# Automated image segmentation and registration of vessel wall MRI for quantitative assessment of carotid artery vessel wall dimensions and plaque composition

Klooster, R. van 't

# Chapter 3

# Carotid wall volume quantification from magnetic resonance images using deformable model fitting and learning-based correction of systematic errors

*This chapter was published in:*

K. Hameeteman, R. van 't Klooster, M. Selwaness, A. van der Lugt, J. C. M. Witteman, W. J. Niessen, S. Klein. Carotid wall volume quantification from magnetic resonance images using deformable model fitting and learning-based correction of systematic errors, *Physics in Medicine and Biology*, Volume 58, Pages 1605–1623, 2013.

**Abstract**

We present a method for carotid vessel wall volume quantification from Magnetic Resonance Imaging (MRI). The method combines lumen and outer wall segmentation based on deformable model fitting with a learning-based segmentation correction step. After selecting two initialization points, the vessel wall volume in a region around the bifurcation is automatically determined. The method was trained on 8 datasets (16 carotids) from a population based study in the elderly for which one observer manually annotated both the lumen and outer wall. Evaluation was done on a separate set of 19 datasets (38 carotids) from the same study for which two observers made annotations. Wall volume and normalized wall index measurements resulting from the manual annotations were compared to the automatic measurements. Our experiments show that the automatic method performs comparably to the manual measurements. All image data and annotations used in this study together with the measurements are made available through the website `http://ergocar.bigr.nl`.

## 3.1 Introduction

The prevalence of cardiovascular diseases (CVD) is rising and heart disease is the leading cause of death in the western world, claiming approximately one out of every five lives [63]. Atherosclerosis, a disease of the vessel wall, is the primary cause of cardiovascular disease. Atherosclerotic wall thickening in the carotid arteries can cause a narrowing or total occlusion of the lumen. Atherosclerotic plaque that does not cause occlusion may still lead to clinical events because of rupture and development of thromboembolism, which may subsequently lead to cerebral ischemia [64]. Consequently, much research is aimed at finding parameters that describe the plaque, and which can be used to improve risk stratification and for monitoring the progression of atherosclerotic disease. One of those parameters is the size of the plaque and its relation to the size of the vessel.

Magnetic Resonance Imaging (MRI) is an important means to monitor and quantify the state of the vessel wall and lumen. Several studies have shown the possibility to visualize the vessel lumen and outer wall on MRI [50, 65–69]. Annotation of both the lumen border and the outer vessel wall is a laborious task. Therefore, several researchers have proposed automatic wall segmentation methods e.g. [18, 31]. Although calculating the volume of the vessel wall is straightforward once a segmentation is made, a comparison between automatic and manual wall volume measurements has, to the best of our knowledge, not yet been done.

In this paper we present an automatic method to measure carotid wall volume from MR images. The contribution is five-fold:

1. The automatic method combines a deformable model approach [18] with a learning-based postprocessing step, in which systematic segmentation errors of the deformable model fitting are corrected. The idea behind this segmentation correction was developed by Wang et al [70]. However, whereas the method described in [70] was designed for brain structure segmentation, we modified it such that it can handle vessel-shaped structures.

2. An intensity inhomogeneity correction method is designed to compensate for the nonuniform sensitivity pattern of the RF surface coils.

3. A training set of eight subjects with manual annotations is used to exhaustively optimize a number of algorithm parameters.

4. The final algorithm is evaluated on a different set of 19 subjects, and the automatic results are compared with manual vessel wall volume measurements and inter-observer variability.

5. All image data used in this study together with the manual and automatic measurements are made available through the website `http://ergocar.bigr.nl` (user/pwd: reviewer/PMBVesselWall).

The remainder of this paper is organized as follows. Firstly, Section 3.2 explains the deformable model fitting and learning-based segmentation correction, and the wall volume quantification. Section 3.3 describes the image data and the preprocessing steps. Section 3.4 describes the experiments that were conducted to optimize the parameters of the method and to evaluate the wall volume quantification. The results of the experiments are given in Section 3.5. Discussion and conclusion are given in Sections 3.6 and 3.7 respectively.

## 3.2 Methods

### Deformable model fitting

The method described by Van 't Klooster et al [18] was used to create an initial segmentation of the lumen and outer wall. This method requires a MR Angiography (MRA) image and a Black Blood (BB) image. The MRA image is used to obtain a robust initial segmentation of the carotid lumen. This segmentation is copied to the BB image and based on this initialization, the lumen and outer wall are segmented. The MRA and BB images are assumed to be co-registered (see Section 3.3). The only required user input consists of two initialization points. Based on these, the method fully automatically fits a deformable surface model of the lumen and the outer wall to the image data. Below, a brief description of this method is given which is summarized in Figure 3.1. For further details we refer to [18].

The method starts with a two-point initialization: one in the common carotid artery (CCA) and one in the internal carotid artery (ICA). These two points are used to initialize a wavefront propagation [51] in the MRA image which results in an approximate centerline with an associated lumen diameter estimate (green curve in frame 1 of Figure 3.1) at each centerline position. This centerline together with the diameters is used to initialize a NURBS surface [51] (blue surface lines in frame 2). The method then performs an optimization step in which the boundaries of the NURBS surface are more precisely fit to the lumen boundaries in the MRA image by searching for the maximum gradient magnitude in an intensity profile with a specified length ($dL_{\mathrm{MRA}}$) perpendicular to the lumen surface (frame 3 shows the location of a profile and frame 4 the corresponding intensities along the profile; frame 5 shows an update of the initial contour). Only those edges are selected which have a high value (bright) inside the model and a low value (dark) outside the model. This optimization step is performed for a given number of iterations ($nL_{\mathrm{MRA}}$).

The lumen NURBS surface found in the MRA image is used to initialize the lumen segmentation in the BB image (frame 6). Here, the same optimization of the surface fit to the lumen boundary is performed as in the MRA image (frame 7 and 8), using again a specified length ($dL_{\mathrm{BB}}$) of the intensity profile and number of iteration steps ($nL_{\mathrm{BB}}$). The edge direction in this step is set from dark inside the model to bright outside the model.

The estimated lumen surface is expanded and used to initialize the search for the outer wall. This is done in the same manner as for the BB lumen segmentation, by searching for the maximum gradient magnitude in an intensity profile with a specified length ($dW_{\mathrm{BB}}$) for a number of iterations ($nW_{\mathrm{BB}}$). For this segmentation step, the edge direction is set from bright inside the model to dark outside the model.

### Learning-based correction of systematic errors

Using a deformable model to acquire an automatic segmentation limits the precision of the final segmentation to the flexibility of the used (NURBS) model. In areas where there is a sudden change in the shape of the vessel, this may lead to errors. Also, the deformable model approach described in the previous section was originally designed and optimized for one particular set of images. To compensate for the limitation of the deformable model and to adjust for the differences in image characteristics when using a different scanner, surface coil, and/or different MR acquisition parameter settings, we propose to use a post-processing step in which the segmentation is tailored to a new set of images. The idea behind this method was developed by Wang et al [70]. We will refer to this method as the

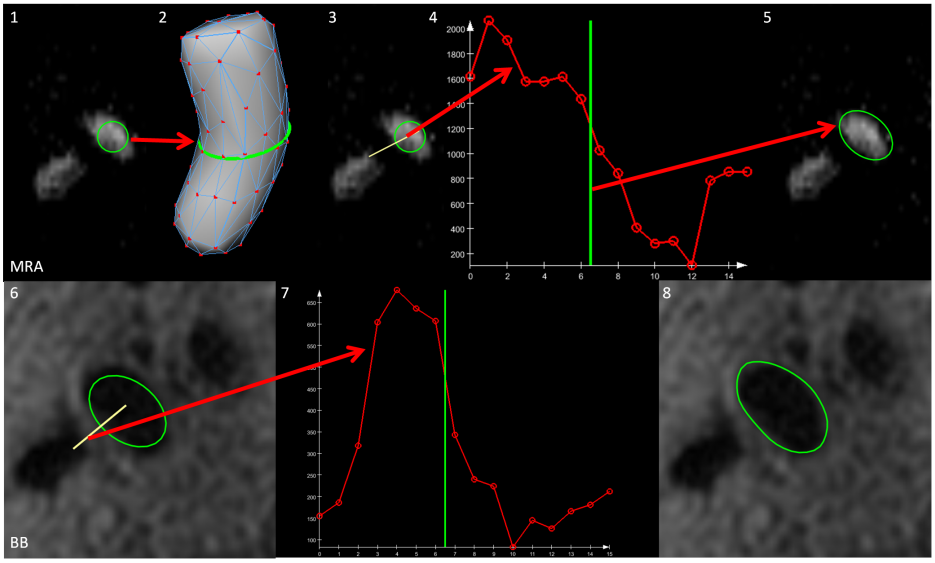



Figure 3.1: Graphical summary of the deformable model fitting method. An MRA segmentation based on wavefront propagation (1) is used to initialize a NURBS surface (2), which is fit the to lumen boundary by searching for a maximum gradient magnitude (4) along an intensity profile (3). The resulting segmentation is used as initialization in the BB image (6) where it is optimized (7,8) in the same manner as is done in the MRA image.

Learning-Based Segmentation Correction (LBSC) method. As the LBSC method was designed for brain structure segmentation, we modified it to be able to handle vessel structures. We will refer to this modified method as the Learning-Based Vessel Segmentation Correction (LBVSC) method.

The LBSC method uses a ground truth segmentation (in our case manual annotations) to train a classifier that modifies a binary segmentation made by a host segmentation method (in our case the result from the deformable model fitting described in Section 3.2). The LBSC method classifies all voxels in a region of interest (ROI), which is created by dilating the host segmentation. For each voxel, a feature vector is computed, consisting of:

- all image values in a 5x5x5 neighborhood of the voxel in the MR image (125 features)
- all values in a 5x5x5 neighborhood of the voxel in the host segmentation (125 features)
- x, y and z coordinate of the voxel relative to the center of mass of the host segmentation (3 features)
- product of the neighborhood and coordinate features ($3 * (125 + 125)$ features)

This leads to a total of 1003 features. The LBSC method trains a classifier on the difference between the host segmentation and the manual annotation. It thus learns to correct the errors made by the host segmentation. AdaBoost is chosen as classifier, which combines many weak classifiers (regression stumps) into a single strong classifier [71]. The

AdaBoost classifier has previously been shown to be successful in the context of medical image segmentation [72, 73].

For our application, a number of modifications are proposed: We use the signed distance to the host segmentation's border and the z-index as spatial features. Since the carotid artery is roughly perpendicular to the transversal plane, this latter feature is an approximation of the position along the centerline. For elongated structures like vessels, these two spatial features seem more appropriate than the distance to the center of mass. In addition to these spatial features we used the intensity of the BB image and its gradient magnitude. Instead of using a 5x5x5 neighborhood we used a 7x7x3 neighborhood. The effect of this neighborhood size, ($fL_x \times fL_y \times fL_z$ and $fW_x \times fW_y \times fW_z$ for the lumen and outer wall feature neighborhood respectively) on the segmentation accuracy was determined experimentally (see Section 3.4). Thus the two spatial features of the LBVSC method are:

- signed distance to the boundary host segmentation (1 feature)
- relative z-index (1 feature)

The 441 appearance features are the 7x7x3 neighborhood voxels of:

- the host segmentation (i.e. 0 or 1) (147 features)
- the BB image (147 features)
- the gradient magnitude of the BB image (147 features)

Similar to the LBSC method we also used the product of the spatial and appearance features as additional features $(2 * (147 + 147 + 147)) = 882$, leading to a total of $2 + 441 + 882 = 1325$ features.

Since most errors in the host segmentation are near the boundary of the segmentation, our LBVSC method only classifies the voxels within a ROI around this boundary. This region is defined by a morphological dilation minus an erosion of the binary host segmentation using a spherical kernel with radius $rL_{\text{dilation}}$ and $rL_{\text{erosion}}$ for the lumen and $rW_{\text{dilation}}$ and $rW_{\text{erosion}}$ for the outer wall segmentation. The binary host segmentation of the outer wall includes the lumen area.

When the classifier is trained on the difference between the host segmentation and the manual annotation (as in the original LBSC method), only the voxels that are not correctly segmented in the host segmentation get a negative label. If the host segmentation has large overlap with the manual annotation, this leads to many positive samples and only a few negative samples. To prevent this unequal class sizes we trained the AdaBoost classifier directly on the label from the manual annotation, which leads to a better balance between the classes. To make the classification tractable, only half of the randomly selected voxels within the ROI are used as samples.

Whereas the output of the deformable model fitting described in Section 3.2 is smooth, the output of the LBVSC method may have holes or isolated voxels. Therefore a morphological closing with a kernel radius of 2 voxels is applied to the output of the LBVSC method and isolated voxels are removed using connected components analysis.

### Vessel wall volume quantification

The vessel wall volume $V_{\text{wall}}$ is quantified in a region of 25 mm in the transversal direction centered at the bifurcation. The bifurcation point is manually annotated and defined as the first transversal plane on which two separate lumens (one of the ICA and one of the External Carotid Artery (ECA) ) are visible. In cases where part of the evaluation region falls outside the scan range, the evaluation region is restricted by the image boundaries.

Application of the LBVSC method results in a binary mask for the lumen and the outer wall. The difference of these two masks is defined as the vessel wall. The vessel wall volume is computed by voxel counting and multiplying the result with the volume of one voxel.

Besides the volume, a clinically used parameter to quantify the vessel wall $V_{\text{wall}}$ is the Normalized Wall Index (NWI) which is defined as:

$$\text{NWI} = \frac{V_{\text{wall}}}{V_{\text{lumen}} + V_{\text{wall}}} \tag{3.1}$$

where $V_{\text{lumen}}$ is the volume of the vessel lumen. Bigger plaques thus lead to a higher NWI.

## 3.3 Data specific preprocessing

### Data description

The image data of this study is taken from a prospective, population-based study among subjects aged 45 years and older. This study has been described in detail elsewhere [74]. All participants having a maximum intima-media thickness of more than 2.5 mm (determined using Ultra Sound [75]) in at least one carotid artery were invited for a carotid MRI exam. In total, 1072 participants were scanned.

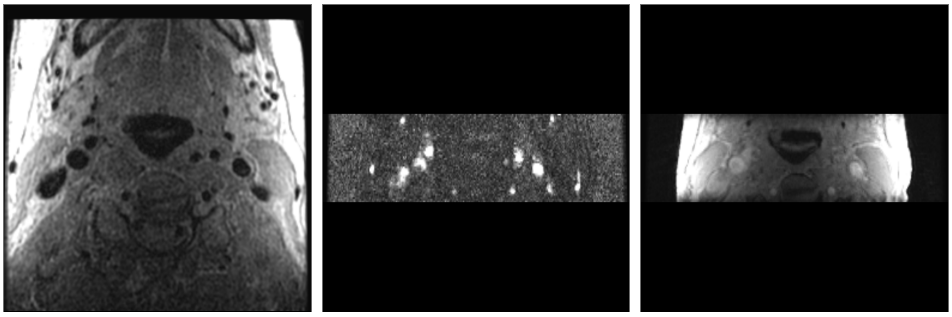
MRI of the carotid arteries was performed on a 1.5-T MR scanner (Signa Excite, GE Healthcare, Milwalkee, USA) with a bilateral phased-array surface coil. To reduce motion artifacts, subjects were stabilized in a custom-designed head holder. The total scanning time was about 30 minutes in which, among others, the following sequences were acquired (see [75] for acquisition details):

- Proton density weighted Fast Spin Echo Black-blood (BB)
- Phase Contrast sequence which consists of an image showing flow in any direction (PC), a magnitude image (PCMag) and 3 images for the flow in the x, y and z direction.

The PC is used as MRA image for the deformable model fitting (see Section 3.2) and the PCMag image is used to register the PC image to the BB image (see Section 3.3). Figure 3.2 shows an example slice of each of these MR sequences.

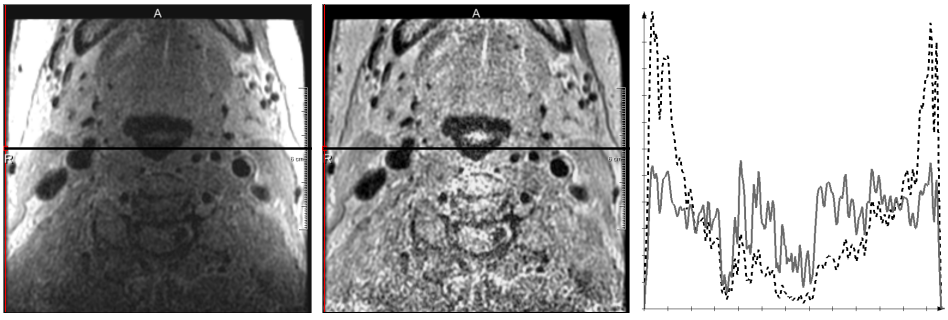
### Inhomogeneity correction

The bilateral phased-array surface coils cause severe intensity inhomogeneity within the neck. The sensitivity near the skin is much higher than in the middle of the neck. Figure 3.3c shows a typical example of the intensity profile across the neck in a BB image. Many popular intensity inhomogeneity correction methods have strong assumptions on



(a) BB (b) PC (c) PCMag

Figure 3.2: Example slice near the bifurcation of the three different MR sequences used in this study.



(a) (b) (c)

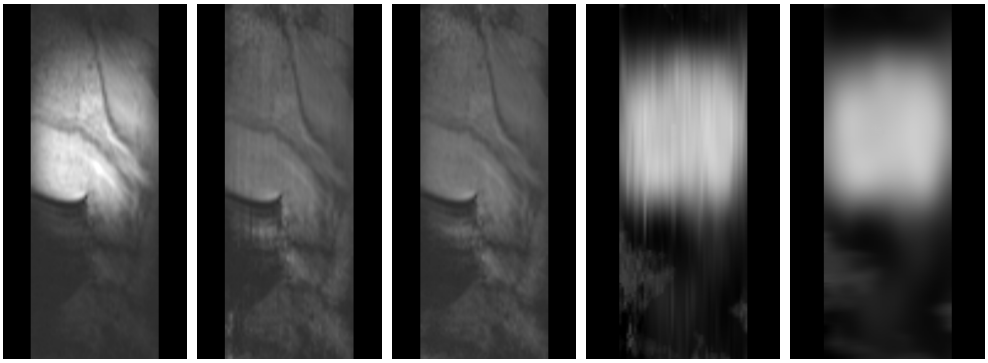
Figure 3.3: Example of the intensity inhomogeneities caused by a surface coil (a), the same image after applying the LEMS method (b) and the intensity profiles of the black line in the original (black, dashed line) and the corrected (gray, solid line) image (c).

the distribution of the modeled bias field. For example, N3 [76] and N4 [77] assume the distribution of the bias field intensities to be log-normal. Other methods like the one proposed by [78] and [79] assume a slowly varying bias field. These methods do not result in a satisfactory correction of the intensity inhomogeneities in the case of phased-array surface coils. The method proposed by [80], called Local Entropy Minimization with a bicubic Spline model (LEMS), was designed to deal with inhomogeneities caused by phased array surface coils.

The original LEMS method was designed for 2D images. To ensure a smooth bias field across slices, we extended the method to 3D with the following modifications.

LEMS tries to identify the voxels that do not have any signal as they cannot be used to estimate the bias field. We estimated the background voxels using the whole 3D image. Furthermore, the 2D slice based estimation of the bias field is smoothed in the slice direction using Gaussian blurring with a $\sigma$ of 3 slices.

This way a spatially smooth bias field is obtained as can be seen in Figure 3.4.



(a) original (b) LEMS (c) LEMS 3D (d) Bias field (e) 3D Bias field

Figure 3.4: Inhomogeneity correction using LEMS and smoothing the estimated correct field. Sagital slice of (a) the original BB image, (b) after correction using standard LEMS (c), using LEMS with smoothed bias field, (d) estimated bias field of the standard LEMS method and (e) estimated bias field with smoothing in the slice direction.

Because the LEMS method uses a multiplicative bias field model, the correction is performed by dividing the original image by the estimated bias field. This can lead to very large intensity values in the corrected image in regions where the estimated bias field has small values. The registration between the BB and MRA image (see Section 3.3) is hampered by a few extremely high intensity values within the image, therefore the highest 1% of the intensities of the corrected image are clamped. Figure 3.3a and 3.3b show an example of the BB image before and after inhomogeneity correction, respectively, and the effect of the correction on the intensity profile can be seen in Figure 3.3c.

### Registration

As stated in Section 3.3, the complete MRI exam takes approximately 30 min. Although the position of the head of the subjects is stabilized using a head-holder, they still have the ability to move their body, which can cause twisting of the neck. Moreover, cardiac and breathing motion can also lead to displacement of the arteries in the neck. This motion leads to small mis-registrations between the different sequences. The lumen and vessel wall segmentation method described in Section 3.2 requires a registered BB and MRA image. We perform an intensity-based registration method to align the images. The PC image contains very little anatomical information as can be seen in Figure 3.2b and is therefore not very well suited for an intensity based registration. The PCMag image (shown in Figure 3.2c) which is simultaneously acquired with the PC image contains much more anatomical information. Therefore the PCMag image was registered to the BB image and the resulting deformation was applied to the PC image.

The registration first performs a rigid registration step to compensate for global displacements. Because the BB image and the PCMag image contain different anatomical regions, a mask was used for both the fixed and the moving image to indicate the regions where there should be overlapping information. The rigid registration was followed by a B-Spline registration [81] using mutual information as similarity metric [82]. For the

optimization we used an adaptive stochastic gradient descent optimizer [83]. The registration was performed in a multi-resolution framework with three resolution levels and was achieved using the ITK [84] based registration toolbox `elastix` [85].

## 3.4 Experiments

For the experiments, 27 subjects were randomly selected from the full database. This evaluation set was split in a training set of 8 subjects and a test set of 19 subjects. The training set was used to optimize the parameters of the deformable model fitting ($dL_{\text{MRA}}$, $nL_{\text{MRA}}$, $dL_{\text{BB}}$, $nL_{\text{BB}}$, $dW_{\text{BB}}$ and $nW_{\text{BB}}$) and the LBVSC method ($rL_{\text{dilation}}$, $rL_{\text{erosion}}$, $rW_{\text{dilation}}$, $rW_{\text{erosion}}$, $fL_{\text{x}} \times fL_{\text{y}} \times fL_{\text{z}}$ and $fW_{\text{x}} \times fW_{\text{y}} \times fW_{\text{z}}$). The test set was used to evaluate the vessel wall volume and NWI quantification.

### Manual annotations

On the complete set of 27 subjects the lumen and outer wall of both the left and the right carotid artery were annotated manually by observer 1 (ob1). On the test set observer 2 (ob2) performed the same annotations. The manual annotation started with an accurate definition of the centerline, after which longitudinal contours along this centerline were drawn in a curved planar reformatted image for both the lumen and the outer wall. These longitudinal contours were used to create cross-sectional contours perpendicular to the centerline. The cross-sectional contours were then adjusted to fit the lumen and the outer wall. More details on the annotation process can be found in [86]. The manual annotations were converted into binary masks for calculating the Dice [87] overlap coefficient and using the masks in the LBVSC step. The mask generation was achieved by fitting a surface through the contour points using variational interpolation [88] and voxelizing this closed surface. The mask of the outer wall contains everything within the outer wall including the lumen.

The manual annotations of the ICA do not incorporate the external part in the bifurcation region. Because the automatic method does not differentiate between ICA and ECA in the bifurcation region, the automatic segmentation may cover a larger part of the lumen area within the bifurcation region. This larger automatic segmentation is not wrong, but would result in errors during the evaluation because the manual ground truth does not contain the ECA. To make sure that this does not influence the evaluation results, the ECA region of the bifurcation was masked out and no evaluation was done in that region.

### Parameter optimization

#### Surface distance

Since there is no ground truth definition in the MRA sequence, the values of profile length $dL_{\text{MRA}}$ and the number of iterations $nL_{\text{MRA}}$ were chosen such that the resulting segmentation visually provided a good initialization for the BB lumen segmentation.

For the parameter optimization of the deformable model fitting in the BB image, a distance measure was computed between the NURBS surface and the manual contours. First, the surface was intersected with the MPR plane (perpendicular to the centerline) in which the contours were drawn. This created a contour for the automatic segmentation.

Then the symmetric average Euclidean distance between the manual and automatic contours was calculated. This distance was calculated for all manual contours and the average was computed for each carotid. The average over all carotid segmentations in the training set (denoted by $\delta L_{\text{surface}}$ and $\delta W_{\text{surface}}$ for the lumen and outer wall segmentation, respectively) was used as optimization objective.

The $dL_{\text{BB}}$, $nL_{\text{BB}}$, $dW_{\text{BB}}$ and $nW_{\text{BB}}$ parameters were selected by an exhaustive search, minimizing $\delta L_{\text{surface}}$ and $\delta W_{\text{surface}}$. First the lumen parameters were optimized and then the outer wall parameters. The value of $dL_{\text{BB}}$ was optimized between 5 and 25 mm in steps of 2 mm. The number of iterations $nL_{\text{BB}}$ was optimized over the following set: {5,10,15,25,50,75,100,150}. The outer wall parameter $dW_{\text{BB}}$ was optimized between 5 and 15 mm in steps of 2 mm and $nW_{\text{BB}}$ over the following set: {0,1,2,3,5,7,10,25,50,75}. Each combination of profile length and number of iterations was tested, leading to 88 lumen experiments and 60 outer wall experiments.

#### Learning-based vessel segmentation correction

The influence of the neighborhood size of the features, $fL_{\text{x}} \times fL_{\text{y}} \times fL_{\text{z}}$ and $fW_{\text{x}} \times fW_{\text{y}} \times fW_{\text{z}}$, on the resulting segmentation was determined for the following values: 3x3x3, 5x5x3, 5x5x5, 7x7x3, 7x7x5, 7x7x7, 9x9x3, 9x9x5, 9x9x7, 9x9x9. For each of these feature neighborhoods the optimal combination of $rL_{\text{dilation}}$ and $rL_{\text{erosion}}$, and $rW_{\text{dilation}}$ and $rW_{\text{erosion}}$ was determined. These LBVSC parameters were optimized with respect to the Dice overlap of the resulting segmentation with the manual segmentation. For each combination of neighborhood size dilation and erosion radius, a new classifier was trained. As the training of each classifier took several hours, performing cross-validation on the training set would take too much time. Therefore each classifier was trained on the complete training set and the evaluation of the resulting segmentation was also performed on the complete training set. This may lead to an overestimation of the performance of the LBVSC method. This is of minor importance, since these experiments are only meant to optimize the dilation and erosion parameters, and the true performance will be evaluated on the test set (see Section 3.5). The dilation and erosion radii were varied over the interval $[0,9]$ mm in steps of 1 mm. The AdaBoost classifier was trained using 400 iterations.

### Evaluation on test set

The algorithm was evaluated on the 38 carotids of the 19 datasets in the test set. We evaluated the final segmentation of the lumen and the outer wall by calculating the Dice coefficient for each. This evaluation was done both without the segmentation correction step, and applying the original LBVC method and our modification, the LBVSC method. The significance of the difference caused by applying the segmentation correction methods was tested using a paired t-test.

The vessel wall volumes and NWI from the manual annotations were computed in the same way as for the automatic segmentation (see Section 3.2). This was done for both observers that annotated the test set.

The manual and automatic wall volumes were compared by calculating the average difference between the two volumes, the average difference of the NWI and the standard deviation of these two difference measures. Also the Pearson and intra-class correlation [89] coefficient were calculated.

| $nL_{BB}$ \ $dL_{BB}$ (mm) | 5 | 7 | 9 | 11 | 13 | 15 | 17 | 19 | 21 | 23 | 25 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 0.84 | 0.77 | 0.73 | 0.70 | 0.69 | 0.69 | 0.71 | 0.90 | 0.91 | 0.92 | 0.93 |
| 10 | 0.76 | 0.62 | 0.54 | 0.51 | 0.50 | 0.50 | 0.51 | 0.72 | 0.73 | 0.74 | 0.74 |
| 15 | 0.70 | 0.54 | 0.46 | 0.44 | 0.43 | 0.43 | 0.45 | 0.66 | 0.66 | 0.66 | 0.67 |
| 25 | 0.65 | 0.47 | 0.40 | 0.39 | **0.39** | 0.39 | 0.42 | 0.63 | 0.63 | 0.64 | 0.65 |
| 50 | 0.70 | 0.53 | 0.49 | 0.46 | 0.45 | 0.43 | 0.46 | 0.52 | 0.63 | 0.71 | 0.87 |
| 75 | 0.68 | 0.52 | 0.48 | 0.45 | 0.46 | 0.45 | 0.48 | 0.56 | 0.78 | 0.82 | 0.94 |
| 100 | 0.67 | 0.50 | 0.48 | 0.45 | 0.46 | 0.47 | 0.49 | 0.59 | 0.78 | 0.92 | 0.96 |
| 150 | 0.65 | 0.50 | 0.47 | 0.45 | 0.47 | 0.47 | 0.51 | 0.64 | 0.80 | 1.30 | 0.96 |

Table 3.1: Average surface distance $\delta L_{surface}$ for the various settings of profile length $dL_{BB}$ and the number of iterations $nL_{BB}$ used in the BB lumen segmentation. The minimum value is shown in bold font.

In these evaluations on the test set both observers were compared to each other and to the automatic method.

The influence of the number of training sets for the LBVSC method was determined by increasing this number from 10 until 18 in steps of 2. Because for the eight datasets in the training set no manual segmentations of the second observer was available, we only used the manual annotations of the first observer for evaluation. For each training set size (10, 12, 14, 16 and 18 datasets) 10 different random selections of the 38 data sets were made and the results of these 10 experiments were averaged. Just as in the other evaluations on the test set we calculated the Dice coefficient between the automatic segmentation of the lumen and the outer wall.

## 3.5 Results

### Parameter optimization

#### Deformable model fitting

The PC MRA images used in this study have a lower contrast compared to the TOF images used in [18]. Especially in the bifurcation area the optimization of the MRA lumen surface generated a small or “half moon” shaped vessel surface leading to a bad initialization for the BB lumen segmentation. To handle this problem we drastically reduced the number of iterations, $nL_{MRA}$, to 1 and kept the profile length $dL_{MRA}$ to the minimum of 5 voxels. This gives the NURBS surface less freedom to adjust and results in a more circular cross-section of the segmented lumen.

Table 3.1 and 3.2 show the results of the optimization experiments for the lumen and outer wall parameters respectively. The bold, most red cell indicates the optimal value, i.e the minimal average surface distance $\delta L_{surface}$ and $\delta W_{surface}$. We did not perform any experiments using a profile length of less than 5 mm as this would require the estimation of a gradient on less than 3 voxels.

| $nW_{BB}$ \ $dW_{BB}$ (mm) | 5 | 7 | 9 | 11 | 13 | 15 |
|---|---|---|---|---|---|---|
| 0 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 |
| 1 | 0.52 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 |
| 2 | 0.51 | 0.50 | 0.49 | 0.49 | 0.49 | 0.49 |
| 3 | 0.49 | 0.49 | 0.50 | 0.54 | 0.58 | 0.62 |
| 5 | **0.48** | 0.50 | 0.57 | 0.67 | 0.75 | 0.84 |
| 7 | 0.49 | 0.53 | 0.66 | 0.81 | 0.93 | 1.05 |
| 10 | 0.51 | 0.61 | 0.82 | 1.02 | 1.19 | 1.34 |
| 25 | 0.62 | 0.90 | 1.27 | 1.59 | 1.87 | 2.11 |
| 50 | 0.74 | 1.12 | 1.55 | 1.95 | 2.28 | 2.57 |
| 75 | 0.81 | 1.23 | 1.69 | 2.11 | 2.50 | 2.83 |

Table 3.2: Average surface distance $\delta W_{surface}$ for the various settings of the profile length $dW_{BB}$ and the number of iterations $nW_{BB}$ used in the BB outer wall segmentation. The minimum value is shown in bold font.

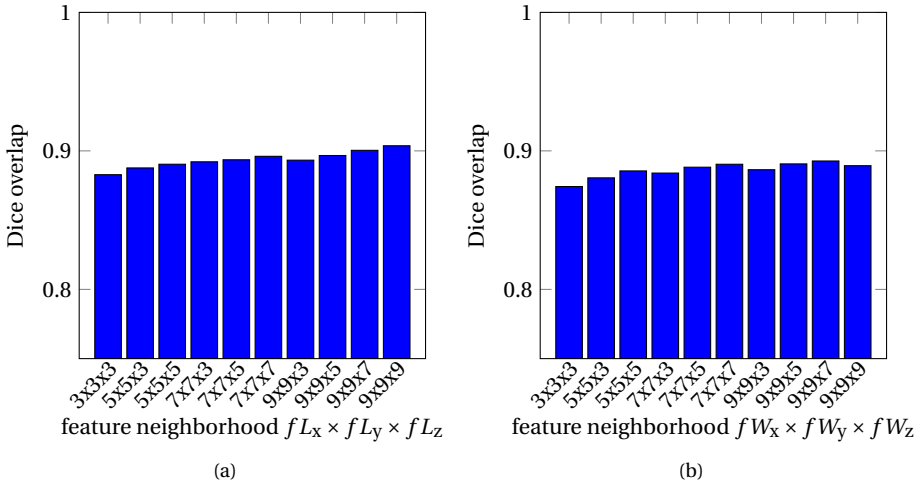



Figure 3.5: Influence of the feature neighborhood size on the final Dice overlap with the manual segmentation for (a) the lumen and (b) the outer wall.

### Learning-based vessel segmentation correction

Figure 3.5 shows for each neighborhood size the maximum Dice overlap (optained with optimum settings of the erosion and dilation parameters, which may be different for each neighborhood size). As can be seen from Figure 3.5 the influence of the feature neighborhood size on the segmentation accuracy is relatively small for both lumen and outer wall. As the computational costs increase significantly with a growing feature neighborhood we chose a neighborhood size of 7x7x3, which is computationally feasible and accounts for the anisotropic voxel size.

Table 3.3 and 3.4 show the Dice overlap values for the 7x7x3 feature neighborhood size for different combinations of dilation and erosion radii, $rL_{dilation}$, $rL_{erosion}$, and $rW_{dilation}$

and $rW_{\text{erosion}}$ for the lumen and outer wall, respectively. The maximum Dice coefficients are shown in bold font. The values for a dilation and erosion with zero radius show the Dice overlap without applying the LBVSC method. The tables for the other feature neighborhood sizes can be found on the website `http://ergocar.bigr.nl` (user/pwd: reviewer/PMBVesselWall). In both Table 3.3 and 3.4 the value for zero dilation and erosion

| $rL_{\text{dilation}}$ (voxels) \ $rL_{\text{erosion}}$ (voxels) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 78.9 | 80.4 | 80.6 | 80.8 | 80.8 | 80.8 | 80.8 | 80.7 | 80.8 | 80.8 |
| 1 | 85.1 | 84.7 | 85.0 | 85.2 | 85.2 | 85.2 | 85.2 | 85.2 | 85.1 | 85.1 |
| 2 | 86.2 | 85.7 | 86.0 | 86.1 | 86.1 | 86.2 | 86.1 | 86.2 | 86.2 | 86.2 |
| 3 | 87.2 | 86.8 | 87.1 | 87.1 | 87.1 | 87.2 | 87.2 | 87.2 | 87.2 | 87.2 |
| 4 | 88.2 | 87.9 | 88.1 | 88.0 | 88.1 | 88.1 | 88.0 | 88.1 | 88.2 | 88.2 |
| 5 | 88.6 | 88.5 | 88.8 | 89.1 | 88.9 | **89.1** | 88.8 | 88.9 | 88.8 | 88.8 |
| 6 | 88.8 | 88.3 | 88.9 | 88.6 | 88.7 | 88.6 | 88.7 | 88.9 | 89.0 | 89.0 |
| 7 | 88.8 | 88.0 | 88.3 | 88.7 | 88.6 | 88.9 | 88.8 | 88.7 | 88.9 | 88.9 |
| 8 | 88.5 | 87.7 | 88.2 | 88.5 | 88.3 | 88.5 | 88.5 | 88.5 | 88.5 | 88.5 |
| 9 | 88.2 | 87.6 | 87.9 | 88.0 | 88.1 | 88.1 | 88.3 | 88.2 | 88.1 | 88.0 |

Table 3.3: Average lumen Dice overlap values of the final segmentation for various sizes of the region around the host segmentation border which is defined by the dilation $rL_{\text{dilation}}$ and erosion $rL_{\text{erosion}}$ radius. The maximum value is in bold font. The value for zero dilation and erosion is the Dice overlap value for the uncorrected segmentation.

| $rW_{\text{dilation}}$ (voxels) \ $rW_{\text{erosion}}$ (voxels) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 81.3 | 82.2 | 82.3 | 82.4 | 82.3 | 82.3 | 82.4 | 82.3 | 82.4 | 82.4 |
| 1 | 85.5 | 85.6 | 85.5 | 85.5 | 85.5 | 85.6 | 85.5 | 85.6 | 85.5 | 85.4 |
| 2 | 86.4 | 86.4 | 86.4 | 86.4 | 86.4 | 86.4 | 86.4 | 86.4 | 86.5 | 86.4 |
| 3 | 87.2 | 87.3 | 87.3 | 87.3 | 87.3 | 87.3 | 87.3 | 87.3 | 87.1 | 87.3 |
| 4 | 87.8 | 87.9 | 88.0 | 87.9 | 87.9 | 87.9 | 88.0 | 87.9 | 87.9 | 88.0 |
| 5 | 88.3 | 88.3 | 88.3 | 88.2 | 88.3 | 88.3 | 88.3 | **88.3** | 88.3 | 88.2 |
| 6 | 88.2 | 88.2 | 88.1 | 88.2 | 88.2 | 88.1 | 88.3 | 88.3 | 88.1 | 88.1 |
| 7 | 87.8 | 87.9 | 87.9 | 87.9 | 87.9 | 87.9 | 88.1 | 88.0 | 88.0 | 88.1 |
| 8 | 87.8 | 87.7 | 87.4 | 87.8 | 87.7 | 87.9 | 87.9 | 87.6 | 87.7 | 87.5 |

Table 3.4: Average outer wall Dice overlap values of the final segmentation for various sizes of the region around the host segmentation border which is defined by the dilation $rW_{\text{dilation}}$ and erosion $rW_{\text{erosion}}$ radius. The maximum value is in bold font. The value for zero dilation and erosion is the Dice overlap value for the uncorrected segmentation.

is the smallest. On the training set the LBVSC method improved the segmentation result of the deformable model fitting: with any setting of $rL_{\text{dilation}}$, $rL_{\text{erosion}}$, $rW_{\text{dilation}}$ and $rW_{\text{erosion}}$ the results was better than $rL_{\text{dilation}} = rL_{\text{erosion}} = 0$ or $rW_{\text{dilation}} = rW_{\text{erosion}} = 0$.

| Lumen | | Outer wall | |
|---|---|---|---|
| parameter | value | parameter | value |
| $nL_{\text{MRA}}$ | 1 | $nW_{\text{BB}}$ | 5 |
| $dL_{\text{MRA}}$ | 5 | $dW_{\text{BB}}$ | 5 |
| $nL_{\text{BB}}$ | 25 | $fW_{\text{x}} \times fW_{\text{y}} \times fW_{\text{z}}$ | $7 \times 7 \times 3$ |
| $dL_{\text{BB}}$ | 13 | $rW_{\text{dilation}}$ | 5 |
| $fL_{\text{x}} \times fL_{\text{y}} \times fL_{\text{z}}$ | $7 \times 7 \times 3$ | $rW_{\text{erosion}}$ | 7 |
| $rL_{\text{dilation}}$ | 5 | | |
| $rL_{\text{erosion}}$ | 5 | | |

Table 3.5: Optimal lumen and outer wall method parameters

| | Lumen | | | | | Outer wall | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | No correction | LBSC | p-value | LBVSC | p-value | No correction | LBSC | p-value | LBVSC | p-value |
| ob1-ob2 | 0.90 | 0.90 | – | 0.90 | – | 0.90 | 0.90 | – | 0.90 | – |
| ob1-auto | 0.83 | 0.84 | 0.45 | 0.86 | 0.03 | 0.85 | 0.86 | 0.02 | 0.87 | $< 0.01$ |
| ob2-auto | 0.84 | 0.83 | 0.64 | 0.84 | 0.86 | 0.83 | 0.85 | 0.11 | 0.86 | 0.01 |

Table 3.6: Dice overlap between the different measurements of the test set without segmentation correction and using the LBSC and LBVSC method.

### Optimal parameters

Table 3.5 summarizes the optimal method parameters for both the lumen and the outer wall segmentation.

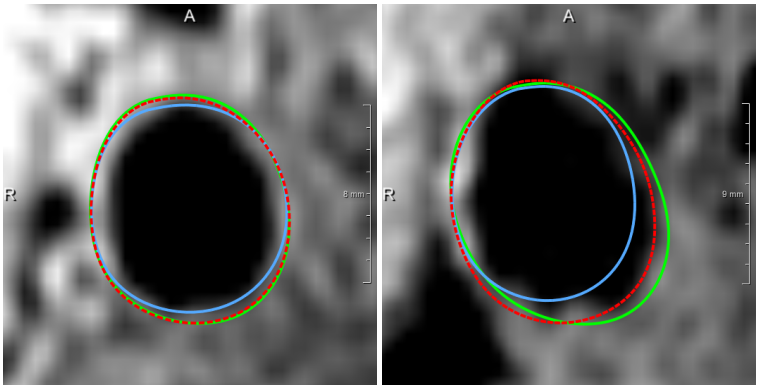
### Evaluation on test set

The average Dice similarity coefficients for the lumen and outer wall are listed in Table 3.6. The average Dice overlap between the automatic segmentation and both observers is comparable to the Dice between the two observers. The overlap with the first observer is higher, which is expected as the automatic method was trained using the segmentations of this observer. The LBSC method only improved the outer wall segmentation of the observer on which it was trained. There was no significant difference for the lumen segmentation. The LBVSC method significantly increased the average lumen overlap for observer 1 whereas it remained the same for observer 2. For the outer wall the LBVSC method has a positive effect on the Dice overlap with respect to both observers.

The difference in effect of the LBVSC method between the training (Table 3.3 and 3.4) and test set (Table 3.6) is explained by the fact that in the former the training set was used to both train the classifier of the LBVSC method and measure the effect of the LBVSC method on the Dice overlap of the resulting segmentation (see also Section 3.4).

Figure 3.6 gives a visual impression of the effect of the LBVSC method on the outer wall segmentation. The LBVSC method (the dashed red contour) here clearly improved

the segmentation of the deformable model fitting method (smallest blue solid contour): it is closer to the manual annotation (solid green contour).




(a) (b)

Figure 3.6: Two examples of the effect of the LBVSC method on the outer wall segmentation. The green solid line is the annotation by observer 1, the blue solid line is the segmentation from the deformable model fitting method, the dashed red line is the segmentation after applying the LBVSC method.

In Figure 3.7 the scatter plots of the volume and NWI measurements by observer 1, observer 2 and the automatic method are shown.

The two large markers in the volume measurements indicate two outliers where observer 1 and observer 2 disagree. Figure 3.8 gives an example slice of these outliers where observer 1 and 2 disagreed.

The contours shown in Figure 3.8 are the manual annotations of the outer wall for both observers. It is obvious from these segmentations that in this dataset it is very hard to determine the location of the outer vessel wall and both observers made a different choice on what to include in the vessel wall. In the subsequent analysis we removed these outliers.

Table 3.7 gives the average vessel wall volume and NWI measurements of both observers and the automatic method without applying the segmentation correction and with the LBSC and LBVSC method. As can be seen, the automatic segmentation slightly overestimated the vessel wall volume when compared to observer 1 and underestimated the volume when compared to observer 2. Observer 2 on average made larger volume measurements than observer 1. The average vessel wall volume of the automatic method is between the volumes estimated by observer 1 and observer 2, but was on average closer to observer 1 on which it was trained. The p-values of the differences between the automatic method with and without the LBVSC method are 0.0005 and 0.1 for $V_{\text{wall}}$ and NWI respectively.

Table 3.8 shows the differences, Pearson and intra-class correlations of the volume and NWI measurements. The average difference in the wall volume measurements of the automatic method with respect to both observers is smaller than the average difference between the observers. Although the Pearson correlation coefficient of the volume measurements between the observers and the automatic method is not as good as the corre-

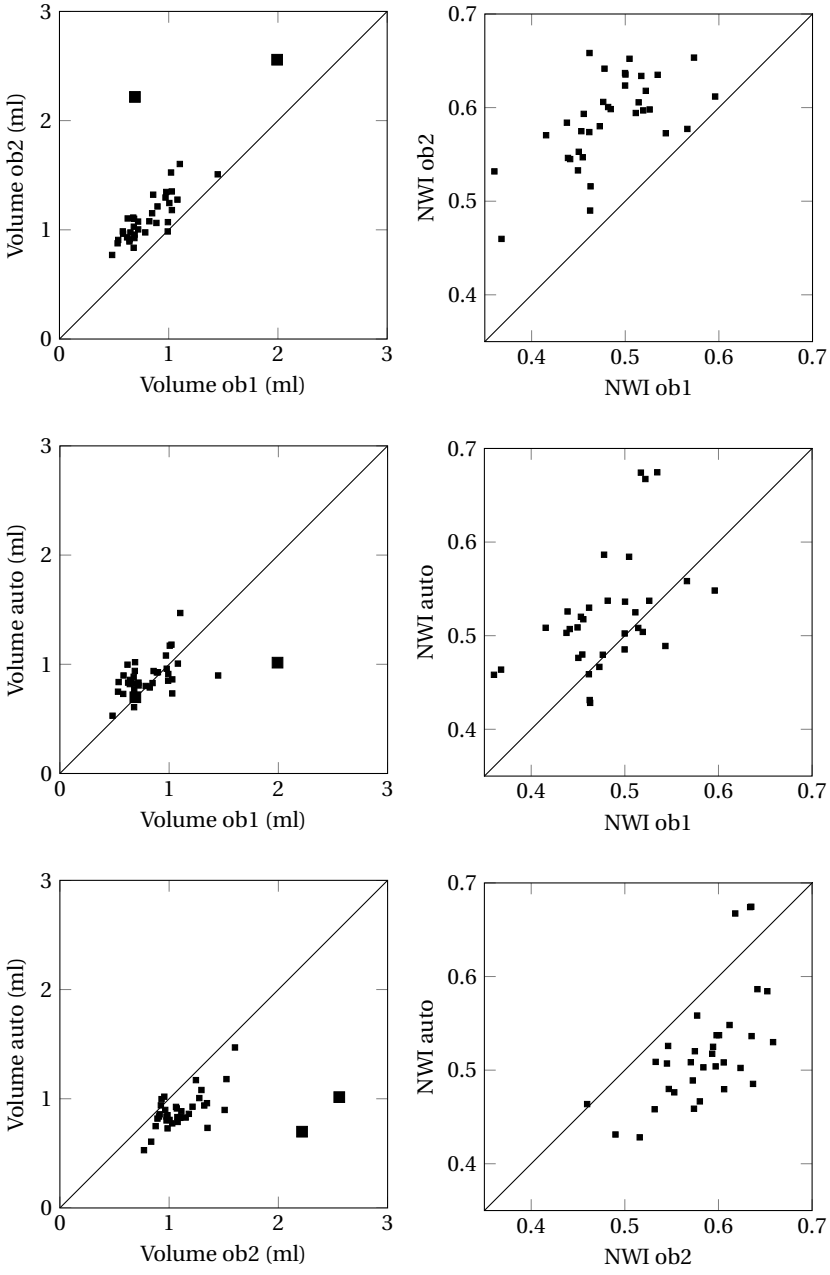
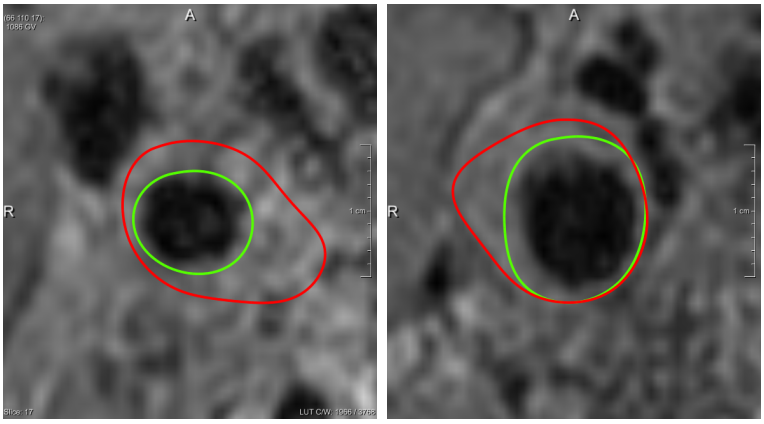



Figure 3.7: Scatter plot of the volume measurements (left column) and normalized wall index (right column) for observer versus observer 2 (top row), observer 1 versus automatic method (middle row) and observer 2 versus automatic method (bottom row). In the volume measurement graphs two outliers are present and shown with big markers. These outliers are not shown in the NWI graphs because they would stretch the scale too much.




(a) (b)

Figure 3.8: Example slice from the two outlier datasets. The green contour is from observer 1 and the red contour from observer 2.

| | ob1 | ob2 | No correction | LBSC | LBVSC |
|---|---|---|---|---|---|
| $V_{\text{wall}}$ (ml) | $0.80 \pm 0.21$ | $1.1 \pm 0.20$ | $0.78 \pm 0.13$ | $0.83 \pm 0.11$ | $0.89 \pm 0.17$ |
| NWI (ml/ml) | $0.48 \pm 0.05$ | $0.59 \pm 0.05$ | $0.51 \pm 0.05$ | $0.51 \pm 0.05$ | $0.54 \pm 0.08$ |

Table 3.7: Average ± standard deviation of vessel wall volume (ml) and normalized wall index measurements (ml/ml).

| | $V_{\text{wall}}$ | | | NWI | | |
|---|---|---|---|---|---|---|
| | $\Delta V_{\text{wall}}$ (ml) | Pearson | icc | ΔNWI (ml/ml) | Pearson | icc |
| ob2 - ob1 | $0.30 \pm 0.19$ | 0.83 | 0.58 | $0.10 \pm 0.04$ | 0.51 | 0.32 |
| auto - ob1 | $0.08 \pm 0.20$ | 0.51 | 0.62 | $0.05 \pm 0.07$ | 0.50 | 0.52 |
| auto - ob2 | $-0.21 \pm 0.15$ | 0.57 | 0.57 | $-0.05 \pm 0.07$ | 0.54 | 0.56 |

Table 3.8: The average volume differences, the Pearson and intra-class correlation coefficients of the volume and NWI measurements between the two observers and the automatic method.

lation between the two observers, the intra-class correlation with both observers is better than (observer 1) or almost the same as (observer 2) between the two observers.

Looking at the clinically used NWI, the average difference between the measurements of automatic method and both observers is again smaller than the average difference between the two observers. Also the correlation coefficients between the automatic method and both observers are better than or almost the same as the correlations between the measurements of the two observers.

Table 3.9 shows the Dice overlap between the manual segmentation of the first observer and the automatic method for different size of the training set. The table show that increasing the size of the training set leads to a small increase in overlap.

| Number of training sets | 10 | 12 | 14 | 16 | 18 |
|---|---|---|---|---|---|
| Dice lumen | 0.84 | 0.84 | 0.84 | 0.85 | 0.86 |
| Dice outer wall | 0.85 | 0.84 | 0.84 | 0.85 | 0.86 |

Table 3.9: Influence of the number of training sets used for training the LBVSC method on the Dice overlap of the lumen and the outer wall, and the NWI

## 3.6 Discussion

The interobserver measurements presented in this study show that the systematic difference between the outer vessel wall annotations of two observer can be considerable (see the scatter plots in Figure 3.7). As we do not have manual annotations of the second observer on the training set, it is impossible to train the automatic method on both observers, or generate a new reference standard based on the combined annotations of these two observers without compromising the size of the test set. However, the annotation protocol for both observers was the same and the variation between the two observers can also be expected in clinical practice.

The method described in this paper was designed in order to analyze the MRI data acquired in the context of a population study. This population study uses a 1.5 T MRI scanner and a BB sequence which may not be the best possible sequence for imaging the vessel wall. Despite these limitations the described automatic method performs comparable to the observers. In future research it would be interesting to study the effect of different acquisition settings (1.5T vs. 3T, different MR sequences, the use of a head stabilizer) on the vessel wall quantification. The evaluation described in this study was performed on 19 subjects. For the future we plan a quantification of the vessel wall volume and normalized wall index on all 1072 subjects that participated in this study. Manual measurements on this number of subjects becomes infeasible.

The deformable model fitting method enforces a NURBS model on the lumen and wall boundary, and uses a single image feature (the maximum gradient magnitude) to fit the model. For the lumen border this is an adequate image feature. The gradient of the outer wall is less strong, which makes the application of this image feature less successful. The LBVSC method is not limited to the use of a single image feature, but instead uses many features to classify the outer wall voxels. This explains the difference in effect of the LBVSC method on the lumen and outer wall segmentation.

As long as there is a training set on which the LBVSC can be trained, the method described in this paper can easily be adapted to other images, acquired on a higher field MR scanner or using different protocols. If there is a training set consisting of a ground truth definition of multiple observers, these could be combined using existing methods like consensus reading or STAPLE [90], but they can also be used to train multiple classifiers and combine these classifiers into a single stronger classifier using methods described in e.g. [91].

## 3.7 Conclusion

In conclusion, we presented a method that can automatically quantify the wall volume and normalized wall index of the carotid artery in a region around the bifurcation. The method consists of a deformable model fitting step and a learning-based correction of

systematic errors. Intensity inhomogeneities in the MR images were reduced using a 3D-extended version of the LEMS method. The parameters of both the deformable model fitting and the LBVSC method have been optimized by extensive experiments using the manual annotations of a single observer on a training set. Evaluation was performed with respect to annotations of two observers on a separate test set. Our experiments justify the conclusion that the automatic method performs comparably to the manual annotations in terms of wall volume and normalized wall index measurements and can therefore be used to replace the manual measurements.

All image data, annotations and results from this study are made available through the website `http://ergocar.bigr.nl` (user/pwd: reviewer/PMBVesselWall). We challenge every one to improve our automatic vessel wall volume and normalize wall index measurements on these datasets and publish their results.