Level set based cerebral vasculature segmentation and diameter quantification in CT angiography

R. Manniesing *, B.K. Velthuis, M.S. van Leeuwen, I.C. van der Schaaf, P.J. van Laar, W.J. Niessen 1

Department of Radiology, Image Sciences Institute, University Medical Center Utrecht, Heidelberglaan 100, Room E01.335, 3584 CX Utrecht, The Netherlands

Received 13 December 2004; received in revised form 9 March 2005; accepted 16 September 2005
Available online 2 November 2005

Abstract

A level set based method is presented for cerebral vascular tree segmentation from computed tomography angiography (CTA) data. The method starts with bone masking by registering a contrast enhanced scan with a low-dose mask scan in which the bone has been segmented. Then an estimate of the background and vessel intensity distributions is made based on the intensity histogram which is used to steer the level set to capture the vessel boundaries. The relevant parameters of the level set evolution are optimized using a training set. The method is validated by a diameter quantification study which is carried out on phantom data, representing ground truth, and 10 patient data sets. The results are compared to manually obtained measurements by two expert observers. In the phantom study, the method achieves similar accuracy as the observers, but is unbiased whereas the observers are biased, i.e., the results are 0.00 ± 0.23 vs. −0.32 ± 0.23 mm. Also, the method’s reproducibility is slightly better than the inter-and intra-observer variability. In the patient study, the method is in agreement with the observers and also, the method’s reproducibility −0.04 ± 0.17 mm is similar to the inter-observer variability 0.06 ± 0.17 mm. Since the method achieves comparable accuracy and reproducibility as the observers, and since the method achieves better performance than the observers with respect to ground truth, we conclude that the level set based vessel segmentation is a promising method for automated and accurate CTA diameter quantification.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Bone masking; Level set; Vessel segmentation; Diameter quantification; CTA; Circle of Willis

1. Introduction

The dramatic increase of the number and sheer size of three dimensional (3D) angiographic data sets (magnetic resonance angiography – MRA and computed tomography angiography – CTA) has lead to an almost overwhelming amount of 3D information to be evaluated by clinicians. Since manual procedures to process these data are often tedious, time consuming and subject to inter- and intra-observer variability, there is a strong demand for segmentation methods that are (semi-) automatic. The methods for the specific task of vessel segmentation have received considerable interest in the last decade. They have been applied for improving visualization, therapy planning, detection of abnormalities (such as aneurysms), in quantification (e.g., of diameters or stenosis grade) and as pre-processing step for 3D vessel modeling, and in the design of computer aided diagnosis systems.

The focus of this work is on segmentation of the cerebral vasculature – in particular the Circle of Willis (CoW) – in CTA data sets, and on diameter quantification of parts of the vessel tree. Fig. 1 shows a maximum intensity projection (MIP) and a schematic drawing of the CoW together with the naming of its most important vessels. Segmentation of these vessels is clinically relevant since it allows...
Several authors have used level set approaches for vascular segmentation. Most methods have focused on vessel segmentation from 3D MRA data. For example, in van Bemmel et al. (2003), a level set framework is proposed for the purpose of artery vein separation in blood pool agents contrast enhanced MRA data. Lorigo et al. (2001) aim at vessel segmentation by a co-dimension two level set evolution in 3D, i.e., evolving line structures in 3D, and the method is applied to cerebral MRA data. The latter is similar to our method, in which we evolve a level set which is constrained by curvature influence of the minimal principal component of the 3D curvature term. Compared to MRA, CTA has the additional problem of presence of bone structures in the image, whose intensity values, especially of the partial volume voxels, overlap with vessel intensities. Besides (Lorigo et al., 2001), the most closely related work is by Suryanarayanan et al. (2004). There a partitioning scheme is proposed which divides the CTA data set into three distinct regions (neck, skull base and skull). Subsequently, for each region a different type of segmentation method is applied based on a variant of a region growing algorithm. However, exact implementation details and details on how to deal with the bone structures, especially at the difficult part of the skull base, are missing. Furthermore, the work by Lorigo et al. (2001) and Suryanarayanan et al. (2004) lack a proper validation. To our knowledge, there is no other work which targets both the segmentation of vessels from 3D CTA data and includes an extensive quantitative validation.

The main contribution of this work is the extensive validation of the results. Visual evaluation of the segmentation results is limited, as it is subject to inter- and intraobserver variability. Moreover, it is an extremely difficult task – especially in the case of the varying and complex morphology that large vessel structures exhibit. Therefore, we choose to validate based on diameter quantification of parts of the vessel tree. This choice is also motivated by the fact that diameter quantification is the first step in other clinical studies such as vasospasm detection or stenosis quantification in patients. Two large studies are conducted, a phantom study and a patient study, both including manual measurements obtained from two expert observers. The manually obtained results are compared to the method’s results.

This paper is organized as follows. In Section 2, the level set based method is explained in detail. We describe the data sets that are used for optimization and validation in Section 3. A major part of this work is devoted to evaluation and validation; Section 4 describes the applied evaluation criteria. Section 5 is on the experimental framework that includes the experiments designated for segmentation,
parameter optimization and diameter quantification. Results are reported in Section 6 and discussed in Section 7.

2. Method

The proposed method can be divided into three successive stages, see Fig. 2 for an overview.

1. Bone masking. The first step is masking of bone tissue voxels in the CTA image. An additional scan of the patient is made at a lower dose and without contrast fluid prior to the CTA acquisition. By registering the scans and segmenting the bone in this low-dose scan, bone tissue can effectively be eliminated from the CTA scan.

2. Speed function construction. The level set evolution (described in stage three) requires an image based speed function. This speed function is a function of the vessel and background intensity distributions of the image. For estimating the vessel intensity distribution, we consider two approaches. The first approach is a global estimation taking into account the complete masked image. In the second approach, local estimations in small volumes of interests (VOIs) around user given seed points are made.

3. Level set evolution. The seed points are used to initialize a single level set for segmentation, and multiple level sets – one for each VOI – for the diameter measurements. The accuracy by which the vessel diameters are finally captured can be fine tuned by weighting the influence of the curvature tension that is imposed on the level set during evolution.

In the following sections, each stage is explained in detail.

2.1. Bone masking

The most important reason to start the method with bone masking – considered from an image processing perspective – is simplifying vessel segmentation in the area where the skull base is situated. This is a challenging area for two reasons: First, the skull base has very thin bone structures giving rise to non-homogeneously distributed intensity values which are close to, and partially overlap with the intensity of vessel voxels. Second, most of the vessels forming the CoW are in close proximity to the skull base. In particular, two feeding arteries of the CoW, the ICAs – the internal carotid arteries, pass through the skull base and at some points; vessel and bone are only separated by less than one millimeter. From a clinical perspective bone masking is justified, since it enables an improved visualization of the data set. However, the occasional failures of registration, and, more importantly, the additional radiation exposed to the patient, may outweigh the benefits. Fortunately, the increased radiation dose is relatively small, approximately one third of the original dose for a contrast enhanced scan (see Section 3 on data acquisition). This can actually be smaller, since the registration process works on a subsampled data, as we will see in this section. In the protocol utilized in our hospital, the additional bone scans are routinely acquired.

The bone masking process is as follows. Two scans are made of the patient: a low-dose scan without contrast, showing bone only, and a high dose contrast enhanced scan, showing bone and vasculature. The time between the scans with and without contrast is about 1 min in order to perform a contrast time delay examination prior to the CTA. The patient’s head is fixed during the scans. To compensate for small remaining motion between the two scans, the low-dose scan is registered to the high dose scan. Registration is performed in 3D to 3D, rigidly and on the basis of the mutual information measure (Maes et al., 1997). Resampling is done by trilinear interpolation. To speed up the registration process, both data sets are subsampled by a factor two in each direction. Furthermore, during registration, voxels with intensity value –1000 Hounsfield Units (HU), representing air, are excluded, thereby circumventing the dominating gradients caused by the transition air-gantry. Then, a threshold is applied to segment the bone in the CT scan without contrast. Since there are no other structures with high intensities, this threshold value is not critical and may be set relatively far below the ‘actual’
threshold value in order to capture all bone structures in the image. The threshold is set to 150 HU. After a morphological opening to remove any noise pixels that may have occurred, followed by a dilation to compensate for inaccuracy of registration, the bone structures are masked in the high dose scan, resulting in an image showing vessels, brain tissue, and background voxels only. In this and in the following stage (speed function construction in Section 2.2), the basic morphological operation, erosion and dilation, are used. These are all implemented with a fixed kernel size of three voxels. The kernel shape (either square or star shape) is selected depending on the accuracy that is needed and depending on the desired effect. For bone masking, the objective is to retain the (curved) shape of the bone as best as possible leading to a star shape kernel, both for opening and dilation. The opening operation is carried out once; compensating for the inaccuracy caused by registration is carried out by two dilations. An illustration of the bone masking process is given in Fig. 3.

The idea of utilizing a non-contrast enhanced CT scan for bone removal, was first published by Venema et al. (2001). There, a 3D-to-3D registration method based on gray value correlation is presented followed by 2D to 3D registration to compensate for possible motion of the patient during acquisition. In our data sets a 3D-to-3D registration already yielded satisfying results. Besides the exclusion of the 2D to 3D registration step, our method differs from Venema et al. (2001) on the information measure that is used (mutual information versus gray value correlation measure) and the intensity range that is selected (full range except HU of air, versus 600–800 HU to specifically select bone tissue voxels). Furthermore, our method significantly reduces the computation time by subsampling the data sets prior to registration.

2.2. Image based speed function

The scalar speed function defines a mapping from the domain of intensity values of the image to the speed values of the propagating interface (or level set, see Section 2.3). This function steers the level set evolution and is therefore of vital importance in the segmentation process: not only does it define the convergence rate, it also has large influence on the final boundaries. A level set that starts to expand from within a vessel (therefore requiring a positive valued speed function), should stop at the transition vessel – background (requiring a zero valued speed function according to some criterion), and possible leaking into background should be prevented by a combination of a negative range of the speed function with additional smoothness constrained by curvature weighting. Thus, our speed function should have range \([-1, \ldots, 1]\). The second requirement, is to let the speed function be a function of the intensity distributions of the background and vessel voxels. In its definition we restrict ourselves to zeroth order image information (i.e., intensity information) only, by using histograms as our building blocks. Including higher order image information and spatial information refines the speed function, and could, potentially, improve the results. However, from experiments, intensity information combined with geometric constraints already yield satisfactory results.

To construct a proper speed function, we first need to estimate the vessel and background intensity distributions. Consider a typical histogram of a raw CT scan of a patient in which all tissue types are included (Fig. 4). The white and gray matter of the brain is dominantly present in the form of a single peak within the range 0 and 200 HU. Left from this peak, within range \(-200–0\) HU, a small increase can be observed which is caused by fat, mostly skin tissue. Vessel voxels have range 200–600 HU, bone structure voxels have range 400–1000 and both are included but not properly visible at this scale.

Background information is extracted from the masked image since in these images bone structures are already suppressed. The complete image is used since it is a reasonable assumption that background intensity will not vary significantly within a patient. What is left are fat tissue voxels that need to be removed which is achieved by first finding the maximum in the histogram, starting the search from \(-900\) HU, and thresholding the masked image at that intensity value. The resulting binary image is cleaned by applying the morphological opening operation, followed by five dilations. For both a square shaped kernel is chosen.

Fig. 3. Zooming in on the skull base and showing the process of bone masking. Shown here is one slice (axial view) from the lower part of the skull base where the feeding cerebral arteries enter the brain. Left and right from the skull base the ICA (Internal Carotid Arteries), denoted by arrows one and two, and in the middle below, the BA (Basilar Artery) denoted by arrow three. (Left) The original contrast enhanced data. Clearly, separating artery from skull base is not a trivial task, since artery and skull base are lying very close together and the intensity values overlap, see for example arrow four. (Middle) Bone masking after registration but without morphological operations. (Right) Bone masking after registration followed by the morphological operations opening and dilation. Notice that in this region, the outer voxels of the arteries can become (partially) masked as well (e.g., the left ICA).
since maintaining the curved shape property is not relevant, and, we would like to have a large extent of these operations. The dilation step ensures that most of the background voxels are included. The mean \( \mu_b \) and standard deviation \( \sigma_b \) of the background is then calculated within the range \([-900, \ldots, 400]\) HU. To speed up background intensity estimation, especially when the morphological operators are applied, the masked image is subsampled by a factor two in all directions.

The vessel information is extracted in two ways; on a global and on a local level giving rise to two speed function \( F_G \) and \( F_L \). First, seed points should be placed in the masked image within the vessel at approximately those locations where diameter measurements should be carried out, see Section 4.1 for more detail. To globally estimate the vessel intensity distribution, all seed points and the complete masked image are taken and an initial threshold is determined by taking the average intensity value of a small neighborhood (three voxels in each direction) around each seed point. This initial threshold value represents the intensity value approximately in the middle of the vessel, and therefore it will be slightly higher than the intensity values found at the boundaries between vessel and background. To include all vessel voxels, the value is lowered by multiplying with a factor \( p \leq 1 \), which also allows for a very accurate adjustment of the bias of the method in a training stage (see Section 6.3). To clean up any single voxels that might have appeared after thresholding, the morphological operation open is applied, with a star shaped kernel to limit the extent of this operation. The mean \( \mu_v \) and standard deviation \( \sigma_v \) of the vessel is then calculated within range \([0, \ldots, 800]\) HU. These are the mean and standard deviation for the global speed function. To locally estimate the vessel intensity distribution, a small cubical VOI around a single seed point is taken, and above steps are repeated for this smaller volume, giving rise to a local mean and a local standard deviation for each seed point.

Given the estimated means of background \( \mu_b \) and vessel \( \mu_v \), and the standard deviations of background \( \sigma_b \), and vessel \( \sigma_v \), the speed function can now be derived. The assumption is that these parameters describe the Gaussian probability density functions of background \( g_b \) and vessel \( g_v \), by

\[
g_v(x) = \frac{1}{\sigma_v \sqrt{2\pi}} e^{-\frac{(x-\mu_v)^2}{2\sigma_v^2}},
\]

\[
g_b(x) = \frac{1}{\sigma_b \sqrt{2\pi}} e^{-\frac{(x-\mu_b)^2}{2\sigma_b^2}}.
\]

Given a threshold \( i \) placed between \( \mu_b \) and \( \mu_v \) (with \( u_v > u_b \) in general), then \( E_v \) and \( E_b \) will be the fraction of voxels erroneously denoted as background and vessel voxels, i.e.

\[
E_v(i) = \int_i^{\infty} g_v(x) \, dx,
\]

\[
E_b(i) = \int_i^{\infty} g_b(x) \, dx.
\]

The total error is given by the sum of \( E_v \) and \( E_b \), and is minimized by setting the derivative to zero. The optimal threshold is found exactly where the Gaussian curves intersect – and this is exactly where \( F_{im} \) is defined to be zero.

Thus, the boundary between vessel and background is defined as the point of minimal classification error. Furthermore, we require the speed function to make a smooth transition on this boundary. This is established by realizing that for a certain intensity value, the functions \( g_b \) and \( g_v \) simply describe the probability that this intensity value belongs to background and vessel. A high intensity value has larger probability of being vessel than being background and vice versa. The speed function \( F_{im} \) is then defined as

\[
F_{im}(x) = \frac{g_v(x) - g_b(x)}{g_v(x) + g_b(x)},
\]

with range \([-1, 1]\] and with \( g_v + g_b \) as a normalization factor. Fig. 5 shows an example of two Gaussians and their corresponding speed function. It is easily seen that the speed function takes value zero at the optimal threshold value, has values approaching one within the vessel and has values approaching minus one outside the vessel. For values much smaller than \( \mu_b \) much greater than \( \mu_v \), \( g_v \) and \( g_b \) are both very small, which can result in unexpected outcomes of \( F_{im} \). Therefore, a domain is defined for which Eq. (5) should hold and it is reasonable to let \( F_{im} \) hold within \([\mu_v - 3\sigma_v, \mu_v + 3\sigma_v]\) and to let \( F_{im} \) equal 0 outside this domain. Since the vessel intensity distribution is estimated in two ways, two kinds of image based speed function are distinguished: \( F_{gi} \) with \( \mu_i \) and \( \sigma_i \) estimated from the complete masked image, and \( F_{li} \) with \( \mu_i \) and \( \sigma_i \) estimated from a small VOI around a seed point.
Minimal classification error is achieved exactly where the Gaussians intersect – at that point, the speed function is defined to be zero.

2.3. Level set evolution

The level set framework (Osher and Sethian, 1988), is an implicit way of describing the propagation of surfaces which move along their normal vector field caused by some speed function acting upon it. Denote \( \tilde{C}(p, t) \) as the current position of the surface, with \( p \) its parametrization and \( t \) denoting time, \( \vec{n} \) as the outward normal on the surface and \( F \) as the speed function. The speed function acts perpendicular to the surface; any tangential component would only affect its parametrization. Thus, the surface moves according to the following equation:

\[
\frac{\partial \tilde{C}}{\partial t} = F \vec{n}.
\]

This \( nD \) surface is embedded as the zero level set of a \((n+1)D\) function \( u \), thereby obtaining an intrinsic, i.e., parameter free representation, and also gaining the flexibility of handling topological changes of the surface, since different topologies for \( \tilde{C} \) correspond to the same topology of \( u \). In general, for function \( u \) the signed distance function is chosen, with negative sign inside the surface and positive sign outside, i.e., \( u(\tilde{C}, t) = \pm d \). The evolution of the level set is described by taking the total derivative of \( u \) to \( t \), giving

\[
\nabla u \cdot \vec{n} + u_t = 0,
\]

where \( u_t \) denotes the partial derivative to time. For any level set function, the gradient vector is perpendicular to the level surfaces, or isophotes, which gives \( \vec{n} = \nabla u / |\nabla u| \) for the normal vector. Together with Eq. (6), this results in

\[
u_t + F |\nabla u| = 0\]

with initial condition \( u_t = u(\tilde{C}, 0) \) given.

For object segmentation the following speed function, first proposed in Osher and Sethian (1988) and Sethian (1989), is often substituted for \( F \)

\[
F = F_{\text{ext}}(c - \epsilon \kappa),
\]

where \( F_{\text{ext}}(I) \) is an external term based on image features, \( c \) a constant advection term (similar to a balloon force – usually \( c = 1 \) is chosen), and \( \epsilon \kappa \) a weighted curvature term assuring smoothness of the surface. The curvature at a point on a (3D) surface can be decomposed in two perpendicular directions in which the curvature takes minimal and maximal value; these extremal curvatures are called the principal curvatures. For vessel structures we prefer a smoothness along the longitudinal direction in which the curvature term is minimal, and therefore we define \( \kappa \) in Eq. (9) as \( \kappa \triangleq \kappa_{\text{min}} \). With the previously defined speed function \( F_{\text{im}} \) (Eq. (5)) substituted for \( F_{\text{ext}} \), the final partial differential equation then reads as

\[
u_t + F_{\text{im}}(1 - \epsilon \kappa_{\text{min}}) |\nabla u| = 0,
\]

obtained by combining Eqs. (5), (8) and (9). The use of the minimal principal curvature has been done before by van Bemmel et al. (2003) and can be considered equivalent to the method proposed by Lorigo et al. (2001) in which ID curves evolves in 3D space.

The level set is continued until convergence, i.e., stopped if there was no change in the volume which is defined as the level set function smaller than zero.

3. Data

The method is applied to phantom and patient data. The phantom that is used is a 3D cerebrovascular flow phantom (Fahrig et al., 1999), with known diameters and is filled with 15 mgI/mL contrast agent which was found to be equivalent to an average of 25 neurovascular CT scans. The patients are examined for acute cerebrovascular events, such as stroke or subarachnoid hemorrhage (bleeding) from a ruptured cerebral aneurysm. CTA scans are also used to screen for cerebral aneurysms in patients with a clipped cerebral aneurysm or to screen family member of patients who have had a subarachnoid hemorrhage. We used scans from the screening group to test our method and we made a refined selection by excluding those patient with clips, since dealing with the resulting clipping artefacts caused by CT at the image processing side is beyond the scope of this work. Both the phantom data and patient data are acquired at the same 16-slice CT scanner (Philips), situated at the radiology department of the University Medical Center (UMC), Utrecht. All data are reconstructed with the same filter type and convolution kernel, have dimensions 512 \( \times \) 512 by approximately 300 slices, have an in plane resolution of 0.3125 \( \times \) 0.3125 mm, and have slice thickness/spacing of 1.0/0.5 mm. The phantom data and contrast enhanced patient data are acquired at 100 kV tube voltage and 440 mA s tube current, the non-contrast enhanced patient scans are acquired at 100 kV and 133 mA s. Since patient dose is linearly related to the tube current (Prokop and Galanski, 2003), it follows that the non-contrast scan is made at approximately one third of the dose of the contrast enhanced scan. In a recent paper on bone masking by van Straten et al. (2004), it is shown
that the tube current of the non-contrast enhanced scan can even be reduced to 65 mA s without a loss of image quality.

4. Evaluation

The evaluation of segmentation methods is a notoriously difficult problem, and this work forms no exception. The complex nature of vessel morphology and the small percentage of voxels contributing to vessel voxels in a typical CTA data set, makes a manual segmentation and a direct evaluation by some comparison measure an infeasible approach. For determining the completeness of the segmentation (in particular, are all vessel present forming the CoW), the depth of the segmentations (are the peripheral vessels present) and the capability of segmenting vessel (segments) that is hampered by e.g., stenosis, calcifications and/or aneurysms, visual evaluation is used. For diameter quantification of segments of the vessel tree, a comparison to ground truth is provided by phantom data and manual measurements are performed on phantom and patient data.

4.1. Diameter quantification

Diameter quantification is carried out on the following vessels. On the phantom data, the Anterior Cerebral Arteries (Al and A2) \{l,r\}, the Middle Cerebral Arteries (Mi) \{l,r\}, the Posterior Communicating Arteries (PCoA) \{l,r\}, the Posterior Cerebral Arteries (P1 and P2) \{l,r\}, the Carotid \{l,r\}, the Vertebral \{l,r\}, and the Basilar Arteries are selected, see Fig. 1 for their locations. This gives a total of 17 vessels. Left out is the Anterior Communicating Artery (ACoA) which is the small vessel segment between the A2s. On the patient data, the following vessels – if present, are selected: the Al and A2 \{l,r\}, Mi and M2 \{l,r\}, PCoA \{l,r\} and P1 and P2 \{l,r\}, see Fig. 1 for their locations, which gives 14 points for each patient.

Manual measurements by expert observers are performed on a Philips MxView clinical workstation. The procedure for obtaining the manual measurements is as follows. First, the window width and level settings are fixed to 150 and 400 HU, respectively. Then several slabs are selected (a slab is a subset of slices, here taken in the z-direction) that encompass all the vessels of interest. The observer draws a line approximately perpendicular to the central vessel axis, the workstation then automatically gives the length of this line. Furthermore, the observer is asked to indicate a point at approximately the middle of the line just drawn, which is used as seed point for our method. The automated procedure to obtain the diameter measurements from the resulting level set images of the method, is as follows, see Fig. 6 for a schematic drawing. The level set image after convergence is used to determine a central vessel axis and together they are used to determine a reformatted segmentation around the seed point. The central vessel axis is approximated by skeletonization of the binary segmentation obtained from the level set image by thresholding at value zero. Skeletonization is carried out by iterative thinning (in 2D, this is done by thinning with the Golay alphabet (Sonka et al., 1998), and in 3D, by thinning with a carefully chosen set of 3D masks (Palágyi and Kuba, 1998)). Thinning deletes border points of the object satisfying certain criteria until the resulting object does not change any more. The final result of the thinning procedure is a connected, topological preserving skeleton of the original object that is a representation of the central vessel axis part of the vascular tree. In Palágyi and Kuba (1998), these center lines are called the medial lines, i.e., 1D structures in 3D space, making this 3D skeleton algorithm particular suitable for our application. Since multiple branches may occur in the area of the seed point, a refined path selection is needed based on this skeleton. The seed point is used to mark a root point on the skeleton, which is either the seed point itself, or the nearest skeleton point, followed by a traversal on both sides of the begin point until an end point (that is a skeleton point with only one neighbor point) is reached. In case of multiple end points on one side, that end point is selected which has the largest distance from the begin point. The result is a sort of ‘stretched’ path, with approximately in the middle the vessel location where diameter measurement should be carried out. Of this path, both sides of the begin point are traversed again, until three skeleton points on both sides are included. The resulting path, with a maximum length of seven voxels (corresponding to approximately 2 mm) is used to reformat the level set image. We assume that along the previously found path, the vessel follows a straight line, i.e., a vector is interpo-

Fig. 6. The steps that are taken before reformatting the vessel segment at the point where the diameter should be measured. Notice that in real data, the sizes of the skeleton points are much smaller than the vessel sizes. The final vector has length of maximum seven voxels corresponding to approximately 2 mm. For more details, see Section 4.1. (a) Vessel segment; (b) skeleton; (c) root point and end points; (d) longest path selection; (e) path selection and vector; and (f) reformatting vessel segment.
lated between the two end points of the path. This seems like a rather crude assumption, but in practice does hold quite well. Along this vector, the level set image is reformatted, and to increase accuracy, reformattting is done on an increased resolution\(^2\) by a factor 10 in all directions. After reformattting, a threshold zero is applied and the diameters are estimated slice by slice by assuming a circular cross-sectional shape. The final diameter is then given as the average diameter of all slices.

4.2. Comparison measure

In order to compare different series of measurements (for instance between the method and observer, or between the method and ground truth), and to assess the accuracy and limits of agreement, a Bland and Altman analysis is performed (Bland and Altman, 1986). Given two corresponding series of measurements on the same subject, for each pair of measurements the difference is plotted against their mean value for visual evaluation, and the limits of agreement (95% confidence interval) is determined by the mean (this is also-called the bias) \(\pm 1.96\) the standard deviation of all the differences of all measurements. For example, given the vectors \(A\) and \(B\) with elements \(A = \{1,2,3,4,5\}^T\) and \(B = \{2,2,4,4,5\}^T\). The mean and standard deviation are then calculated of the difference of these vectors, i.e., \(A - B = -0.4 \pm 0.55\).

The limits of agreements then becomes \([-1.5,0.7]\). When presenting the results in Section 6, above notation will usually be combined, resulting in \(A - B = -0.4\pm 0.55\).[−1.5,0.70].

5. Experiments

In Section 5.1, the implementation details of the level set based method are explained. We then discuss the experimental details related to segmentation of the patient data (Section 5.2), the parameter optimization study concerning diameter quantification (Section 5.3) and the experiments related to the performance assessment study (Section 5.4).

5.1. Implementation details

To speed up the level set evolution, a narrow-band implementation is used (Sethian, 1999), which only updates the values in a small band inside and outside the zero level set during evolution. A bandwidth of three voxels on either side is selected. The time step is set to \(t = 0.1\) and re-initialization occurs every two iterations. Second, order derivatives, needed for curvature calculation, are obtained by convolutions with a Gaussian kernel, carried out in a neighborhood of seven voxels (that is twice the band width plus one for the zero level set) in each direction and at scale of two voxels. The segmentation of the complete cerebral vasculature is carried out with a total number of \(n = 10,000\) iterations; in case of diameter quantification a small VOI is selected around each seed point of 25 voxels in each direction in which the upper bound of iterations is set at \(n = 600\). This VOI is taken from the masked image for the local speed function \(F_L\) and is taken from the speed function image for the global speed function \(F_G\).

5.2. Segmentation

The method is applied to 10 patient data sets for cerebral vasculature segmentation. The level set is steered by the global speed function \(F_G\), and is initialized by the user given seed points. The parameters which need to be filled in are the fraction \(p\) and curvature weighting \(\epsilon\). Selected values are \(p = 0.79\) and \(\epsilon = 0.5\), based on the optimization study related to diameter quantification as described in Section 5.3. After convergence of the level set, the binary segmentation is extracted by thresholding at value zero, which then is multiplied with the original masked image to obtain the final grey value segmentation.

5.3. Parameter optimization \(p, \epsilon\)

The diameter quantification study requires the optimization of the remaining parameters of the method; the fraction \(p\) and the curvature weighting \(\epsilon\). As it turns out, \(p\) allows for a very accurate bias adjustment of the method, and together with \(\epsilon\) enables a fine tuning of the method towards the specific data sets. Diameter quantification is carried out for both the phantom and patient data, and with both types of speed functions – the global and local speed functions \(F_G\) and \(F_L\), as was defined in Section 2.2. This yields a total of four optimization experiments.

In case of the phantom data, the bone masking stage can be skipped. The seed points are placed on the vessel segments that are mentioned in Section 3. The diameters of the phantom vary in the range of \([1.0, \ldots ,3.5]\) mm. Given the measurement (vector) \(M\) and the phantom specifications, or ground truth information (vector) \(P\), we plot the scalar \(P - M(p, \epsilon)\), i.e., the mean difference between ground truth and measurements as function of \(p\) and \(\epsilon\), for a wide range of settings of \(p\) and \(\epsilon\). After finding the optimal value for \(p\), the limits of agreements are plotted for this particular value of \(p\) as a function of \(\epsilon\). This pair of experiments is carried out once for the local speed function \(F_L\) and once for the global speed function \(F_G\). In case of the patient data, five patients are randomly selected to serve as training set. Seed points are placed on the vessel segments that are mentioned in Section 3. Ground truth \(P\) is provided by the measurements of one observer, i.e., \(P \equiv O\). The method’s measurements \(M\) are optimized to \(P\) by plotting

---

\(^2\) Strictly speaking, simply resampling the level set image by some interpolation scheme, e.g., trilinear interpolation as is done here, is invalid, since that can cause the zero level set to move (although minimally) seen from world coordinates. A possible solution would be to resample the input image before constructing the speed function, but that would result in such an increase in running time of the level set, that the small loss in accuracy is taken for granted.
\[ P - M(p, \epsilon), \text{i.e., the mean difference between ground truth and measurements. Again, after finding the optimal value for } p, \text{ the limits of agreements are plotted for this particular } p \text{ as function of } \epsilon. \text{ And again, these experiments are carried out twice, once for the local speed function } F_L \text{ and once for the global speed function } F_G. \]

5.4. Performance study

After defining the optimal values for \( p \) and \( \epsilon \) for both the phantom data and the patient training set, and after selecting the best performing type of speed function, a performance assessment study is conducted. First, for the phantom study. The method is initialized twice by the same observer (results \( M_1 \equiv M \) and \( M_2 \)), manual measurements are carried out by two radiologists (\( O_1 \) and \( O_2 \)), of which one is repeated (\( O_1^\prime \equiv O_1 \) and \( O_2^\prime \)), the mean measurement by the observers is given by \( O_{\text{mean}} \equiv \frac{1}{2} (O_1 + O_2) \). The agreements with ground truth \( P - M \) and \( P - O_{\text{mean}} \), the method and observers agreements \( M - O_{\text{mean}} \), the inter-observer agreement \( O_1 - O_2^\prime \), the intra-observer agreement \( O_1^\prime - O_2^\prime \) and finally, the method's reproducibility \( M_1 - M_2 \) are determined. For the patient study, the optimal parameters are applied to the remaining five patients. The diameter measurements are carried out twice by different observers (\( O_1 \) and \( O_2 \)) and carried out twice by the automated method (\( M_1 \) and \( M_2 \)) which is initialized by the same observer. The agreements between observer and method \( O_1 - M_1 \) and \( O_2 - M_2 \), the inter-observer variability \( O_1 - O_2 \), and the method's reproducibility \( M_1 - M_2 \) are determined.

6. Results

6.1. Bone masking

All masked images are evaluated by visual inspection, with special attention given to the skull base. One out of 10 shows that some small regions of the skull base are not properly masked, which is caused by the low intensity values of that region, lower than the bone threshold \( T_{\text{bone}} \). Three out of 10 results show that some vessel voxels near the skull base are also masked, mainly caused by the dilation step to compensate for inaccuracy by registration in the masking stage (see for an example Fig. 3). However, these are minor errors, and in general we have found that bone masking yields good results. The results of bone masking segmentation for three patients are shown in Fig. 7, first column.

6.2. Segmentation

The segmentation results for three patients are shown in the second and third columns in Fig. 7. Column two show the segmentation results when the level set is evolved with the optimal parameters found for diameter quantification, i.e., \( p = 0.79 \) and \( \epsilon = 0.5 \). Although these settings are optimal for diameter quantification, this curvature weighting is too strong if the goal is to segment a large part of the vascular tree. The same experiments are run again, but now without curvature, and the results are shown in column three. Note that running a level set without curvature influence, i.e., without internal geometrical forces, can be considered equivalent to a region growing algorithm. It is actually a refined approximation since region growing works voxel based, while the level set has sub voxel accuracy. Observe that now more vessels of the cerebral vasculature are captured, including the much smaller ones and the ones with degraded intensity values. Still, the method has difficulties in segmenting the vessels distal to the CoW, i.e., at the region of the peripheral cerebral vasculature.

6.3. Parameter optimization \( p, \epsilon \)

The optimization results of the phantom study with local speed functions are shown in Fig. 8. Plotted in (a) is the mean difference (bias) between the phantom specifications (ground truth) \( P \) and the method measurements \( M \), for the ranges \( p \in [0.71, 0.75] \) and \( \epsilon \in [0.0, 3.0] \), with step sizes 0.01 and 0.1, respectively. Each point on a graph corresponds to 17 measurements on the phantom. Parameter \( p \) allows for very accurate bias adjustment of the method, and for increasing curvature a small decreasing trend in bias can be observed. It can be observed that for a large range of settings, the bias is extremely small, much smaller than the in plane voxel size of 0.3125 mm. Based on this graph \( p \) is set \( p = 0.73 \). Plotted in (b) are the limits of agreements for this \( p \) as function of \( \epsilon \). The limits of agreement remain constant, except for the extreme value \( \epsilon = 3.0 \). Based on both graphs, \( \epsilon \) is set \( \epsilon = 0.4 \), and for these settings \( P - M = 0.00 \pm 0.23 [-0.44, 0.44] \).

The optimization results of the patient training set with local speed functions are shown in Fig. 9. Plotted in (a) is the bias between the method \( M \) and one observer \( O \) which is considered ground truth \( P \equiv O \). The curvature varied in range \([0.0, 3.0]\) with step size 0.1, the fraction \( p \) varied in range \([0.74, 0.78]\) with step size 0.01; each point on a graph corresponds to a total of 60 diameter measurements. Fraction \( p \) allows for bias adjustment, and for \( \epsilon < 2.3 \) the bias is smaller than 0.1 mm, which is smaller than the in plane resolution of 0.3125 mm. For values \( \epsilon > 2.0 \) greater fluctuations are observed in the bias. Based on this graph, \( p \) is set \( p \equiv 0.76 \). Plotted in (b) are the limits of agreements for this \( p \) as function of \( \epsilon \). For \( \epsilon < 2.2 \) the limits of agreements remain constant, for \( \epsilon > 2.2 \) the limits of agreements dramatically increases; the minimum was found for \( \epsilon \equiv 0.4 \). The final optimization result is \( P - M \equiv O - M = 0.00 \pm 0.11 [-0.56, 0.56] \).

These series of optimization experiments for both the phantom study as well as the patient training set are repeated for the global speed function \( F_G \).

We refrain from giving all the corresponding plots for readability and fulfill by mentioning the optimal results...
of the final settings only, which are summarized in Table 1. Clearly, the best findings, for both the phantom as well as the patient training set, are with the local speed functions $F_L$. Together with the settings for $p$ and $\epsilon$ mentioned above, they will be applied in the performance assessment study in Section 6.4.
After masking, an image based speed function needs to be defined by Venema et al. (2001). Masking is used instead of subtraction to preserve the signal to noise ratio of the image. The major disadvantage of masking is the additional acquisition of a non-contrast enhanced scan, increasing radiation dose for the patient by approximately one third, i.e., the tube current settings are 440 mA s for the contrast enhanced versus 130 mA s for the non-contrast enhanced scan. However, in a recent paper by van Straten et al. (2004) it is shown that the mA s setting of the non-contrast enhanced scan. However, in a recent paper by van Straten et al. (2004) it is shown that the mA s setting of the non-contrast enhanced scan can even be reduced to 65 mA s without a loss of image quality.

After masking, an image based speed function needs to be defined that is used for steering the level set evolution. This speed function is completely based on parameters derived from the intensity histogram. We distinguish two speed functions, a local speed function $F_L$ which is constructed based on a local VOI in the masked image, and a global speed function $F_G$ which is constructed based on the complete masked image. In our approach we separate the extraction of background and vessel information, and the complete masked image. In our approach we separate the extraction of background and vessel information, and

6.4. Performance study

From the results from the optimization study (Table 1), we learn that for both the phantom and patient data the local speed function achieves higher accuracy than the global speed function. Therefore, in the performance assessment study, the local speed function has been selected. The performance results are summarized in Table 2. Corresponding Bland and Altman plots for the phantom study are given in Fig. 10, and for the patient study in Fig. 11.

7. Discussion and conclusion

(Semi-)automated segmentation of cerebral vasculature in CT angiography data sets is an important task. It enables visualization and diameter quantification, which are both important for diagnostic purposes. In this work a level set based method is proposed. We believe that the level set framework is well suited for vessel segmentation, since it has the flexibility of handling the large morphological variations these vessel structures exhibit. In CTA data sets vessel segmentation is mainly hampered by the presence of bone structures, especially in the area around the skull base where the CoW is situated. In this work, a solution is found by bone masking, a technique that was introduced by Venema et al. (2001). Masking is used instead of subtraction to preserve the signal to noise ratio of the image. The major disadvantage of masking is the additional acquisition of a non-contrast enhanced scan, increasing radiation dose for the patient by approximately one third, i.e., the tube current settings are 440 mA s for the contrast enhanced versus 130 mA s for the non-contrast enhanced scan. However, in a recent paper by van Straten et al. (2004) it is shown that the mA s setting of the non-enhanced scan can even be reduced to 65 mA s without a loss of image quality.

After masking, an image based speed function needs to be defined that is used for steering the level set evolution. This speed function is completely based on parameters derived from the intensity histogram. We distinguish two speed functions, a local speed function $F_L$ which is constructed based on a local VOI in the masked image, and a global speed function $F_G$ which is constructed based on the complete masked image. In our approach we separate the extraction of background and vessel information, and

Table 1
Optimization of the method’s measurements $M$ to ground truth $P$ concerning the phantom data and the patient training set

<table>
<thead>
<tr>
<th></th>
<th>Phantom data (ground truth)</th>
<th>Patient training set</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P - M$</td>
<td>$-0.01 \pm 0.33 [-0.67, 0.64]$</td>
<td>$-0.01 \pm 0.32 [-0.63, 0.62]$</td>
</tr>
<tr>
<td>$P - \hat{M}$</td>
<td>$0.00 \pm 0.23 [-0.44, 0.44]$</td>
<td>$0.00 \pm 0.28 [-0.56, 0.56]$</td>
</tr>
</tbody>
</table>

Ground truth for the patient data are given by the measurements of one observer, $P \equiv O$. A global speed functions $F_G$ and local speed functions $F_L$ are compared.

Table 2
Performance study

<table>
<thead>
<tr>
<th></th>
<th>Phantom study</th>
<th>Patient study</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P - M$</td>
<td>$0.00 \pm 0.23 [-0.44, 0.44]$</td>
<td>$0.13 \pm 0.32 [-0.50, 0.76]$</td>
</tr>
<tr>
<td>$P - \hat{M}$</td>
<td>$-0.32 \pm 0.23 [-0.77, 0.14]$</td>
<td>$0.23 \pm 0.32 [-0.39, 0.85]$</td>
</tr>
<tr>
<td>$M - \hat{M}$</td>
<td>$-0.31 \pm 0.34 [-0.98, 0.36]$</td>
<td>$-0.04 \pm 0.17 [-0.36, 0.29]$</td>
</tr>
<tr>
<td>$O_1 - O_2$</td>
<td>$-0.14 \pm 0.19 [-0.51, 0.24]$</td>
<td>$0.03 \pm 0.09 [-0.16, 0.21]$</td>
</tr>
<tr>
<td>$M_1 - M_2$</td>
<td>$0.02 \pm 0.14 [-0.26, 0.31]$</td>
<td>$0.06 \pm 0.17 [-0.26, 0.39]$</td>
</tr>
</tbody>
</table>

These results are obtained by running the level set with the local speed functions (see Table 1). For both the phantom and patient data, two series of measurements are obtained by two expert observers ($O_1$ and $O_2$). The method is initialized twice by seed points given by the same observers ($M_1$ and $M_2$).
as a consequence, we implicitly circumvent the modeling of partial volume voxels. Modeling the partial volume voxels becomes an increasing burden especially when the prior selection of vessel and background voxels results in only a few voxels while the voxels at the transition vessel – background are dominant. Examples of how to model partial

Fig. 10. Bland and Altman plots concerning the phantom study. (a) Method’s reproducibility; (b) agreement between method and observers; (c) agreement between ground truth and method; and (d) agreement between ground truth and observers.

Fig. 11. Bland and Altman plots concerning the patient study. (a) Agreement between first observer and method; (b) agreement between second observer and method; (c) inter-observer variability; and (d) method’s reproducibility.
volume voxels can be found in Glasbey and Robinson (1999, 2002) for CT, and Laidlaw et al. (1998) and Shattuck et al. (2001) for MRI.

Also, the speed function depends on the seed points and therefore the seed points influence indirectly the diameter measurements. This influence is minimal, as can be learned from Table 2. The method’s reproducibility is 0.03 ± 0.09 mm for the phantom data and 0.06 ± 0.17 mm for the patient data. Note that these numbers are obtained by asking different observers to point at the location of the vessel segment denoted only by the vessel segment’s name (e.g., ICA, PcoA, etc.), leaving much room for variability for the seed point locations. Thus, in spite of this variability, the method achieves fairly good and acceptable reproducibility. Still, an improvement can be made by making the initial threshold guess more robust, by e.g., choosing a larger region of interest around the seed point, finding maximum intensity values along the central vessel axis.

For both phantom data and patient data, local speed functions \( F_L \) around each seed point – which marks a point within the vessel where a diameter measurement should be carried out, outperform one global speed function \( F_G \) with respect to accuracy. It conforms to the hypothesis that intensity values in CTA data sets vary locally, influencing the accuracy by which diameter quantification can be performed. This observation is not always clear by a visual inspection of the data sets. For instance consider the MIPs of the masked images in Fig. 7; at first glance the intensity values of the vessels forming the CoW seem almost constant. Therefore, region based level set methods, compensating for intensity variations along the vessels, such as described in Chan and Vese (2001) and Paragios and Deriche (2002) are required.

A second important aspect which can be observed from the optimization experiments concerns the relevance of curvature weighting during level set evolution. As the graphs in Figs. 8 and 9 show, the influence on the bias and limits of agreements is minimal, except for extreme high values of \( \epsilon \). Running the level sets with zero curvature would yield roughly the same results. For segmentation purposes of the cerebral vasculature, including curvature makes the level set to miss the smaller vessels, distal and proximal, see Fig. 7, second column. It is clear that the optimal parameters for diameter quantification do not necessarily yield the best segmentation results. The motivation for using curvature lies in its ability to prevent the occurrence of sharp corners, the ability to steer the level set in a preferred direction of tube like structures by taking the minimal principal curvature component of the 3D curvature, and preventing leakage into background.

However, if the influence of curvature seems negligible on the accuracy, it is tempting to see what accuracy can be obtained by running a mere region growing algorithm, after bone masking. For comparison, we implemented the seeded region growing algorithm by Adams and Bischof (1994). The algorithm starts from a set of seed points, and includes those voxels if the difference between the intensity value of the voxel and the average intensity value of the segmented region is smaller than some threshold. We extended the method by including a range test which simply includes only those voxels if their value lies between the range 200 and 600 HU. We made the lower bound depend on the average intensity value that is found locally (around one initial seed point) or globally (around multiple initial seed points). Again, this intensity value is lowered by some fraction \( p \). We did experiments on the patient data sets – first on the training set of five patients to find \( p \), then on the remaining five patients. The results are summarized in Table 3 and put next to their corresponding level set results. The level set method gives better results. This is due to the fact the level set has sub voxel accuracy, while region growing works voxel based. Sub voxel accuracy is another important motivation for the use of level set based methods.

Optimization of the level set method measurements to ground truth yields the following results: 0.00 ± 0.23 mm for the phantom data and 0.00 ± 0.28 mm for the training set, which is smaller than the voxel size in the \( z \)-direction of 0.5 mm and can therefore be considered as good results. The optimal settings are then applied in a performance assessment study.

In the performance assessment study, the goal is to gain insight in the method performance with respect to observer variability and reproducibility. The phantom data, representing ground truth, tells us that the method and an expert observer achieves the same level of accuracy, except that the method is non-biased whereas the observer is. Although the inter- and intra-observer variability are remarkably good \((−0.14 ± 0.19 \text{ and } 0.02 ± 0.14 \text{ mm, respectively})\), the method reproducibility is even better \(0.03 ± 0.09 \text{ mm}\). Turning our focus towards the remaining patient data, the following observations are made.

<table>
<thead>
<tr>
<th>Region growing</th>
<th>Training</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>−0.01 ± 0.38 [−0.75,0.74]</td>
<td>0.12 ± 0.40 [−0.66,0.90]</td>
</tr>
<tr>
<td>Local</td>
<td>0.00 ± 0.38 [−0.75,0.74]</td>
<td>0.11 ± 0.37 [−0.61,0.84]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level set</th>
<th>Training</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>−0.01 ± 0.12 [−0.63,0.62]</td>
<td>–</td>
</tr>
<tr>
<td>Local</td>
<td>0.00 ± 0.28 [−0.56,0.56]</td>
<td>0.13 ± 0.32 [−0.50,0.76]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.23 ± 0.32 [−0.39,0.85]</td>
</tr>
</tbody>
</table>

Table 3 The results of the region growing algorithm applied to patient data and compared to the level set results

The region growing algorithm is a modified version of the seeded region growing algorithm (Adams and Bischof, 1994) such that an intensity range test is included. The lower bound of this range can be set depending on the initial average intensity value, which can be found locally (around one seed point) or globally (around multiple seed points). During training, the local speed function gave the best results (Table 1) and hence this speed function is selected for the performance study. In the performance study for the region growing method, the measurements are carried out by one observer. In the performance study for the level set based method, the measurements are carried out by two observers – see also Table 2. The level set method gives better accuracy.
The agreement between the method and observers is similar to the agreement between the method and observer for the phantom data. However, by the lack of ground truth we cannot tell which measurements are correct or wrong. A critical view at the observer’s procedure for manually measuring the diameters reveals the following possible sources of errors: the usage of a 2D slab MIP may introduce an error if the vessel is non-circular, the line which should be perpendicular to the vessel axis, the end points of the line on both sides of the vessels and the fixed window width and level setting. In spite of these potential pitfalls, the expert observers are quite consistent with each other, as the inter-observer variability –0.04 ± 0.17 mm indicated us. The inter-observer variability is very similar to the method’s reproducibility (0.06 ± 0.17 mm). Because of this, and because of the good performance of the method with respect to ground truth, we conclude that the proposed method has the potential to replace the manual procedure for diameter quantification.

References


