

AUTOMATIC LABELLING OF CORONARY ARTERIES

Akin Akinyemi ^a, Sean Murphy ^b, Ian Poole ^b, Colin Roberts ^b

^a Institute for System Level Integration
Alba Campus, Livingston, Scotland, EH54 7EG, UK
phone: +447859004144, email: akina@sli-institute.ac.uk

^b Toshiba Medical Visualization Systems, Europe, Ltd
2 Anderson Place, Bonnington Bond, Edinburgh, Scotland, EH6 5NP, UK

ABSTRACT

Automatically assigning the correct anatomical labels to coronary arteries is an important task that would speed up work flow times of radiographers, radiologists and cardiologists, and also aid the standard assessment of coronary artery disease. However, automatic labelling faces challenges resulting from structures as complex and widely varied as coronary anatomy. A system has been developed which addresses this requirement and is capable of automatically assigning correct anatomical labels to pre-segmented coronary artery centrelines in Cardiac Computed-Tomography Angiographic (CCTA) images with 84% accuracy. The system consists of two major phases: 1) training a multivariate gaussian classifier with labelled anatomies to estimate mean-vectors for each anatomical class and a covariance matrix pooled over all classes, based on a set of features; 2) generating all plausible label combinations per test anatomy based on a set of topological and geometric rules, and returning the most likely based on the parameters generated in 1).

1. INTRODUCTION

Coronary heart disease is a major life-threatening disease in Europe and America at the moment; the rate of heart disease is increasing in countries within Africa and Asia every year at a rate almost proportional to the development rate of these countries [1]. In a computer-assisted cardiac analysis application, the radiologist will be able to assess each coronary artery for stenoses and other defects, then report findings for diagnosis or recommendation; therefore, being able to label these arteries by their standard anatomical names will facilitate the workflow, thus speeding the whole process of cardiac analysis, and reducing the time to detect coronary artery disease.

1.1 Previous Work

The problem is viewed as a combined image analysis and classification task. Previous work has been carried out in attempts to solve the same problem, with some attempts aimed at different imaging modalities. Ezquerro et al [3] address the problem of coronary artery labelling for X-ray angiograms by using a graph-matching approach. The method introduced in [3] addresses the challenges posed by the presence of noise, artefacts, and competing structures. In this system, a model graph of a generic coronary vasculature is used as a reference, incoming datasets are then matched to this using features collected by Dogde et al [6] as extra matching criteria. The method assumes that segments of vessel

in-between bifurcations can still possess characteristic features of the entire vessel, and hence use vessel length, lumen diameter and branch angle as similarity features in the cost function; although this may not be the case, as the individual segments do not accurately represent the entire vessel especially in terms of length and branch angles. Haris et al [5] and Chalopin et al [4] also address the problem for 2-D X-ray angiograms using graph-matching methods. The method in [4] stores a list of reference graphs and performs an initial matching to find the best reference graph; with similarity based on the inertia axis orientation, inertia, and the eccentricity of the nodes and arcs, without using any physical features of the arteries as matching criteria.

The cost functions used in the above methods do not take into account the multivariate nature of the features, which is why the method proposed in this paper uses a multivariate Gaussian Classifier in order to exploit this property. Furthermore, the graph-matching approaches as used above assume a generic vasculature, but in clinical datasets, anomalies do exist that will lead to a failure of such methods; hence this paper proposes a less strict model-based approach to the problem.

The method proposed in this paper takes into account the tree structure of the coronary vasculature, and exploits knowledge of its topology to generate a set of plausible labelling candidates. It then exploits knowledge of the spatial properties of each anatomical vessel class to reduce the number of candidate labelled trees. These are then scored using the physical features of each vessel in a trained classifier to find the most likely labelling.

In this paper, the individual segments of a vessel in-between bifurcations are referred to as topological segments; the concatenation of such contiguous topological segments make up a vessel segment as shown in figure 2.

2. METHOD

The data-flow is as shown in figure 1.

2.1 Label Pool

The label pool is a list of all the classes (labels) to be considered, which are *LM* (Left Main), *LAD* (Left Anterior Descending), *LCX* (Left Circumflex), *LMG* (Left Obtuse Marginal), *LRI* (Left Ramus-Intermedius), *LD* (Left Diagonal), *Linsig*, *Rinsig* (All segments not required in specification, left and right prefixes), *RCA* (Right Coronary Artery), *RPD* (Right Posterior Descending Artery), *RPLA* (Right Postero-lateral Artery), and *RAM* (Right Acute Marginal

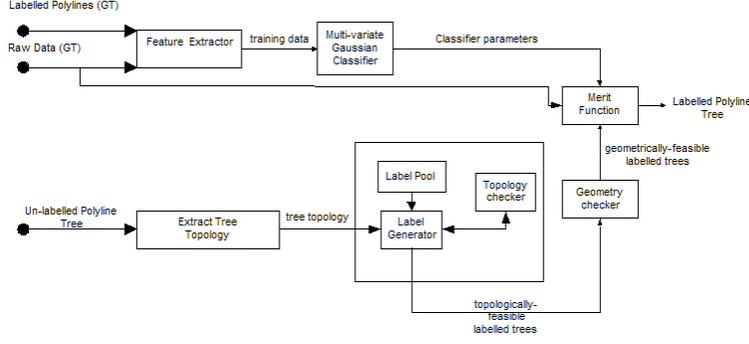


Figure 1: System Flow Diagram
Shows the modules and their connections within the algorithm

Artery). These labels are used to supply the Label Generator.

2.2 Training Data

The training data is collected by manually segmenting individual vessels from Cardiac CT Angiography data of 42 patients, using commercial vessel-tracking software to track their centrelines (skeletons). The vessel segments are stored as polylines:

$$polyline = \{x_0, x_1, \dots, x_{n-1} | x_i \in \mathfrak{R}^3\} \quad (1)$$

In this context, a polyline is an ordered list of points along the centreline of a single coronary artery. The entire coronary artery vasculature is represented as a polyline-tree, see figure 2.

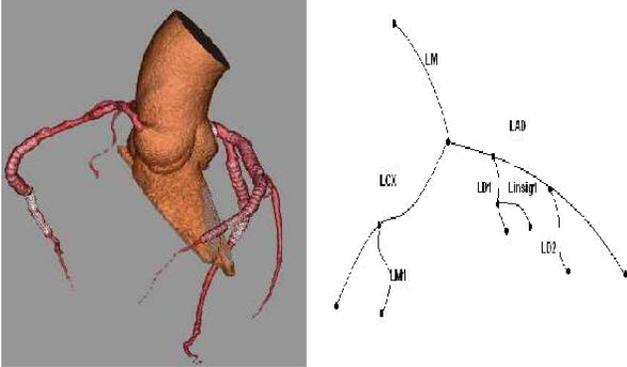


Figure 2: Ground-Truth format for the left coronary
The LAD for instance, is stored as a single vessel rather than as three individual segments. The image on the left shows a 3-D volume rendering of the entire coronary vasculature, including the aorta.

2.3 Feature Extraction

This module takes in labelled coronary vessel segments as polylines, and with the raw data in context, up to 14 features are measured and saved. The most significant features are defined below:

- $Length_v = \sum_{i=0}^{N-1} length_i$, where N is the number of topological segments making up the vessel v .
- $LogLengthProportion_v = \text{Log}(\frac{Length_v}{\sum_{v=0}^{V-1} Length_v})$, where there are V vessel segments in the tree.
- $BranchAngle(degrees) = 180 - \cos^{-1}(\frac{\vec{u}_v}{u_{v-1}})$, where \vec{u}_v is the unit vector in the direction of vessel v . This is the angle made between a vessel segment and its parent segment, belonging to another vessel class.
- $LogTortuosity_v = \text{Log}(\frac{Length_v}{v_{P-1}-v_0})$ where P points make up a vessel segment v . This is a measure of curvature of the vessel segment; other measures exist [2].
- **Diameter**: the measured diameter of the cylinder created by dilating the topological segment within the data context using a spherical structuring element with radius greater than the maximum radius of a coronary artery, and thresholding at the mean intensity along the segment.
- $AvgDiameter_v = \frac{1}{N} \sum_{i=0}^{N-1} Diameter_i$,
- $Volume_v = \sum_{i=0}^N Diameter_i * length_i$
- $LogVolumeProportion_v = \text{Log}(\frac{Volume_v}{\sum_{v=0}^{V-1} Volume_v})$
- Direction of Vessels represented by 3D unit vectors

These features, collected from manually labelled coronary vessel segments, are used to train the multivariate Gaussian classifier; in the application, these features when measured on novel vessel segments, form the measured feature vector for each candidate.

2.4 Multi-Variate Gaussian Classifier

The MVG classifier operates in a training phase and a classification phase.

Training phase: The classifier takes as input the feature vector for each class computed by the feature extractor on training data, and computes the parameters of the multivariate Gaussian distribution:

- The mean vector for each class k , μ_k and
- The Covariance matrix, C pooled over all classes

Classification phase:

- The classifier takes as input, the measured feature vector for each vessel, \mathbf{z} and uses the above parameters to calculate the likelihood for each class;

$$P(\mathbf{z}|\omega_k) = \frac{1}{|C|\sqrt{2\pi}} \exp\left(-\frac{(\mathbf{z} - \mu_k)^T C^{-1} (\mathbf{z} - \mu_k)}{2}\right) \quad (2)$$

- By applying Bayes theorem, the posterior probability, i.e. the probability of a vessel belonging to class ω_k , given its measurement vector \mathbf{z} is given by:

$$P(\omega_k|\mathbf{z}) = \frac{P(\omega_k)P(\mathbf{z}|\omega_k)}{P(\mathbf{z})} \quad (3)$$

$P(\omega_k)$ is the prior probability of observing the class ω_k .

- Empirical priors are estimated in this case; based on the relative occurrences in the training data.
- The Posterior probability in equation 3 is used as the measure of merit in the Merit Function module.

2.5 Label Generator

The label generator module is a recursive function which attempts to assign to the topological segments of the coronary centreline tree all possible label combinations from the label pool. For any function of this sort, the implied computational cost is apparent, especially for trees as seen in figure 2, almost 2 billion (7^{11}) different trees produced for scoring. For efficiency, only topologically legal candidates are generated. Furthermore, only geometrically plausible labelled candidates are allowed to be scored by the classifier. The module works on each level/segment of the tree as follows:

For each vessel class in the label pool

- assign the class to the current tree level
- if the assignment is topologically legal (section 2.6)
 - if all levels of the tree have been labelled
 - ◊ if the labelled tree is geometrically legal (section 2.7)
 - send to Merit module (section 2.8) for scoring
 - record best scoring labelled tree
 - else
 - ◊ Proceed to next level (recursive call)

2.6 Topology Checker

The label generator generates a sequence of legal labelled trees. Certain rules pertinent to the syntax of the coronary anatomy are encoded in this module to generate legal labels based on the topology of the un-labelled polyline-tree. For instance, on the left coronary tree, some of the rules are as follows:

- The root segment of the left coronary artery is the *LM*;
- An *LM* segment can not have any siblings;
- Segments labelled *RCA*, *RPD*, *RAM*, *RPLA* and *Rinsig* are not permitted;
- An *LAD* segment can only be a child of *LM* or an extension of itself;
- An *LCX* segment can only be a child of *LM* or an extension of itself;

- An *LD* segment can only be a child of *LAD* or an extension of itself;
- An *LMG* segment can only be a child of *LCX* or an extension of itself.

These rules exist for both left and right coronary vessels and are applied within the label generator to limit the number of candidates. In the case of the 12 segment tree (figure 2) with 7 label classes, the number of candidates can be reduced to as low as 511. This module effectively applies syntactic approaches to pruning the search space; other syntactic approaches to pattern recognition have been defined in [7].

2.7 Geometry Checker

Vessel geometry in this case refers to the physical and spatial characteristics of the coronary vessel tree. In this algorithm, only the attributes pertaining to the centrelines of the vessels are considered here (e.g. diameter is not considered), and therefore the polyline-tree representation is sufficient. This module takes in a fully labelled polyline-tree as input, and exploits certain spatial geometric rules unique to individual classes and other rules pertaining to fully labelled trees, resulting in the elimination of more candidates. An example labelled coronary vasculature is shown in 3D in figure 3. Some of the rules applied are as follows:

- For the *LAD*
 - Looking in the direction going down the vessel, all *LD* segments must exist on the right side of the vessel, while *Linsig* segments (actually called Septal) must be on the left.
- For the *LCX*
 - The vessel must traverse a direction posterior to its initial sibling (normally the *LAD*);
 - Looking down the *LCX* vessel; all *LMG* vessels must be on the left side, and any vessels coming off the right side must be *Linsig*;
 - The *LCX* always starts within an average distance from the root of the left coronary tree.

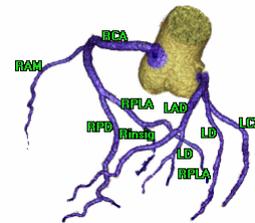


Figure 3: Labelled Coronary Tree 3D
An example labelled coronary vasculature showing the spatial properties of each vessel class.

This step in the algorithm is able to reduce the number of segments in the 12-segment tree (figure 2) from 511 to as low as 71 in some cases.

2.8 Merit Function

This module calculates the merit of each considered labelled tree. The module takes in the labelled polyline tree as input and sends it along with the raw data to the feature extractor module, which computes a measured feature vector \mathbf{z} for

each labelled vessel. The merit of each classification is the sum of the normalised posterior probabilities of the assigned classes, calculated by the MVG classifier in equation 3 for each vessel.

$$Merit = \sum_{v=0}^{V-1} P(\omega_k | \mathbf{z}_v) \quad (4)$$

where V is the number of vessels in the tree.

The posterior probability is normalised to account for labelling configurations having a different number of vessels; this is done by computing a posterior probability per unit length for each vessel. Therefore the merit becomes:

$$Merit' = \sum_{v=0}^{V-1} P(\omega_k | \mathbf{z}_v) * \text{Log}^{-1}(\text{LogLengthProportion}_v) \quad (5)$$

The classified tree with the highest merit is returned as a labelled polyline tree.

3. RESULTS

The MVG classifier (section 2.4) is trained using vessel feature data from 42 datasets of varying image quality, although most cases were of above average quality. The validation environment employed performs leave-one-out (LOO) experiments on each of the 42 datasets in two modes.

In one mode; the manually segmented coronary centrelines are passed into the labelling algorithm, and their results validated against manual labelling of the same centrelines by medical experts. In the other mode, output from an in-house automatic coronary centreline tracking algorithm is passed into the labelling algorithm, and the resulting labels are validated against automatically transferred labels from the corresponding manually segmented centrelines.

The first mode is aimed at assessing the performance of the algorithm in isolation; whilst the second mode assesses the algorithm as if it were in a real application.

The validation metrics are presented using a confusion matrix, hence the overall accuracy can be calculated by the trace divided by the sum over the entire matrix. These confusion matrices have been summarised in table 1.

Vessel Class	Accuracy Mode 1(%)	Accuracy Mode 2(%)
LM	100.0	93.2
LAD	97.4	91.7
LCX	91.7	62.7
LRI	62.5	28.6
LD	80	87.0
LMG	78.9	50.0
Linsig	63.6	53.3
RCA	98.9	86.7
RAM	67.56	36.8
RPD	65.0	33.3
RPLA	86.5	69.2
Rinsig	59.7	81.4
Overall	84.4	76.5

Table 1: Results for both modes of validation
A summary of the per-class accuracy computed from the confusion matrix.

In mode 2, a label-transfer algorithm is used to automatically transfer labels from the manually-labelled/manually segmented vessels to the automatically segmented ones using the overlapping volume as criteria. This results in more *Linsig* and *Rinsig* segments, as non-overlapping segments are labelled as such. A popular source of error here is the *LRI* segments being misclassified as *LD* or *LMG*. This is because the topological rule, which states that the *LRI* has to occur at a trifurcation, has to be relaxed for this experiment configuration because the tree-tracking algorithm is unable to detect trifurcations. Without this constraint, the *LRI* and the *LD* are easily confused.

Subjective validation was carried out by an experienced radiologist on datasets that were not included in the training data; this is performed as an end-to-end implementation, using multi-planar reconstructions (MPR) and 3D volume renderings (figure 3). The algorithm was accurate in cases where the coronary centreline tracking algorithm was successful in tracking correct trees, but returned more *Linsig* and *Rinsig* in cases where deformed trees were produced.

4. CONCLUSION

A method has been developed for labelling segmented coronary vessels in CCTA images. The approach can be seen as a two-step process: 1) knowledge-based assignment, whereby all plausible labelled trees are generated and 2) statistical classification, whereby the most likely labelling is chosen based on closeness to parameters defined by a multivariate Gaussian classifier. The method is tested on 42 CCTA datasets of varying coronary anatomy using two different sources as input: 1) manually-segmented vessels 2) automatically-segmented vessels. In mode 1, the method yields an overall level of agreement of 84.4% (table 1) between the labels assigned by the human expert and those assigned by the system. In mode 2, the system yields an overall level of agreement of 76.47% between the labels automatically transferred to the tracked vessels and those assigned by it.

Further subjective validation by a radiologist found the labelling tool to be useful in clinical situations.

5. ACKNOWLEDGMENT

Image data courtesy of Frank J. Rybicki, MD, PhD, Director, Applied Imaging Science Laboratory Brigham and Women's Hospital and Harvard Medical School Boston, MA, USA 02115.

REFERENCES

- [1] *The Global Burden of Disease*. World Health Organisation, 2008.
- [2] Eze C.U., Gupta R., and Newman D.L. A comparison of quantitative measures of arterial tortuosity using sine wave simulations and 3d wire models. *Physics in Medicine and Biology*, 45:2593–2599(7), 2000.
- [3] N. Ezquerro, S. Capell, L. Klein, and P. Duijves. Model-guided labeling of coronary structure. In *Medical Imaging, IEEE Transactions on*, volume 17 of 3, pages 429–441, 1998.
- [4] C. Chalopin; I.E. Magnin; G. Finet. Automatic labelling of the coronary tree using a three dimensional reference prior model. *Computers in Cardiology*, 5, 1998.

- [5] K. Haris, S.N. Efstratiadis, N. Maglaveras, C. Pappas, J. Gourassas, and G. Louridas. Automatic labelling of the coronary tree using a three dimensional reference prior model. In *Medical Imaging, IEEE Transactions on*, volume 18, pages 1003 – 1015, 1999.
- [6] JT Dodge Jr, BG Brown, EL Bolson, and HT Dodge. Lumen diameter of normal human coronary arteries: Influence of age, sex, anatomic variation, and left ventricular hypertrophy or dilation. In *Circulation, American Heart Association*, volume 86, pages 232–246, 1992.
- [7] Guangzheng Yang. A syntactic approach for building a knowledge-based pattern recognition system. *Pattern Recognition, 1988., 9th International Conference on*, pages 1236–1238 vol.2, Nov 1988.