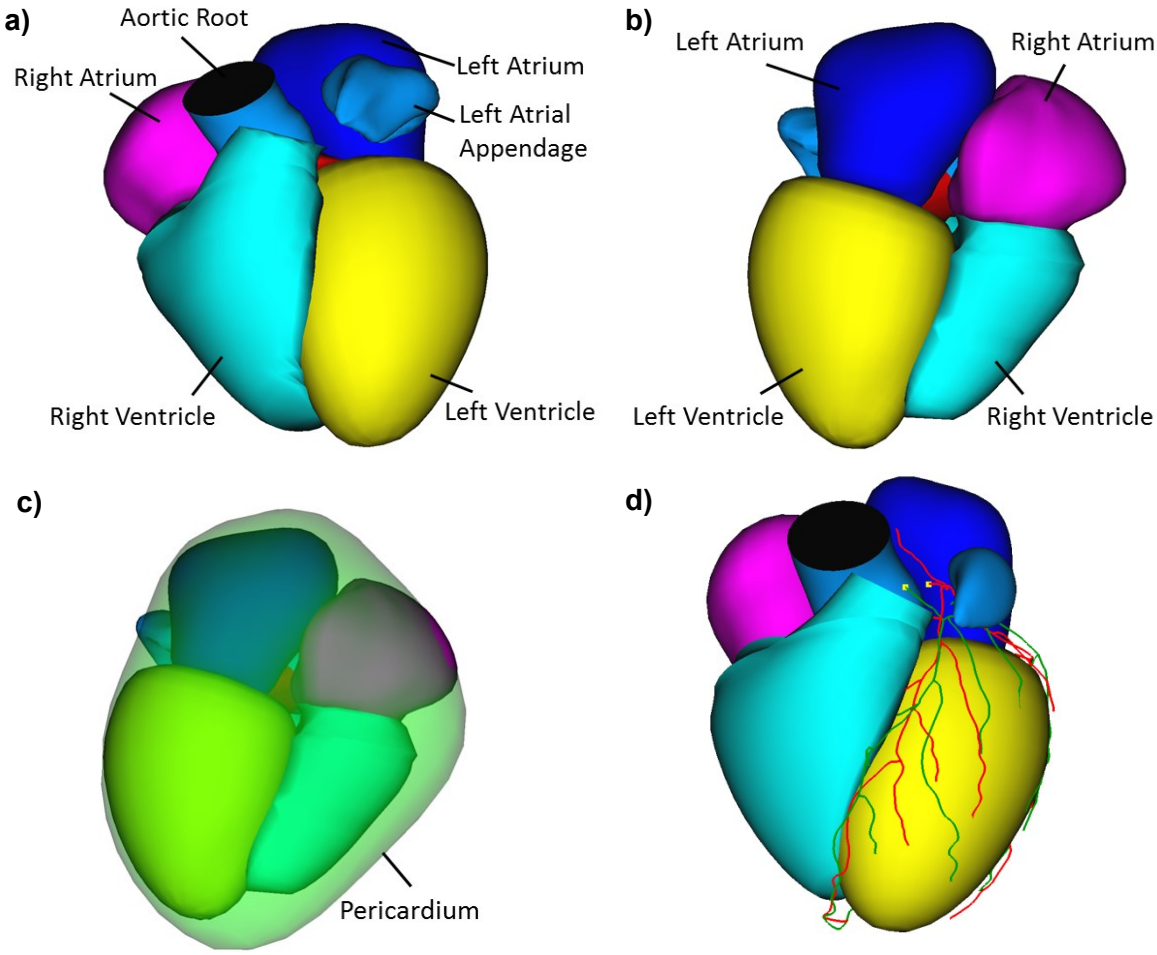


Figure 1: Automatically detected heart models visualized in (a) anterior and (b) posterior views. (c) Detected pericardium mesh. (d) Average heart model with two aligned left coronary trees overlaid.

We first explain how we compute coronary territory prior which is used in various steps of our method. Following the coronary territory section, we explain the traditional QED metric, its incorporation into our coronary geodesic formulation and our heuristic method to handle missing side branches. Finally, we present our coronary atlas model and explain how we address computational difficulties in applying QED for labeling large complex coronary trees.

2.2 Statistic On Coronary Territories

Coronary arteries lie roughly on the surface of the heart between the heart chambers and the pericardium. Our labeling method uses the spatial distribution of coronary branches in 3D space as anatomy priors for computational efficiency and improved labeling results. We compute the spatial distribution of coronary branches based on the average density of coronary branches over the population.

The shape and size of coronary trees may vary depending on the patient specific anatomy of the heart. Therefore, it is necessary to first align the coronary trees in a canonical coordinate system. We segment the heart structures in patient cardiac CTA dataset that is the same volume used for the extraction of coronary arteries. Specifically, we automatically detect the patient heart model including cardiac chambers, aortic root, left atrial appendage and pericardium mesh models in patient CTA dataset using the method [16, 17], Figure 1a, b, c. The method in [17] is a two-step approach where it first localizes heart structures by predicting their position, orientation and scale using marginal space learning, and then estimates their 3-D mesh models through a learning-based boundary delineation algorithm. In addition, we compute a reference heart model by first segmenting the patient specific heart structures from 50 CTA datasets in our training set using the same method in [17] and then averaging them. Each mesh point of these heart structures has correspondence through the training set. We achieve alignment of a patient specific coronary tree by using a thin-plate spline (TPS) interpolation [18] defined between the patient specific heart model and our

reference heart model as data points. TPS is a spline-based technique for two dimensional data interpolation that has been widely used for non-rigid image alignment and shape matching. Given a set of data points, TPS fits an interpolation function that passes through the points exactly while minimizing the smoothness (bending) energy defined as the integral of the square of second derivatives of interpolation function. The smoothing variant of TPS uses a regularization parameter in order to relax the exact data fit requirement. In our work, no smoothness regularization is used such that the TPS interpolation function passes through the points exactly. Figure 1d shows two aligned coronary trees overlaid on our reference heart model.

Aligned coronary trees are used to compute average density map of coronary branches in a discrete volume bounding the reference heart model. Specifically, for each aligned coronary tree in the training set, we first compute a coronary tree distance map D by running Dijkstra's algorithm on a graph $G = (N, E)$ with nodes N that correspond to discrete voxels and edges E that connects each voxel to its 26 neighbors and are assigned with a weight as the Euclidean distance to its neighboring voxels. The distance map is initialized to infinity for all voxels in the volume except the voxels intersecting with coronary tree points and the distance map for intersecting voxels are initialized with their distance to closest coronary tree point. The labels along the branch points are propagated during the Dijkstra's algorithm in order to assign a label to each voxel in the distance map, Figure 2. The density value $\rho_l(x)$ of label l at voxel x is then computed from Gaussian weighting of distance value $D(x)$ with standard deviation σ in a local

Coronary Distance Map Computation

Input: 3D coronary centerlines and discrete volume bounding the heart model.

Let $G = (N, E)$ be an undirected graph with nodes N that correspond to discrete voxels and edges E that connect each node to its 26 neighbors with Euclidean distance weight.

1) Initialize distance map d with infinity and label map l with blank values.
2) Find nodes u_{source} intersecting input coronary centerline points and initialize their distance map $d(u_{source})$ with their distance to closest centerline point and their label map $l(u_{source})$ with the label of closest centerline point.
3) Run Dijkstra's algorithm:
    Let Q be a minimum priority queue
    Add all nodes N in graph G to Q with priority d
    While Q is not empty:
        Let u be the node in Q with minimum distance
        Remove u from Q
        For every neighbor of v of u:
            If $d(v) > d(u)$ + Euclidean length between u and v:
                $d(v) = d(u)$ + Euclidean length between u and v // update neighbor's distance
                $l(v) = l(u)$ // update neighbor's label
                Update priority of v in Q with $d(v)$ // update priority queue Q

Figure 2: Pseudo code of our algorithm for computing coronary distance map using Dijkstra.

neighborhood. Average density values $\bar{\rho}_l(x)$ are obtained over N coronary samples in the population:

$$\rho_l(x) = e^{\left(-\frac{D^2(x)}{\sigma^2}\right)} \qquad \bar{\rho}_l(x) = \frac{1}{N}\sum_{i=1}^{N} \rho_l(x)$$

Major side branches could be missing in some coronary samples due to anatomical variation or image acquisition issues, and using samples with missing branches in averaging lowers density values. Therefore, we normalize the average density map of each label by the 90 percentile of the densities. Specifically, for each label l, we consider its non-zero average densities in all discrete voxels in order to compute a histogram. We

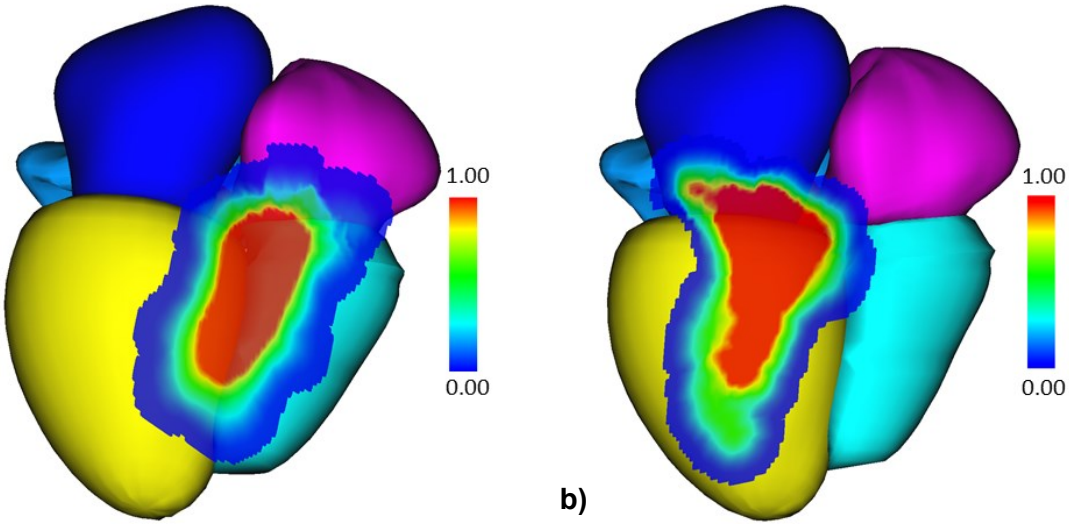


Figure 3: Spatial distribution of (a) right posterior descending artery and (b) left circumflex posterolateral branches.

then use the 90 percentile of this histogram to normalize average density map of that label.

Let L and X be discrete and continuous random variables and take labels and voxel location as values, respectively. We assume that labels at different voxel locations are independent. We obtain the conditional distribution of L given X based on its average density as follows where the density is linearly interpolated at continuous location x. Note that, we normalize the density at a location x by the sum of densities from all labels at that location so that their values sum to one in order to have a valid probability distribution:

$$P(L = l|X = x) = \overline{\rho}_l(x) / \sum_{l \in L} \overline{\rho}_l(x)$$

Figure 3 illustrates the computed probability map from a few sample labels.

2.3 QED Geodesics

2.3.1 Tree-Shape Representation

As in [13], our vessel tree representation consists of a description of the branch topology and the shape of the branches which together correspond to a point in high dimensional space X. We represent a tree-shape as a pair (T, f) consisting of first, an ordered binary tree $T = (B, r)$ with branches B and a root point r which together represent the tree topology, and second, branch attributes f that map each branch in B to n landmark points sampled along the branch geometry. In this tree-shape model, tree-shapes with non-binary topology are represented via binary trees with collapsible zero-length internal branches. Figure 4c, d makes it clear that the same tree-shape may correspond to different points in our space.

2.3.2 Quotient Space and QED Metric

We are interested in finding the geodesic path between two tree-shape points defined in the Euclidean space X. However, geodesic paths in such a high dimensional space do not necessarily give continuous and natural deformations, since these geodesic paths do not contain equivalent tree-shapes with different binary representations, Figure 4a. Therefore, it is important to identify all the points in this space representing the same tree-shape. In [13], a quotient space $\tilde{X}$ is defined to glue all points in the space X that correspond to the same tree-shape. In the quotient space, two tree-shape representations are considered to be identical, after collapsing their zero-length

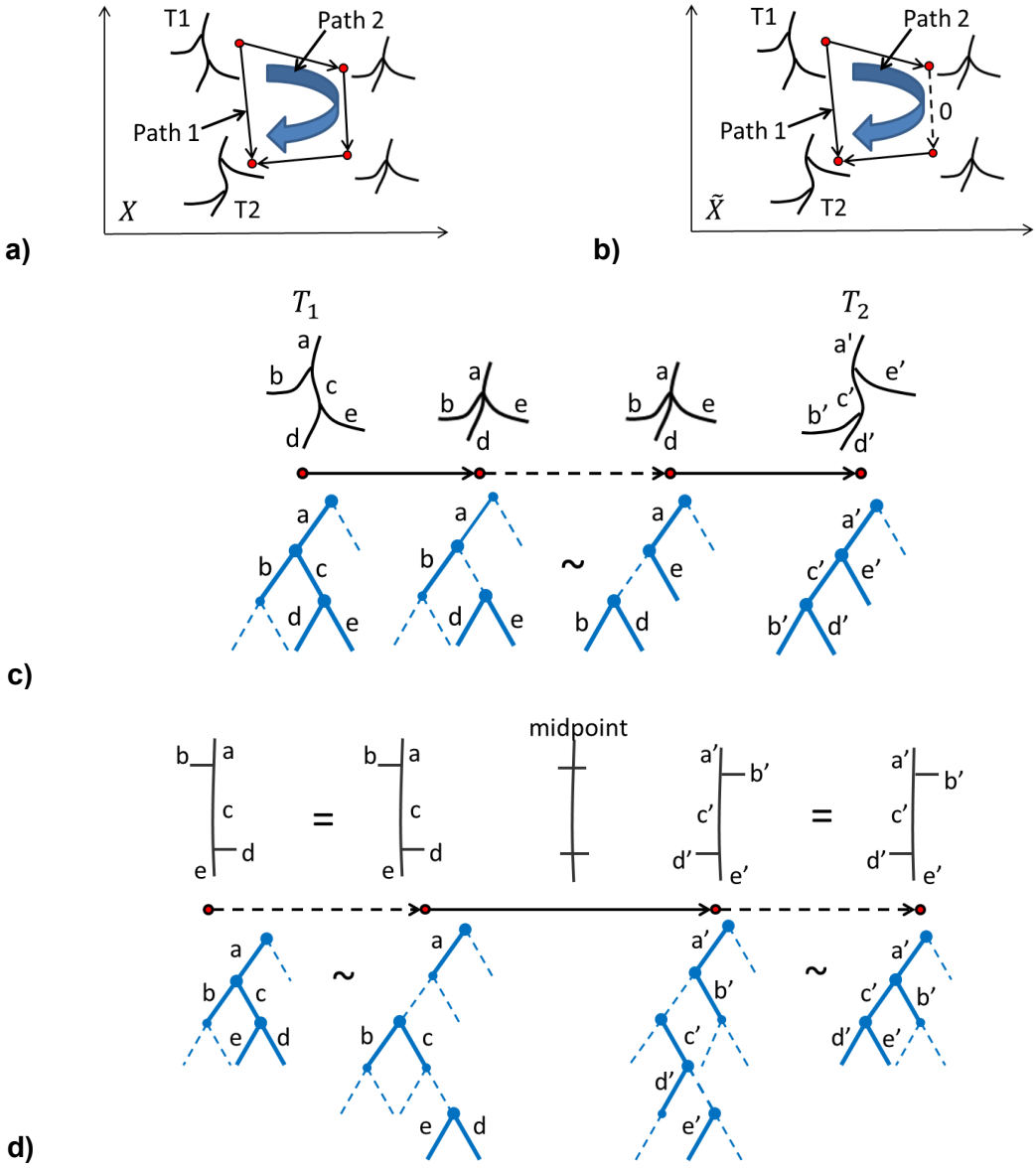


Figure 4: (a) T1 and T2 are tree-shape points in a high dimensional space. A third tree shape can be obtained by removing the internal branch of T1. Path 1 is the geodesic path in the original space X. Path 2 is a continuous deformation but not a minimum cost path. (b) In Quotient space $\tilde{X}$, Path 2 is a feasible geodesic path because two different representations of the same tree-shape are allowed along the path with zero cost between them. Note that Path 2 goes through an internal topology change. (c) Tree-shapes (top) and their representations (bottom) along the geodesic Path 2. Note that the side branches are switched along the geodesic path. (d) Another example of QED geodesic computation. In this example, the side branches are grown along the geodesic path. Midpoint along the geodesic path is also illustrated.

branches, if the corresponding combinatorial, ordered, rooted and attributed trees are exactly the same, Figure 4c,d. The geodesic metric in this space is called the Quotient Euclidean Distance (QED) where the distance between glued points corresponding to same tree-shape is considered to be zero.

The QED geodesic between two tree-shape points in quotient space is the path with minimum distance cost over all possible paths, Figure 4b, c, d. It was shown in [13] that the geodesic path in quotient space follows the tree-shapes which only differ from each other with internal topological changes, Figure 4c. The distance cost $d_T(x, y)$ between two non-identical tree-shape representations x and y is defined as the L2 norm of the deformation costs between corresponding branches in two tree-shapes. This branch deformation cost is computed as the Euclidean distance between landmark points of the two branches with their first points aligned. The cost of removing an internal branch is same as deforming it to a zero length branch.

In general, finding geodesic paths in QED space is computationally very expensive. One algorithm for approximating QED geodesic is to explore all paths following all possible internal topology changes and choose the one with minimum cost among them, Figure 5. However, the number of all possible internal topology changes grows exponentially with the total number of internal branches of the tree-shape. In fact, some sequences of internal topology changes may have a cost larger than the geodesic cost. In order to eliminate redundant paths containing such sequences, we propose to use Dijkstra's shortest path algorithm which is guaranteed to compute the shortest path in a graph

QED Geodesic Approximation

Input: Source tree-shape T_1 and target tree-shape T_2.

Let $\widetilde{\boldsymbol{T}}_{\boldsymbol{1}}$ and $\widetilde{\boldsymbol{T}}_{\boldsymbol{2}}$ be set of tree-shapes in QED equivalent classes of T_1 and T_2, respectively.
Let $\boldsymbol{T}_{\boldsymbol{i}}$ be set of tree-shapes including all possible combinations of internal topology changes (removal of internal branches) in T_1 and including T_1 itself.

Let $\widetilde{\boldsymbol{p}}$ and $\mathrm{d}(\widetilde{\boldsymbol{p}})$ be geodesic path and its distance, respectively.
Initialize $\widetilde{\boldsymbol{p}} = \{\}$ and $\mathrm{d}(\widetilde{\boldsymbol{p}}) = \infty$
For each tree-shape x_1 in $\boldsymbol{T}_{\boldsymbol{i}}$:
  Let $\widetilde{\boldsymbol{x}}_{\boldsymbol{1}}$ be tree-shapes in QED equivalent class of x_1
  For each tree-shape x_2 in $\widetilde{\boldsymbol{x}}_{\boldsymbol{1}}$:
    For each tree-shape y in $\widetilde{\boldsymbol{T}}_{\boldsymbol{2}}$:
      Set $\boldsymbol{p} = \{x1,\ x2,\ y\}$
      Compute path length $d(\boldsymbol{p}) = \|T_1 - x_1\|_2 + \|x_2 - y\|_2$
      If $d(\boldsymbol{p}) < \mathrm{d}(\widetilde{\boldsymbol{p}})$:
        $\mathrm{d}(\widetilde{\boldsymbol{p}}) = d(\boldsymbol{p})$ // update geodesic distance
        $\widetilde{\boldsymbol{p}} = \boldsymbol{p}$ // update geodesic path

Figure 5: Pseudo code of an exhaustive search algorithm for approximating QED geodesic.

with non-negative metric. We construct a graph $G = (N, E)$ where each node N in this graph corresponds to a tree-shape and are connected with edges E. An edge connecting two nodes with identical tree-shape representations is assigned with zero cost and an edge connecting two nodes with different tree-shape representations is assigned with the distance cost between them, Figure 6. The cost of an internal topology change is the removal cost for the corresponding internal branch. The edge costs are computed only when they are needed in Dijkstra's propagation for efficiency. Note that our Dijkstra based geodesic computation approach searches for "cone paths" that first go through internal topology transitions and then deform to match the target tree-shape. Therefore, our approach approximates QED geodesic rather than exactly computing it.

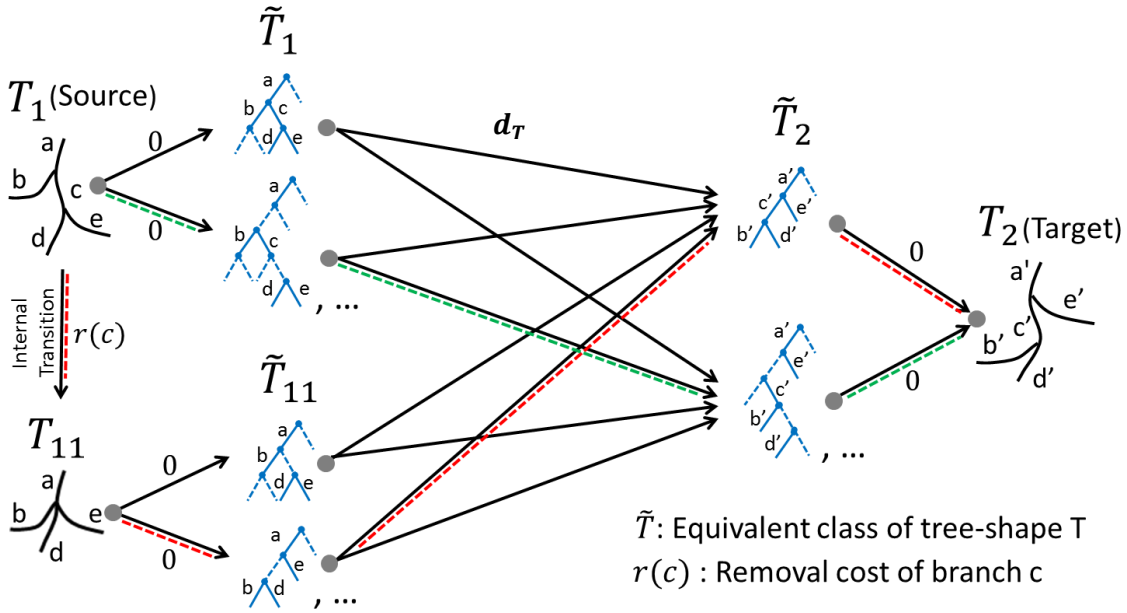
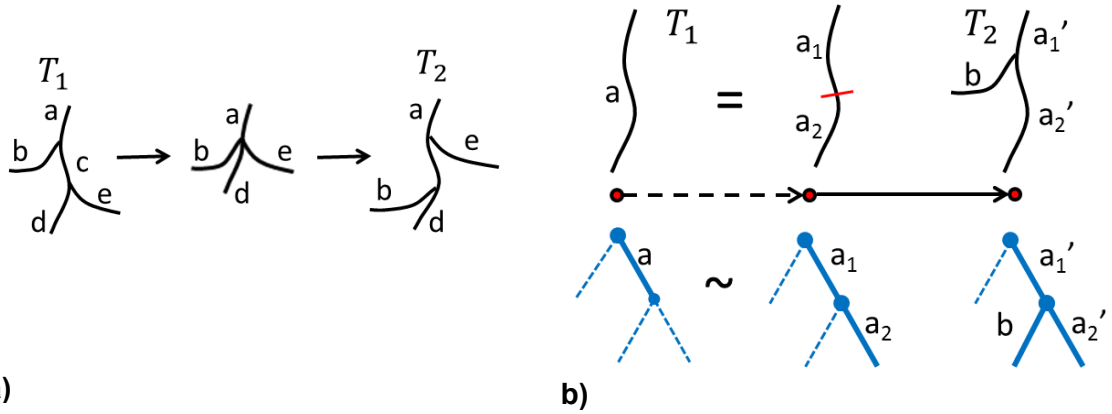


Figure 6: Our constructed graph for Dijkstra's algorithm to approximate QED geodesic from T1 to T2. An edge connecting two identical tree-shape representations is assigned with zero cost where an edge connecting two different tree-shape representations is assigned with the distance cost d_T between them. The path depicted with red color goes through internal topology change whereas the path depicted with green color only shrinks and grows side branches along the deformation

2.4 Coronary Geodesic and Labeling

The QED geodesic was applied to airway tree labeling in [15] by propagating labels from a labeled tree to an unlabeled tree along the geodesic deformation path where the labels on collapsed branches are dropped, Figure 7a.

In this work, we know the underlying cardiac anatomy, so we can use it to compute the likelihood of where a particular coronary branch tends to lie. Therefore, we do not directly use QED geodesic deformation for labeling but rather use our coronary geodesic which combines traditional QED metric with a second term that accounts for the cost of assigning a label to a specified branch on the target tree based on its location over the



a) **b)**

Figure 7: (a) Labels in T1 are propagated to T2 along the geodesic path. Label c on collapsed branch along the deformation is dropped. (b) T1 has a missing side branch. Its main branch is partitioned and mapped onto the equivalent second representation (middle) with two consecutive branches. This allows the side branch to grow naturally along the geodesic path.

heart territory. Specifically, our coronary geodesic searches for a path from tree-shape $T_1$to tree-shape T_2 among all possible paths with internal topological transitions in QED space that minimizes the total deformation cost. This problem can be formulated as an optimization problem as follows:

$$d_c(T_1, T_2) = \min_{p \in \tilde{P}} d_P(T_1, T_2) + \lambda^{-1} L_p(T_2)^{-1} \qquad \text{eq. 1}$$

where

$d_c(T_1, T_2)$: Coronary geodesic cost between $T_1$and T_2

$\tilde{P}$: Set of all possible paths with internal topological transitions between T_1 and T_2 in QED space

λ: Regularization term

$d_P(T_1, T_2)$: Length of path p between $T_1$and T_2 in QED space

$L_p(T_2)$: Likelihood for assigning labels to T_2 that are propagated from T_1 along path p

The labeling cost term $L_p(T_2)^{-1}$ is computed from the log likelihood of labels assigned to each point x in the coronary tree T_2 based their joint probability:

$$L_p(T_2) = \log\left(\prod_{x \in T_2} P(L = l_x | X = x)\right)$$

where l_x is the label assigned to coronary tree point x.

It should be clear that the solution to optimization problem in eq. 1 without the labeling cost term is the approximated QED geodesic. In order to solve this problem with the labeling cost term, we modify our Dijkstra graph for computing tree-shape geodesic in Figure 6 by adding the weighted labeling cost term $\lambda^{-1}L_p(T_2)^{-1}$ to all incoming edges of target tree-shape equivalent class. Note that, with this added cost term, the length of each path from tree-shape T_1 to tree-shape T_2 becomes a weighted sum of QED distance and the labeling cost.

Along the deformation between a coronary reference model and a target coronary tree, some of the target coronary branches may be grown without any assigned labels. This is possible, especially, when some branches in the target coronary tree are missing in the reference atlas model. One way of accounting for unlabeled branches is to assign zero labeling cost to them. However, this could bias the coronary geodesic computation to favor growing branches in order to minimize the total labeling cost. Therefore, for an unlabeled point x of the target coronary tree, we first find the label that has the maximum likelihood at that particular location among all possible labels and use this

label for computing the labeling cost term. Let $T_{2\_labeled}$ and $T_{2\_unlabeled}$ be the set of labeled and unlabeled points in T_2, respectively. Our labeling cost is then:

$$L_p(T_2) = \log\left(\prod_{x \in T_{2\_labeled}} P(L = l_x | X = x) \prod_{x \in T_{2\_unlabeled}} \max_{l \in L} P(L = l | X = x)\right)$$

where l_x is the label assigned to coronary tree point x in $T_{2\_labeled}$.

2.4.1 Handling Missing Side Branches

Major side branches of the coronaries could be missing in a sample due to anatomical variation, pathology, image acquisition issues, or segmentation errors. One major contribution of this work is to consider equivalent partitioned representations of source and target tree-shapes to account for missing side branches, Figure 7b. Given two tree-shape representations, we collapse all zero-length branches and merge the branch geometry of consecutive branches without bifurcations to form a single branch in a new tree representation. If the two new representations are exactly the same, we consider these two tree-shape representations to be equivalent. This definition of identical tree-shapes allows for partitioning a tree-shape branch into one or more consecutive branches in B to naturally account for missing side branches, Figure 7b. Specifically, we compute coronary geodesic between all possible partitioning of source and target tree-shape branches into at most three equal length polylines. Note that the tree-shape partitioning is a heuristic step of our algorithm that uses the QED metric.

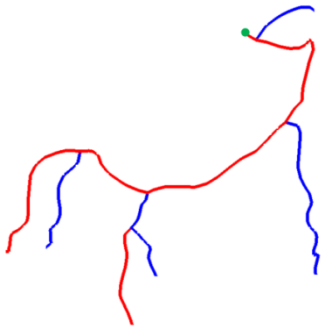


Figure 8: This figure illustrates the computed ordering of right coronary branches for a sample coronary tree. Branches with right and left ordering are shown in red and blue colors, respectively. The ostium is depicted as green point.

2.4.2 Coronary Tree-shape Representation

Our coronary trees are given as a connected set of points in 3D space. However, it is not practical to apply QED to arbitrary trees embedded in 3D space because it requires considering all possible orderings of the child branches. Based on the fact that coronaries lie roughly on the surface of the heart, we can order the branches based on their position relative to the surface of the heart. Specifically, we use the automatically detected heart chambers to find branch orderings of right and left coronary trees relative to the right and left ventricles, respectively.

For a given coronary bifurcation point, we compute the order of child branches as follows: Let $\vec{v}_1$, $\vec{v}_2$and $\vec{v}_3$ represent the vectors from the chamber center to bifurcation point and from the bifurcation point to the mean point of proximal part of first and second child branches, respectively. If $\vec{v}_1 . (\vec{v}_2 x \vec{v}_3) > 0$ then first and second child branch orders are right and left, respectively, otherwise the orders are vice versa. Note that our input coronary trees do not have trifurcation points and it is trivial to convert a

trifurcation point to two bifurcation points by adding a small intermediate branch. Figure 8 illustrates the detection of branch orders on a sample RCA coronary tree.

2.4.3 Coronary Atlas Model

Our coronary atlas model has a fixed topology described by [4] and we use mean branch shapes computed over the most complete 10 coronary trees in the training set. Figure 9 shows our coronary model in 3D space. Our coronary model is an instance of both right-dominant and co-dominant circulation which are seen in 92% of the population [4]. This is why we excluded the CX posterior descending artery in our coronary model. A coronary tree typically has only one ramus intermediate branch. However, as it will be described in a latter section, we label LAD and CX tree separately. Therefore, the ramus intermediate branch appears in both trees in our atlas model.

2.5 Efficient Computation of Coronary Geodesic

The computational cost of QED metric exponentially increases with the number of generations of tree-shapes and therefore, coronary arteries may have more generations than what is practical with the QED metric. In our method, we exploit the topology of coronary arteries in order to reduce computational cost and label more complete coronary trees. We automatically preprocess the input coronary trees by splitting LCA into LAD/CX subtrees, partitioning RCA and simplifying the topology of trees based on a constraint on the number of bifurcations along their side branches.

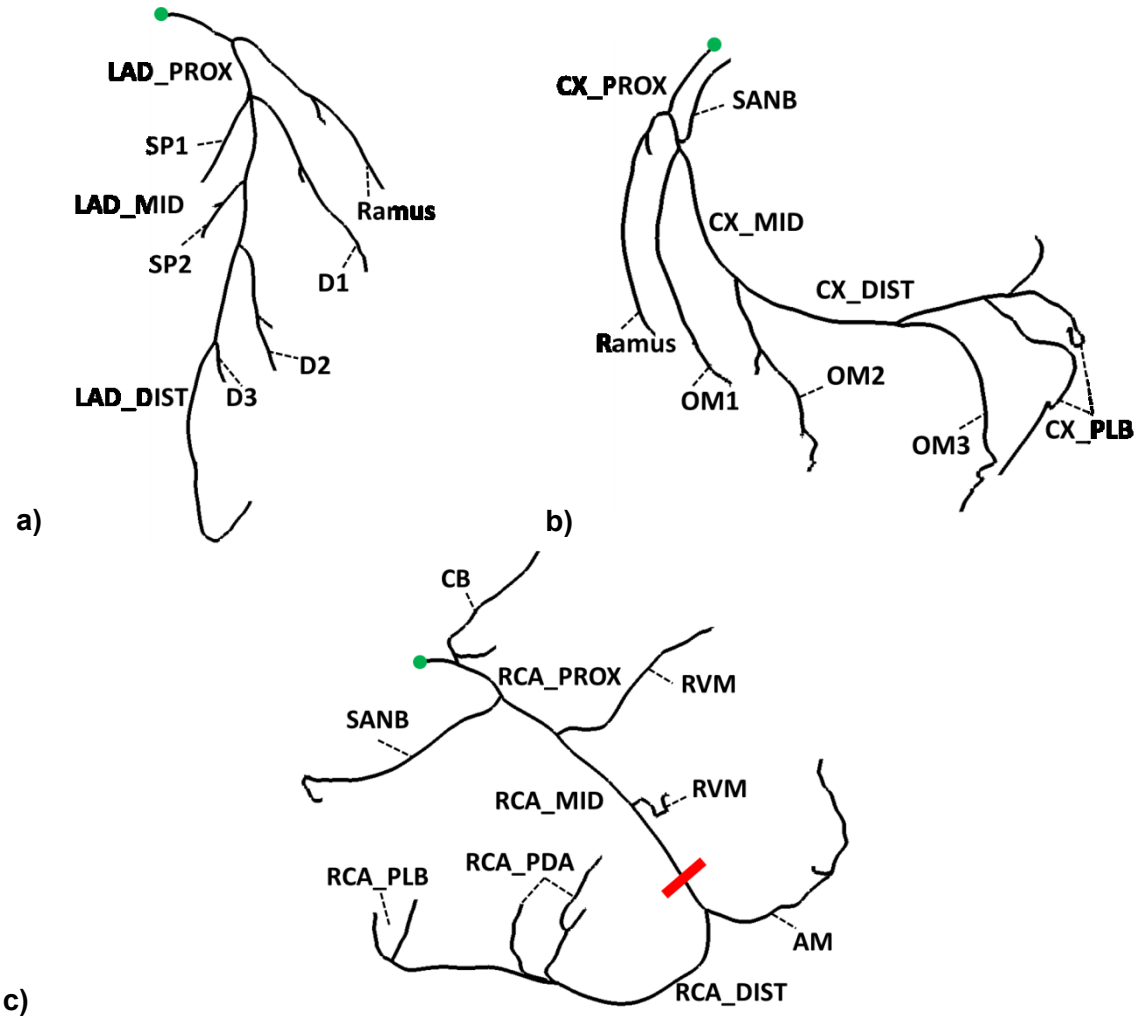


Figure 9: The left coronary artery (LCA) has 19 labels: left main (LM), (a) three segments of the left anterior descending (LAD): proximal (LAD PROX), mid (LAD MID), and distal (LAD DIST), Ramus, diagonals (D1, D2, D3), septal perforating (SP1, SP2), (b) three segments of the circumflex (CX): proximal (CX PROX), mid (CX MID), and distal (CX DIST), Ramus, sinoatrial nodal (SANB), posterolateral (CX PLB), and obtuse marginals (OM1, OM2, OM3). (c) The right coronary artery (RCA) has 9 labels: three segments of the right coronary artery: proximal (RCA PROX), mid (RCA MID), and distal (RCA DIST), sinoatrial nodal (SANB), conus (CB), right ventricle marginal (RVM), acute marginal (AM), posterolateral (RCA PLB), and posterior descending (RCA PDA). Proximal and distal subtrees of RCA are separated by red lines.

2.5.1 LAD/CX Tree Extraction

The LM branch of LCA tree originates from left ostium and bifurcates into LAD and CX trees. The LM branch is short in length and does not have side branches. Therefore, we can label LAD and CX trees separately. Although the LM bifurcation point can be

detected as the first bifurcation in LCA tree, this may not be robust for detected coronary trees with false positive side branches along the LM. For this reason, we detect the LM bifurcation point using our coronary territory. Specifically, for each bifurcation point in LCA, we consider the downstream tree of right child branch as LAD candidate tree T_{LAD}, the downstream tree of left child branch as CX candidate tree T_{CX} and the remaining points in LCA as LM candidate tree T_{LM}. Let l_{LM}, l_{LAD} and l_{CX} be a set of labels belonging to LM, LAD and CX trees, respectively, as shown in our coronary atlas model, Figure 9. We then compute a score based on the log likelihood of assigning labels l_{LM}, l_{LAD} and l_{CX} to candidate trees T_{LM}, T_{LAD} and T_{CX}:

$$log\left(\prod_{x \in T_{LM}} \max_{l \in l_{LM}} P(L = l|X = x) \prod_{x \in T_{LAD}} \max_{l \in l_{LAD}} P(L = l|X = x) \prod_{x \in T_{CX}} \max_{l \in l_{CX}} P(L = l|X = x)\right)$$

We choose the bifurcation point with highest log likelihood score as the LM bifurcation point.

2.5.2 RCA Tree Partitioning

We partition the RCA into proximal and distal subtrees by maximizing the likelihood of their locations on the coronary territory, and label each subtree separately. We found that it is non-trivial to partition CX and LAD tree due to large variation of the location of side branch bifurcations. Therefore, we consider labeling them without partitioning.

Let l_{prox} and l_{dist} be a set of labels for proximal and distal subtrees of RCA, respectively. For each point in the tree, we consider the downstream tree as distal subtree candidate T_{dist} and the remaining points as proximal subtree candidate T_{prox}. We then compute a score based on the log likelihood of assigning labels l_{prox} and l_{dist} to candidate trees T_{prox} and T_{dist}:

$$log\left(\prod_{x \in T_{prox}} \max_{l \in l_{prox}} P(L = l|X = x) \prod_{x \in T_{dist}} \max_{l \in l_{dist}} P(L = l|X = x)\right)$$

We choose the point with highest log likelihood score for splitting the RCA tree into proximal and distal subtrees. Figure 9c shows the labels for proximal and distal subtrees of RCA separated by red lines.

2.5.3 Simplification of Topology Complexity

Even though partitioning coronary trees into subtrees reduces their complexity, it is still computationally expensive to compute coronary geodesic between them due to complexity of side branches. In addition, we do not partition CX and LAD trees which may have more than five generations. Here, we explain how we simplify a subtree by reducing its topological complexity.

A typical coronary tree has a main branch that originates from the ostium and travels along the atrioventricular and interventricular sulcus of the heart. The main coronary branch gives off side branches which can further bifurcate into secondary branches in order to reach to the heart muscles, Figure 10a, b, c. Assigning a distinct label to the

downstream vessels originating from side branches is of less clinical importance. Therefore, we simplify our input coronary trees by reducing the complexity of side branches.

Let l_T and l_{main} be the set of all possible labels in tree T and along its main branch, respectively. For each branch B_{main} in T from the root point to distal points, we compute a score based on the likelihood of assigning labels l_{main} and its complement $l_T \backslash l_{main}$ to B_{main} and its complement $T \backslash B_{main}$, respectively:

$$log\left(\prod_{x \in B_{main}} \max_{l \in l_{main}} P(L = l|X = x) \prod_{x \in T \backslash B_{main}} \max_{l \in l_T \backslash l_{main}} P(L = l|X = x)\right)$$

We choose the branch that maximizes this log likelihood score as the main branch. We consider proximal, mid and distal segments of the coronary tree as belonging to the main branch. In addition, we consider PLB as a label along the main branch of RCA and CX trees.

We fix the detected main branch and extract the subtree with maximum size that satisfies the topology requirement with constraint on the number of bifurcations along side branches. For RCA, we allow only one bifurcation along side branches. Since the side branches of LAD and CX are usually not complex, we do not allow any bifurcation along their side branches, Figure 10a, b, c.

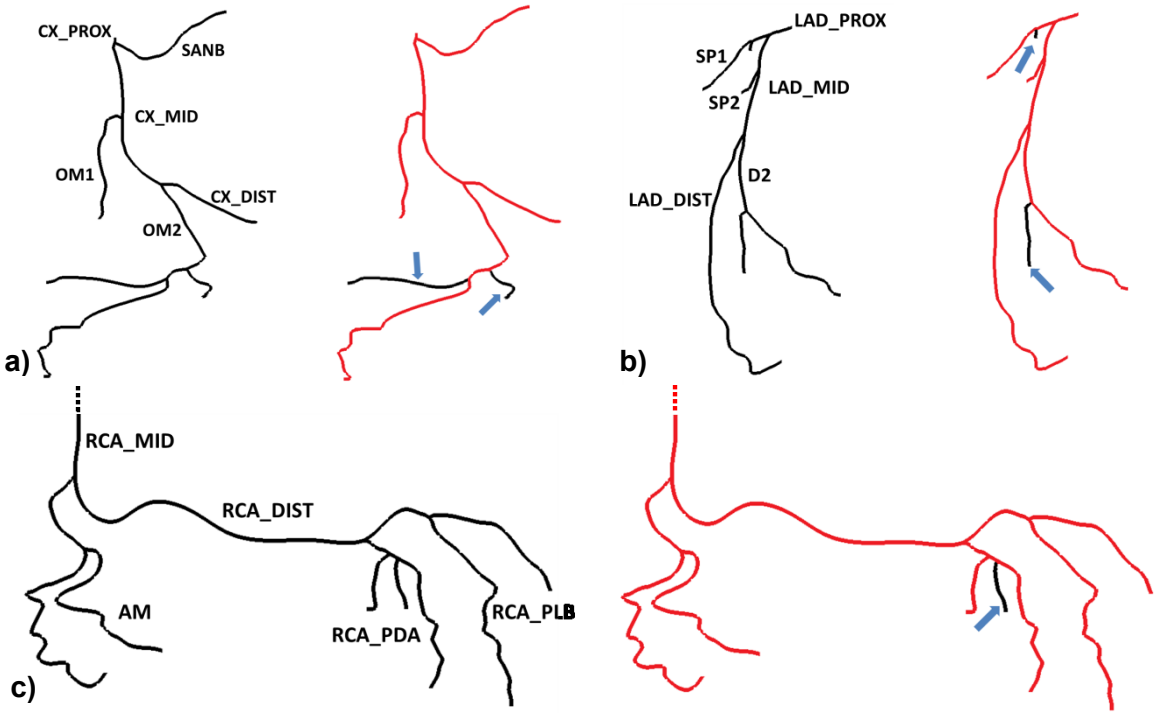


Figure 10: This figure illustrates our method for simplifying input coronary trees. Red trees show the simplified maximal subtrees where the pruned branches are depicted with arrows. (a) Second obtuse marginal branch of CX gives off multiple secondary branches. CX is simplified to satisfy a topology without any bifurcation along side branches. (b) First septal and second diagonal branches of LAD give off one secondary branch. LAD is simplified to satisfy a topology without any bifurcation along side branches. (c) Posterior descending branch of RCA gives off multiple secondary branches. RCA is simplified to satisfy a topology with at most one bifurcation along side branches.

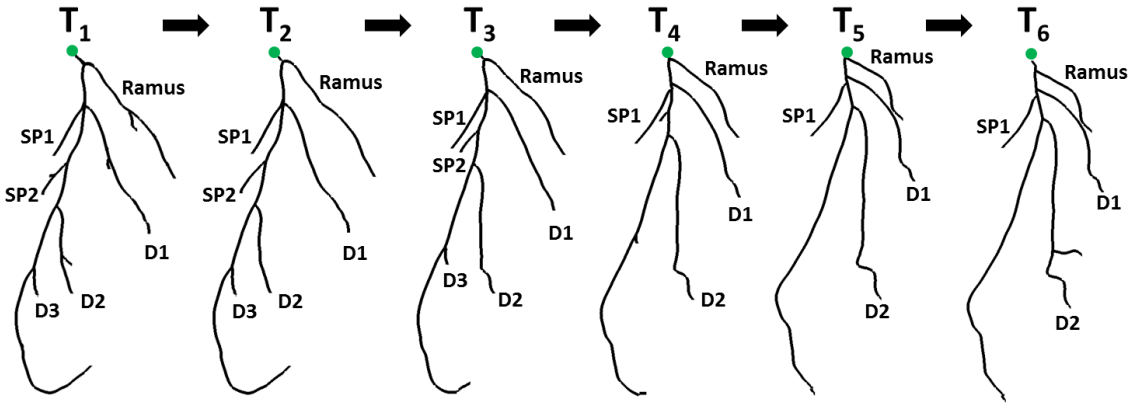


Figure 11: Our LAD model (T1) matched to a sample LAD tree (T6). Trees are first simplified such that they have single side branches (T2, T5). T3 and T4 are intermediate trees along the coronary geodesic path from T2 to T5. Labels in model tree are propagated along the path except SP2 and D3 which are dropped.

3 Results

Our coronary geodesic implementation supports 4 generations for trees that have side branches with at most one bifurcation and 6 generations for trees that have side branches without any bifurcation. We keep a table for quick look-up of equivalence classes of tree shapes for computational efficiency.

We consider searching for geodesic path in QED space among all possible paths that allow up to 2 internal topological transitions. We cache computed branch deformation and label assignment costs. Our parallelized implementation with OpenMP labels both left and right coronary trees in 15 seconds on average on an Intel Core i7 2.8GHz processor and 32 GB RAM.

We assume that the left main (LM) and the proximal branches of the RCA, LAD and CX always exist. To make the input centerlines suitable for use in our QED implementation, input trees were preprocessed by iteratively pruning their shortest branches until they had a maximum 4 generations for proximal and distal subtrees of RCA and 6 generations for CX and LAD.

We evaluated our method for right and co-dominant dominant cases using leave-one-out cross-validation where the coronary spatial distribution is re-computed for each training set. We experimentally chose the best weighting term between QED distance and label assignment likelihood terms. We used expert annotated coronary centerlines in 50 CTA datasets. We automatically detected the pericardium mask and four chamber

models to define our coordinate system [16, 17]. Each of the ground truth centerlines was labeled according to [4] where the same label of the side branch is used to label its children.

For evaluation, labels corresponding to pruned branches are recorded as part of the missed results. Figure 11 illustrates the matching between the LAD model and a test LAD coronary tree. We calculated the overlap measures between automatic and expert annotated labels for each branch that exists in our coronary model. Branches that exist in our model but that are not labeled by the algorithm are counted as mislabeled. RCA was incorrectly partitioned below AM branch in 2 of the 50 datasets. Figure 12 compares our results to those of [10]. Note that the two test datasets are different and so the results are suggestive rather than being directly comparable.

We evaluated our labeling method on 20 patient cases of automatically detected coronary centerlines using [7] that combines a model driven approach for the three major coronary arteries with a data driven approach for the side branches and distal parts of the main branches. Because the method of [7] knows which of the three main coronary branches it has recovered, the label of these branches is known as part of the segmentation. However, in order to show the generality of our approach to labeling, these known labels are ignored and we present a method that can label the coronary tree without any known labels.

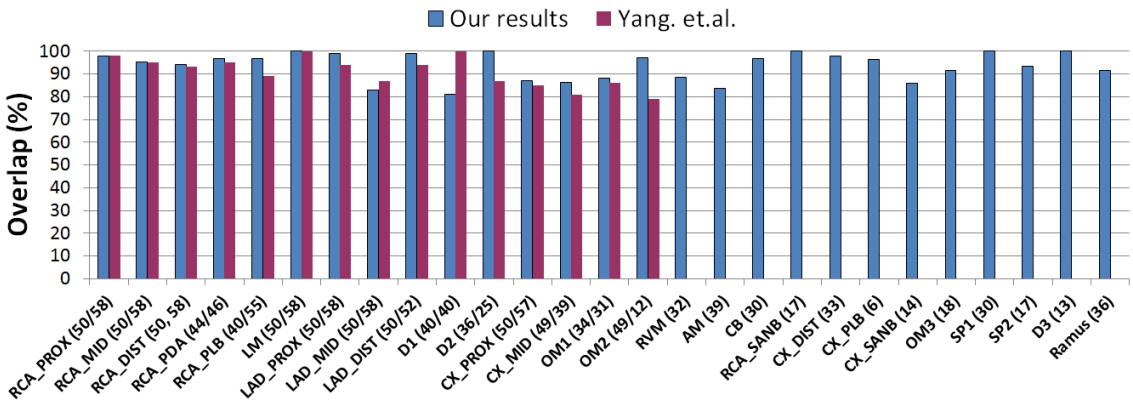


Figure 12: Comparison of our overlap measures to Yang et. al. [9]. Total count of each branch is depicted next to its label for our and their test set, respectively. Our results for branches that were not part of their results appear as single columns with the total count of branch is depicted next to its label.

RVM	RVM (89%), CB (4%), AM (2.7%), None (4.3%)
AM	AM (84%), PDA (6%), RVM (2%), None (8%)
D1	D1 (81%), Ramus (19%)
OM1	OM1 (88%), Ramus (12%)
CX_SANB	CX_SANB (86%), None (14%)

Table 1: Label assignments as a percentage for those branches with the most mislabels.

Our success rate is 96% for the right coronary and 97% for the left coronary for hand annotated centerlines, and 92% for the right coronary and 87% for the left coronary for detected centerlines. Without the anatomical location prior, the results for the hand annotated centerlines decrease to 93% for the right coronary and 94% for the left coronary. Table 1 shows the accuracy of our approach for branches that were assigned with the most mislabels.

In the case of the automatically detected coronary branches, there are 14 instances of false positive detections which explain our reduced labeling performance. Specifically, in 39% of these cases, the false positives are incorrectly given a label and the assigned label depends on the location.

4 Discussion

We have presented an efficient method that determines a geodesic path between tree shapes in order to propagate labels from a standard model of the coronary arteries to unlabeled coronary centerlines. Our approach adapts the framework of [15] to work relative to the detected structures of the heart. For most of the labels assigned by [10] our approach produces equivalent or better results which are not directly comparable since the two methods were tested on different datasets. However, our approach can handle more general tree structures and we have shown results on 12 additional labels. Also, we show results for coronary trees produced by an automatic detection algorithm. Additional contributions include the efficient approximation of the QED geodesic using Dijkstra's algorithm, the use of a coronary territory location prior, and an explicit description of how to support missing side branches in the QED framework. Our approach with coronary tree partitioning and topology simplification allows us to label unseen coronary trees using a complete coronary atlas model without need to reduce its size. However, the computational complexity of QED metric for trees with large number of generations is the major limitation of our approach for applying it to other vascular anatomy.

In future work, we will investigate how our QED geodesic approximation through cone paths affects the labeling results. We will consider having multiple coronary models in order to handle left dominant circulation (approximately 8% of the population [4]). Some branches such as CX obtuse marginal or LAD diagonal have similar shape to ramus

with a large overlap between their territories. In order to improve labeling results for such confusing branches, we will use other geometric features such as diameter and curvature information in our geodesic metric. False positive centerlines in automatic coronary detections reduce the accuracy of labeling. Detection probabilities can be incorporated into our metric for encouraging the algorithm to grow branches for centerlines with low probabilities. In addition, we can generate confidence of labeling results based on the likelihood of assigned labels over the coronary territory. Our method for computing labeling likelihood assumes independence between labels at different locations in the tree. Another future work would be to investigate the correlation of labels along different branches.

References

1. D. Lloyd-Jones, R. J. Adams, T. M. Brown, M. Carnethon, S. Dai, G. De Simone, and et al. Executive summary: Heart disease and stroke statistics–2010 update: A report from the American Heart Association. Circulation, 121(7), 2010.

2. W. G. Austen, J. E. Edwards, R. L. Frye, and et al. A reporting system on patients evaluated for coronary artery disease: Report of the ad hoc committee for grading of coronary artery disease, council on cardiovascular surgery, American Heart Association. Circulation, 51:5-40, 1975.

3. G. L. Raff, A. Abidov, S. Achenbach, and et al. SCCT guidelines for the interpretation and reporting of coronary computed tomographic angiography. J. of Cardiovascular Computed Tomography, 3(2):122-136, 2009.

4. D. M. Fiss. Normal coronary anatomy and anatomic variations. Applied Radiology, 36(1):14, 2007.

5. R. M. Setser, T. P. O'Donnell, N. G. Smedira,. J. F. Sabik, S. S. Halliburton, A. E. Stillman, and R. D. White. Coregistered MR imaging myocardial viability maps and multi-detector row CT coronary angiography displays for surgical revascularization planning: Initial experience. Radiology 2005 237 (2).

6. P. Beliveau, R. Setser, F. Cheriet, and T. O'Donnell. Patient-specific coronary territory maps. Proceedings of the SPIE 2007; 6511: 65111J.

7. Y. Zheng, H. Tek, and G. Funka-Lea. Robust and accurate coronary artery centerline extraction in CTA by combining model-driven and data-driven approaches. In K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, editors, MICCAI 2013, volume 8151 of LNCS, pages 74-81. Springer Berlin Heidelberg, 2013.

8. N. Ezquerra, S. Capell, L. Klein, and P. Duijves. Model-guided labeling of coronary structure. IEEE Trans. on Medical Imaging, 17(3):429-441, 1998.

9. K. Haris, S. N. Efstratiadis, N. Maglaveras, C. Pappas, J. Gourassas, and G. Louridas. Model-based morphological segmentation and labeling of coronary angiograms. IEEE Trans. on Medical Imaging, 18(10):1003-1015, 1999.

10. G. Yang, A. Broersen, R. Petr, and et al. Automatic coronary artery tree labeling in coronary computed tomographic angiography datasets. In Computing in Cardiology, volume 38, pages 109-112, 2011.

11. P. Bille. A survey on tree edit distance and related problems. Theor. Comput. Sci., 337:217–239, June 2005.

12. W. H. Tang and A. C. S. Chung. Cerebral vascular tree matching of 3d-ra data based on tree edit distance. Medical Imaging and Augmented Reality, page 116123, 2006.

13. A. Feragen, F. Lauze, P. Lo, M. de Bruijne, and M. Nielsen. Geometries on spaces of treelike shapes. In R. Kimmel, R. Klette, and A. Sugimoto, editors, Proc. of the 10th Asian Conference on Computer Vision - Part II, ACCV'10, pages 160-173, Berlin, Heidelberg, 2011. Springer-Verlag.

14. A. Feragen, S. Hauberg, M. Nielsen, and F. Lauze. Means in spaces of tree-like shapes. In ICCV, pages 736-746, 2011.

15. A. Feragen, P. Lo, V. Gorbunova, and et al. An airway tree-shape model for geodesic airway branch labeling. In Third MICCAI Workshop on Mathematical Foundations of Computational Anatomy, 2011.

16. Y. Zheng, H. Tek, G. Funka-Lea, and et al. Efficient detection of native and bypass coronary ostia in cardiac CT volumes: Anatomical vs. pathological structures. In G. Fichtinger, A. Martel, and T. Peters, editors, MICCAI 2011, volume 6893 of LNCS, pages 403-410. Springer Berlin Heidelberg, 2011.

17. Y. Zheng, A. Barbu, B. Georgescu, and et al. Four-chamber heart modeling and automatic segmentation for 3-D cardiac CT volumes using marginal space learning and steerable features. IEEE Trans. on Medical Imaging, 27(11):1668-1681, 2008.

18. F. L. Bookstein. Principal warps: Thin-Plate splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Mach. Intell., 11(6):567-585, 1989

Curriculum Vita

Mehmet Akif Gulsun was born on December 1, 1982 in Konya, Turkey. He received his B.Sc. (2005) in Electrical Engineering from Bilkent University, Turkey. He enrolled in the Computer Science M.Sc. program at The Johns Hopkins University in Spring 2011. He has been working as a full-time Research Scientist at Siemens Corporation, Corporate Technology in Princeton, USA since 2009. His research focuses on vascular image segmentation and analysis in MR and CT angiography images.