Hierarchical Learning for Tubular Structure Parsing in Medical Imaging: A Study on Coronary Arteries Using 3D CT Angiography

Le Lu Jinbo Bi Shipeng Yu Zhigang Peng Arun Krishnan Xiang Sean Zhou

CAD and Knowledge Solutions, Siemens Healthcare, Malvern, PA 19355, USA
le-lu@siemens.com; firstname.lastname@siemens.com

Abstract

Automatic coronary artery centerline extraction from 3D CT Angiography (CTA) has significant clinical importance for diagnosis of atherosclerotic heart disease. The focus of past literature is dominated by segmenting the complete coronary artery system as trees by computer. Though the labeling of different vessel branches (defined by their medical semantics) is much needed clinically, this task has been performed manually.

In this paper, we propose a hierarchical machine learning approach to tackle the problem of tubular structure parsing in medical imaging. It has a progressive three-tiered *classification process at volumetric voxel level, vessel segment level, and inter-segment level. Generative models are employed to project from low-level, ambiguous data to class-conditional probabilities; and discriminative classifiers are trained on the upper-level structural patterns of probabilities to label and parse the vessel segments. Our method is validated by experiments of detecting and segmenting clinically defined coronary arteries, from the initial noisy vessel segment networks generated by low-level heuristics-based tracing algorithms. The proposed framework is also generically applicable to other tubular structure parsing tasks.*

1 Introduction

Coronary artery disease is the number one killer in the developed world. Therefore, automatic coronary artery extraction from 3D CT Angiography (CTA) has a significant clinical impact. The majority of previous vessel tracing/tracking work target to ("blindly") construct the coronary artery tree system, by running classical *minimum-cost path optimization* [3, 7, 11, 5, 1] on a graph in a breadth-first-search fashion. In 2D/3D medical imaging, this cost is often formulated by a numerical term called "vesselness" that defines the magnitude of how likely each voxel (on the path) is on a tubular structure. Further more, various "vesselness" metrics based on *Match Filters* [8], *Hessian Matrix* [4], *Medialness* [7], and many other assumptions of low-level image observation have been proposed. However this *vesselness + shortest path* formula lacks semantic or class-conditional information, rather than builds a breadth-first tree. The clinically important vessel segments can not be distinguished from many other branches. In practice, proper labeling of coronary arteries can significantly improve clinical workflows of disease diagnosis, assessment, reporting, and image-guided vessel intervention procedures.

In this paper, we propose a hierarchical, supervised learning approach to label and segment generic tubular structures in medical imaging. In particular, three categories of coronary arteries, including right coronary artery (RCA), left anterior descending artery (LAD) and left circumflex artery (LCX[1]), are studied examples, based on an annotated vessel set in CTA by experts. Annotating and labeling vessels only by hands, is labor intensive (as browsing through volume slides) and can be unreliable or have large between-reader variation [16]. Our purpose is developing a trainable system upon a database of verified expert annotations to provide a quantitatively stable reference for improved clinical usage. The method can be generalized for other tubular structure parsing tasks, e.g. vessels in liver and brain, lung airway and colon, in 3D medical imaging. Specifically, the mathematical notation of our problem is as follows. Assume that a set of vessel segments as 3D curves $\{\mathcal{V}_i\}_{i=1,...,n}$ are generated from each CTA volume as different branches of the RCA/LAD/LCX vessel tree, based on a variation of above-mentioned vessel tracking techniques[1]. Given the annotated curve $\mathcal{A}$ per volume, the optimal output is $\max\{sim(\mathrm{V}, \mathcal{A})\}$, $\mathrm{V} \in \{\mathcal{V}'_i\}_{i=1,...,m}$, where $\{\mathcal{V}'_i\}_{i=1,...,m}$ is a super-set of the orig-

[1]The main branches of the three vessel trees all have a proximal, a mid, and a distal segment. The differences of these main branches with other outgrowths can be very ambiguous, due to large structural variation of coronary arteries across patients and races. Our goal is to automatically separate the three main branches from all other side branches within its corresponding tree.

2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

inal $\{\mathcal{V}_i\}_{i=1,...,n}$ after *expansion* during our process and $m > n$, and $sim(\cdot,\cdot)$ is a similarity function between two 3D curves. Therefore, our problem is indeed converted as classification issue with the *set-expansion* process as discussed in section 3.

Our method is partially inspired by statistical edge/road modeling in natural or satellite images [10, 6]. However our image observation model is totally different from [10, 6], due to the different nature of data resources and modalities. A novel set of *geodesic distance-indexed local geometry features* is fed into both generative and discriminative models and evaluated for their descriptive capacity in capturing different outgrowth patterns of a 3D tubular structure. Other distance metrics are also explored. More importantly, our application demands a higher accuracy of classification and segmentation than the general-purpose statistical edge learning [10], and requires the parsing of more specific and semantic information, in order to achieve clinical relevance and usefulness. Therefore, we propose a novel three-tiered classification hierarchy on the top of likelihood ratio testing [10, 6], at volumetric voxel level , vessel segment level and inter-segment level. It projects low-level noisy image observations as class-conditional probability values and propagates them to further infer high-level knowledge. This local-to-global scaled hierarchical inference structure can be considered as an analogy to the increasingly complex cognitive vision system and its computational model [14].

In summary, we have three major contributions: (i) a hierarchical classification architecture progressively extracting and inferring clinical semantics from medical images, (ii) effective joint (geodesic) distance and geometry feature sets describing 3D tubular structures with extensive evaluation, and (iii) the first complete system, to our best knowledge, addressing automatic coronary artery branch labeling and segmentation , as shown in Figure 2.

2 Problem Formulation & Overview

In this paper, we present a *three tiered* hierarchical learning framework to address the tubular structure parsing problem in 3D medical imaging. Though our work primarily focuses on coronary arteries, it is straightforward to extend to lung airways, colon or other tubular structures. The diagram of our proposed algorithm is shown in Figure 1.

Coronary arteries of RCA, LAD and LCX are learned in the same way and processed independently, so we only use LAD as an example in the following, unless otherwise stated. Both the annotated vessel $\mathcal{A}$ and each of traced vessel segments $\mathcal{V}_i$ are mathematically represented as vectors of 3D point coordinates $[P_i^1; P_i^2; ...; P_i^j; ...]$. We first fit each array using a cubic spline and compute 11 distances + geometric features on each point P_i^j analytically. After this step, our data format converts from vectors of 3D point coordinates into arrays of 11 combined features. Note that all $\{\mathcal{V}_i\}$ and $\mathcal{A}$ rooted from the same artery ostium. Next

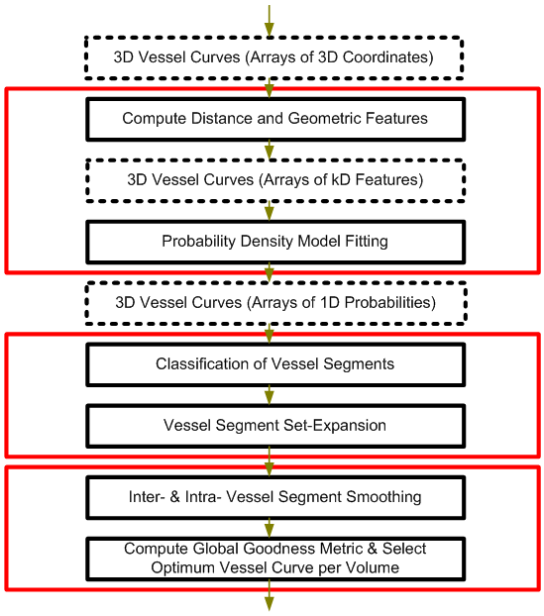


Figure 1. *The framework of our proposed three tiered classification hierarchy.*

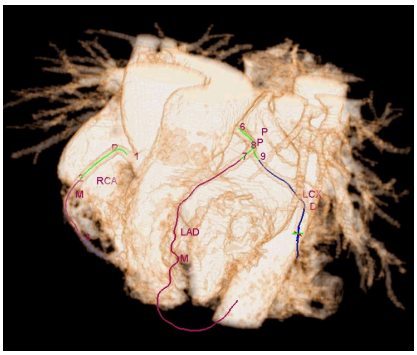


Figure 2. *A screenshot of our vessel parsing system using CTA volume.*

all identities in $\{\mathcal{V}_i\}$ are manually labeled as one of four groups according to their spatial overlapping relationships with $\mathcal{A}$: $C_1: \mathcal{V}_i \subset \mathcal{A}$, $C_2: \mathcal{V}_i \bigcap \mathcal{A} \neq 0$, $C_3: \mathcal{A} \subset \mathcal{V}_i$ and $C_4: \mathcal{V}_i \bigcap \mathcal{A} = 0$. By performing the piecewise alignment and matching between $\mathcal{A}$ and $\mathcal{V}_i$, e.g., dynamic time warping [13], we can construct the positive and negative datasets of classes "on-$\mathcal{A}$" and "off-$\mathcal{A}$" using features of P_i^j from the matched portion and unmatched portion of $\{\mathcal{V}_i\}$ from four above sub-classes respectively.

The voxel level feature classification and modeling algorithms are extensively evaluated using three generative models of Kernel Density Estimator [9], marginally optimized histogram and jointly optimize histogram on Chernoff Information of "on-$\mathcal{A}$" and "off-$\mathcal{A}$" class distributions, and probabilistic discriminative Relevance Vector Machine (RVM) [15] with feature selection [12]. The output of this step is converting $\{\mathcal{V}_i\}$ from vectors of k-features into sequences of normalized "on-$\mathcal{A}$" class-conditional probabilities ρ_i^j. Then at the vessel segment level, we de-

sign discriminative classifiers to distinguish the following three classes as C_1, $(C_2 \cup C_3)$ and C_4. Note that any C_2 or C_3 labeled $\mathcal{V}_i$ shows the same syntactic pattern as $[+, ..., +, -, ...-]$ and is treated the same way in testing. We further run *set-expansion* on $(C_2 \cup C_3)$ using a greedy bi-partitioning method to find the probable cutting points on $[\rho_i^1; \rho_i^2; ...; \rho_i^j; ...]$. Each $\mathcal{V}_i \in (C_2 \cup C_3)$ is split into two new segments which are then added into $\{\mathcal{V}_i\}$ to form a super-set $\{\mathcal{V}_i'\}$. After intra- and inter- vessel segment ρ smoothing on $\{\mathcal{V}_i'\}$ that is a specific 2D Markov Random Field formulation and solved using *Belief Propagation*, the best $V \in \{\mathcal{V}_i'\}$ (per volume or patient) is selected as the final system output, according to a global goodness metric using both $\mathcal{V}_i'$'s ρ and length statistics.

3 Algorithm

Our *three tiered* hierarchical learning algorithm for vessel tree structure parsing is described in detail as follows.

3.1 Voxel Level Processing and Modeling

Voxel Level Feature Computation: Figure 4 shows an illustrative example of a coronary artery tree $\{\mathcal{V}_i\}$ along with its vessel annotation and $\mathcal{A}$. Each curve of $\mathcal{V}_i$ or $\mathcal{A}$ is a sequence of 3D point world coordinates in CT scanner. First we fit each sequence using a 3D cubic spline and resample the spline uniformly to get $[P_i^1; P_i^2; ...; P_i^j; ...]$, under a fixed spatial resolution $\varepsilon = 0.5$. The resulting sequence lengths ranges from 16 to 865 in a pool of 82 patients. The cubic spline representation permits us to compute the local geometric features as the normalized tangent vector $T = [\tau_1, \tau_2, \tau_3]$ where $\|T\| = 1$, tangent orientations $\theta = [\theta_1, \theta_2]$, curvature κ and torsion ψ analytically at each sampling knot P_i^j. We also obtain the geodesic distance[2] $\mathcal{G}$ and Euclidean displacement $\Delta = (P_i^j - \mathcal{O})$ from P_i^j to the annotated LAD ostium $\mathcal{O}$. $\mathcal{G}$ is uniquely determined as the curve length, traced from any point P_i^j on the vessel tree back to the ostium $\mathcal{O}$. In total, we obtain 11 features $[T; \theta; \kappa; \psi; \mathcal{G}; \Delta]$ for each vessel knot point P_i^j. In testing, coronary artery ostia are detected using boosted 3D Haar features [17] in a multi-scale implementation.

3.1.1 Training Data Construction of "on-$\mathcal{A}$" and "off-$\mathcal{A}$" Classes

Each of $\{\mathcal{V}_i\}$ and $\mathcal{A}$ per coronary artery in a 82 patient pool, generates dozens or hundreds of feature vectors $\{T; \theta; \kappa; \psi; \mathcal{G}; \Delta\}$ at all sampling knots. Our task is to label and segment vessel fragments of $\{\mathcal{V}_i\}$ and approximate $\mathcal{A}$ as close as possible. Thus we first construct the binary-class feature datasets of $\{T; \theta; \kappa; \psi; \mathcal{G}; \Delta\}$ for "on-$\mathcal{A}$" and "off-$\mathcal{A}$", according to the alignment criterion if P_i^j is spatially close enough to $\mathcal{A}$. Recall that $\{\mathcal{V}_i\}$ is manually labeled as four subcategories C_1, C_2, C_3, C_4 and all $\{\mathcal{V}_i\}$ stem from the same ostium as $\mathcal{A}$ per volume, which forms a vessel tree as shown in Figure 4. We save all features computed from $\mathcal{V}_i \in C_1$ into $\mathcal{F}_{on}$, and all features from $\mathcal{V}_i \in C_4$ into $\mathcal{F}_{off}$. Then each $\mathcal{V}_i \in (C_2 \bigcup C_3)$ is piecewise matched against its corresponding annotated curve $\mathcal{A}$ to obtain $[d_i^1; d_i^2; ...; d_i^j; ...]$, using partial sequence matching. e.g., modified dynamic time warping (DTW) [13]. In Figure 3, we permit the matching from an input sequence to partial of the template by searching the element with the minimal value along the last column of DTW cost graph. Thus the overall minimum-cost path may not be traced back from the top-right element as in the original DTW algorithm. This is because only the ostium end correspondence is guaranteed for different branches on a vessel tree. A semi-automatic anomaly/changing point detection scheme[3] is employed onto the sequence of

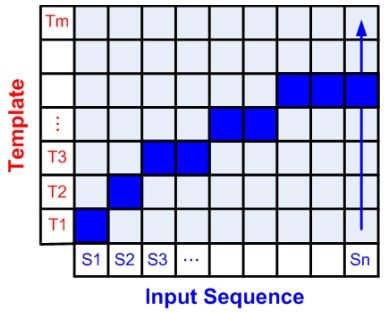


Figure 3. *An illustration of our modified Dynamic Time Warping which allows partial matching from input sequence against template.*

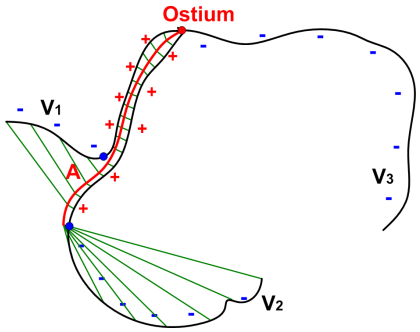


Figure 4. *An illustrative drawing of left anterior descending artery (LAD) with its simplified neighboring vessel branches.* **A** *is the annotation curve (Red), while V_1, V_2, V_3 are vessel segments traced by computer (Black). We further label V_1 as a bifurcated branch (C_2), V_2 as an over-grown branch (C_3), and V_3 as an opposite branch (C_4). By locating the cutting points (as blue dots) on V_1, V_2, voxel features on the matched portions (with* **A***) are used to construct the "on-$\mathcal{A}$" set (+); and we use the unmatched portions of V_1, V_2 and the whole range of V_3 for the "off-$\mathcal{A}$" set (-).*

[2] In training, we need to align $\mathcal{V}_i$ to $\mathcal{A}$ by allowing partial sequence matching, e.g., dynamic time warping. Given that the first point on $\mathcal{A}$ is the vessel ostium $\mathcal{O}$, all knot points on $\mathcal{V}_i$ can be traced back to $\mathcal{O}$ through the aligned path and consequently their geodesic distances can be calculated. Details are omitted due to space limit.

[3] It is expected that the sequential distance pattern is constantly low for matched vessel segment portion, and gradually increasing for unmatched, bifurcated or outgrowth portion in C_2 or C_3 respectively.

$[d_i^1; d_i^2; ...; d_i^j; ...]$ to find the splitting point $d_i^{j'}$ and $P_i^{j'}$. Finally the voxel level feature vectors $\{T; \theta; \kappa; \psi; \mathcal{G}; \Delta\}$ computed at $[P_i^1; P_i^2; ...; P_i^{j'}]$ are pooled into $\mathcal{F}_{on}$ and other features at $[P_i^{j'+1}; P_i^{j'+2}; ...]$ are stored as $\mathcal{F}_{off}$. The final $\mathcal{F}_{on}$ and $\mathcal{F}_{off}$ is obtained by repeating above process for all volumes in the training patient set. A drawing example of how to handle the structural matching between the annotated left anterior descending artery (Red) and its neighboring vessel tree branches (Black) during training and construct the "on-$\mathcal{A}$" and "off-$\mathcal{A}$" feature sets of $\mathcal{F}_{on}$ and $\mathcal{F}_{off}$, is illustrated in Figure 4.

3.1.2 Generative Density Modeling and RVM

Once the $\mathcal{F}_{on}$ and $\mathcal{F}_{off}$ feature sets are generated, in theory, any classification algorithms can be applied. However low-level visual data is often noisy and ambiguous, we argue that generative density models are preferred to give smooth class-conditional outputs. We first investigate three density modeling algorithms of Kernel Density Estimation [9] and two variations of the multi-dimensional histogram model by maximizing the Chernoff Information [2, 10] on two classes $\mathcal{F}_{on}$ and $\mathcal{F}_{off}$. Our feature set is over-complete because it contains 11 distances + geometric features which are not independent. Moreover, density models also generally favor low-dimensional feature space. We initially hand pick four features out of 11 as $\hat{f} = [\mathcal{G}, \theta; \kappa]$ as *global geodesic distance-indexed local tangent orientations and curvature pattern*. Then different feature selection settings are compared and evaluated empirically. Finally each $\mathcal{V}_i$ is presented as $[\hat{f}_i^1; \hat{f}_i^2; ...; \hat{f}_i^j; ...]$ by default in feature space.

Our method has the following three steps:

1. Build density functions $\mathcal{M}_{on}()$ and $\mathcal{M}_{off}()$ of Kernel Density Estimator (KDE), marginally optimized Histogram ($Histogram_1$), or jointly optimized Histogram ($Histogram_2$) for $\mathcal{F}_{on}$ and $\mathcal{F}_{off}$ respectively.

2. For a new feature vector f_t for testing, obtain its normalized "on-$\mathcal{A}$" probability $\rho(\hat{f})$ by

$$\rho(f_t) = \frac{\mathcal{M}_{on}(\hat{f}_t) + \epsilon}{\mathcal{M}_{on}(\hat{f}_t) + \mathcal{M}_{off}(\hat{f}_t) + 2 \times \epsilon} \quad (1)$$

   where $\mathcal{M}_{on}(\hat{f}_t)$ and $\mathcal{M}_{off}(\hat{f}_t)$ are the density values outputted by either of three generative models; $\rho(\hat{f}_t)$ ranges 0 to 1 through normalization; and ϵ is a small constant to avoid divide-by-zero.

3. Output $[\rho_i^1; \rho_i^2; ...; \rho_i^j; ...]$ for each vessel segment $\mathcal{V}_i$.

We employ a publicly available toolbox [9] for KDE implementation. The Chernoff Information [2] in the discrete distribution format of histograms H_{on} and H_{off} is

$$C(H_{on}, H_{off}) = -\min_{0 \le \lambda \le 1} \log(\sum_{k=1}^{K} H_{on}^{\lambda}(k) H_{off}^{1-\lambda}(k)) \quad (2)$$

which describes the magnitude of separatability of two distributions. K is the number of histogram bins. To make less ambiguous classification of "on-$\mathcal{A}$" and "off-$\mathcal{A}$", we optimize $C(H_{on}, H_{off})$ while constructing histograms $Histogram_1$ and $Histogram_2$ by exhaustively searching adaptive bin boundaries in the marginal or joint feature space. The final $Histogram_2$ model is shown in Figure 8 for RCA (Left), LAD (Center) and LCX (Right) arteries. Refer [10] for more details. Through our evaluation, three density models have shown similar performances (as seen their ROC curves in Figure 9) where KDE is slightly advantageous but with much higher model complexity than histogram models. Under the similar number of bins, $Histogram_2$ outperforms $Histogram_1$. Thus we use the density outputs from $Histogram_2$ for the following steps by default.

We also investigate discriminative modeling approaches as opposed to generative models. Relevance vector machine (RVM) [15] has been a widely-used *Bayesian* approach for classification. It follows the same functional form as support vector machines (SVM) but provides probabilistic predictions for class labels. We employ a Matlab implementation of RVM publicly available to construct a discriminative classifier on eight of the original 11 features $[T; \theta; \kappa; \psi; \mathcal{G}; \Delta]$ by the extended RVM feature selection scheme [12]. The resulting classifier is then used to provide probabilistic estimates of ρ as the positive "on-$\mathcal{A}$" class confidences or probabilities. Receiver operating characteristic (ROC) curves of Kernel Density Estimator (KDE), marginally optimized histogram ($Histogram_1$), jointly optimized Histogram ($Histogram_2$) and RVM are given in Figure 9, using three-fold cross validation.

3.2 Vessel Segment Level Classification and Set-Expansion

In testing, if the class information of any vessel segment $\mathcal{V}_i \in C_1$, $(C_2 \cup C_3)$, or C_4 is known, we can discard C_4 vessels, detect and segment the high ρ portions of $(C_2 \cup C_3)$, and combine them with C_1 curves in each volume, to infer the best output per volume/patient. Vessel segment level classification is treated as a general multi-class problem. Because computer traced vessel segments can have various lengths, the dimensions of their resulting ρ-sequences are also different. This problem is tackled by our three step processing as follows: For each $\mathcal{V}_i$,

1. Sub-sample the ρ-sequence $[\rho_i^1; \rho_i^2; ...; \rho_i^j; ...]$ to make them of equal length (say L) as $S_i(\rho)$. This is done by splitting the sequence into L equal fragments, and choosing a representing point (e.g., *Median*) for each segment.

2. Represent features from $S_i(\rho)$. Two sets of features are extracted: directly treating $S_i(\rho)$ after sub-sampling as a L-dimensional vector, and its gradient feature vector $\Delta S_i(\rho)$.

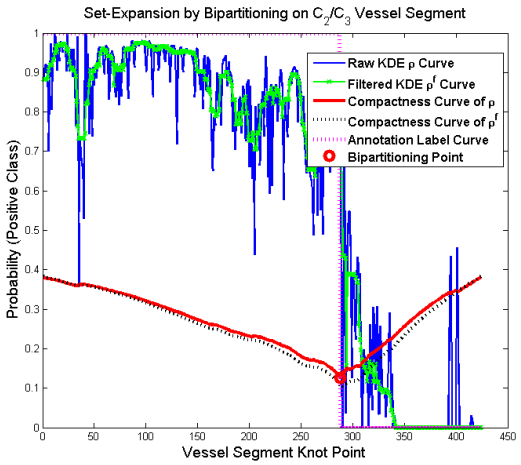


Figure 5. Set-Expansion *by bipartitioning each C_2 or C_3 classified vessel segment into two new portions/curves. Though* Median filter *can make the raw normalized "on-$\mathcal{A}$" probability $\rho()$ curve more smooth, the estimated "bipartitioning" point is highly insensitive with respect to* Median filtering. *The bipartitioning point (Red Circle) coincides closely with our annotation label.*

3. Classify each sequence into the three classes. Since C_4 is not of any use for us, it's more important to separate C_4 sequences from all the other sequences. Therefore we train a binary classifier first to separate C_4 from the others, and then train another binary classifier to distinguish C_1 from $(C_2 \cup C_3)$.

Two classifiers, the linear discriminate analysis (LDA) and the support vector machines (SVM), are exploited and they yield similar results. We measure the performance in AUC (Area Under ROC Curve) values.

For each vessel curve $\mathcal{V}_i \in (C_2 \cup C_3)$, we run *Set-Expansion* operation by bipartitioning it into two pieces with higher ρs for matched "on-$\mathcal{A}$" or lower ρs for unmatched "off-$\mathcal{A}$". The ρ compactness metric is first computed for every point P_i^j on $[\rho_i^1; \rho_i^2; ...; \rho_i^j; ...]$ of $\mathcal{V}_i$ as

$$\Phi(i,j) = std([\rho_i^1; \rho_i^2; ...; \rho_i^j]) \times (j/N_i) + std([\rho_i^{j+1}; \rho_i^{j+2}; ...\rho_i^{N_i}]) \times (N_i - j)/N_i \quad (3)$$

where $std()$ returns the standard deviation from a one-dimensional array of ρs and represents its compactness, and N_i is the length of curve $\mathcal{V}_i$. $\Phi(i,j)$ is the length-weighted average of the standard deviations from two split subsequences. We then select the point $P_i^{j'}$ as the bipartitioning boundary of any $\mathcal{V}_i \in (C_2 \cup C_3)$ where

$$\Phi(i,j') = \min_j\{\Phi(i,j)\} \quad (4)$$

Optionally, multiple bipartitioning boundaries can be computed. Refer to Figure 5 for an example of vessel curve ρ-bipartitioning process. Finally, two new vessel segments split from each $\mathcal{V}_i \in (C_2 \cup C_3)$ are added into $\{\mathcal{V}_i\}$ to form a super-set $\{\mathcal{V}_i'\}$ for each volume.

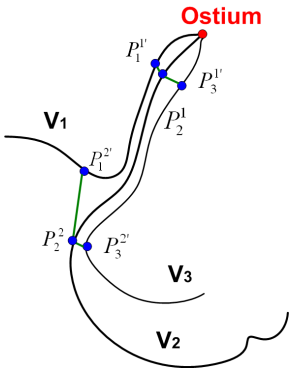


Figure 6. *Example of star-graph linking P_2^j to $P_1^{j'}$ and $P_3^{j'}$ for message-passing based ρ smoothing ($j = 1, 2$).*

3.3 Inter- Vessel Segment Level Smoothing and Goodness Measure

For each $\mathcal{V}_i \in \{\mathcal{V}_i\}$ of per patient/volume, we match it against the rest of vessel curves using dynamic time warping [13] by permitting partial pointwise correspondences. See the dark green line segments between any two curves in Figure 6 as piecewise point matching results. We then build a star-graph for every knot point P_i^j on $\mathcal{V}_i$ to its corresponding points of all other vessel segment $\mathcal{V}_k \in \{\mathcal{V}_i\}$, $k \neq i$, which results an asymmetric, loopy graph in the volume space. Message-passing based smoothing on ρ_i^j is optionally performed, as a specific case of Markov Random Field (MRF) solved by Loopy Belief Propagation (LBP) [18]. The belief is initialized with ρ_i^j output by our generative density functions, and all messages are initialized as 1. Moreover, the pairwise potential function between P_i^j and its correspondence $P_k^{j'}$ is defined as

$$\psi(P_i^j, P_k^{j'}) \propto \exp\{-\frac{d(P_i^j, P_k^{j'})}{\sigma_1}\} \quad (5)$$

where $d(\cdot,\cdot)$ is the Euclidean distance function between two points in 3D coordinates and σ_1 is the estimated standard deviation of Euclidean distances from pairwise point matches between "on-$\mathcal{A}$" and $\mathcal{A}$ during training. The ρ smoothing process iterates on every star-graph on $\{P_i^j, P_k^{j'}\}_{k \neq i}$ for several rounds until convergence. We denote the updated ρ_i^j as $\tilde{\rho}_i^j$. Notice that we only perform inter-vessel segment level smoothing for $\{\mathcal{V}_i\}$ per volume because $\{\mathcal{V}_i'\}$ produce only new curve groupings of $\{\rho_i^j\}$ or $\{\tilde{\rho}_i^j\}$, but not unseen graph structure.

At last, we describe an overall goodness measure for each $\mathcal{V}_i' \in \{\mathcal{V}_i'\}$ and choose the system output as the vessel curve $\mathrm{V} \in \{\mathcal{V}_i'\}$ with the maximum goodness value

$$\Upsilon(\mathcal{V}_i') = \exp\{\frac{\mu_i(\tilde{\rho}) - \bar{\mu}(\rho)}{\sigma(\rho)}\} + \lambda \exp\{\frac{\mid \mathcal{L}_i - \bar{\mu}(\mathcal{L}) \mid}{\sigma(\mathcal{L})}\} \quad (6)$$

where $\mu_i(\tilde{\rho})$ is the mean of $\{\tilde{\rho}_i^j\}$ values on $\mathcal{V}_i'$ and $\mathcal{L}_i$ is the curve length of $\mathcal{V}_i'$. $\bar{\mu}(\rho), \sigma(\rho)$ and $\bar{\mu}(\mathcal{L}), \sigma(\mathcal{L})$ are the estimated means and standard deviations of ρ and curve

length on the annotation vessel set per artery category of RCA, LAD and LCX, as prior information. This goodness measure balances the overall $\tilde{\rho}$ score and the vessel length fitness term against the prior model of annotation dataset. $\lambda = 0.35$ is determined by optimizing $sim(\mathrm{V}, \mathcal{A})$ using cross-validation. It is probably due to the relatively large variation of vessel lengths which makes the second term in equation 6 less weighted, i.e., the vessel length prior is more flat.

4 Experiments

Data: In each CTA volume, we implement and present an interactive graphical user interface which allows radiologist to annotate the three RCA, LAD and LCX arteries. In order to avoid missing of main branches, a multi-hypothesis vessel centerline detection algorithm [1, 5] is implemented to extract as many branches as possible in RCA, LAD and LCX vessel trees. As shown in Figure 7, the distributionof computer traced vessel segments (Right) from 82 patients demonstrates significantly larger variation than the artery annotation dataset (Left), for all three categories. Our CT angiography datasets were acquired from two hospitals under their normal imaging protocols, including both Caucasian and Asian populations.

We first evaluate the performance of generative and discriminative model on the voxel-level feature sets $\mathcal{F}_{on}$ and $\mathcal{F}_{off}$. Figure 9 draw ROC curves of probability density values based on KDE; $Histogram_1$ and $Histogram_2$ models on the domain knowledge selected $[\mathcal{G}, \theta; \kappa]$ features; RVM_A on the full 11 $[T; \theta; \kappa; \psi; \mathcal{G}; \Delta]$ features where a subset of eight features is actually selected by [12]; and RVM_S on the same $[\mathcal{G}, \theta; \kappa]$. In our experiments, the generative density modeling apparently demonstrates superior classification accuracies than the sophisticated discriminative model of Relevance Vector Machine [15, 12] on our noisy low-level imaging data. The AUC measures are $0.9179, 0.9069, 0.9020, 0.8019, 0.8018$ respectively for these five models. The performance difference between RVM_A and RVM_S is negligible which implies our manually picked feature set is empirically as good as the theoretical optimal feature set in RVM_A [12]. To exploit the effects of introducing different distance metrics into the joint distance and geometry feature space, we train four KDE models (as KDE_Δ, KDE, KDE_E and KDE_G) on $[\Delta, \theta; \kappa]$, $[\mathcal{G}, \theta; \kappa]$, $[\mathcal{D}, \theta; \kappa]$, or Geometry-only feature set $[\theta; \kappa]$. Δ is the Euclidean distance displacement and $\mathcal{D} = \mid \Delta \mid$. The AUC measures are $0.9621, 0.9179, 0.9091, 0.8499$, as shown in Figure 10. To get direct insights on ρ's performance at vessel level, we simply use the mean value of $\{\rho\}$ on each curve as its similarity to the "on-$\mathcal{A}$" class. This is executed for all traced vessel segments $\{\mathcal{V}_i\}$ and the annotation $\mathcal{A}$ per artery per volume. Then, if the curve with the highest similarity score is the annotation $\mathcal{A}$ or belongs to C_1, the classification is correct; otherwise, it is incorrect. Using KDE, $Histogram_1$, $Histogram_2$, RVM_A and KDE_G models and LAD artery as example, the classification error rates are $2, 3, 4, 25, 17$ out of 82. Note that the above vessel level evaluation is not applicable in final testing since $\mathcal{A}$ is not available. A Plot of ρ Curves, using KDE, $Histogram_1$, $Histogram_2$ and RVM on a LAD vessel segment $\in C_3$, is also illustrated in Figure 11.

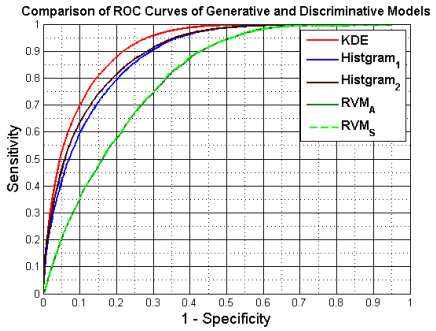


Figure 9. *Receiver operating characteristic curves for learning the binary "on-$\mathcal{A}$" and "off-$\mathcal{A}$" classification using five models: KDE, $Histogram_1$ and $Histogram_2$ on the domain knowledge picked feature set $[\mathcal{G}, \theta; \kappa]$; RVM_A of automatic feature selection on the original* 11 *features $[T; \theta; \kappa; \psi; \mathcal{G}; \Delta]$ and RVM_S on $[\mathcal{G}, \theta; \kappa]$.*

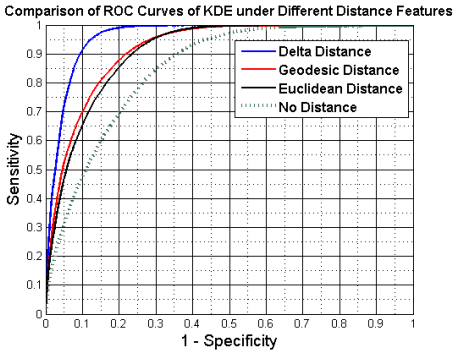
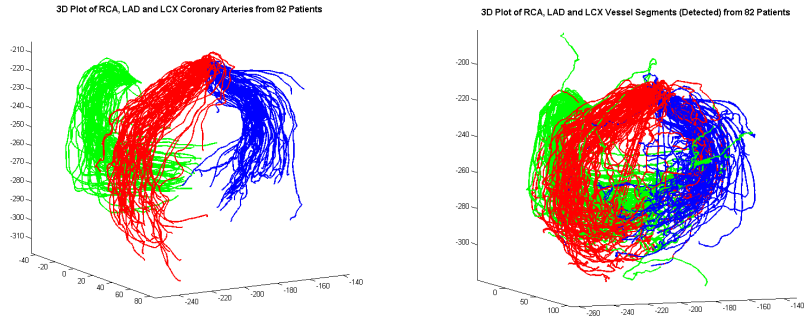


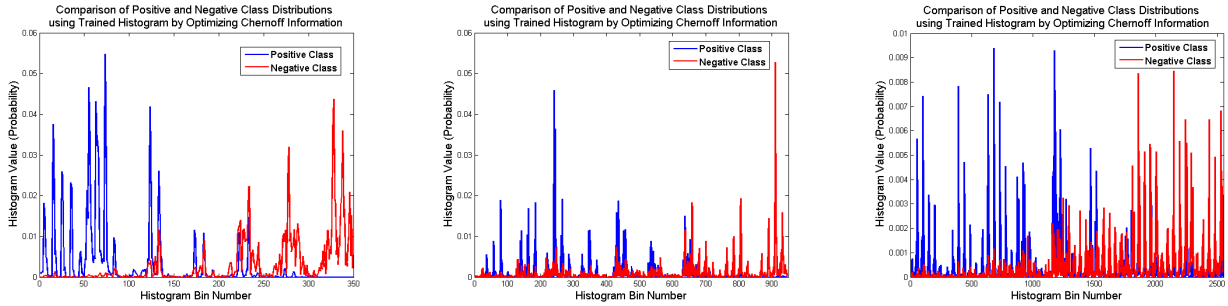
Figure 10. *Receiver operating characteristic curves for learning the binary "on-$\mathcal{A}$" and "off-$\mathcal{A}$" classification using four KDE models under different distance metrics with Geometric features: Distance Displacement $[\Delta, \theta; \kappa]$; Geodesic distance $[\mathcal{G}, \theta; \kappa]$; Euclidean distance $[\mathcal{D}, \theta; \kappa]$; and Geometry-only feature set $[\theta; \kappa]$.*

At the vessel segment level classification of $C_1, (C_2 \cup C_3), C_4$, we employ the linear discriminate analysis classifier (LDA) and the support vector machines (SVM) on $S_i(\rho)$ or $\Delta S_i(\rho)$, as described in section 3.2. The performance differences of LDA and SVM on both feature sets are negligible, and $S_i(\rho)$ offers slightly better accuracy than $\Delta S_i(\rho)$. The AUC values of using LDA on $S_i(\rho)$ features are 0.9898 and 0.9122, for the first classifier to discard C_4 and the second classifier to distinguish $(C_2 \cup C_3)$ from C_1 respectively. After this step vessel segments classified $\in C_4$, with 100% accuracy, are removed from $\{\mathcal{V}_i\}$. The final system has the detection rate of 98% for $(C_2 \cup C_3)$ with 10% false positive rate of C_1, by tuning the operat-



(a) Annotation (b) Vessel Segments

Figure 7. *The 3D Plots of all* annotation *curves (Left) and initially* traced *vessel segment curves [1] (Right), integrated from a population of 82 patients. RCA is shown in Green and LAD in Red and LCX in Blue. The ostia of RCA, LAD, or LCX vessel curves are respectively aligned for display clarity.*



(a) $Histogram_2$ RCA (b) $Histogram_2$ LAD (c) $Histogram_2$ LCX

Figure 8. *Jointly optimized* $Histogram_2$ *model of RCA (Left), LAD (Center) and LCX (Right) arteries by optimizing Chernoff Information in the joint* $\hat{f} = [\mathcal{G}, \theta; \kappa]$ *feature space. The final* $C(H_{on}, H_{off}) = 1.1492, 0.4108, 0.4042$; $\lambda = 0.5$ *for RCA, LAD and LCX.* $Histogram_2$ *has total* 350*,* 945 *or* 2592 *bins in four dimensions from RCA, LAD to LCX, which indicates increasingly more complex models.*

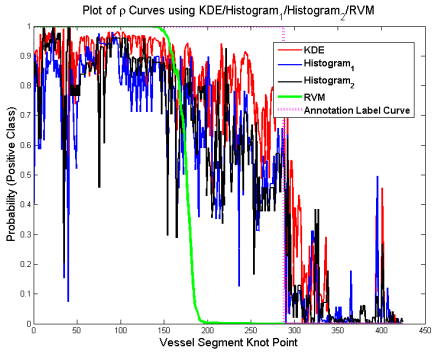


Figure 11. *Plot of* ρ *Curves using KDE,* $Histogram_1$*,* $Histogram_2$ *and RVM on a LAD vessel segment* $\in C_3$*. RVM shows low variance but high bias against the annotation label curve (in magenta), compared with its generative counterparts.*

ing point on ROC curve. Our *Set-Expansion* process by ρ-bipartitioning is highly robust and accurate, as demonstrated in Figures 5 and 12, though the raw ρ-curves appear very bumpy and noisy. We especially test the repeatability of bipartitioning boundaries using *Median filter* smoothing on ρ-curves with various window sizes. In our experiments, the effect of *Median filter* is inappreciable.

The accuracy of our whole vessel structure parsing system depends on the similarity of the annotation artery curve $\mathcal{A}$ and the optimal vessel output V from the expanded set $\{\mathcal{V}'_i\}_{i=1,...,m}$ as $sim(\mathrm{V}, \mathcal{A})$ per artery per volume. Assuming $Cov(\mathcal{V}_1, \mathcal{V}_2)$ is an asymmetric function that gives the length of the portion of curve $\mathcal{V}_1$ overlapped with $\mathcal{V}_2$, we define

$$sim(\mathrm{V}, \mathcal{A}) = \{\frac{Cov(\mathrm{V}, \mathcal{A})}{\mathcal{L}(\mathrm{V})} + \frac{Cov(\mathcal{A}, \mathrm{V})}{\mathcal{L}(\mathcal{A})}\}/2 \quad (7)$$

where $\mathcal{L}(\cdot)$ returns the length of curve V or $\mathcal{A}$ and $sim(\mathrm{V}, \mathcal{A})$ is symmetric and ranges from 0 to 1. In practice, $Cov(\mathcal{V}_1, \mathcal{V}_2)$ is implemented by first computing the distances of all sampling points on $\mathcal{V}_1$ to their correspondence on $\mathcal{V}_2$, and counting the number of points with distance $\leq \mathcal{T} = (\mu_1 + \lambda_1\sigma_1)$. μ_1 and σ_1 are the estimated mean and standard deviation of Euclidean distances from pairs of point matches between "on-$\mathcal{A}$" and $\mathcal{A}$ during training. λ_1 is set as 0.5. $\mathcal{T} = (\mu_1 + 0.5\sigma_1) = 1.96mm$ value is the same for all three coronary arteries and coincides closely to $2mm$ as the averaged vessel width in [16]. The estimated $sim(\mathrm{V}, \mathcal{A})$ accuracies of three artery categories (RCA, LAD and LCX), using ρ produced by three density models of KDE, $Histogram_1$ or $Histogram_2$, and with *Median filtering* of various window sizes, are presented in

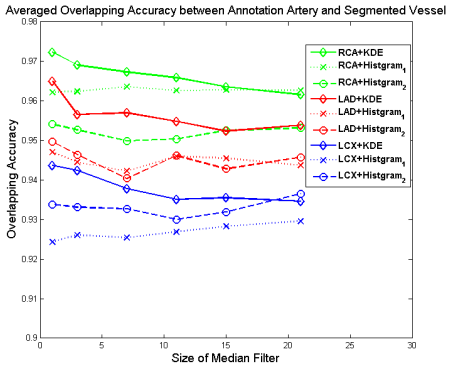


Figure 12. *Averaged overlapping accuracy between annotation artery and our final output vessel segment. Our results are stable and robust to no* Median filtering *on* $\rho()$ *and* Median filtering *with various window sizes.*

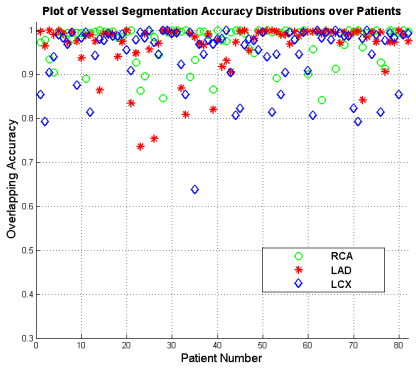


Figure 13. *Plot of Vessel Segmentation Accuracy Distributions over Patients for RCA, LAD and LCX.*

Figure 12. These accuracy measures are averaged from a dataset of 82 patients by three-fold cross validation. In summary, we achieve accurate and stable results as the averaged $sim(\mathrm{V}, \mathcal{A}) = 0.9497 \sim 0.9716$ for RCA, $0.9328 \sim 0.9653$ for LAD and $0.9219 \sim 0.9444$ for LCX. LBP smoothness can slightly improve the final vessel segmentation accuracies at 0.0057 for RCA, 0.0096 for LAD and 0.0126 for LCX in average. Plot of vessel segmentation accuracy distributions over 82 patients for RCA, LAD and LCX is shown in Figure 13. The statistics of mean, standard deviation (Std) and percentages (%) of vessel segmentation accuracy above certain thresholds ($> 0.95, 0.90, 0.85, 0.80$) over patients is also given in Table 1. Robust and accurate coronary vessel labeling and segmentation results are achieved for majority or high majority of patients. For cases with lower accuracies $<= 0.80$, we observe that $\{\mathcal{V}_i\}$ often covers only partial of $\mathcal{A}$ using the vessel tree tracer [1] under poor imaging quality. Our algorithm performs at per patient and per artery category level in run time.

	Mean	Std	0.95	0.90	0.85	0.80
RCA	0.972	0.042	80.5	90.2	97.6	100
LAD	0.965	0.056	82.9	90.2	92.7	97.6
LCX	0.944	0.070	64.6	81.7	87.8	96.3

Table 1. *Table of mean, standard deviation (Std) and percentages (%) of vessel segmentation accuracy above certain thresholds over patients.*

5 Conclusion

In this paper, we present a hierarchical machine learning approach for the problem of tubular structure parsing in 3D medical imaging. It has a progressive *three-tiered* classification process at volumetric voxel level, vessel segment level, and inter-segment level. Our new *joint distance and local geometry feature sets* are shown to be capable of describing and discriminating complex 3D tubular structures/branchs. In future work, we plan to integrate the class-conditional densities ρ as an additional cost into the vessel tracing algorithm for more efficient object-specific tracing, and investigate extensions of our framework for other 2D/3D vessels or tubular structures in medical imaging.

References

[1] S. R. Aylward. Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. *IEEE Trans. on Medical Imaging*, 21:61–75, 2002.

[2] T. M. Cover and J. A. Thomas. Elements of information theory, 1991.

[3] E. W. Dijkstra. A note on two problems in connections with graphs. *Numerische Mathematic*, 1:269–271, 1959.

[4] A. Frangi, W. Niessen, K. Vincken, and M. Viergever. Multiscale vessel enhancement filtering. In *MICCAI*, volume 1, 1998.

[5] O. Friman, C. Kuhnel, and H. Peitgen. Coronary artery centerline extraction using multiple hypothesis tracking and minimal paths. In *MICCAI*, 2008.

[6] D. Geman and B. Jedynak. An active testing model for tracking roads in satellite images. *IEEE Trans. Pat. Anal. Mach. Intell.*, 18:1–14, 1996.

[7] M. Gulsun and H. Tek. Robust vessel tree modeling. In *MICCAI*, 2008.

[8] A. Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans. on Medical Imaging*, 19:203–210, 2000.

[9] A. Ihler. Kernel density estimation toolbox for matlab, 2004.

[10] S. Konishi, A. Yuille, J. Coughlan, and S. Zhu. Statistical edge detection: Learning and evaluating edge cues. *IEEE Trans. Pat. Anal. Mach. Intell.*, 25:57–74, 2003.

[11] H. Li and A. J. Yezzi. Vessels as 4-d curves: Global minimal 4-d paths to extract 3-d tubular surfaces and centerlines. *IEEE Trans. on Medical Imaging*, 26:1213–1223, 2007.

[12] V. Raykar, B. Krishnapuram, J. Bi, M. Dundar, and B. Rao. Bayesian multiple instance learning: automatic feature selection and inductive transfer. In *ICML*, 2008.

[13] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 26:43–49, 1978.

[14] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pat. Anal. Mach. Intell.*

[15] M. E. Tipping. Sparse bayesian learning and the relevance vector machinen. *Journal of Machine Learning Research*, 1:211244, 2001.

[16] T. van Walsum, M. Schaap, C. Metz, A. van der Giessen, and W. Niessen. Averaging center lines: Mean shift on paths. In *MICCAI*, 2008.

[17] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *CVPR*, 2001.

[18] J. Yedidia, W. T. Freeman, and Y. Weiss. Understanding belief propagation and its generalizations. In *IJCAI*, 2001.