



## Evaluation framework for carotid bifurcation lumen segmentation and stenosis grading

K. Hameeteman<sup>a,\*</sup>, M.A. Zuluaga<sup>b,c</sup>, M. Freiman<sup>d</sup>, L. Joskowicz<sup>d</sup>, O. Cuisenaire<sup>e</sup>, L. Flórez Valencia<sup>f</sup>, M.A. Gülsün<sup>g</sup>, K. Krissian<sup>h</sup>, J. Mille<sup>i</sup>, W.C.K. Wong<sup>k</sup>, M. Orkisz<sup>b</sup>, H. Tek<sup>g</sup>, M. Hernández Hoyos<sup>c</sup>, F. Benmansour<sup>j</sup>, A.C.S. Chung<sup>k</sup>, S. Rozie<sup>l</sup>, M. van Gils<sup>l</sup>, L. van den Borne<sup>l</sup>, J. Sosna<sup>m</sup>, P. Berman<sup>m</sup>, N. Cohen<sup>m</sup>, P.C. Douek<sup>n,b</sup>, I. Sánchez<sup>n</sup>, M. Aissat<sup>n</sup>, M. Schaap<sup>a</sup>, C.T. Metz<sup>a</sup>, G.P. Krestin<sup>l</sup>, A. van der Lugt<sup>l</sup>, W.J. Niessen<sup>a,o</sup>, T. van Walsum<sup>a</sup>

<sup>a</sup> Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics, Erasmus MC, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

<sup>b</sup> CREATIS, Université de Lyon, Université Lyon 1, INSA-Lyon, CNRS UMR5220, INSERM U630, F-69621 Villeurbanne, France

<sup>c</sup> Grupo IMAGINE, Grupo de Ingeniería Biomédica, Universidad de los Andes Bogotá, Colombia

<sup>d</sup> Computer-Aided Surgery and Medical Image Processing Laboratory, School of Engineering and Computer Science, The Hebrew University of Jerusalem, Israel

<sup>e</sup> Philips Healthcare, Medisys Research Laboratory, 33 rue de Verdun, B.P. 313, F-92156 Suresnes Cedex, France

<sup>f</sup> Pontificia Universidad Javeriana, Facultad de Ingeniería, Departamento de Ingeniería de Sistemas, Bogotá, Colombia

<sup>g</sup> Imaging and Visualization Department, Siemens Corporate Research, Princeton, NJ 08540, United States

<sup>h</sup> GIMET, Dept. de Informática y Sistemas, Univ. Las Palmas de Gran Canaria, Spain

<sup>i</sup> Université de Lyon, CNRS, Université Lyon 1, LIRIS, UMR5205, F-69622, France

<sup>j</sup> CEREMADE, UMR CNRS 7534, Université Paris Dauphine, F-75775, France

<sup>k</sup> Lo Kwee-Seong Medical Image Analysis Laboratory, Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong

<sup>l</sup> Department of Radiology, Erasmus MC, Rotterdam, The Netherlands

<sup>m</sup> Hadassah Hebrew University Medical Centre, Jerusalem, Israel

<sup>n</sup> Hôpital Louis Pradel, Bron, France

<sup>o</sup> Imaging Science and Technology, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands

### ARTICLE INFO

#### Article history:

Received 7 May 2010

Received in revised form 4 February 2011

Accepted 10 February 2011

Available online 17 February 2011

#### Keywords:

Evaluation framework

Carotid

CTA

Lumen segmentation

Stenosis grading

### ABSTRACT

This paper describes an evaluation framework that allows a standardized and objective quantitative comparison of carotid artery lumen segmentation and stenosis grading algorithms. We describe the data repository comprising 56 multi-center, multi-vendor CTA datasets, their acquisition, the creation of the reference standard and the evaluation measures. This framework has been introduced at the MICCAI 2009 workshop *3D Segmentation in the Clinic: A Grand Challenge III*, and we compare the results of eight teams that participated. These results show that automated segmentation of the vessel lumen is possible with a precision that is comparable to manual annotation. The framework is open for new submissions through the website <http://cls2009.bigr.nl>.

© 2011 Elsevier B.V. All rights reserved.

### 1. Introduction

Cardiovascular diseases account for 30% of all deaths worldwide (Abegunde et al., 2005). Atherosclerosis, a disease of the vessel wall, is the major cause of cardiovascular diseases such as stroke (Frostegård, 2005). Atherosclerosis may lead to stenosis (luminal narrowing), but it is also possible for atherosclerotic plaque to build up without narrowing the lumen (Glagov et al., 1987).

Cardiovascular imaging is an important means to monitor and quantify the state of the vessel wall and lumen.

Manual lumen segmentation and stenosis quantification is laborious and suffers from inter and intra rater variabilities (Scherl et al., 2007). Consequently much work has been performed on (semi)-automated cardiovascular image processing, of which the majority focuses on lumen quantification and assessment of the severity of luminal stenosis in various imaging modalities.

In 2004, Kirbas et al. (Kirbas and Quek, 2004) published a review paper on vessel lumen segmentation algorithms. More recently, an extensive review paper on the same subject was published by Lesage et al. (Lesage et al., 2009). Both papers provide an overview of the currently published vessel segmentation

\* Corresponding author. Tel.: +1 31107044124; fax: +1 31107044722.

E-mail address: [K.Hameeteman@Erasmusmc.nl](mailto:K.Hameeteman@Erasmusmc.nl) (K. Hameeteman).

URL: <http://www.bigr.nl/people/reinhardhameeteman> (K. Hameeteman).

methods, categorized according to the technology (Kirbas and Quek, 2004) or the combination of method characteristics such as the used vessel models, image features and extraction schemes (Lesage et al., 2009). However, as explicitly indicated by Lesage et al., direct performance comparisons of different approaches are lacking because of the wide range of applications, and more generally, because of the lack of standard image databases and validation criteria for most vascular segmentation applications. Moreover, many of the reviewed algorithms are not publicly available which hampers an objective and fair comparison by third parties.

The goal of the framework described in this paper (CLS2009 framework) is to provide a standardized evaluation framework for carotid artery lumen segmentation and stenosis grading in Computed Tomography Angiography (CTA). We focus on the carotid bifurcation (Fig. 1), where the Common Carotid Artery (CCA) splits into the External Carotid Artery (ECA) and Internal Carotid Artery (ICA). The latter is one of the major blood supplying arteries to the brain. In 25% of all stroke patients, the stroke is caused by atherosclerotic disease in the carotid artery bifurcation. As shown by the North American Symptomatic Carotid Endarterectomy Trial (NASCET Collaborators, 1991) and the European Carotid Surgery Trial (ECST Collaborators, 1998) the stenosis grade is an important clinical measure in deciding whether or not to perform carotid endarterectomy (surgical procedure to remove the atherosclerotic plaque from the vessel). Carotid endarterectomy is indicated in case of a stenosis of 50–99% (Rothwell et al., 2003). So the accuracy of stenosis analysis influences clinical decision making. Traditionally stenoses around the carotid bifurcation have been assessed with intra-arterial catheter angiography, *i.e.* Digital Subtraction Angiography (DSA), which is still considered the gold standard. DSA has a 0.3–1% risk of neurological deficits such as stroke (Vaughn et al., 1992; Hankey et al., 1990) which makes the use of less invasive diagnostic imaging preferable. CTA is a good, less invasive alternative to DSA for the assessment of stenosis in the carotid bifurcation (Binaghi et al., 2001; Koelemay et al., 2004).

Besides the assessment of the stenosis caused by atherosclerotic plaque, imaging also allows for the evaluation of the atherosclerotic plaque itself. Much research is currently focused on the question whether specific plaque features like composition or morphology are associated with an increased risk for clinical events. However, in current practice the degree of stenosis is still the only biomarker which influences clinical decision making.

Besides CTA, MRI and US are able to visualize the plaque. Currently it is not clear which technique is most suitable in providing clinical useful information on both stenosis and the atherosclerotic plaque.

Our choice for lumen segmentation of the carotid bifurcation is motivated by the clinical relevance of this structure. Additionally, the carotid is a medium sized vessel which makes it suitable as a test case for a wide range of vessel segmentation algorithms, while still having the typical challenges in segmenting diseased vessels, such as calcifications and severe stenosis. Also the presence of a bifurcation and nearby bone and veins imposes additional challenges on the segmentation. Moreover, lumen segmentation is a first step in the assessment of stenosis, which is the clinical relevant parameter that is assessed in the CLS2009 framework.

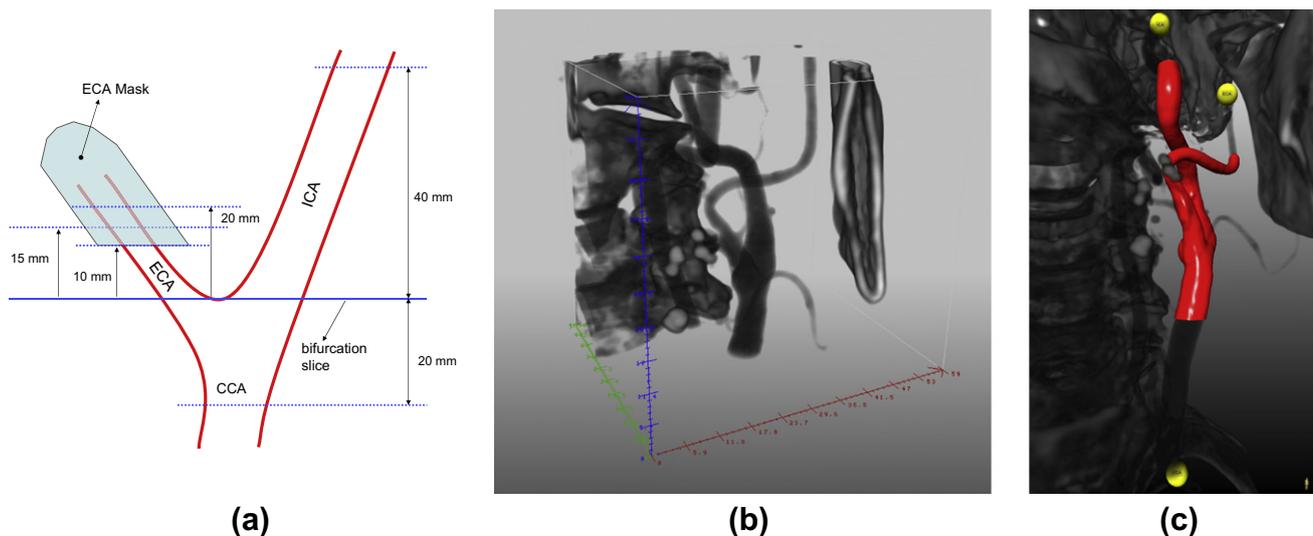
The remainder of this paper is organized as follows: in Section 2 previous work is discussed, followed in Section 3 by a description of the CLS2009 framework for the segmentation of the carotid artery lumen and its stenosis grading. Section 4 gives a description of the first use of the framework during the MICCAI workshop, including a short description of the methods that were tested and the results of these methods as produced by the framework. This is followed by some concluding remarks in Section 5. This paper is an extended version of the editorial of the carotid challenge workshop proceedings (Hameeteman et al., 2009a).

## 2. Previous work

There is a growing number of initiatives that set up a publicly available data repository and standardized evaluation framework. This demonstrates an increasing interest in standardized evaluation and the possibility to compare methods to each other. In this section we give a few examples and we briefly discuss some published lumen segmentations and stenosis grading methods.

In the field of computer vision the following frameworks can be used: the Range Image Segmentation Comparison (Hoover et al., 1996), the Retrospective Image Registration Evaluation Project (West et al., 1997), the Berkeley Segmentation Dataset and Benchmark (Martin et al., 2001), and the Middlebury Stereo Vision evaluation (Scharstein and Szeliski, 2002).

Several workshops involving the setup of an evaluation framework in the field of medical imaging, have been organized at the MICCAI and SPIE conferences. Reports on some of these



**Fig. 1.** Region around carotid bifurcation: (a) schematic depiction of the region of interest; (b) a rendering of this region for one of the datasets; (c) visualization of a dataset with the three initialization points and the reference segmentation.

frameworks have appeared recently (Heimann et al., 2009; Schaap et al., 2009; Niemeijer et al., 2010; Van Ginneken et al., 2010). More initiatives can be found at the website <http://www.grand-challenge.org/>.

The carotid bifurcation segmentation and stenosis grading framework in CTA images (CLS2009 framework) was initially presented at the *3D Segmentation in the Clinic: A Grand Challenge III* workshop during the MICCAI 2009 conference in London. There are two challenges that are related to CLS2009: the Rotterdam Coronary Artery Algorithm Evaluation Framework which does not address lumen segmentation, but centerline extraction and the EXACT09 framework, which compares submitted segmentations of the airways to each other and generates a reference standard from these submitted segmentations. In contrast to the CLS2009 framework the EXACT09 framework does not have a gold standard obtained from manual segmentations.

Some CTA vessel lumen segmentation methods have been evaluated on carotid arteries around the bifurcation (Manniesing et al., 2007; Baltaxe et al., 2007; Cuisenaire et al., 2008). All these methods use different evaluation measures, and different datasets. Manniesing reports a Dice similarity index (Dice, 1945) of 0.77 for 14 out of 20 carotid arteries for which a path through the carotid lumen could be obtained. Algorithm parameters were trained on a separate set of 10 CTA datasets. Milwer reports a sensitivity of 0.85 and a positive predictive value of 0.80 (equivalent to a Dice similarity of 0.82) on a set of 65 2D cross-sections of 13 CTA datasets. The algorithm parameters were determined empirically. Cuisenaire performs a visual assessment on failure or success by an expert and reports a 93% success score on extraction of six modeled head and neck vessels in 28 CTA datasets.

Results of semi-automatic stenosis grading methods for carotid arteries in CTA images have been reported by Berg et al. (2005), Scherl et al. (2007) and Wintermark et al. (2008). The level of automation varies, and only Scherl gives an extensive description of the lumen segmentation and stenosis grading algorithm. Different evaluation measures are used, although they all report an agreement measure, correlation (Berg), difference (Scherl), regression and Bland Altman analysis (Wintermark), with DSA (Berg) or manual (Scherl and Wintermark) measurements. Both Scherl and Wintermark use the minimal diameter as a quantity to measure the stenosis, but neither of them explicitly defines this measure. The number of included patients for the evaluation differs widely: Berg uses 36 CTA datasets, Scherl 10, and Wintermark 125.

### 3. Evaluation framework

The CLS2009 framework comprises a publicly accessible data repository, a set of standardized evaluation measures and an online evaluation system. This section starts with a description of the segmentation tasks, followed by a description of the datasets and their acquisition protocols. Next we describe the procedure that was used to generate manual segmentations and stenosis gradings, followed by a description of the used procedure to merge the different manual segmentations into a reference standard. The section concludes with the description of the evaluation measures.

#### 3.1. Lumen segmentation and stenosis grading

The evaluation framework focuses on the CCA and ICA (see Fig. 1), the clinically most relevant parts of the carotid bifurcation. A small part of the ECA is also included, to allow evaluation of the segmentation at the location where the ECA bifurcates from the ICA.

The segmentation task focuses on the lumen of the CCA, ICA and the ECA around the carotid bifurcation in a CTA dataset. The exact

region of which the lumen must be segmented is defined around the bifurcation slice, which is defined as the first (caudal to cranial) slice where the lumen of the carotid artery appears as two separate lumens: the lumen of the ICA and the lumen of the ECA. The segmentation must include the CCA, starting at least 20 mm caudal of the bifurcation slice, the ICA, up to at least 40 mm cranial of the bifurcation slice, and the ECA, up to between 10 and 20 mm cranial of the bifurcation slice, see also Fig. 1.

The performance measures are computed in the region of interest. The bifurcation slice is not revealed to the participants, who must ensure that their segmentation at least includes this region. Our definition of the bifurcation slice and the region of interest should be sufficient to determine a suitable region of interest for the segmentations.

The segmented lumen of the ECA should be cut between 10 and 20 mm cranial of the bifurcation slice. To allow for some flexibility in cutting of the ECA lumen, the region around the ECA between 10 and 20 mm cranial of the bifurcation slice is a “masked” region, where the evaluation measures are not evaluated, see also Fig. 1a.

The lumen segmentation must be represented as a partial volume segmentation, *i.e.* an image where each voxel value represents the occupancy of the voxel by the vessel lumen: a value of 0 means no lumen present, and a value of 1 means fully occupied with lumen.

The second use of the CLS2009 framework is the evaluation of stenosis gradings for the ICA. Two stenosis grades must be determined for each ICA: an area-based and a diameter-based stenosis grade.

We use the following NASCET-like (NASCET Collaborators, 1991) definition for the area-based stenosis grade  $S_a$ :

$$S_a = 100\% \times \left(1 - \frac{a_m}{a_r}\right) \quad (1)$$

where  $a_m$  is the minimal cross-sectional area along the CCA and ICA, and  $a_r$  the average cross-sectional area over a distal reference part of the ICA. For the diameter-based stenosis grade  $S_d$  we similarly use:

$$S_d = 100\% \times \left(1 - \frac{d_m}{d_r}\right) \quad (2)$$

where  $d_m$  and  $d_r$  are the minimal and average reference cross-sectional diameter, respectively. We define the diameter of a cross-section as the shortest straight line that divides the contour in two equally-sized areas. Using this definition, the diameter of non circular or non elliptical cross sections is well defined. See Fig. 2 for examples of the diameter for several contour shapes. The reference measure in Eqs. (1) and (2) should be based on a vessel section, distal to the location of the stenosis and should reflect the normal luminal diameter or area.

We distinguish two categories for the lumen segmentation and the stenosis grading task: a category for fully automatic methods,

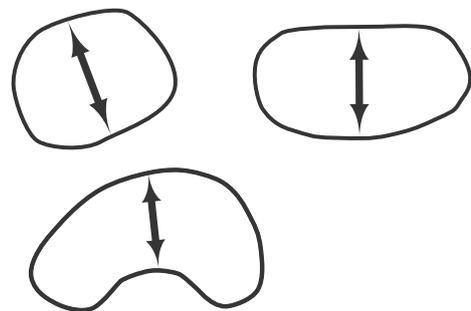


Fig. 2. Examples of our definition of the diameter of a cross-sectional contour.

which do not require user interaction, and one for semi-automatic methods where three initialization points may be used. To eliminate dependency on these initialization points, they are incorporated in the available data repository. The image data for both tasks (lumen segmentation and stenosis grading) is the same: the CTA dataset (including header information such as voxel size and world coordinate system). The three points are located within the carotid artery proximal and distal to the above defined region of interest (see Fig. 1c):

1. A point in the CCA, at the level of the cranial side of the thyroid gland.
2. A point in the ICA, just before the artery enters the skull base.
3. A point in the ECA, where the artery is close to the mandible.

### 3.2. CTA data

The dataset repository consists of 56 CTA datasets from three different medical centers: the Erasmus MC (Rotterdam, The Netherlands), Hôpital Louis Pradel (Bron, France) and the Hadassah Hebrew University Medical Centre (Jerusalem, Israel). All datasets were routinely acquired CTA datasets of the carotid bifurcation. The medical center in which a dataset was acquired is made known to the participants. Example slices of the datasets from the three research centers are shown in Fig. 3.

The datasets were selected such that they contain a large range of stenosis degrees and all stenosis categories (as listed in Table 1) contained an equal number of datasets. The initial stenosis degrees were determined by visual inspection. Creating the reference standard and measuring the stenosis based on lumen area slightly changed this distribution. The distribution, based on the measurements from the reference standard, is shown in Table 1.

The details of the scanning protocols for each of the medical centers are provided in Sections 3.2.1, 3.2.2, 3.2.3. The CTA scanning parameters are summarized in Table 2.

#### 3.2.1. Erasmus MC protocol

The CTA data from Erasmus MC was acquired on a 16-row CT scanner (Sensation 16 – Siemens Medical Solutions, Forchheim, Germany) with a standard scan protocol using the following parameters: 120 kV, 180 mAs, collimation  $16 \times 0.75$  mm, table feed per rotation 12 mm, pitch 1.0, rotation time 0.5 s and scan time 10–14 s. The CTA scan range is from the ascending aorta to the intra cranial circulation (2 cm above the sella turcica). All patients received 80 ml contrast material (Iodixanol 320 mg/ml, Visipaque Amersham Health, Little Chalfont, UK), followed by 40 ml saline bolus chaser, both with an injection rate of 4 ml/s. Synchronization between the passage of contrast material and data

**Table 1**

Distribution of the datasets over the five stenosis categories based on the reference standard.

Stenosis category	Stenosis degree (%)	Number of datasets
0	0	12
1	0–30	10
2	30–50	6
3	50–70	10
4	70–99	18

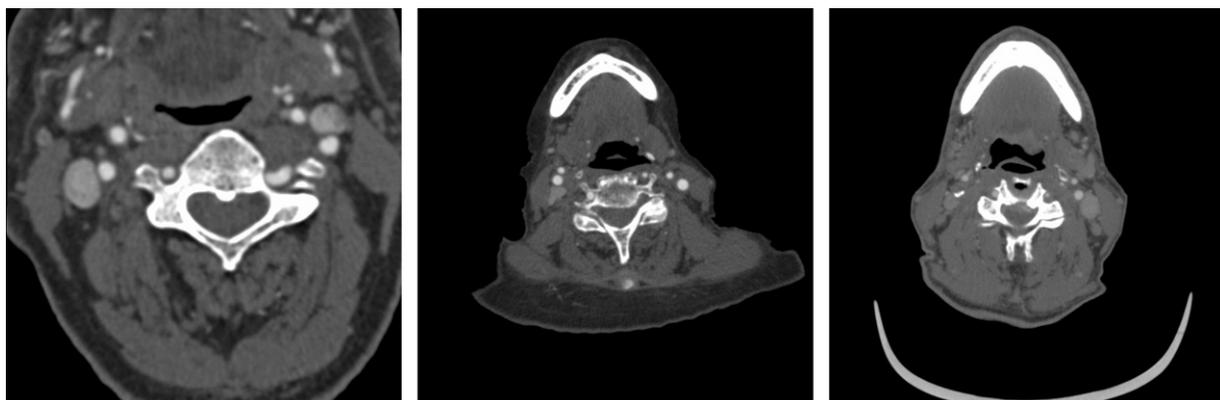
acquisition was achieved by real time bolus tracking at the level of the ascending aorta. The trigger threshold was set at an increase in attenuation of 75 Hounsfield Units (HU) above baseline attenuation (approximately 150 HU in absolute value). Image reconstructions were made with in-plane pixel sizes of  $0.23 \times 0.26 \times 0.23\text{--}0.26$  mm<sup>2</sup>, matrix size of  $512 \times 512$  (real in-plane resolution  $0.6 \times 0.6$  mm), slice thickness of 1.0 mm, increment of 0.6 mm and with an intermediate reconstruction kernel (B30f).

#### 3.2.2. Hadassah protocol

The CTA data from the Hadassah Hebrew University Medical Centre was acquired on a 64-row CT scanner (Brilliance 64 – Philips Healthcare, Cleveland, OH) with a standard scan protocol using the following parameters: 120 kV, 251 mAs, collimation  $64 \times 0.625$  mm, pitch 1.20, rotation time 0.75 s and scan time 7.30 s. The CTA scan range was from the ascending aorta to the intra cranial circulation (2 cm above the sella turcica). All patients received 75 ml contrast material (Iopamiro, Bracco Diagnostics, Milano, Italy), with an injection rate of 3.5 ml/s. Image reconstructions were made with in-plane pixel sizes of  $0.55 \times 0.55$  mm<sup>2</sup>, matrix size of  $512 \times 512$ , slice thickness of 1.0 mm, increment of 0.5 mm and with an intermediate reconstruction kernel (B).

#### 3.2.3. Louis Pradel protocol

The CTA data of Hôpital Louis Pradel was acquired on a 64-row CT scanner (Brilliance 64 – Philips Healthcare, Cleveland, OH) with a standard scan protocol using the following parameters: 120 kV, 300 mAs, collimation  $52 \times 1.5$  mm, rotation time 0.35 s and scan time 10–14 s. The CTA scan range was from ascending aorta to the intra cranial circulation (2 cm above the sella turcica). All patients received 80 ml contrast material (Iomeron 4000 mg/ml, BRACCO, Milano, Italy) followed by 40 ml saline bolus chaser, both with an injection rate of 4 ml/s. Synchronization between the passage of contrast material and data acquisition was achieved by real time bolus tracking at the level of the ascending aorta. The trigger threshold was set at an increase in attenuation of 75 HU above baseline attenuation. Image reconstructions were made with



**Fig. 3.** Examples of a slice of the CTA datasets. From left to right a slice from Erasmus MC, the Hadassah Hebrew University Medical Centre and Hôpital Louis Pradel.

**Table 2**

Overview of scanning parameters of CTA datasets, EMC = Erasmus MC, Hd = Hadassah, LP = Louis Pradel.

	Scanner	In plane voxels	Slices	Pixel size (mm)	Z-spacing (mm)	Slice thick. (mm)	Kernel
EMC	Sensation 16	512 × 512	395–579	0.23–0.26	0.6	1	B30f
Hd	Brilliance 64	512 × 512	750	0.55	0.5	1	B
LP	Brilliance 64	512 × 512	636–827	0.414–0.547	0.45	0.9	B

in-plane pixel sizes of  $0.414\text{--}0.547 \times 0.45 \text{ mm}^2$ , matrix size of  $512 \times 512$  (real in-plane resolution  $0.6 \times 0.6 \text{ mm}$ ), slice thickness of  $0.9 \text{ mm}$ , increment of  $0.45 \text{ mm}$  with an intermediate reconstruction kernel (B).

### 3.2.4. Datasets for training and testing

Fifteen of the 56 datasets are made available for training. The reference standard for these training datasets is also available for download. The training datasets were chosen such that they were equally distributed over the five stenosis categories. The reference standard of the remaining 41 datasets is not made available for download. These datasets are used for testing the performance of the algorithms. The distribution of the number of datasets over the training and testing sets and over the different centers is shown in Table 3.

### 3.3. Reference standard

The reference standard was created by averaging three manual segmentations of the lumen in the CTA datasets. This section describes the manual annotation process, the creation of a partial volume representation from each observers' annotations and the averaging of the observers' segmentations to obtain the reference standard.

#### 3.3.1. Manual annotations

Three different observers annotated the carotid lumen boundary and graded the stenosis in the ICA for each dataset. Each contributing center performed the annotations on the data it provided, thus the three observers for each of the three centers were different and there were nine observers in total.

The manual annotations for the lumen segmentation and stenosis grading were performed with a custom made tool, based on MeVisLab (MeVis Research, Bremen, Germany). The annotation procedure was as follows:

1. The bifurcation point was identified and manually selected.
2. Positions along the centerlines for both the ICA and ECA were clicked, starting in the CCA, 20 mm caudal of the bifurcation slice, and extending to 40 mm cranial of the bifurcation slice for the ICA and 20 mm cranial of the bifurcation slice for the ECA. This way two centerlines were defined: CCA-ICA and CCA-ECA. The part in the CCA thus was annotated twice.
3. Resampled centerlines were used to generate Curved Multi Planar Reformatted images (CMPRs), in which longitudinal contours were drawn for three different orientations (each  $60^\circ$

apart) of the CMPRs. Fig. 4 shows an example of one of the three orientations with two longitudinal contours.

4. Cross-sectional contours orthogonal to the centerline were created at 1 mm intervals along the centerline. These contours were created by fitting a Catmull–Rom spline (Catmull and Rom, 1974) through the (six) positions where the cross-sectional plane intersects the longitudinal contours. Fig. 5 shows a visualization of a subset of these contours together with the longitudinal contours, the centerline and a cross-sectional image plane that is perpendicular to the centerline. The longitudinal contours of Figs. 4 and 5 are the same. The contours were edited and updated if they did not match the luminal boundaries.

The contours were drawn on a display with fixed window level settings (center = 176 HU, width = 800 HU).

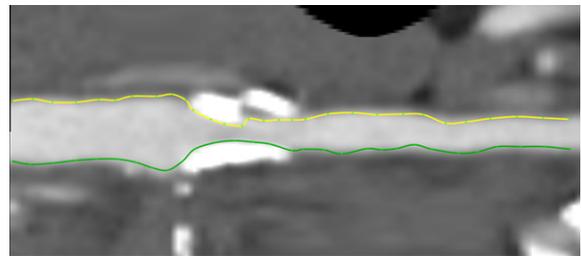


Fig. 4. Example of two longitudinal contours drawn on a CMPR image.

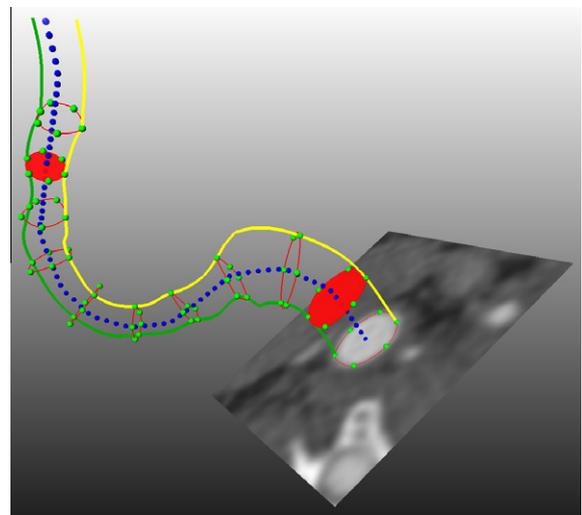


Fig. 5. Example of two longitudinal contours (yellow and green, which correspond to the ones shown in Fig. 4) with cross-sectional contours that are spline interpolations of the six intersection points (green dots) of longitudinal contours with the plane (shown image plane) that is perpendicular to the centerline (dark blue dots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**

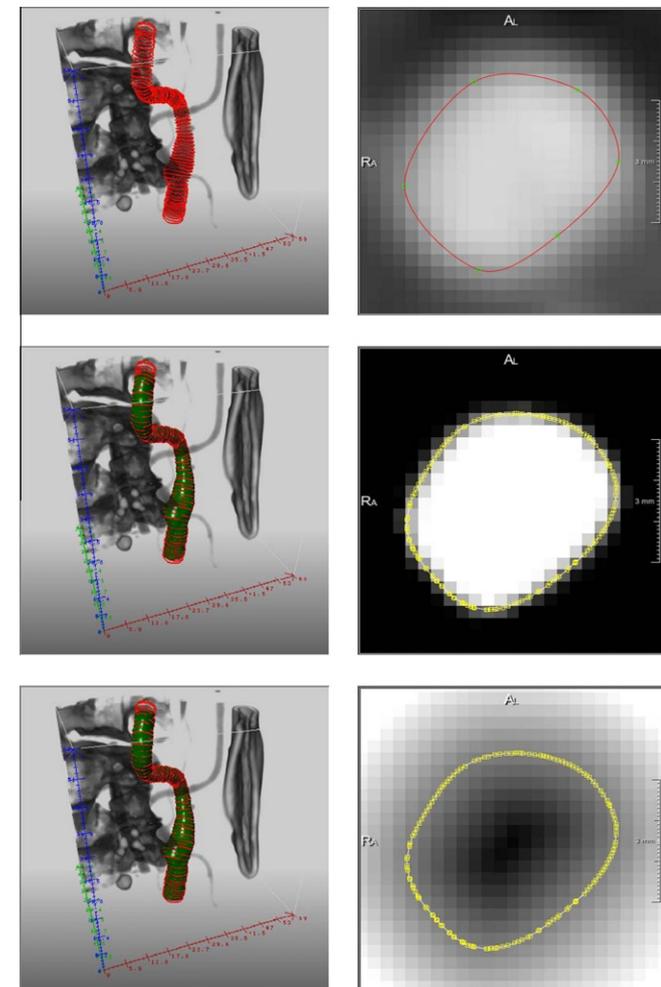
Datasets per center, and distribution over training and testing sets.

Center	Training	Testing	Total
Erasmus MC	9	27	36
Hadassah	3	7	10
Louis Pradel	3	7	10
Total (#)	15	41	56

### 3.3.2. Observers' contours processing

The contours of each observer were separately processed to obtain the partial volume segmentations (see Fig. 6):

1. The contours (both for the ICA and the ECA) were converted to partial volume segmentations, using a Thin Plate Spline interpolation between the contour points (Turk and O'Brien, 1999).
2. Signed distance maps (both for the ICA and the ECA) were generated from the partial volume segmentations of the ICA and ECA.
3. The ICA and ECA signed distance maps were combined, where the resulting value was the signed minimum absolute value if both distances had the same sign and the minimum value if the distances differed in sign. Thus a voxel that is inside the vessel in either the ICA or the ECA segmentation is also part of the vessel in the combined segmentation and a voxel outside the vessel in both segmentations will get the minimum distance of both distance maps. The result is a signed distance map for the complete bifurcation.
4. The partial volume (pv) segmentations of the ICA and ECA were also combined by taking the voxel-wise maximum to obtain the partial volume segmentation of the bifurcation. In this way we effectively combine the two parts, since at least the maximum



**Fig. 6.** Processing of observer annotations, left column shows a 3D visualization and right column a 2D visualization. From top to bottom: initial contours, partial volume from contours (left: isosurface at 0.5) and signed distance map from partial volume (left: isosurface at 0.0).

of both pv segmentations is part of the combined segmentation. This partial volume segmentation is used to rate the observer in the same way as the contestants' segmentations.

All signed distance maps used world distances in mm.

### 3.3.3. Combining observers' lumen segmentations

The three segmentations of the observers were used to generate the data of the reference standard: bifurcation slice number, evaluation region of interest and the lumen segmentation.

The bifurcation slice was computed by averaging the locations, as indicated by the observers. This slice, and the bounding boxes of the contours, were used to determine the region of interest for the evaluation.

The region of interest was the bounding box of the contours, extended with 15 mm both in x- and in y-direction. The z-range was determined from the reference bifurcation slice, and ranged from 20 mm caudal of the bifurcation slice to 40 mm cranial of the bifurcation slice.

The lumen segmentation of the reference standard contains three representations: a signed distance map, a surface representation and a partial volume segmentation. It also contains the mask for the distal part of the ECA. These lumen representations and the mask were constructed in the following way:

1. The signed distance map was obtained by averaging the observers' signed distance maps, and subsequently performing a signed distance transform on the zero-level set.
2. The surface representation was obtained by determining the isosurface at zero of the reference signed distance map, obtained in the previous step.
3. The partial volume representation was generated from the reference standard signed distance map by interpolating the distance map on super-resolution, and determining the fraction of subvoxels that have a negative distance (*i.e.* that were inside the lumen).
4. For the separate branches of the ICA and ECA average distances maps of the three observers were created. These average maps of the two separate branches were used to create the mask of the distal part of the ECA. This mask contains all voxels that satisfied each of the following three criteria:
  - the voxel is in the 10–20 mm range cranial of the bifurcation, and
  - the ECA signed distance map value of the voxel is less than 2 mm, *i.e.* the voxel is inside or close to the ECA, and
  - the ECA signed distance map value is less than the ICA signed distance map value, *i.e.* the voxel is closer to the ECA than to the ICA.

### 3.3.4. Stenosis values

The ICA and CCA contours were used for stenosis grading. The ECA contours were not used as atherosclerosis and luminal stenosis in this artery is of less clinical relevance. Graphs of the contour area and diameter (see Section 3.1) along the centerline were created, based on the corrected cross-sectional contours.

In the graphs, the longitudinal position of the minimal area or diameter could be selected, after which a default reference segment was shown 20 mm distal to the selected stenosis and 10 mm in length. The position and length of the reference segment could be manually edited, subject to the constraint that the reference area should remain distal of the minimal area location, and should not extend outside the segmented region, *i.e.* beyond 40 mm cranial of the bifurcation slice. The stenosis grade was determined using the values from these graphs. The three observer values for the stenosis were averaged to obtain the reference standard stenosis values.

### 3.3.5. Initialisation points

The three initialization points for the semi-automatic methods (see Section 3.1) were annotated by one of the observers.

## 3.4. Evaluation measures and ranking

### 3.4.1. Lumen segmentation

The partial volume lumen segmentations as supplied by a participant are evaluated using the following three performance measures<sup>1</sup>:

- The Dice similarity index  $D_{si}$ :

$$D_{si} = \frac{2 \times |pv_r \cap pv_p|}{|pv_r| + |pv_p|} \quad (3)$$

where  $pv_r$  and  $pv_p$  are the reference and a participant's partial volumes, respectively, the intersection operation is the voxel-wise minimum operation, and  $|\cdot|$  is the volume, i.e. the integration of the voxel values over the complete image.

- The mean surface distance  $D_{msd}$ :

$$D_{msd} = \frac{1}{2} \times \left( \frac{\int_{S_r} |sdm_p| ds}{A_r} + \frac{\int_{S_p} |sdm_r| ds}{A_p} \right) \quad (4)$$

where  $|sdm_p|$  and  $|sdm_r|$  are the absolute signed distance maps of the reference and a participant's segmentation, respectively,  $S_r$  and  $S_p$  are the lumen boundary surfaces (isosurfaces of the signed distance map at the value 0) and  $A_i$  is the surface area of surface  $S_i$ , i.e.  $A_i = \int_{S_i} ds$ .

- The Hausdorff<sup>2</sup> distance  $D_{hd}$ :

$$D_{hd} = \max \left( \max_{x \in S_r} |sdm_p(x)|, \max_{x \in S_p} |sdm_r(x)| \right) \quad (5)$$

Both distance measures are symmetric, and all measures are only evaluated in the region of interest that is specified in Section 3.1.

### 3.4.2. Stenosis grading

The stenosis grade error is defined as the absolute difference between the reference standard value and the value determined by a participant. Because revealing the (exact) error per dataset also reveals the reference stenosis grades, the stenosis errors are not communicated per dataset, but only per ensemble (training or testing). Providing the ranking per dataset may potentially be used to determine the reference values by using multiple submissions. Hence we also do not show the stenosis ranking per dataset. The final ranking, however, is determined by averaging the (hidden) errors per dataset and stenosis grade (diameter and area).

Stenosis grading is often performed after a lumen segmentation is obtained. Thus, we developed a simple standard stenosis grading algorithm and applied it to all segmentation results. The algorithm takes the partial volume lumen segmentation of the participants as input and determines the desired stenosis measure. The algorithm first determines the vessel surface. Based on this surface a center line is created using the vmtk package (Antiga and Steinman, 2004). Along this centerline the vessel surface is cut at 1 mm intervals with a plane perpendicular to the centerline using the default vtkCutter class (Schroeder et al., 2006). The intersection points are

then used to calculate the cross-sectional area and minimal diameter of the vessel along the centerline. The resulting area and diameter curves are smoothed using a Gaussian kernel with  $\sigma = 3$  mm. The minimal value of these smoothed curves is found, which is defined as the location of the stenosis. Using the found position, the degree of stenosis is then determined from the un-smoothed curve using both the minimum value and the average value of the reference area which is defined as 2 cm distal to the bifurcation and 1 cm long.

The stenosis values of this standard algorithm are determined for each submission where there is a successful lumen submission, but are not used in ranking the stenosis submissions.

### 3.4.3. Ranking

The evaluation measures lead to one performance value for each participant, for each dataset and for each evaluation measure. Per dataset and per evaluation measure a ranking of the participants is made, by ordering them from best to worst and giving a rank according to their position in the list: 1 for the best submission and  $P$  for the worst (with  $P$  the number of participants). If there are no results for a participant for a certain dataset the ranks for all evaluation measures for that dataset are set to the highest value ( $P$ ).

In this way, we obtain  $N \times M$  rankings (for  $N$  datasets and  $M$  measures, with  $N = 41$  for testing,  $M = 3$  for lumen segmentation and  $M = 2$  for stenosis grading). The overall ranking for a participant is obtained by averaging the ranks of all these  $N \times M$  rankings.

We also report the average values of the evaluation measures per participant. The values of evaluation measures for missing or failed datasets are not taken into account in this averaging, so these values represent the results for the datasets that were successfully processed. We also report the number of datasets on which a method was successful.

### 3.4.4. Availability of the framework

The evaluation framework is publicly available through the website: <http://cls2009.bigr.nl>. All CTA datasets, the reference standard for the training set as well as binaries and source code of the evaluation software can be downloaded from this website. This software is the same as which is used by the framework to evaluate the submitted results with the testing data. Participants can use it together with the training set reference standard to determine the evaluation measures without using the website. The framework provides the possibility to submit results. The evaluation measures will be determined for submitted results and, after confirmation by the registrant, included in the ranking of all the methods. The website also provides the possibility to create custom queries on the datasets based on the research center and stenosis grading.

## 4. MICCAI workshop

The evaluation framework was used in a competition setup at the MICCAI 2009 workshop *3D Segmentation in the Clinic: A Grand Challenge III*. Around 100 groups from academia and industry were invited by e-mail to participate in the workshop. Thirty-one teams registered at the website, 21 of which sent in the data confidentiality form, which was required to download the data, 13 teams downloaded the data and 9 teams submitted results. For the lumen segmentation, there was one submission from a commercial company, which did not participate in the workshop. Eight teams participated in the category for semi-automatic methods and one team submitted results of a fully automatic method. Only three teams submitted results for the stenosis grading, none of these methods were fully automatic.

<sup>1</sup> We used a fourth measure at the MICCAI Workshop: the root mean squared distance, which is in between the other two distance measures. We decided to remove this fourth measure, as it did not add valuable information.

<sup>2</sup> During the workshop we used the average instead of the maximum of the two maximum surface distances. This measure gives similar rankings, but is considered to be more informative.

The set of testing CTA datasets was divided in two parts for the workshop: a set to be processed prior to the workshop (31 datasets), and an on-site set to be processed by the participants during the morning session of the workshop (10 datasets). The results presented here are the aggregate results over all 41 datasets of the test set. Each team had the opportunity to resend a submission after the workshop to eliminate practical problems (e.g. erroneous file format submissions) during the on-site competition.

#### 4.1. Evaluated algorithms

Below we provide a very brief description of the algorithms involved in the MICCAI workshop challenge. For a more extensive description we refer to the respective papers that are available through the workshop issue of the Midas Journal (Hameeteman et al., 2009a,b).

##### 4.1.1. Krissian and Arencibia-García

The algorithm by Krissian and Arencibia-García (2009) is divided into two main steps: first, two minimal cost paths are tracked between the CCA and both the ECA and the ICA. The cost functions are based on a multi-scale vesselness response. Second, after detecting the junction position and cutting or extending the paths based on the requested lengths, a level set segmentation is initialized as a thin tube around the computed paths and evolves until reaching the vessel wall or a maximal evolution time.

##### 4.1.2. Freiman et al.

The algorithm by Freiman et al. (2009) uses a two-phase graph-based energy minimization approach. Its inputs are start and end seed points inside the vessel. The two-step graph-based energy minimization method starts by computing the weighted shortest path between the vessel seed endpoints based on local image and seed intensities and vessel path geometric characteristics. It then automatically defines a region of interest from the shortest path and the estimated vessel radius, and extracts the vessels boundaries by minimizing the energy on a corresponding graph cut.

##### 4.1.3. Zuluaga et al.

The algorithm by Zuluaga et al. (2009) consists of two stages: intensity-based preprocessing and model-based lumen delineation. The intensity-based preprocessing contains denoising by a Dual Tree Complex Wavelet Transform filter (Doré and Chérier, 2009), pre-segmentation using typical HU ranges refined by a Fuzzy C-means classifier, enhancement using gradient magnitudes and an exponential function within the pre-segmented regions. The model-based delineation extracts one centerline per seed-point by use of an elastic model and multi-scale eigen-analysis of the inertia matrix (Orkisz et al., 2008), which is followed by a distance-based identification of the bifurcation and surface pruning along the centerlines.

##### 4.1.4. Wong et al.

The algorithm by Wong et al. (2009) first generates a rough centerline using a vesselness measure and Minimal Cost Path search. This centerline is then simplified and used as initialization of an algorithm to find the principal curve (Kégl et al., 2000). This algorithm uses a 1D intensity model and results in a centerline with associated vessel radii. By assuming circular lumen cross-sections, both the segmentation and stenosis grading can then be determined.

##### 4.1.5. Cuisenaire

The algorithm by Cuisenaire (2009) is a fully automated method that segments not only the carotid bifurcation but also the vessels

from the aortic arch to the circle of Willis (COW), as well as the external branch and the vertebral arteries (Cuisenaire et al., 2008). The centerline of each vessel is first extracted using a local adaptive fast marching algorithm that is both seeded and constrained by an anatomical model. These constraints are adapted to the individual patient using both registration of the brain and segmentation of the brain and spine. Seeds are automatically placed in the COW and the lower part of the neck. The vessel lumen is segmented using 3D active objects initialized as a tube around the centerline.

##### 4.1.6. Florez Valencia et al.

In the algorithm of Flórez Valencia et al. (2009) the lumen is modeled by right generalized cylinders (RGC) with piece-wise constant parameters (Azencot and Orkisz, 2003). Their parameters are identified from planar contours extracted along the minimal paths between the end-points. The contours are extracted by use of 2D Fast Marching. Identification of the parameters (observations) is based on a geometrical analogy with piece-wise helical curves (axis) and on Fourier series decomposition of the contours (surface). Consistency from one piece to another is managed by a Kalman optimal estimator (Flórez Valencia et al., 2006).

##### 4.1.7. Gülsün and Tek

The algorithm by Gülsün and Tek (2009) uses graph-cuts optimization technique together with centerline models for segmenting the carotid arteries. It first detects the centerline representations between user placed seed points. This centerline extraction algorithm is based on a minimal path detection method which operates on a medialness map. The lumen of carotid arteries is then extracted by a graph-cut optimization technique using the detected centerlines as input.

##### 4.1.8. Mille et al.

The algorithm by Mille et al. (2009) is based on an extension of the minimal path method that models the vessel as a centerline and boundary. This is done by adding one dimension for the local radius around the centerline. The crucial step is the definition of the local metrics to minimize. The tubular structure of the vessels is exploited by designing an anisotropic metric. This metric is well oriented along the direction of the vessel, admits higher velocity on the centerline, and provides a good estimate of the vessel radius. In order to deal with carotid stenosis or occlusions, the segmentation is refined using a region-based level set method derived from the Chan–Vese model (Chan and Vese, 2001).

#### 4.2. Workshop results

Table 4 shows the average results and ranking for all lumen evaluation measures, for both the automatic and semi-automatic algorithms. The ‘Total success’ column shows the number of datasets over which the average values were calculated (see Section 3.4.3). The only fully automatic method submitted by Cuisenaire has similar distance and Dice measures as the team of Gülsün and Tek, but because it is only successful in 33 of the 41 datasets (the automatic method to find the bifurcation apparently did not work in eight of the cases) it is ranked lower than the team of Gülsün and Tek and the team of Krissian and Arencibia-García.

The table also contains the results of the observers, which have been scored in the same ways as the other participants. Observer A is the overall best observer (for each of the three clinical sites), Observer C is the overall worst observer, and Observer B is the remaining observer.

Fig. 7 gives a visual impression of the relation between the Dice measure and the segmentation quality. The figure shows a CMPPR image of a region around a calcified bifurcation overlaid with three

**Table 4**

Lumen segmentation performance of all submissions. The average value and rank is shown for each of the three similarity measures as defined in Section 3.4.1. The algorithm by Cuisenaire (in **bold**) is the only automatic method.

Method name	Total success	Dice		Msd		Hausdorff		Processing time	Total rank
		%	rank	mm	rank	mm	rank		
Observer A	41	95.1	2.61	0.10	2.68	0.65	2.73	–	2.67
Observer B	41	94.6	3.34	0.11	3.37	0.83	3.07	–	3.26
Observer C	41	94.4	3.46	0.12	3.32	0.97	3.85	–	3.54
M.A. Gülsün and H. Tek	41	91.8	5.95	0.18	5.68	1.5	5.27	50 s.	5.63
K. Krissian et al.	41	87.3	6.05	0.54	7.02	4.4	6.83	15 min	6.63
<b>O. Cuisenaire</b>	33	89.6	8.05	0.17	6.90	1.7	7.00	1–2 min	7.32
J. Mille et al.	41	83.5	8.66	0.74	8.68	10	8.90	90 s	8.75
M. Freiman et al.	41	82.9	8.85	0.75	9.05	9.2	8.90	2 min	8.94
W.C.K. Wong et al.	41	77.5	9.34	1.1	9.73	11	9.07	90 s	9.38
M.A. Zuluaga et al.	41	80.9	9.63	0.82	9.68	10	9.07	4 min	9.46
L. Florez Valencia et al.	37	53.6	11.1	3.4	10.9	12	10.2	2 min	10.7

different segmentations. The Dice measures are calculated over the whole evaluation region and not just at the shown region.

We also determined the overlap and distance measures of the three observers with respect to each other. Together with their averages, which is the inter-observer variability, they are shown in Table 5. To make these inter-observer measures better comparable to the results of the participants, we only determined them on the test data.

These results show that protocolized manual carotid artery lumen segmentation in CTA can be performed with an average precision of around 0.19 mm, which is smaller than the voxel size. The Dice and Msd of the observers with respect to each other are comparable to the results of the best semi-automatic method by M.A. Gülsün and H. Tek. Only the Hausdorff distance of the last mentioned method is substantially worse. The method of Cuisenaire also shows comparable results, but this fully automatic method fails on eight cases, which makes comparison less straightforward as these may be the more difficult cases.

Table 4 shows a substantial difference in the distance measures between the best three methods and the others. This can be explained by the fact that some of those other methods use a circular or elliptic model for the vessel cross-sectional area (Mille et al., 2009; Zuluaga et al., 2009; Wong et al., 2009), whereas the best three methods are based on graph cut (Gülsün and Tek, 2009), level set (Krissian and Arencibia-García, 2009) and active surface (Cuisenaire, 2009) algorithms, respectively, and do not have such hard vessel model constraints. Since the segmented vessel contains a bifurcation, the cross-sectional area is expected to deviate from the circular model, which makes methods that use such a model less successful. The lower scores of Florez Valencia et al. can be explained by failures in the initialization of the fast marching contour extraction. This initialization is based on the extraction of a centerline, which is not robust enough in the presence of complex lesions.

**Table 5**

Lumen measure of the three observers with respect to each other.

Observer	AB	BC	CA	Average
Dice	0.92	0.91	0.91	0.92
Msd	0.18	0.20	0.19	0.19
Hausdorff	1.23	1.39	1.41	1.34

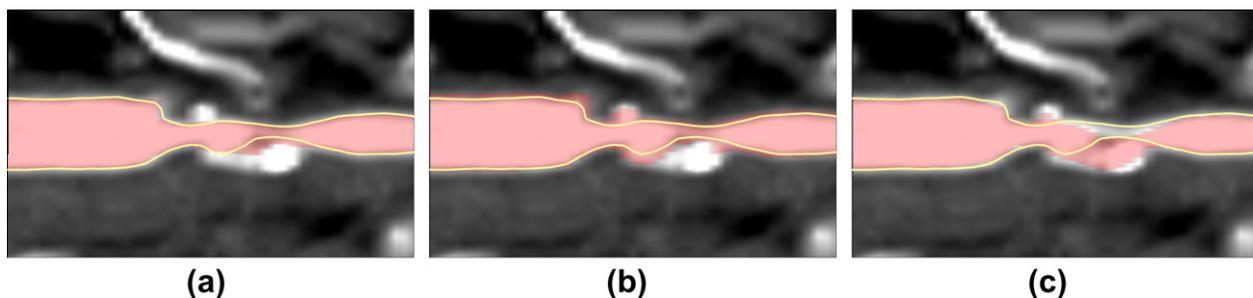
**Table 6**

The average rank of all the methods specified for the different stenosis categories.

Method name	Stenosis category				
	None	0–30	30–50	50–70	70–99
Observer A	2.58	2.73	2.57	2.50	2.96
Observer B	3.58	3.30	3.43	3.12	2.88
Observer C	3.62	3.63	3.19	3.88	3.33
M.A. Gülsün and H. Tek	6.08	5.90	5.09	5.62	5.33
K. Krissian et al.	7.46	6.03	6.38	6.29	7.12
<b>O. Cuisenaire</b>	7.12	6.50	6.71	8.54	7.83
J. Mille et al.	8.62	8.47	9.81	8.42	8.62
M. Freiman et al.	8.46	8.87	8.95	9.00	9.42
W.C.K. Wong et al.	9.50	10.2	9.52	8.75	8.79
M.A. Zuluaga et al.	8.62	9.83	9.57	9.46	9.75
L. Florez Valencia et al.	10.8	10.7	11.3	10.6	10.3

Note that the results of Zuluaga et al. in Table 4 reveal that the procedure to calculate the ranks (as described in Section 3.4.3) makes it possible for method A to have better average measures than method B, while still having a worse average rank. If e.g. for a certain dataset the measures of all methods are approximately the same, this may lead to much higher rank for method A compared to method B although their measures do not differ much.

Table 6 lists the average ranking for lumen segmentation in each of the stenosis categories. It shows that the top three ranked methods perform consistently best over the various categories.



**Fig. 7.** Visual impression of reference standard (yellow line) and user segmentations (red) with different Dice measures: (a) 94.5, example from M.A. Gülsün and H. Tek, (b) 88.4, example from O. Cuisenaire and (c) 88.1, example from L. Florez Valencia et al. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 7**  
Stenosis measures (as defined in Section 3.4.2) for all submissions. For both stenosis measures, the average difference with the reference stenosis and the rank is shown. The values in italics have been calculated with the standard stenosis grading algorithm and have not been supplied by the participants. The lumens column shows the number of lumen segmentations from which the standard values are calculated.

Method Name	Total Success	Diameter		Area		Standard algorithm			Total Rank
		$\Delta\%$	Rank	$\Delta\%$	Rank	Diameter*, $\Delta\%$	Area*, $\Delta\%$	Lumens*	
Observer A	41	3.40	2.15	2.90	2.02	6.70	8.20	41	2.09
Observer B	41	5.40	2.51	4.30	2.32	9.00	8.70	41	2.41
Observer C	41	5.70	2.66	5.00	2.68	11.2	10.3	41	2.67
M.A. Gülsün and H. Tek	41	-	-	-	-	9.70	12.8	41	-
K. Krissian et al.	41	-	-	-	-	14.3	19.8	41	-
<b>O. Cuisenaire</b>	33	-	-	-	-	16.7	16.5	33	-
J. Mille et al.	41	-	-	-	-	31.2	35.7	41	-
M. Freiman et al.	41	-	-	-	-	24.5	29.4	41	-
W.C.K. Wong et al.	41	31.4	4.20	25.2	4.17	27.8	27.5	41	4.18
M.A. Zuluaga et al.	41	17.0	3.56	16.9	3.71	30.0	31.4	41	3.63
L. Florez Valencia et al.	41	43.3	4.73	37.0	4.39	36.2	39.7	37	4.56

**Table 8**  
Pearson correlation coefficient between the standard stenosis measures and the lumen measures.

Stenosis measure	Dice	Msd	Hausdorff	Total rank
Diam	-0.87	0.78	0.97	0.93
Area	-0.87	0.79	0.97	0.92

Table 7 shows the average stenosis results on the test set. The ranking of these methods does not change for the different stenosis categories.

The difference in stenosis grading performance between the observers and the participants is large when compared to the results of the lumen segmentation. This could be explained by the fact that the teams that participated in the stenosis grading competition have a relatively large error in the lumen segmentation. As the stenosis grades are determined from lumen segmentations, inaccurate segmentations may lead to inaccurate stenosis grades. This is also confirmed by looking at the standard stenosis grading in relation to the lumen measures. Table 8 provides the correlation between the lumen measures and the standard stenosis measures and Fig. 8 shows a scatter plot of the area based standard stenosis error and the Hausdorff distance of the used segmentation. As can be seen there is a strong correlation between the Hausdorff lumen distance and the error in both the diameter and area based stenosis measures. The correlation between the standard stenosis measures

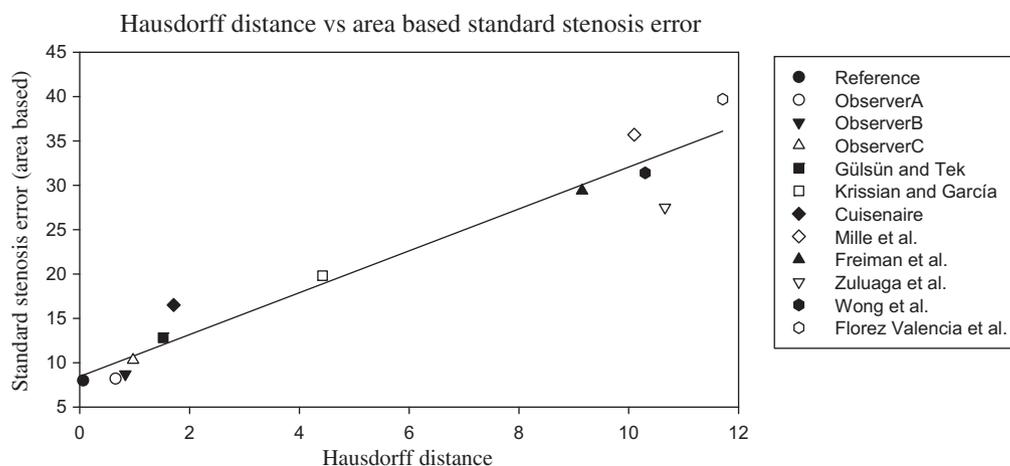
and the other lumen measures (Dice and Msd) is less strong. This can be explained by noting that both Dice and Msd are averages over the whole segmentation while a good stenosis measure requires a good segmentation on one specific location: the location of the stenosis. A large segmentation error on this location hardly influences the Dice and Msd, while it will directly influence on both the Hausdorff distance and the stenosis measures.

Note that in the standard stenosis algorithm, the orientation of the centerline determines the cross-sectional planes from which the area and minimal diameter are calculated. Thus the calculated measures highly depend on the extracted centerline. These standard centerlines are defined using the radius of the maximum inscribed sphere. Some of the submitted methods extract a centerline as an intermediate result, which may be a better centerline for stenosis measurement than the one extracted by the standard stenosis algorithm.

## 5. General discussion and conclusion

We presented an evaluation framework for carotid lumen segmentation and stenosis grading and made this framework publicly available via the website <http://cls2009.bigr.nl>.

The evaluation framework focuses on the effectiveness of lumen segmentation and stenosis gradings. Another important aspect for getting a novel technology accepted in daily clinical routine is efficiency. We decided not to take the computational costs into ac-



**Fig. 8.** Average area based standard stenosis error with respect to the average Hausdorff distance for all methods. The line shows a linear regression fit. The Pearson correlation coefficient is 0.97.

count in our framework. The main reasons are the lack of a standard performance parameter, and the variety of hardware and software platforms that is being used by the various groups participating in the challenge. Additionally, several research groups do not optimize their algorithms on computational costs, and performance estimates on current hardware and software platforms will be outdated soon because of rapid developments and increases in computational power. However, to get an indication in the computational expenses associated with the compared algorithms, we asked the current users to provide us with a rough estimate of the computational time for the processing of one dataset. Those values are listed in Table 4.

The framework we presented contains data from three medical centers and two different CT scanner vendors (Siemens Healthcare and Philips Healthcare), which prevents bias of the algorithms towards a specific scanner or acquisition protocol. However, other CT vendors (e.g. GE Healthcare, Toshiba Medical Systems) are still lacking, and thus the variety of the CTA datasets could still be improved. If other clinical centers want to contribute CTA data to our framework we will support them. From the current participants, only Wong et al. (2009) uses the provided scanner information to resample the image data of Erasmus MC.

Quantitative evaluation of medical imaging algorithms is essential for progress in the field of medical imaging. An objective, unbiased comparison of quantitative results is only feasible if the same evaluation measures are used, and if the datasets used in the evaluation are similar. We intend to contribute with our framework to the growing need for quantitative evaluations of medical image processing techniques. New lumen segmentation and stenosis grading algorithms can be evaluated using the described framework.

The framework proved to be an effective tool in the comparison of vessel lumen segmentation and stenosis grading techniques at the 2009 MICCAI workshop *3D Segmentation in the Clinic: A Grand Challenge III*, and is available for new evaluations of lumen segmentation and stenosis grading algorithms.

## Acknowledgments

Reinhard Hameeteman, Wiro Niessen and Theo van Walsum are supported by the Stichting voor de Technische Wetenschappen (STW) of The Netherlands Organization for Scientific Research (NWO). Aad van der Lugt is supported by Netherlands Organization for Health Research and Development (NWO-KF Grant 907-00-122) and The Dutch Heart Foundation (Grant 2007B161).

## References

- Abegunde, D., Beaglehole, R., Durivage, S., Epping-Jordan, J., Mathers, C., Shengelia, B., Strong, K., Tukuitonga, C., Unwin, N., 2005. Preventing chronic disease: a vital investment. World Health Organization. <[http://www.who.int/chp/chronic\\_disease\\_report/full\\_report.pdf](http://www.who.int/chp/chronic_disease_report/full_report.pdf)>.
- Antiga, L., Steinman, D.A., 2004. Robust and objective decomposition and mapping of bifurcating vessels. *IEEE Trans. Med. Imag.* 23 (6), 704–713.
- Azencot, J., Orkisz, M., 2003. Deterministic and stochastic state model of right generalized cylinder (RGC-sm): application in computer phantoms synthesis. *Graph. Models* 65 (6), 323–350. <[http://dx.doi.org/10.1016/S1524-0703\(03\)00073-0](http://dx.doi.org/10.1016/S1524-0703(03)00073-0)>.
- Baltaxe, M., Flórez Valencia, L., Hernández Hoyos, M., Magnin, I.E., Orkisz, M., 2007. Fast-marching contours for the segmentation of vessel lumen in CTA cross-sections. In: *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 791–794. <<http://dx.doi.org/10.1109/IEMBS.2007.4352409>>.
- Berg, M., Zhang, Z., Ikonen, A., Manninen, H., Vanninen, R., 2005. Carotid stenosis assessment with ct angiography using advanced vessel analysis software. *International Congress Series* 1281, 322–327, CARS 2005: Computer Assisted Radiology and Surgery. <<http://dx.doi.org/10.1016/j.ics.2005.03.331>>.
- Binaghi, S., Maeder, P., UskT, A., Meuwly, J.Y., Devuyt, G., Meuli, R.A., 2001. Three-dimensional computed tomography angiography and magnetic resonance angiography of carotid bifurcation stenosis. *Eur. Neurol.* 46 (1), 25–34. <<http://dx.doi.org/10.1159/000050752>>.
- Catmull, E., Rom, R., 1974. A class of local interpolating splines. In: Barnhill, R.E., Riesenfeld, R.F. (Eds.), *Computer Aided Geometric Design*, pp. 317–326.
- Chan, T., Vese, L., 2001. Active contours without edges. *IEEE Trans. Image Process.* 10 (2), 266–277. <<http://dx.doi.org/10.1006/jvci.1999.0442>>.
- Cuisenaire, O., 2009. Fully automated segmentation of carotid and vertebral arteries from CTA. In: *The Midas Journal*. <<http://hdl.handle.net/10380/3100>>.
- Cuisenaire, O., Virmani, S., Olszewski, M.E., Ardon, R., 2008. Fully automated segmentation of carotid and vertebral arteries from contrast enhanced cta. In: Reinhardt, J.M., Pluim, J.P.W. (Eds.), *Medical Imaging 2008: Image Processing*, 6914. SPIE, p. 69143R. <<http://dx.doi.org/10.1117/12.770481>>.
- Dice, L.R., 1945. Measures of the amount of ecologic association between species. *Ecology* 26 (3), 297–302. July. <<http://www.jstor.org/stable/1932409>>.
- Doré, V., Cheriet, M., 2009. Robust NL-means filter with optimal pixel-wise smoothing parameter for statistical image denoising. *IEEE Trans. Signal Process.* 57 (5), 1703–1716. <<http://dx.doi.org/10.1109/TSP.2008.2011832>>.
- ECST Collaborators, 1998. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the mrc european carotid surgery trial (ecst). *Lancet* 351 (9113), 1379–1387.
- Flórez Valencia, L., Azencot, J., Orkisz, M., 2009. Carotid arteries segmentation in CT images with use of a right generalized cylinder model. *Midas J.* <<http://hdl.handle.net/10380/3106>>.
- Flórez Valencia, L., Azencot, J., Vincent, F., Orkisz, M., Magnin, I., 2006. Segmentation and quantification of blood vessels in 3D images using a right generalized cylinder state model. In: *Proceedings of the IEEE International Conference on Image Process*, pp. 2441–2444. <<http://dx.doi.org/10.1109/ICIP.2006.312770>>.
- Freiman, M., Frank, J., Weizman, L., Nammer, E., Shilon, O., Joskowicz, L., Sosna, J., 2009. Nearly automatic vessels segmentation using graph-based energy minimization. *Midas J.* <<http://hdl.handle.net/10380/3090>>.
- Frostegård, J., 2005. Sle, atherosclerosis and cardiovascular disease. *J. Intern. Med.* 257 (6), 1365–2796. <<http://dx.doi.org/10.1111/j.1365-2796.2005.01502.x>>.
- Glagov, S., Weisenberg, E., Zarins, C.K., Stankunavicius, R., Koletis, G.J., May, 1987. Compensatory enlargement of human atherosclerotic coronary arteries. *N. Engl. J. Med.* 316 (22), 1371–1375. <<http://dx.doi.org/10.1056/NEJM198705283162204>>.
- Gülsün, M.A., Tek, H., 2009. Segmentation of carotid arteries by graph-cuts using centerline models. *Midas J.* <<http://hdl.handle.net/10380/3107>>.
- Hameeteman, K., Freiman, M., Zuluaga, M.A., Joskowicz, L., Rozie, S., Van Gils, M.J., Van den Borne, L., Sosna, J., Berman, P., Cohen, N., Douek, P., Sanchez, I., Aissat, M., Van der Lugt, A., Krestin, G.P., Niessen, W.J., Van Walsum, T., 2009a. Carotid lumen segmentation and stenosis grading challenge. *Midas J.* <<http://hdl.handle.net/10380/3128>>.
- Hameeteman, K., Freiman, M., Zuluaga, M.A., Joskowicz, L., van Walsum, T. (Eds.), 2009b. *Grand Challenge III: Carotid Lumen Segmentation and Stenosis Grading*. *Midas J.* <[www.midasjournal.org/browse/journal/48](http://www.midasjournal.org/browse/journal/48)>.
- Hankey, G.J., Warlow, C.P., Sellar, R.J., Feb, 1990. Cerebral angiographic risk in mild cerebrovascular disease. *Stroke* 21 (2), 209–222. <<http://stroke.ahajournals.org/cgi/reprint/21/2/209>>.
- Heimann, T., van Ginneken, B., Styner, M.A., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., Bello, F., Binnig, G., Bischof, H., Bornik, A., Cashman, P.M.M., Chi, Y., Crdova, A., Dawant, B.M., Fidrich, M., Furst, J., Furukawa, D., Grenacher, L., Hornerger, J., Kainmüller, D., Kitney, R.I., Kobatake, H., Lamecker, H., Lange, T., Lee, J., Lennon, B., Li, R., Li, S., Meinzer, H.-P., Nlmet, G., Raicu, D., Rau, A.-M., van Rikxoor, E.M., Rousson, M., Rusk, L., Saddy, K.A., Schmidt, G., Seghers, D., Shimizu, A., Slagmolen, P., Sorantin, E., Soza, G., Susomboon, R., Waite, J.M., Wimmer, A., Wolf, I., 2009. Comparison and evaluation of methods for liver segmentation from ct datasets. *IEEE Trans. Med. Imag.* 28 (8), 1251–1265. <<http://dx.doi.org/10.1109/TMI.2009.2013851>>.
- Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P.J., Bunke, H., Goldgof, D.B., Bowyer, K., Eggert, D.W., Fitzgibbon, A., Fisher, R.B., 1996. An experimental comparison of range image segmentation algorithms. *IEEE Trans. Pattern Anal. Machine Intell.* 18 (7), 673–689. <<http://dx.doi.org/10.1109/34.506791>>.
- Kégl, B., Krzyzak, A., Linder, T., Zeger, K., 2000. Learning and design of principal curves. *IEEE Trans. Pattern Anal. Machine Intell.* 22 (3), 281–297. <<http://dx.doi.org/10.1109/34.841759>>.
- Kirbas, C., Quek, F., 2004. A review of vessel extraction techniques and algorithms. *ACM Comput. Surv.* 36 (2), 81–121. <<http://doi.acm.org/10.1145/1031120.1031121>>.
- Koelmay, M.J.W., Nederkoorn, P.J., Reitsma, J.B., Majoie, C.B., 2004. Systematic review of computed tomography angiography for assessment of carotid artery disease. *Stroke* 35, 2306–2312. <<http://dx.doi.org/10.1161/01.STR.0000141426.63959.cc>>.
- Krissian, K., Arencibia-García, S., 2009. A minimal cost path and level set evolution approach for carotid bifurcation segmentation. In: *Midas J.* <<http://hdl.handle.net/10380/3103>>.
- Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G., 2009. A review of 3d vessel lumen segmentation techniques: models, features and extraction schemes. *Med. Image Anal.* 13 (6), 819–845. <<http://dx.doi.org/10.1016/j.media.2009.07.011>>.
- Manniesing, R., Viergever, M.A., Niessen, W.J., 2007. Vessel axis tracking using topology constrained surface evolution. *IEEE Trans. Med. Imag.* 26 (3), 309–316. <<http://doi.acm.org/10.1109/TMI.2006.891503>>.
- Martin, D., Fowlkes, C., Tal, D., Malik, J., 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proceedings of the Eighth IEEE International*

- Conference on Computer Vision ICCV 2001, vol. 2, pp. 416–423. <<http://dx.doi.org/10.1109/ICCV.2001.937655>>.
- Mille, J., Benmansour, F., Cohen, L.D., 2009. Carotid lumen segmentation based on tubular anisotropy and contours without edges. *Midas J.* <<http://hdl.handle.net/10380/3101>>.
- NASCET Collaborators, 1991. Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade carotid stenosis. *N. Engl. J. Med.* 325 (7), 445–453. <<http://content.nejm.org/cgi/content/abstract/325/7/445>>.
- Niemeijer, M., van Ginneken, B., Cree, M.J., Mizutani, A., Quellec, G., Sanchez, C.I., Zhang, B., Hornero, R., Lamard, M., Muramatsu, C., Wu, X., Cazuguel, G., You, J., Mayo, A., Li, Q., Hatanaka, Y., Cochener, B., Roux, C., Karray, F., Garcia, M., Fujita, H., Abramoff, M.D., Jan, 2010. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *IEEE Trans. Med. Imag.* 29 (1), 185–195. <<http://dx.doi.org/10.1109/TMI.2009.2033909>>.
- Orkisz, M., Flórez Valencia, L., Hernández Hoyos, M., 2008. Models, algorithms and applications in vascular image segmentation. *Machine Graph. Vis.* 17 (1/2), 5–33.
- Rothwell, P.M., Eliasziw, M., Gutnikov, S.A., Fox, A.J., Taylor, D.W., Mayberg, M.R., Warlow, C.P., Barnett, H.J.M., Collaboration, 2003. Analysis of pooled data from the randomised controlled trials of endarterectomy for symptomatic carotid stenosis. *Lancet* 361 (9352), 107–116. <[http://dx.doi.org/10.1016/S0140-6736\(03\)12228-3](http://dx.doi.org/10.1016/S0140-6736(03)12228-3)>.
- Schaap, M., Metz, C.T., van Walsum, T., van der Giessen, A.G., Weustink, A.C., Mollet, N.R.A., Bauer, C., Bogunovic, H., Castro, C., Deng, X., Dikici, E., O'Donnell, T., Frenay, M., Friman, O., Hernández Hoyos, M., Kitslaar, P.H., Krissian, K., Kühnel, C., Luengo-Oroz, M.A., Orkisz, M., Smedby, O., Styner, M., Szymczak, A., Tek, H., Wang, C., Warfield, S.K., Zambal, S., Zhang, Y., Krestin, G.P., Niessen, W.J., 2009. Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Med. Image Anal.* (13/5), 701–714. <<http://dx.doi.org/10.1016/j.media.2009.06.003>>.
- Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* 47 (1–3), 7–42. <<http://dx.doi.org/10.1023/A:1014573219977>>.
- Scherl, H., Hornegger, J., Prnmmer, M., Lell, M., Feb, 2007. Semi-automatic level-set based segmentation and stenosis quantification of the internal carotid artery in 3d cta data sets. *Med. Image Anal.* 11 (1), 21–34. <<http://dx.doi.org/10.1016/j.media.2006.09.004>>.
- Schroeder, W., Martin, K., Lorensen, B., December, 2006. Visualization Toolkit: An Object-Oriented Approach to 3D Graphics, 4th ed. Kitware.
- Turk, G., O'Brien, J.F., 1999. Shape transformation using variational implicit functions. In: SIGGRAPH '99: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, pp. 335–342. <<http://doi.acm.org/10.1145/311535.311580>>.
- Van Ginneken, B., Armato III, S.G., de Hoop, B., van Amelsvoort-van de Vorst, S., Duindam, T., Niemeijer, M., Murphy, K., Schilham, A., Retico, A., Fantacci, M.E., Camarlinghi, N., Bagagli, F., Gori, I., Hara, T., Fujita, H., Gargano, G., Bellotti, R., Tangaro, S., Bolaos, L., Carlo, F.D., Cerello, P., Cheran, S.C., Torres, E.L., Prokop, M., 2010. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the ANODE09 study. *Med. Image Anal.* 14 (6), 707–722. <<http://dx.doi.org/10.1016/j.media.2010.05.005>>.
- Waugh, J.R., Sacharias, N., Jan, 1992. Arteriographic complications in the DSA era. *Radiology* 182 (1), 243–246.
- West, J., Fitzpatrick, J.M., Wang, M.Y., Dawant, B.M., Maurer, C.R., Kessler, R.M., Maciunas, R.J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Suetens, P., Vandermeulen, D., van den Elsen, P.A., Napel, S., Sumanaweera, T.S., Harkness, B., Hemler, P.F., Hill, D.L., Hawkes, D.J., Studholme, C., Maintz, J.B., Viergever, M.A., Malandain, G., Woods, R.P., 1997. Comparison and evaluation of retrospective intermodality brain image registration techniques. *J. Comput. Assist. Tomogr.* 21 (4), 554–566.
- Wintermark, M., Glastonbury, C., Tong, E., Lau, B.C., Schaeffer, S., Chien, J.D., Haar, P.J., Saloner, D., 2008. Semi-automated computer assessment of the degree of carotid artery stenosis compares favorably to visual evaluation. *J. Neurol. Sci.* 269 (1–2), 74–79. <<http://doi.acm.org/10.1016/j.jns.2007.12.023>>.
- Wong, W.C.K., So, R.W.K., Chung, A.C.S., 2009. Principal curves: a technique for preliminary carotid lumen segmentation and stenosis grading. *Midas J.* <<http://hdl.handle.net/10380/3096>>.
- Zuluaga, M.A., Orkisz, M., Delgado Leyton, E.J.F., Doré, V., Morales Pinzón, A., Hernández Hoyos, M., 2009. Adaptations of MARACAS algorithm to the segmentation of the carotid bifurcation and stenosis quantification in CTA images. *Midas J.* <http://hdl.handle.net/10380/3102>.