Automated aortic calcium scoring on low-dose chest computed tomography

Ivana Išgum\textsuperscript{1}, Annemarieke Rutten\textsuperscript{2}, Mathias Prokop\textsuperscript{2}, Marius Staring\textsuperscript{1}, Stefan Klein\textsuperscript{1}, Josien P.W. Pluim\textsuperscript{1}, Max A. Viegever\textsuperscript{1} and Bram van Ginneken\textsuperscript{1}

\textsuperscript{1}Image Sciences Institute, University Medical Center Utrecht, Utrecht, the Netherlands
\textsuperscript{2}Department of Radiology, University Medical Center Utrecht, Utrecht, the Netherlands

June 17, 2009

Abstract

Thoracic computed tomography (CT) scans provide information about cardiovascular risk status. These scans are non-ECG-synchronized, thus precise quantification of coronary calcifications is difficult. Aortic calcium scoring is less sensitive to cardiac motion, so it is an alternative to coronary calcium scoring as an indicator of cardiovascular risk. We developed and evaluated a computer-aided system for automatic detection and quantification of aortic calcifications in low-dose non-contrast-enhanced chest CT. A computer-aided system for automatic detection and quantification of aortic calcifications was trained and tested on scans from participants of a lung cancer screening trial. A total of 433 low-dose, non-ECG-synchronized, non-contrast enhanced 16-detector row examinations of the chest were randomly divided into 340 training and 93 test data sets. A first observer manually identified aortic calcifications on training and test scans. A second observer did the same on the test scans only. First, a multi-atlas-based segmentation method was developed to delineate the aorta. Subsequently, the training data was used to train the system based on statistical pattern recognition theory to automatically identify calcifications in the aortic wall. Calcium volume scores for computer system and first observer and for the two observers were com-
pared using descriptive statistics and Spearman rank correlation coefficients. The computer system correctly detected on average 768 mm$^3$ out of 871 mm$^3$ calcified plaque volume in the aorta with an average 61 mm$^3$ of false positive volume per scan. Spearman rank correlation coefficient was $\rho=0.97$ between system and first observer compared to $\rho=0.99$ between the two observers. Automatic calcium scoring in the aorta appears feasible with good correlation between manual and automatic scoring.

Keywords: automatic calcium scoring, aorta, aortic segmentation, atlas based segmentation

1 Introduction

Lung cancer screening programs with computed tomography (CT) mainly focus on lung cancer detection although cardiovascular risk in the typical screening population of smokers or former smokers is substantially increased as well [1]. CT scans of the thorax also provide information about the presence and the extent of atherosclerotic disease, a fact that only recently has led to efforts to quantify these changes on lung cancer screening scans [2].

Presence of coronary calcifications is a strong predictor of cardiovascular risk [3, 4]. Coronary calcium scoring is typically performed a non-contrast enhanced cardiac CT scan acquired with ECG-synchronization. However, scans from lung cancer screening programs are not ECG-triggered, and therefore a substantial amount of cardiac motion may be present in the images. Pulsation-induced errors can be significant [5] and make it difficult to determine the amount of coronary calcium in these scans with high precision [6]. Atherosclerosis, however, is a generalized process, which makes it possible to use calcifications in the aorta as a marker of cardiovascular disease. Many studies have investigated relationship and prognostic value of aortic calcium and cardiovascular disease, e.g. [7–12]. Association between aortic calcification and other diseases has been studied as well [13, 14]. Moreover, pre-operative quantification of atherosclerotic burden may be a predictor for the occurrence of aortic emboli during surgery [15]. The advantage of quantifying aortic calcium is the fact that pulsation affects the aorta much less than the coronaries.

In the aforementioned studies calcified lesions in the aorta were identified manually. In manual calcium scoring a region of interest in a scan is selected and all clusters of voxels above a certain threshold are considered. Usually, the threshold value of 130 Hounsfield units (HU) is used. The clusters representing arterial calcifications are manually identified. The calcium
is then quantified and expressed in terms of calcium scores \cite{16, 17}. Often, small clusters of voxels above the threshold are ignored, because they are considered to represent noise.

The amount of aortic calcium in scans from smokers can be considerable and will require a lot of user interaction to complete the scoring procedure. Manual scoring is further demanding because in lung cancer screenings the scans are acquired with low-dose. This means that the amount of noise that exceeds the threshold for calcification extraction is high. Moreover, screening programs include large cohorts, which makes it costly to include labor-intensive manual calcium scoring in the routine evaluation of these scans.

Very few methods for automatic arterial calcification detection have been developed, and commercial software packages do not provide it. To our knowledge there are only two published automatic methods, which have concentrated on calcification detection and quantification in the aorta. De Bruijne \cite{18} presented a pixel classification based method for automated detection of calcifications in the lumbar aorta in radiographic images. In our previous work \cite{19} an automated method for the detection of the aortic calcifications in CTA scans of the abdomen was presented. That method considered all high density areas in the complete scan as potential calcifications. Because of the contrast present in the images, these areas were extracted using a high threshold value of 220 HU. Subsequently, an object based classification method was employed to separate true aortic calcifications from other high density areas based on features derived from the potential calcifications and their contextual information.

In this work we present a system for automatic detection and quantification of aortic calcifications in low-dose, non-contrast, non-ECG triggered CT scans of the chest. This method does not analyze all potential calcifications in the complete scan, but first segments the aorta and considers only high density areas extracted within the segmented volume. Both the ascending and the descending aorta and the aortic arch are analyzed. Because of the low-dose used at acquisition and the standard threshold level for calcification extraction, level of extracted noise in the images is very high. Pattern recognition based analysis is employed to identify true aortic calcifications from other high density objects based on the features describing the potential calcifications and their contextual information. The system is evaluated with the data from a lung cancer screening cohort.

The paper is organized as follows. Section 2 describes the data. Section 3 gives a detailed overview of the method. Next, results are presented in Section 4. Section 5 provides a discussion and Section 6 gives a conclusion.
2 Materials

2.1 Data

In this study low-dose, non-ECG synchronized, non-contrast enhanced CT scans of the chest that were acquired during a population-based randomized lung cancer screening trial were used [20]. The study was approved by the Ministry of Health and by the Ethics Committee of each participating hospital.

The scans were acquired on a 16 detector-row scanner (Mx8000 IDT or Brilliance 16P, Philips Medical Systems, Cleveland, OH, USA). All scans were realized in about 12 seconds in spiral mode with 16 x 0.75 mm collimation. Axial images of 1.0 mm thickness at 0.7 mm increment were reconstructed with a moderately soft kernel (Philips "B"), using the smallest field of view to include the outer rib margins at the widest dimension of the thorax. The peak voltage was 120 kVp - 140 kVp depending on patient weight, with tube current 30 mAs. Scans were performed in inspiration after appropriate instruction of the subjects, without spirometric control or respiratory belt. No intravenous contrast-injection was induced. A detailed description of the inclusion criteria and scanning protocol is provided elsewhere [21].

Between April 2004 and March 2005, 1684 male subjects received baseline screening with low-dose CT. From these participants, we randomly selected 436 subjects for calcium scoring in the aorta. Two scans with metal implants and one scan with extreme amounts of noise were excluded, thus 433 scans could be used in the experiments. The age of the subjects included in our study was 60.8±6.2 years.

To control image noise and to obtain data comparable with previously published results [4, 22], scans were subsampled to 3.1 mm thick sections by averaging four consecutive sections.

2.2 Reference standard

Identification of aortic calcium is usually a relatively simple task. Potential problems may occur when the noise level in the scan is high, or in areas where arteries are branching off the aortic arch. Therefore, the following approach was chosen: two medical students (first and second observer) were trained by a medical investigator with calcium scoring experience in approximately 500 scans, laying special emphasis on the above mentioned critical factors. The students used a software tool developed specifically for this purpose that automatically marked all voxels with a CT number above 130 HU using a color overlay. The students were instructed to select calcifications in the...
ascending aorta, aortic arch, and the descending aorta not lower than the level of the apex of the heart. By clicking on one of the voxels of such a calcification, all voxels that were spatially connected to it were automatically identified and indicated by a color change.

Because of the noise in the scans, some calcifications in the descending aorta appeared connected to the spine. This could easily be seen because the color overlay spread into the adjacent bony structures. Observers were instructed not to mark those lesions. The scans were randomly divided into a training set of 340 scans and a test set of 93 scans. The training set was used to develop and train our computer system. The test set was used for testing the system’s performance.

Calcifications in the training set were identified by the first observer only. For the test set, calcifications were independently identified by the first and second observer. The results of the first observer were used as a reference for testing our method. The scores of the second observer were used to estimate inter-observer variability, which was compared to the results of the automatic system.

Calcium volume scores were computed for calcifications identified during manual segmentation [23].

3 Methods

The detection and quantification of aortic calcifications started with the automatic segmentation of the aorta. Subsequently, calcifications in the aortic wall were detected within the segmented volume. Finally, a calcium volume score was computed for each subject.

3.1 Aortic segmentation

Automatic delineation of the aorta in a non-contrast enhanced CT scans is a very complex task due to absence of gradient information especially in the ascending aorta. The only publication presenting fully automatic aortic segmentation in such scans is our multi-atlas based segmentation approach [24]. This method was employed in this work and here we provide a short description.

In atlas segmentation approaches an atlas (image in which the aorta has been manually delineated) is deformed in such a way that it aligns with the target image (image that needs automatic delineation of the aorta). Since the location of the aorta is known for the atlas (because it had been manually
delineated), the location of the aorta in the target image can be estimated from the position of the aorta in the deformed atlas (Figure 1).

We used multi-atlas segmentation approach where the process is the same, except that multiple atlases were used (Figure 2). Each atlas provided an estimation of the position of the aorta in the target image. By combining the positions of the aorta in all deformed atlas images, a probabilistic segmentation of the aorta was obtained that provided the likely position of the aorta in the target image. To account for differences in how well the deformed atlas matched the target image, we used a weighting factor that depended on the difference between deformed atlas and target image. The resulting probabilistic segmentation was then thresholded to obtain the final segmentation result (computed location of the aorta in the target image). The threshold was chosen conservatively, so that not only the aorta but possibly also a small amount of surrounding tissue in the target image was included. This made sure that the complete aorta was included in the subsequent analysis. Details of the implementation and settings used in our experiments can be found in the Appendix.

3.2 Candidate extraction

Thresholding at 130 HU was applied within the segmented aortic volume to extract candidate voxels that might represent aortic calcifications [1, 25, 26]. Three-dimensional connected component labeling of the candidate voxels was used to determine candidate objects for aortic calcifications [27]. All objects which have at least one voxel contained in the binary segmentation were considered candidate objects.

3.3 Feature computation

Each candidate object was represented by a number of numerical characteristics (features). Generally, there are many features that can be computed and in computer-aided detection systems it has been shown that a certain combination of for example size, appearance, or position characteristics can distinguish between positive and negative candidate objects. Based on our experience and previous work we have computed the following features: volume of the candidate object expressed in mm$^3$; average and maximum gray value within the candidate object; the ten Gaussian derivatives [28] through the second order ($L, L_x, L_y, L_z, L_{xx}, L_{yy}, L_{zz}, L_{xy}, L_{xz}, L_{yz}$) at scales $\sigma = 1, 2, 4, 8, 16$ voxels calculated in x, y and z directions at center of mass in the candidate object; x, y and z coordinates of the center of mass of the candidate object; average probability that the candidate object belongs to the
Figure 1: The principle of atlas based segmentation: In the atlas the aorta has been manually delineated and it needs to be delineated in the target. First, the atlas is deformed to match the target and subsequently, the manual segmentation of the aorta from the atlas is deformed by applying the same transformation. This deformed aortic segmentation provides an estimation of the aortic position in the target.

segmented volume; the percentage of the candidate object’s volume within the binary aortic segmentation; the value in a 3-dimensional signed distance transform map of the aortic segmentation [27] at the center of mass of the candidate object; the minimum value in the distance transform map within the candidate object; the distance in the z-direction from the candidate object to the lowest point of the segmented part of the descending aorta; the distance in the z-direction from the candidate object to the top of the aortic arch; the minimum of the previous two values. This resulted in a total 63 features were computed and they are listed in Table 1.
Figure 2: Multi-atlas segmentation with weighted decision fusion algorithm. N atlases with their manual segmentations are matched to the target image. Subsequently, the atlases and their manual segmentations are transformed using the corresponding transformations obtained by registration. The transformed manual segmentations are combined by local decision fusion based on the difference images between each deformed atlas and the target. In this way a probabilistic segmentation in the target image is obtained. Finally, thresholding the probabilistic segmentation results in a binary segmentation of the target.

3.4 Classification

The extracted candidate objects not only represented calcifications in the wall of the aorta, but also calcifications of the aortic valve, noise, and calcifications in the vicinity of the aorta extracted due to a somewhat erroneous segmentation result. Because of the noise present in the scans, it is possible that parts of the large bony structures, such as the spine became candidate objects. Those could easily be discarded based on their size, without a risk of discarding any calcifications. Therefore, an upper threshold on the vol-
Table 1: List of features describing candidate objects.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Volume</td>
</tr>
<tr>
<td>2</td>
<td>Average gray value within the candidate object</td>
</tr>
<tr>
<td>3</td>
<td>Maximum gray value within the candidate object</td>
</tr>
<tr>
<td>4-8</td>
<td>$L$ at scales $\sigma = 1, 2, 4, 8, 16$ voxels</td>
</tr>
<tr>
<td>9-13</td>
<td>$L_x$ at scales $\sigma = 1, 2, 4, 8, 16$ voxels</td>
</tr>
<tr>
<td>14-18</td>
<td>$L_y$ at scales $\sigma = 1, 2, 4, 8, 16$ voxels</td>
</tr>
<tr>
<td>19-23</td>
<td>$L_z$ at scales $\sigma = 1, 2, 4, 8, 16$ voxels</td>
</tr>
<tr>
<td>24-28</td>
<td>$L_{xx}$ at scales $\sigma = 1, 2, 4, 8, 16$ voxels</td>
</tr>
<tr>
<td>29-33</td>
<td>$L_{yy}$ at scales $\sigma = 1, 2, 4, 8, 16$ voxels</td>
</tr>
<tr>
<td>34-38</td>
<td>$L_{zz}$ at scales $\sigma = 1, 2, 4, 8, 16$ voxels</td>
</tr>
<tr>
<td>39-43</td>
<td>$L_{xy}$ at scales $\sigma = 1, 2, 4, 8, 16$ voxels</td>
</tr>
<tr>
<td>44-48</td>
<td>$L_{xz}$ at scales $\sigma = 1, 2, 4, 8, 16$ voxels</td>
</tr>
<tr>
<td>49-53</td>
<td>$L_{yz}$ at scales $\sigma = 1, 2, 4, 8, 16$ voxels</td>
</tr>
<tr>
<td>54-56</td>
<td>$x$-, $y$- and $z$-coordinate of the center of mass of the candidate object</td>
</tr>
<tr>
<td>57</td>
<td>Average probability that the candidate object belongs to the aorta</td>
</tr>
<tr>
<td>58</td>
<td>Percentage of candidate objects within the aortic segmentation</td>
</tr>
<tr>
<td>59</td>
<td>Distance from the center of mass of the candidate object to the aortic border</td>
</tr>
<tr>
<td>60</td>
<td>Minimum distance from the candidate object to the aortic border</td>
</tr>
<tr>
<td>61</td>
<td>Distance from the candidate object to the lowest point of the descending aorta</td>
</tr>
<tr>
<td>62</td>
<td>Distance from the candidate object to the top of the arch</td>
</tr>
<tr>
<td>63</td>
<td>Minimum of features 61 and 62</td>
</tr>
</tbody>
</table>

The volume was set to 5000 mm$^3$, and all candidate objects above that size were discarded.

A statistical classifiers were used to separate the remaining candidate objects. In our experiments two-stage classification was performed. In the first stage the most obvious negatives were discarded and in the subsequent stages the remaining candidate objects were separated. This gave better performance because in each stage a classifier specifically designed for a given task was used instead of a single one that separated all candidate objects.

Before classification, all features were scaled to zero mean and unit variance to account for differences in the ranges of values different features might had. In both stages sequential forward floating selection (SFFS) of features
was employed to select the best features for the task [29, 30]. This feature selection uses a given classifier to select a subset of features giving the best classification result. The algorithm is based on a "plus 1, take away r" strategy. At each iteration, the algorithm adds the best single feature to an initially empty feature set and then removes features as long as that improves performance. In this way nested groups of good features can be found. The training set was divided into feature selection training set and feature selection test set. The feature selection training set contained 75% of the objects from the original training set and the feature selection test set contained the remaining 25% of candidate objects from the training set. In pilot experiments with several supervised classifiers (linear discriminant, quadratic discriminant, k-nearest neighbor) the best results were obtained with a k-nearest neighbor (kNN) classifier [31]. Therefore, in both classification stages a kNN classifier with Euclidean metric and maximum of 15 features to be selected was employed. Classification performance was evaluated by accuracy over all candidate objects.

In the first classification stage the number of neighbors’ k was set to 10. All objects with a posterior probability for a negative class larger than 0.75 were discarded. Thus, only candidate objects with a high probability to be negative were removed from both training and test set. The remaining objects were further classified. In the second stage the number of neighbors was set to three. A threshold on the posterior probability was set to 0.5. This means that all objects with a posterior probability for a positive class larger than or equal to 0.5 were classified as calcifications and others as non-calcifications. These parameter settings were determined in pilot experiments that used only the training data.

3.5 Calcium score

After calcifications in the aorta had been identified, the detected amount of calcification was quantified. A standard for quantification of calcified lesions in the aorta is lacking. In previous studies different approaches have been used and they are described by Jayalath et al. [32]. Our choice was to compute the volume of the detected calcifications scores, implemented as described in [23].

3.6 Evaluation

Descriptive statistics were used to summarize the number of aortic calcifications and the corresponding volume scores. Numbers of calcifications and their volumes were calculated per scan and in total for both observers and
the automated system. Medians, 25th and 75th percentiles and total range are presented since the results did not show a normal distribution. For the automated system also the number of detected objects other than aortic calcifications ("non-aortic calcifications") and their volume score in the segmented volume was determined to allow for an analysis of the performance of the system. The number and volume of true positive (identified as aortic calcification by reference observer and automated system), true negative (identified as non-aortic calcification by the automated system and by the reference observer), false positive (identified as aortic calcification by the automated system, but not by the reference observer) and false negative (identified as non-aortic calcification by the automated system, but selected as aortic calcification by the reference observer) calcifications were calculated.

To evaluate the performance of the system relative to the performance of human observers, we plotted the volume scores from the system against the volume scores of the first observer (reference) and the volume scores from both observers against each other. We used Spearman rank correlation to assess the degree of correlation.

4 Results

In 93 test scans the first observer selected 1029 aortic calcifications (positives). Distribution per scan was: median=7; 25th percentile=3; 75th percentile=14; range: 0 - 51. In terms of volume these scans contained 80963 mm$^3$ of aortic calcium (per scan: median=279 mm$^3$; 25th percentile=56 mm$^3$; 75th percentile=1013 mm$^3$; range: 0 - 8563 mm$^3$).

The automatic method extracted all these calcifications as candidate objects and 1014 other candidate objects (negatives). Distribution of the negatives per scan was: median=3; 25th percentile=2; 75th percentile=5; range 0-266. In terms of volume this corresponded to 27,394,133 mm$^3$ (per scan: median=343,780 mm$^3$; 25th percentile=225,755 mm$^3$; 75th percentile=396,711 mm$^3$; range: 0 - 690,941 mm$^3$).

By discarding large candidate objects, 71 objects (on average 0.8 objects per scan) were eliminated from the test set without removing any calcifications in the aorta. In total 943 (11200 mm$^3$) negatives remained.

The computer system correctly identified 84.1% of all aortic calcifications and 88.2% of the calcified aortic plaque volume. For an average number of 9.3 (768 mm$^3$) true positive calcifications, an average of 1.8 (61 mm$^3$) false positive calcifications per scan were detected. Also, on average there were 1.8 (103 mm$^3$) false negative objects.

The system assigned a zero calcium score to 1/93 subjects (1%) in whom
both observers assigned a positive score. The system assigned a positive score to 1/93 subjects (1%) in whom both observers assigned a zero score. The two observers did not agree about zero scores in three patients.

The detailed numbers for false positive and false negative errors are given in Table 2. The main causes of false positives were noise and motion artifacts (61%), calcifications detected in the descending aorta below the level of the apex of the heart (17%), calcifications in the arteries branching out of the aortic arch (6%) and calcifications in the coronary arteries branching out of the ascending aorta (6%). Figure 3 shows examples of false positive errors and Figure 4 examples of false negative errors.

Figure 5, left correlates the volume scores assigned to each subject by the automatic system to that of the first observer. Figure 5, right correlates the scores of the first and second observer. Spearman rank correlation demonstrated excellent agreement with a correlation coefficient $\rho = 0.97$ between automatic system and reference, and $\rho = 0.99$ between observers. Identical calcium scores for the computer system and the first observer were found in 20 (21.5%) scans. Identical calcium scores for first and second observer were also seen in 20 (21.5%) scans. These scans were not identical to the ones in which the computer system and the first observer fully agreed.

Figure 3: Examples of false positive objects. Swallowing artifact around the esophagus identified as aortic calcium by the computer system (left). Left main coronary artery calcification selected as aortic calcification by the computer system (right). Both objects are blurred due to the motion artifacts.
Table 2: Number and volume of all candidate objects averaged over 93 test scans. The candidates are divided in true negatives, false positives, true positives and false negatives as detected by the computer system. False positive and false negative errors of the system are divided in sub-categories.

<table>
<thead>
<tr>
<th>Category</th>
<th>Objects/scan</th>
<th>Volume(mm$^3$)/scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>True negatives</td>
<td>9.1</td>
<td>294500</td>
</tr>
<tr>
<td>False positives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noise and artifacts (motion, swallowing)</td>
<td>1.1</td>
<td>8</td>
</tr>
<tr>
<td>Calcium in</td>
<td></td>
<td></td>
</tr>
<tr>
<td>descending aorta below the heart</td>
<td>0.3</td>
<td>15</td>
</tr>
<tr>
<td>arteries above the arch</td>
<td>0.1</td>
<td>12</td>
</tr>
<tr>
<td>coronary arteries</td>
<td>0.1</td>
<td>13</td>
</tr>
<tr>
<td>bypass artery</td>
<td>0.06</td>
<td>3</td>
</tr>
<tr>
<td>mediastinum</td>
<td>0.05</td>
<td>4</td>
</tr>
<tr>
<td>tracheal wall</td>
<td>0.04</td>
<td>1</td>
</tr>
<tr>
<td>aortic valve</td>
<td>0.04</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>1.8</td>
<td>61</td>
</tr>
<tr>
<td>Total negatives</td>
<td>10.9</td>
<td>294561</td>
</tr>
<tr>
<td>True positives</td>
<td>9.3</td>
<td>768</td>
</tr>
<tr>
<td>False negatives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Descending aorta</td>
<td>0.8</td>
<td>36</td>
</tr>
<tr>
<td>Ascending aorta</td>
<td>0.7</td>
<td>9</td>
</tr>
<tr>
<td>Aortic arch</td>
<td>0.3</td>
<td>58</td>
</tr>
<tr>
<td>Total</td>
<td>1.8</td>
<td>103</td>
</tr>
<tr>
<td>Total positives (reference standard)</td>
<td>11.1</td>
<td>871</td>
</tr>
</tbody>
</table>

5 Discussion

The presented system automatically segmented the aorta and correctly identified the vast majority of aortic calcifications. Results from the computer system correlated only slightly less with the results of a human reference standard ($\rho = 0.97$) than two human observers with each other ($\rho = 0.99$). Even for the presence or total absence of calcium in the aorta, the system performed well: there were only two subjects in which the system disagreed with both observers while observers disagreed in three cases. The majority (61%) of false positive objects represented noise. Such objects are difficult to discriminate from small calcifications in the aortic wall. Although these errors occurred frequently, they were small in size and only caused a minor volume error. Most remaining errors can be attributed to problems at the aortic root, the branching vessels in the aortic arch and the definition of the distal cutoff point in the descending aorta. Some of these errors are a question of judgment (e.g., calcification located in the aortic root or in the proximal coronary (see Figure 3, right) or precise definition of the aortic con-
Figure 4: Examples of false negative objects. Calcification in the ascending aorta (left). Because of the motion this lesion seems positioned inside the aortic lumen. Calcification in the aortic arch at the point where an artery is branching off the aorta (right).

Figure 5: Volume scores assigned to subjects by the first observer (x-axis), vs. (left) the score assigned by the computer system and (right) the second observer (y-axis). Equal volumes are demonstrated by the solid line. The data is plotted with the logarithmic scale (x+1).
tour (calcification located in the aortic arch or extending into a branching vessel). No general agreement exists where to place such a point for the thoracic aorta. We have chosen the apex of the heart as the reference. By using a different reference point, e.g., the celiac artery, some of this variability might be reduced but the algorithm would have to be adapted. Moreover, the celiac artery is not always displayed on a chest CT scan. False negative objects appeared in various loci within the aorta. Some of them might have been caused by the fact that aortic calcifications distal to the cutoff point were counted as non-aortic calcifications for the purpose of our score. However, these calcifications had similar features to aortic calcifications located just above the cutoff point, which may have contributed to excluding such calcifications from the score. This hypothesis is supported by the relatively high number of false negatives in the distal descending aorta.

Inspection of the errors per scan showed the following. In the three scans where the disagreement between the automatic and reference calcium score was the largest, the results were dominated by false positive arterial calcifications. Those calcifications were located in the aorta below the apex of the heart (in two scans), and in the artery branching of the aortic arch (one scan). Furthermore, the zero scan to which the computer system assigned a positive score was caused by a calcification in the descending aorta below the apex of the heart. This indicates that the largest errors were caused by the automatic aortic segmentation and suggests that possibly a less conservative threshold for the probabilistic aortic segmentation could have been chosen.

Aortic calcifications have not widely been used as a cardiovascular risk marker, but a recent publication suggests that they are a risk marker for cardiovascular mortality [12]. There are large patient cohorts being screened for lung cancer now, and assessment of their cardiovascular risk has just been started. Coronary calcium scoring on such non-gated scans is less precise because of cardiac motion artifacts. Aortic calcium may therefore be a marker for cardiovascular risk that can more readily be accessed in these patients. Since scoring is time-consuming, an automated technique would be very welcome, and could make cardiovascular risk assessment an integrated part of such screening studies. Our system has the potential to be adapted to standard-dose scans but would require slightly bigger adjustments to accommodate contrast-enhanced scans as well.

In this study the manual segmentations were performed by medical students without previous experience in calcium scoring. We compensated their inexperience by providing them intensive training for this study. The high correlation between the scores of the two observers reflects that this approach was reasonable.

Future research should focus on the definition of the volume in which
the calcium scoring is performed. As already discussed our analysis of the results showed that major false positive and false negative errors occurred around the cutoff points in the descending aorta and the aortic arch. Also, calcifications that were partly contained in the aorta and partly in adjacent structures such as arch vessels or spine could be split in a part in the aorta and a part outside it.

Future research could also investigate if the presented method could be applied to calcium scoring in the aorta in different types of CT scans in which the aorta is visible, like scans of the abdomen or cardiac scans, either with or without contrast enhancement. The segmentation method for the delineation of the aorta is general, but it is likely that different atlas sets would need to be provided. The registration of atlas scans to the target scans needs to be successful, but both CT and CTA scans are normally of high enough quality. In CTA scans the segmentation result might even be better. Once the aorta is segmented, candidate objects can be extracted following the same approach, but in the case of CTA data the threshold value would need to be increased, as was done in [19]. All features used in the presented method for the candidate objects description can be computed. We do expect however, that the pattern recognition system would need a new training data set.

It would also be interesting to investigate the effect of the position of the calcifications within the artery. Once the aorta is segmented, it would be possible to automatically separate the ascending aorta, the descending aorta and the aortic arch. This would enable investigation of the risks depending on the loci of calcifications.

6 Conclusion

A method for automatic detection and quantification of calcifications in low-dose, non-ECG synchronized, non-contrast enhanced CT scans of the chest has been presented. The system correctly detected 84.1% of all aortic calcifications with on average 1.8 false positive calcifications per scan. Correlation with human observers was very high ($\rho = 0.97$). By providing automatic quantification of calcium burden in the aorta from thoracic CT scans, a risk marker for cardiovascular disease becomes available in these subjects at no additional radiation burden to the patient and no additional work for the radiologist.
Appendix: Multi-atlas based segmentation

In all scans the aorta was segmented automatically using a multi-atlas based segmentation approach. Initially an affine transformation was used to get a global alignment of the atlas and target image. Subsequently, a nonrigid registration was applied to account for local differences between the atlas and the target image. This nonrigid transformation was modeled by B-splines [33, 34]. For the optimization of the cost function (negative mutual information) an iterative stochastic gradient descent optimizer was used. In each iteration a step was taken towards the minimum. The direction of this step was based on the derivative of the cost function to the transformation parameters. The derivative was calculated based on a small subset of the image samples, randomly chosen every iteration, in order to speed up the registration. A multi-resolution strategy was taken to avoid local minima in the cost function. A Gaussian pyramid was employed, using a sub-sampling factor of two in each dimension. Also, a multi-grid approach was used for the nonrigid registration: the registration started with a coarse B-spline control point grid, which was refined in subsequent resolutions [35]. The experiments were performed using the following settings: For the affine registration four resolutions were used, in each of which 512 iterations of the stochastic gradient descent optimizer were performed. The derivative of the mutual information was calculated based on 2048 image samples, randomly chosen every iteration. For the nonrigid B-spline registration five resolutions were used. The B-spline grid spacing used in these resolutions was 64, 64, 32, 16, and 8 voxels respectively. The optimizer performed 256 iterations in each resolution. To estimate the derivative of the mutual information 4096 image samples were used, again randomly chosen every iteration. For both affine and nonrigid registration 32 histogram bins were used.

The implementation of the registration algorithm applied can be found at http://elastix.isi.uu.nl. This software package is based on the Insight Segmentation and Registration Toolkit (ITK), which can be found at www.itk.org.

To determine how much a propagated label of each transformed atlas image should contribute to the segmentation of the target, local weights were assigned based on the absolute difference image (D) between the transformed moving atlas and the target image. First, D was convolved with a Gaussian kernel at a width of one voxel to obtain a smoothed local estimate of the registration success. The assigned weights were inversely proportional to the voxel value in D. Thus, large values in D resulted in small weights, and vice versa. Adding all weighted propagated segmentations resulted in the probabilistic segmentation of the target. A binary segmentation was obtained by first blurring the probabilistic segmentation with a Gaussian kernel with
a width of 0.5 voxels, and subsequently thresholding at 0.5.

In our experiments eight atlas images were used, and they were not contained in the training nor in the test set of this study.

References


