
3D Segmentation in the Clinic: A Grand Challenge II - Coronary Artery Tracking

C.T. Metz^{1,*}, M. Schaap^{1,*}, T. van Walsum¹, A.G. van der Giessen²,
A.C. Weustink³, N.R. Mollet³, G. Krestin³ and W.J. Niessen^{1,4}

August 25, 2008

*Both authors contributed equally to this paper.

¹Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics,
Erasmus MC, Rotterdam, The Netherlands.

²Department of Biomedical Engineering, Erasmus MC, Rotterdam, The Netherlands.

³Department of Radiology, Erasmus MC, Rotterdam, The Netherlands.

⁴Imaging Science & Technology, Faculty of Applied Sciences,
Delft University of Technology, The Netherlands.

Abstract

In this paper the Coronary Artery Tracking competition, which was part of the workshop: "3D Segmentation in the Clinic: A Grand Challenge II" is described. This workshop was held during the 2008 Medical Image Computing and Computer Assisted Intervention (MICCAI) conference. An introduction is given to underline the importance of (semi-)automatic coronary artery centerline extraction methods and the advantages of an online framework facilitating a fair comparison of these methods. Furthermore, information is provided about the set-up of the workshop, the evaluation measures used and the online framework. Results for the algorithms, submitted by both industrial and academic research institutes, are presented as well.

Contents

1	Introduction	2
2	Workshop	2
3	Evaluation framework	4
3.1	Cardiac CTA data	4
3.2	Reference standard	5
3.3	Evaluation	5
	Definition of true positive, false positive and false negative points	6
	Overlap measures	6
	Accuracy measures	7
3.4	Comparing the algorithms	7
4	Results for <i>testing 1</i> data	7

5 Discussion	9
A Inter-observer variability and scores	9
A.1 Overlap measures	9
A.2 Accuracy measures	10

1 Introduction

(Semi-)automatic coronary artery centerline extraction from CTA data is important in clinical practice, where visualization techniques and measurement tools often rely on these centerlines. Until now, several CTA coronary artery centerline extraction and segmentation methods have been described in literature. In these papers promising results are generally reported. However, they often lack an extensive quantitative evaluation, making it difficult to judge the performance of the methods on clinical data. Moreover, authors use their own datasets and measures for the evaluation, hampering a fair comparison of the performance of different algorithms.

We are convinced that a common image database, a corresponding reference standard and good evaluation measures are a prerequisite for a fair comparison of different algorithms. Making these available through an easily accessible web-based interface would provide a continuous benchmark of state-of-the-art techniques. The usefulness of such a framework has already been proven by similar frameworks in the field of medical imaging and computer vision [10, 13]. Next to the continuous benchmark provided by the framework, the understanding of algorithm performance that results from the standardized evaluation can also steer the clinical implementation and utilization, as a system architect can use objective measures to choose the best tracking algorithm for a specific task.

For this workshop, which is the successor of the liver and caudate segmentation workshop at MICCAI 2007 [13], such an evaluation framework was initiated. Details about the workshop can be found in section 2. Section 3 gives a description of the new evaluation framework and in section 4 results for thirteen submitted coronary artery tracking methods are listed.

2 Workshop

The MICCAI 2008 workshop "3D Segmentation in the Clinic: a Grand Challenge II" consisted of competitions for three different applications: coronary artery tracking, liver tumor segmentations and MS lesion segmentation. This document describes the first competition. Teams that were willing to join could register on the website of the challenge and after signing a letter of intent¹, the image data could be downloaded. For the training data, a corresponding reference standard could be downloaded as well. Besides that, source code of the evaluation software (C++) was available for download.

The imaging data was divided into three sets. The first set of eight datasets could be used for training and was made publicly available, together with a corresponding reference standard that was created from centerlines of three observers. The second set with sixteen other datasets (*testing I*) was provided without

¹ In this letter of intent participants stated that they were intending to submit tracking results to the workshop and that they would only use the data in the context of the introduced evaluation framework. Furthermore they stated to not distribute the data to other people.

corresponding reference standard and was used for performance assessment of the algorithms before the workshop. Submissions of the teams consisted of the results of their methods on the *testing 1* data and an article describing the method and giving a standardized overview of the results. The set with the remaining eight datasets (*testing 2*) was used for the on-site competition during the workshop. This on-site competition is important, because it prevents (un)intentional adaptation of the method or parameter settings based on the testing datasets. Furthermore, it can give an indication of how well the methods can produce results in limited time.

An online submission system was available for evaluation of the training data and for the final submission of the testing data. Results on the training data could be submitted and inspected by the teams unlimitedly, but only inspection of one *testing 1* submission was allowed. This policy prevented unwanted training of the methods on the testing data.

Approximately 100 authors of related publications and the major medical imaging companies were invited to submit their results on the *testing 1* datasets. Fifty-three teams from all over the world² showed their interest by registering for the challenge, 36 teams downloaded the training and testing data, and thirteen teams submitted results: five fully-automatic methods, three minimally interactive methods and five interactive methods. A wide variety of approaches for centerline extraction was applied by the participants and both industrial and academic research institutes submitted results.

Submission of results was possible in three different challenge categories: automatic tracking, tracking with minimal user-interaction and interactive tracking.

Challenge 1: automatic tracking

Methods submitted for the automatic tracking challenge had to find the coronary artery centerlines without any user-interaction. In order to evaluate the performance of the tracking method, two points were provided per vessel to select the coronary artery of interest:

- Point A: a point inside the distal part of the vessel; this point unambiguously defines the vessel to be tracked;
- Point B: a point approximately 3 cm (measured along the centerline) distal of the start point of the centerline.

Point A had to be used for selecting the appropriate centerline. If the automatic tracking result did not contain centerlines near point A, point B could be used instead to select the appropriate centerline. Point A and B were only meant for selecting the right centerline, and it was **not allowed** to use these points as input for the tracking method.

Challenge 2: tracking with minimal user-interaction

Methods submitted for the tracking with minimal user-interaction challenge were allowed to use **one point per vessel** as input for the algorithm. This could be either:

- Point A or B, as defined above;
- Point S: the start point of the centerline;
- Point E: the end point of the centerline;
- Point U: any other manually defined point.

²Europe: 20, North America: 16, Asia: 12, South America: 3, Africa: 1, Australia: 1

Table 1: Image quality of the training and testing datasets.

	Poor	Moderate	Good	Total
Training	2	3	3	8
Testing 1	2	7	7	16
Testing 2	2	1	5	8

Table 2: Presence of calcium in the training and testing datasets.

	Low	Moderate	Severe	Total
Training	3	4	1	8
Testing 1	6	8	2	16
Testing 2	3	4	1	8

Points A, B, S and E were provided with the data. Furthermore, in case the method obtains a vessel tree from the initial point, point A or B could be used after centerline extraction to select the appropriate centerline for evaluation.

Challenge 3: interactive tracking

All methods that required more user-interaction than one point per vessel as input were part of the interactive tracking challenge. Methods could use e.g. both points S and E from challenge 2, a series of manually clicked positions, or one point and a user-defined threshold.

3 Evaluation framework

3.1 Cardiac CTA data

Cardiac CTA datasets of 32 patients who were referred for CT coronary angiography at Erasmus MC, Rotterdam, The Netherlands, were randomly selected. CTA scans were acquired between June 2005 and June 2006 using a 64 slice (20 cases) and dual-source (12 cases) computed tomography scanner (Somatom Sensation 64 and Definition, Siemens Medical Solutions, Forchheim, Germany). Diastolic reconstructions were used, with reconstruction intervals varying from 250 ms to 400 ms before the R-peak. Three datasets were reconstructed using a sharp (B46f) kernel, all others were reconstructed using a medium-to-smooth (B30f) kernel.

To ensure representative training and testing sets, the image quality and presence of calcium in each dataset was visually assessed by a radiologist with three years experience in cardiac CT. Image quality was classified as good (defined as absence of any image-degrading artifacts related to motion and noise), moderate (presence of artifacts but evaluation possible with moderate confidence) or poor (presence of image-degrading artifacts and evaluation possible with low confidence). Presence of calcium was scored as absent, modest (presence of calcified spots) or severe (presence of extensive calcifications). Based on these scorings the data was distributed equally over the sets with training and testing data. The patient and scan parameters were assessed by the radiologist to be representative for the clinical situation. Table 1 and 2 describe the distribution of image quality and the presence of calcium in the datasets.

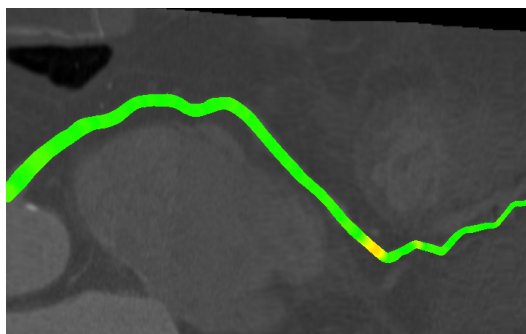


Figure 1: An example of the color-coded curved-planar-reformatted images used to detect possible annotation errors.

3.2 Reference standard

We define the centerline of a coronary artery in a CTA scan as a curve that passes through the center of gravity of the lumen in each cross-section. Furthermore, we define the start point of a centerline as the center of the coronary ostium (i.e. the point where the coronary artery originates from the aorta), and the end point as the most distal point where the artery is still distinguishable from the background. The centerline is smoothly interpolated if the artery is partly indistinguishable from the background, e.g. in case of a total occlusion or imaging artifacts.

This definition was used by three trained observers to annotate in all 32 datasets the centerlines of four coronary arteries, namely the RCA, LAD, LCX and a large side branch of one of the main coronary arteries. First, the observers annotated points along the center of the lumen. In a second step, the observers specified the radius of the lumen at least every 5 mm, where the radius was chosen such that the enclosed area of the annotated circle matched the area of the lumen. Subsequently, the paths of the three observers were combined to one centerline per vessel using a mean shift algorithm for open curves. The centerlines were averaged while taking into account the possibly varying accuracy of the observers. This is done by jointly estimating the reference standard and the accuracy of the observers [14].

After creating this first weighted average, the observers compared their centerlines with this average centerline. This comparison was performed utilizing curved-planar-reformatted images displaying the annotated centerline color-coded with the distance to the reference standard and vice-versa (see Figure 1). The observers used these images to detect and subsequently correct possible annotation errors.

The corrected centerlines were subsequently used to create the reference standard, using the same mean shift algorithm. The radius at every point of the reference standard was calculated by averaging the annotated radii of the observer paths. Note that the uncorrected centerlines were used to calculate the inter-observer variability. The points where for the first time the centerlines of two observers lie within the radius of the reference standard when traversing over this centerline from respectively the start to the end or vice versa were designated as the start- and end point of the reference standard.

3.3 Evaluation

Because coronary artery tracking algorithms can be used for different applications, different evaluation measures may apply. In the evaluation we discern between tracking capability and tracking accuracy. Tracking capability is assessed using three overlap measures and three distance measures are used to determine the accuracy of the centerline tracking. For each of these measures a score is calculated by relating the mea-

measures to the inter-observer variability. These scores range from 100 to 0: 100 points implies that the result of the method is perfect, 50 points implies that the performance of the method is similar to the inter-observer variability and 0 points implies a complete failure. The exact mechanism for calculating the inter-observer variabilities and scores is explained in appendix A.

Definition of true positive, false positive and false negative points

The evaluation measures are based on a point-to-point correspondence between the reference standard and the evaluated centerline. This correspondence is determined as described in [14].

Quality measures for semi-automatic centerlines are in this study based on a labeling of points on the centerlines as true positive, false negative or false positive. This labeling, in its turn, is based on the correspondence between the reference standard centerline and an evaluated centerline.

A point of the reference standard is marked as true positive TPR_{ov} if the distance to at least one of the connected points on the evaluated centerline is less than the local radius and false negative FN_{ov} otherwise.

A point on the evaluated centerline is marked as true positive TPM_{ov} if there is at least one connected point on the reference standard at a distance less than the radius defined at that reference point, and it is marked as false positive FP_{ov} otherwise. See also Figure 2 for a schematic explanation of these terms and the terms mentioned in the next section.

Overlap measures

Three different overlap measures are used in our evaluation framework.

Overlap (OV) represents the ability to track the complete vessel annotated by the observers. It is defined as:

$$OV = \frac{TPM_{ov} + TPR_{ov}}{TPM_{ov} + TPR_{ov} + FN_{ov} + FP_{ov}}. \quad (1)$$

Overlap until first error (OF) is the ratio of the number of true positive points on the reference before the first error (TPR_{of}) and the total number of reference points ($TPR_{of} + FN_{of}$). The first error is defined as the first FN_{ov} point when traversing from the start of the reference standard to its end.

$$OF = \frac{TPR_{of}}{TPR_{of} + FN_{of}} \quad (2)$$

Overlap with the clinically relevant part of the vessel (OT) gives an indication of how well the method is able to track those parts of the vessels that have a diameter of 1.5 mm or larger and are therefore assumed to be clinically relevant [8]. The point closest to the end of the reference standard with a radius larger than or equal to 0.75 mm is determined. Only points on the reference standard between the start of the reference standard and this point, and points on the (semi-)automatic centerline connected to these reference points are taken into account when defining the true positives (TPM_{ot} and TPR_{ot}), false negatives (FN_{ot}) and false positives (FP_{ot}).

$$OT = \frac{TPM_{ot} + TPR_{ot}}{TPM_{ot} + TPR_{ot} + FN_{ot} + FP_{ot}}.$$

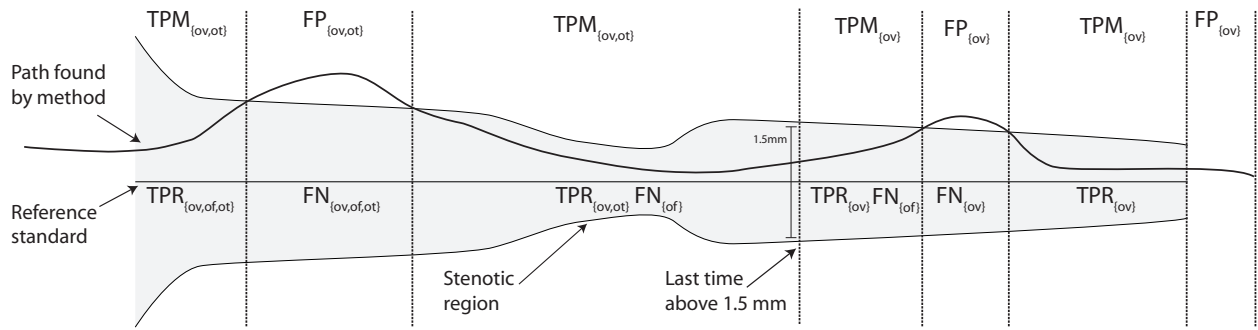


Figure 2: An illustration of the terms used in the overlap measures (section 3.3). The reference standard with annotated radius is depicted in gray. The terms on top of the figure are assigned to points on the centerline found by the method. The terms below the reference standard line are assigned to points on the reference standard.

Accuracy measures

We also use three different measures to assess the accuracy of coronary artery tracking algorithms:

Average distance (AD) is the average distance between the reference standard and the automatic centerline. The average distance is defined as the summed length of all the connections between the two equidistantly sampled centerlines, divided by the number of connections.

Average distance inside vessel (AI) represents the accuracy of tracking, provided that the tracking results are inside the vessel. The measure is calculated in the same way as AD, except that connections that have a length larger than the annotated radius at the connected reference point are excluded.

Average distance to the clinically relevant part of a vessel (AT) represents how well the method can track vessel parts that are clinically relevant, i.e. those parts of the vessels having a diameter larger than or equal to 1.5 mm. The difference with the AD measure is that instead of all connections, only connections that connect TPM_{ot} , TPR_{ot} , FN_{ot} , and FP_{ot} points are averaged.

3.4 Comparing the algorithms

In order to compare different coronary artery tracking algorithms the different measures have to be combined. We do this by ranking the resulting scores of all the methods for each measure and vessel. The average rank of all the 6 (measures) \times 64 (vessels) = 384 ranks defines the quality of each method; the method with the lowest average rank is assumed to be the best.

4 Results for *testing 1* data

Results for the thirteen methods that participated in the competition on the *testing 1* data are shown in Table 3, 4 and 5. Table 3 shows the results for the three overlap measures, Table 4 shows the results for the three accuracy measures, and Table 5 shows the final ranking.

Table 3: Overlap measures

Method	Challenge			OV			OF			OT		
	1	2	3	%	score	rank	%	score	rank	%	score	rank
Friman et al. [4]			×	98.2	86.6	1.78	85.1	77.3	2.48	98.3	84.1	2.07
Zambal et al. [16]	×			86.1	48.9	7.41	60.2	46.8	5.23	88.0	58.9	6.23
Szymczak [11]			×	95.5	70.1	3.83	60.2	47.0	5.84	96.0	63.1	4.64
Bauer and Bischof [1]	×			94.0	56.0	5.78	71.7	56.8	4.61	96.7	67.6	4.77
Krissian et al. [7]		×		90.9	72.1	3.84	70.9	61.7	4.14	91.4	69.7	3.95
Tek et al. [12]	×			87.7	49.5	7.92	50.6	33.8	7.30	89.1	50.3	7.61
Kitslaar et al. [6]	×			79.2	43.6	8.92	57.2	37.3	6.44	81.0	46.9	7.96
Metz et al. [9]			×	91.6	65.5	4.91	57.1	48.1	5.77	92.3	62.9	5.27
Wang and Smedby [15]	×			79.4	41.4	10.16	46.1	31.2	7.58	82.9	46.5	8.86
Dikici et al. [3]		×		90.9	56.7	6.58	23.5	17.6	9.38	91.6	51.2	7.72
Zhang et al. [17]			×	91.4	53.2	7.05	51.2	33.6	7.77	92.3	50.0	7.36
Hernández Hoyos et al. [5]			×	80.2	43.4	9.56	39.3	25.2	8.59	82.1	46.7	9.08
Castro et al. [2]		×		68.9	35.7	11.28	34.3	21.7	8.61	70.8	36.4	10.34

Table 4: Accuracy measures

Method	Challenge			AD			AI			AT		
	1	2	3	mm	score	rank	mm	score	rank	mm	score	rank
Friman et al. [4]			×	0.27	48.4	1.23	0.23	48.6	1.47	0.27	48.6	1.25
Zambal et al. [16]	×			2.59	37.4	3.48	0.28	42.8	2.72	2.33	38.1	3.50
Szymczak [11]			×	0.44	31.1	5.81	0.36	32.0	6.81	0.43	31.3	5.94
Bauer and Bischof [1]	×			0.58	30.2	6.33	0.37	31.5	7.16	0.45	30.9	6.13
Krissian et al. [7]		×		2.77	27.7	7.52	0.39	29.9	8.27	3.01	27.7	7.66
Tek et al. [12]	×			1.92	33.5	5.08	0.34	37.6	4.13	1.66	34.1	4.92
Kitslaar et al. [6]	×			5.39	31.8	5.11	0.31	38.1	3.84	4.71	32.6	5.03
Metz et al. [9]			×	1.30	27.0	7.77	0.48	28.7	8.67	1.29	27.0	8.03
Wang and Smedby [15]	×			4.30	25.8	8.36	0.39	31.9	7.27	3.66	26.9	8.05
Dikici et al. [3]		×		0.92	25.2	9.20	0.48	26.9	9.69	0.90	25.1	9.41
Zhang et al. [17]			×	0.60	23.9	10.02	0.51	25.3	10.78	0.60	24.0	10.11
Hernández Hoyos et al. [5]			×	5.05	24.5	8.91	0.41	29.8	8.19	4.58	25.0	8.94
Castro et al. [2]		×		7.35	15.8	12.19	0.62	20.7	12.02	6.87	16.5	12.05

Table 5: Overall

Method	Challenge			Overlap rank	Accuracy rank	Avg. rank
	1	2	3			
Friman et al. [4]			×	2.07	1.32	1.69
Zambal et al. [16]	×			6.23	3.23	4.73
Szymczak [11]			×	4.64	6.19	5.41
Bauer and Bischof [1]	×			4.77	6.54	5.65
Krissian et al. [7]		×		3.95	7.82	5.88
Tek et al. [12]	×			7.61	4.71	6.16
Kitslaar et al. [6]	×			7.96	4.66	6.31
Metz et al. [9]			×	5.27	8.16	6.71
Wang and Smedby [15]	×			8.86	7.89	8.38
Dikici et al. [3]		×		7.72	9.43	8.57
Zhang et al. [17]			×	7.36	10.30	8.83
Hernández Hoyos et al. [5]			×	9.08	8.68	8.88
Castro et al. [2]		×		10.34	12.09	11.21

5 Discussion

The large number of registrations (~50) and teams that downloaded the data (36) demonstrates the large interest from the community for such a publicly available database and comparative evaluation framework. Despite the successful launch of the database through the MICCAI challenge, there is still room for ample improvement of the evaluation framework. Currently, all available data is acquired using scanners of the same vendor and in the same institution. We expect to cooperate with other institutions around the world to make the collection more varied. Other extensions to the framework, such as lumen segmentations, stenose gradings and calcium scores, can be relevant as well and it would be great if such extensions can be added in the future. After this workshop, the evaluation framework will remain available under the name "Rotterdam Coronary Artery algorithm evaluation framework" and the website (<http://coronary.bigr.nl>) will serve as a continuous benchmark for coronary artery tracking algorithms.

Acknowledgements

This work was supported by the Stichting voor de Technische Wetenschappen (STW) of The Netherlands Organization for Scientific Research (NWO).

A Inter-observer variability and scores

In this study, inter-observer variabilities are used to derive a (relative) score from the absolute overlap and accuracy values. This appendix explains how these inter-observer variabilities are calculated for each of the six evaluation measures and how scores are created from the evaluation measures by relating the measures to the inter-observer variabilities.

A score of 100 points implies that the result of the method is perfect, 50 points implies that the performance of the method is similar to the inter-observer variability, and 0 points implies a complete failure.

A.1 Overlap measures

The inter-observer variabilities for the overlap measures are calculated by comparing the uncorrected paths with the reference standard. The three overlap measures were calculated for each uncorrected path and the true positives, false positives and false negatives for each observer were combined into the inter-observer variability per centerline as follows:

$$\begin{aligned} OV_{io} &= \frac{\sum (TPR_{ov}(i) + TPM_{ov}(i))}{\sum (TPR_{ov}(i) + TPM_{ov}(i) + FP_{ov}(i) + FN_{ov}(i))} \\ OF_{io} &= \frac{\sum TPR_{of}(i)}{\sum (TPR_{of}(i) + FN_{of}(i))} \\ OT_{io} &= \frac{\sum (TPR_{ot}(i) + TPM_{ot}(i))}{\sum (TPR_{ot}(i) + TPM_{ot}(i) + FP_{ot}(i) + FN_{ot}(i))}, \end{aligned}$$

where $i = \{0, 1, 2\}$ indicates the observer.

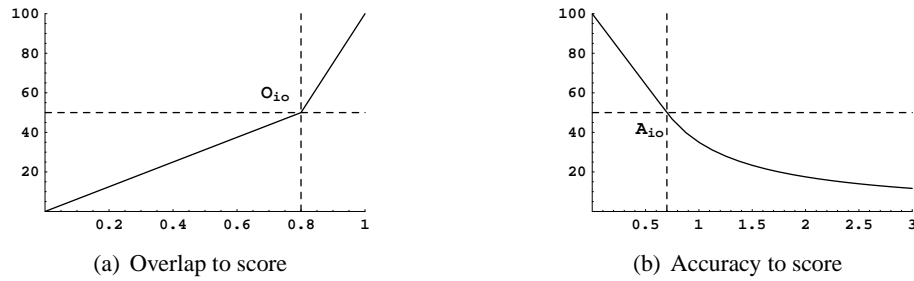


Figure 3: (a) shows how overlap measures are transformed into scores and (b) shows this transformation for the accuracy.

After calculating the inter-observer variabilities, the performance of the method is scored with a measure related to the inter-observer variability. For methods that perform better than the observers the OV, OF, and OT measures are converted to scores by linearly interpolating between 100 and 50 points, respectively corresponding to an overlap of 1.0 and an overlap similar to the inter-observer variability. If the method performs worse than the inter-observer variability the score is obtained by linearly interpolating between 50 and 0 points, with 0 points corresponding to an overlap of 0.0.

$$\text{Score}_O = \begin{cases} (O_m/O_{io}) * 50 & O_m \leq O_{io} \\ 50 + 50 * \frac{O_m - O_{io}}{1 - O_{io}} & O_m > O_{io}, \end{cases}$$

where O_m and O_{io} define the OV, OF or OT performance of respectively the method and the observer. An example of this conversion is shown in Figure 3(a).

A.2 Accuracy measures

The inter-observer variability for the accuracy measures is defined at every point of the reference standard as the expected error that an observer locally makes while annotating the centerline. It is determined as the root mean squared distance between the uncorrected annotated centerline and the reference standard:

$$A_{io}(x) = \sqrt{1/n \sum (d(p(x), p_i))^2}$$

where $n = 3$ (number of observers), and $d(p(x), p_i)$ is the average distance from point $p(x)$ on the reference standard to the connected points on the centerline annotated by observer i .

The tracking accuracy of the method is related per connection to the observer performance. A connection is worth 100 points if the distance to the reference standard is 0 mm and 50 points if the distance is equal to the inter-observer variability at that point. Methods that perform worse than the inter-observer variability are rewarded per connection 50 points times the fraction of the inter-observer variability and the method accuracy.

$$\text{Score}_A(x) = \begin{cases} 100 - 50(A_m(x)/A_{io}(x)) & A_m(x) \leq A_{io}(x) \\ (A_{io}(x)/A_m(x)) * 50 & A_m(x) > A_{io}(x), \end{cases} \quad (3)$$

where $A_m(x)$ and $A_{io}(x)$ define the distance from the method centerline to the reference centerline and the inter-observer accuracy variability at point x . An example of this conversion is shown in Figure 3(b).

The average score over all connections yields the AD score for the centerline, the average over all connections that connect TPR and TPM points yields the AI score, and the AT score is defined as the average score over all connections that connect a point in the clinically relevant parts of the vessel.

References

- [1] C. Bauer and H. Bischof. Edge Based Tube Detection for Coronary Artery Centerline Extraction. *The Insight Journal*, 2008.
- [2] C. Castro, M.A. Luengo-Oroz, A. Santos, and M.J. Ledesma-Carbayo. Coronary Artery Tracking in 3D Cardiac CT Images using Local Morphological Reconstruction Operators. *The Insight Journal*, 2008.
- [3] E. Dikici, T. O'Donnell, L. Grady, R. Setser, and R.D. White. Coronary Artery Centerline Tracking using Axial Symmetries. *The Insight Journal*, 2008.
- [4] O. Friman, C. Kühnel, and H. Peitgen. Coronary Artery Centerline Extraction using Multiple Hypothesis Tracking and Minimal Paths. *The Insight Journal*, 2008.
- [5] M. Hernández-Hoyos, M.A. Zuluaga, M. Lozano, J.C. Prieto, P.C. Douek, I.E. Magnin, and M. Orkisz. Coronary Centerline Tracking in CT Images with use of an Elastic Model and Image Moments. *The Insight Journal*, 2008.
- [6] P. Kitslaar, M. Frenay, E. Oost, J. Dijkstra, B. Stoel, and J.H.C. Reiber. Automated Coronary Tree Segmentation and Path line Detection in MSCT. *The Insight Journal*, 2008.
- [7] K. Krissian, H. Bogunovic, J.M. Pozo, M.C. Villa-Uriol, and A.F. Frangi. Minimally Interactive Knowledge-based Coronary Tracking in CTA using a Minimal Cost Path. *The Insight Journal*, 2008.
- [8] S. Leschka, H. Alkadhi, A. Plass, L. Desbiolles, J. Grünenfelder, B. Marincek, and S. Wildermuth. Accuracy of MSCT Coronary Angiography with 64-Slice Technology: First Experience. *Eur Heart J*, 26(15):1482–1487, Aug 2005.
- [9] C. Metz, M. Schaap, T. van Walsum, and W. Niessen. Two Point Minimum Cost Path Approach for CTA Coronary Centerline Extraction. *The Insight Journal*, 2008.
- [10] D. Scharstein and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47:7–42, 2002.
- [11] A. Szymczak. Vessel Tracking by Connecting the Dots. *The Insight Journal*, 2008.
- [12] H. Tek, M. A. Gulsun, S. Laguitton, L. Grady, D. Lesage, and G. Funka-Lea. Automatic Coronary Tree Modeling. *The Insight Journal*, 2008.
- [13] B. van Ginneken, T. Heimann, and M. Styner. 3D Segmentation in the Clinic: A Grand Challenge. In *3D Segmentation in the Clinic: A Grand Challenge*, pages 7–15, 2007.
- [14] T. van Walsum, M. Schaap, C.T. Metz, A.G. van der Giessen, and W.J. Niessen. Averaging Center Lines: Mean Shift on Paths. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2008*, 2008.

-
- [15] C. Wang and Ö. Smedby. An Automatic Seeding Method For Coronary Artery Segmentation and Skeletonization in CTA. *The Insight Journal*, 2008.
- [16] S. Zambal, J. Hladuvka, A. Kanitsar, and K. Bühler. Shape and Appearance Models for Automatic Coronary Artery Tracking. *The Insight Journal*, 2008.
- [17] Y. Zhang, K. Chen, and S. Wong. 3D Interactive Centerline Extraction. *The Insight Journal*, 2008.