Automated aortic calcium scoring on low-dose chest computed tomography

Ivana Išgum[a)]
Image Sciences Institute, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands

Annemarieke Rutten and Mathias Prokop
Department of Radiology, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands

Marius Staring, Stefan Klein, Josien P. W. Pluim,
Max A. Viergever, and Bram van Ginneken
Image Sciences Institute, University Medical Center Utrecht, Utrecht 3584, The Netherlands

(Received 17 June 2009; revised 19 October 2009; accepted for publication 11 December 2009; published 21 January 2010)

Purpose: 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. The authors developed and evaluated a computer-aided system for automatic detection and quantification of aortic calcifications in low-dose noncontrast-enhanced chest CT.

Methods: The system was trained and tested on scans from participants of a lung cancer screening trial. A total of 433 low-dose, non-ECG-synchronized, noncontrast-enhanced 16 detector row examinations of the chest was 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 multiatlas-based segmentation method was developed to delineate the aorta. Segmented volume was thresholded and potential calcifications (candidate objects) were extracted by three-dimensional connected component labeling. Due to image resolution and noise, in rare cases extracted candidate objects were connected to the spine. They were separated into a part outside and parts inside the aorta, and only the latter was further analyzed. All candidate objects were represented by 63 features describing their size, position, and texture. Subsequently, a two-stage classification with a selection of features and *k*-nearest neighbor classifiers was performed. Based on the detected aortic calcifications, total calcium volume score was determined for each subject.

Results: The computer system correctly detected, on the average, 945 mm^3 out of 965 mm^3 (97.9%) calcified plaque volume in the aorta with an average of 64 mm^3 of false positive volume per scan. Spearman rank correlation coefficient was $\rho=0.960$ between the system and the first observer compared to $\rho=0.961$ between the two observers.

Conclusions: Automatic calcium scoring in the aorta thus appears feasible with good correlation between manual and automatic scoring.
[DOI: 10.1118/1.3284211]

Key words: automatic calcium scoring, aorta, aortic segmentation, atlas based segmentation

I. 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 in lung cancer screening scans.[2]

The presence of coronary calcifications is a strong predictor of cardiovascular risk.[3,4] Coronary calcium scoring is typically performed on a noncontrast-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 the relationship and prognostic value of aortic calcium and cardiovascular disease, e.g., Refs. 7–12. Association between aortic calcification and other diseases has been studied as well.[13,14] Moreover, preoperative 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.[16,17]

The amount of aortic calcium in scans from smokers can be considerable and the scoring procedure will therefore require a lot of user interaction. Manual scoring is further demanding because in lung cancer screening 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.

Only a few methods for automatic arterial calcification detection have been developed, and commercial software packages do not provide it. We previously presented an automatic method for the detection of the aortic calcifications in computed tomography angiography (CTA) scans of the abdomen,[18] and an automatic method for the detection of coronary calcifications in noncontrast-enhanced, prospectively ECG-triggered cardiac CT scans.[19] Both methods considered all high density areas in the complete scan as potential calcifications. Because of the contrast present in the CTA images, these areas were extracted using a high threshold value of 220 HU. Subsequently, object-based classification methods were employed to separate true aortic or cardiac calcifications from other high density areas based on features derived from the potential calcifications and their contextual information. Saur *et al.*[20] proposed a method for both coronary and aortic calcium scoring, where a combination of noncontrast- and contrast-enhanced data was used. Coronary arteries and the aorta were segmented in contrast-enhanced data, and both scans were used for calcium detection. Brunner *et al.*[21] utilized unsupervised classification for the detection of arterial calcifications in cardiac CT scans. de Bruijne[22] presented a pixel classification-based method for automated detection of calcifications in the lumbar aorta in projection radiographs.

In this work, we present a system for automatic detection and quantification of aortic calcifications in low-dose, noncontrast, 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, the level of 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 data from a lung cancer screening cohort.

This paper is organized as follows. Section II describes the data and the reference standard. Section III gives a detailed overview of the method. Next, the results are presented in Sec. IV. Section V provides a discussion and Sec. VI concludes.

II. MATERIALS

II.A. Data

In this study, low-dose, non-ECG-synchronized, noncontrast-enhanced CT scans of the chest were used. They were acquired during a population-based randomized lung cancer screening trial.[23] 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). All scans were realized in about 12 s in spiral mode with 16×0.75 mm^2 collimation. Axial images 1.0 mm thick 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 120–140 kVp depending on the patient weight, with tube current of 30 mA s. The pitch was 1.3. 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.[24]

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 yr.

To control image noise and to obtain data comparable with the previously published results,[4,25] scans were subsampled to 3.1 mm thick sections by averaging four consecutive sections.

II.B. 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 more than 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.

The descending aorta and the spine can be in very close proximity. Because of the limited image resolution and noise in the scans, calcifications in the descending aorta occasionally appeared connected to the spine. When such a calcification was identified, the color overlay would spread into the adjacent bony structures. In such a situation, the observers were instructed to outline the area containing the aortic calcification using a circle tool: The center of the circle was determined by a mouse click and the appropriate radius was determined by dragging the mouse away from the center. After the circle was drawn, the area above the threshold of 130 HU was shown in a color overlay (aortic calcification). This was performed on a per section basis.

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 observers. The results of the first observer were used as a reference in this study. The scores of the second observer were used to estimate interobserver variability, which was compared to the results of the automatic system.

In this work results are expressed in terms of calcium volume scores following the algorithm described by Ohnesorge *et al.*[26]

III. 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. A flow diagram of the proposed method is shown in Fig. 1.

III.A. Aortic segmentation

Automatic delineation of the aorta in a noncontrast-enhanced CT scans is a very complex task due to the absence of sharp edges (gradient information) between different anatomical structures adjacent to the aorta. This is especially the case in the ascending aorta. Our previous work[19] utilized a Hough transform-based segmentation method applied to the aorta, but although this provides an estimate of its location, it does not yield an accurate segmentation. Kurkure *et al.*[27] presented a method for segmentation of the ascending and descending aorta based on the Hough transform. Recently, Feuerstein *et al.*[28] presented a method for the segmentation of the aortic arch and its branches in the thoracic CT scans. Their method is based on Hough and Euclidean distance transforms, followed by probability weighting. Kitasaka *et al.*[29] and Taeprasartsit and Higgins[30] presented a model based segmentations of the aorta. We have developed a

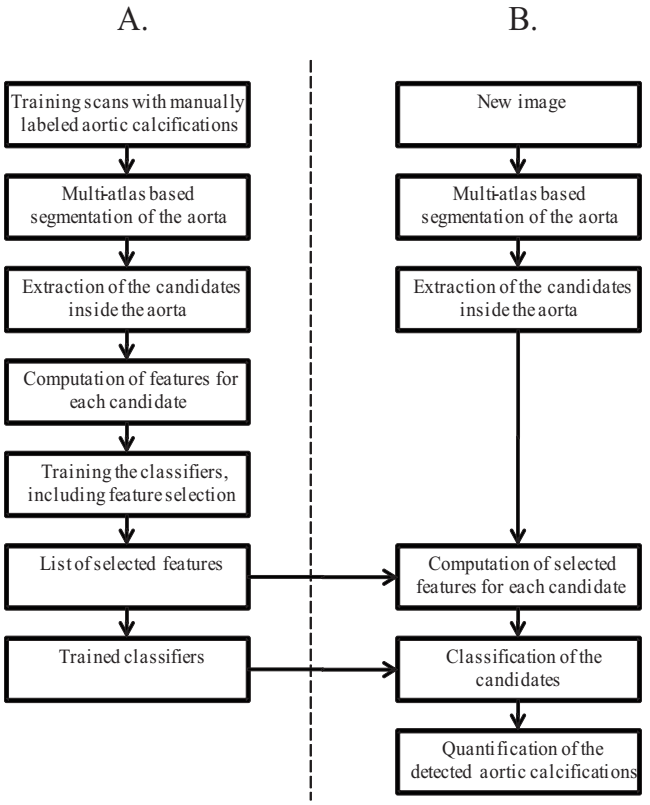

FIG. 1. Flowchart of the system for the automatic detection and quantification of aortic calcifications. (A) Training phase. (B) Testing phase.

multiatlas-based segmentation method.[31] This method was employed here for the aortic segmentation and here we provide a short description.

In registration, the spatial correspondence between voxels in two images, the fixed and the moving image, is determined. The moving image is warped so that it appears to be as similar as possible to the fixed image. The similarity between the fixed image U and moving image A is maximized with respect to the transformation $\boldsymbol{u}$,

$$\hat{\boldsymbol{u}} = \arg\min_{\boldsymbol{u}} \mathcal{C}[\boldsymbol{u}; U(\boldsymbol{p}), A(\boldsymbol{p})], \tag{1}$$

where $\hat{\boldsymbol{u}}$ is the optimal transformation making $A(\boldsymbol{u}(\boldsymbol{p}))$ spatially aligned to $U(\boldsymbol{p})$, $\boldsymbol{p}=(x,y,z)$ denotes a voxel in the image, and $\mathcal{C}$ is an appropriate cost function. Extensive reviews on the subject of image registration are given in Refs. 32–36.

In this work, we used negative mutual information as the cost function.[37] The fixed and moving images were first roughly aligned by an affine registration, compensating for differences in global translation, rotation, and scaling. To achieve a fine alignment, nonrigid (elastic) registration was subsequently applied. The registrations were performed using an open source software package, called ELASTIX (http://elastix.isi.uu.nl). The registration framework is largely based on the papers by Rueckert *et al.*[38] and Mattes *et al.*[39]

ELASTIX optimizes the cost function by iterative stochastic gradient descent: In each iteration a step is taken toward the minimum and the direction of this step is based on the derivative of the cost function to the transformation

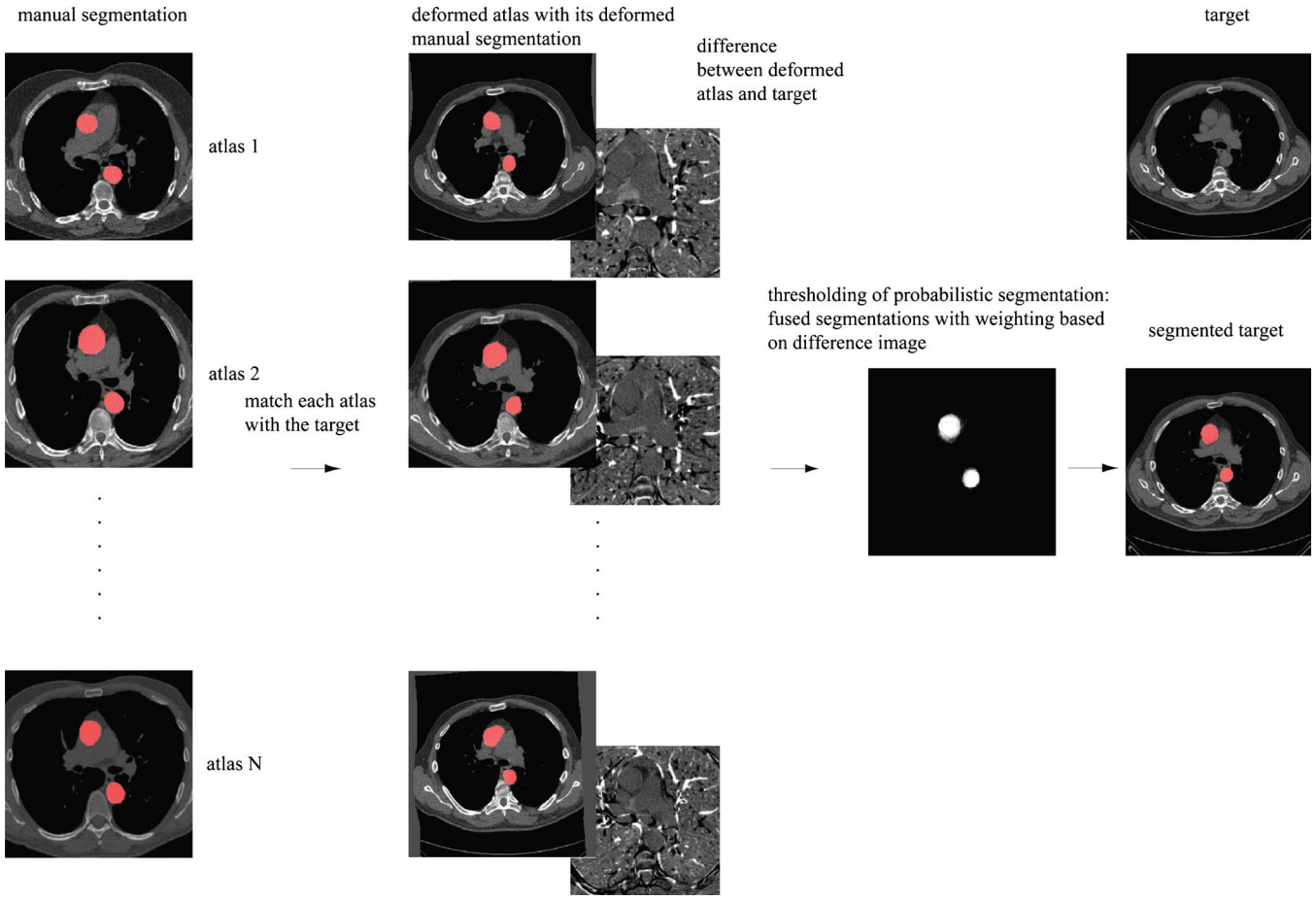

FIG. 2. Multiatlas segmentation with weighted decision fusion algorithm. N atlases (images with their manual segmentations) are matched to the target image. Subsequently, the atlases and their manual segmentations are transformed using the corresponding transformations obtained by registration. The transformed 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.

parameters. The derivative is calculated based on a small subset of the image samples, randomly chosen every iteration, in order to speed up the registration. A multiresolution strategy was taken to avoid local minima in the cost function. A Gaussian pyramid was employed using a subsampling factor of 2 in each dimension. Also, a multigrid approach was used for the nonrigid registration: The registration started with a coarse B-spline control point grid, which was refined in subsequent resolutions.[40] 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 spacings used in these resolutions were 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.

In the remainder of the text, we will refer to an image data set in which the aorta has been manually delineated as an atlas, and to an image data set that needs automatic delineation of the aorta as a target image. In atlas segmentation approaches an atlas is always chosen to be the moving image. This means that it is deformed in such a way that it aligns with the target image. Because the segmentation of the object of interest (the aorta in our case) is known for the atlas the segmentation of the target image can be obtained by applying the transformation obtained when aligning atlas and the target image to the atlas segmentation. This process is called label propagation.[41] Our method utilizes a multiatlas-based segmentation approach illustrated in Fig. 2. Each atlas provided an estimation of the position of the aorta in the target image. In our experiments eight atlas images were used, and they were not contained in the training nor in the test set of this study. These multiple aortic segmentations of the target image were combined to obtain a probabilistic segmentation of the aorta. This was performed using a weighted averaging to account for the differences in how well the deformed atlas matched the target image. The quality of the registration was evaluated from a absolute difference image D_i between the ith transformed moving image and the target. Local weight at point $\boldsymbol{p}$ was determined according to

$$\lambda_i(\boldsymbol{p}) = \frac{1}{D_i(\boldsymbol{p}) * g_{\sigma_1}(\boldsymbol{p}) + \epsilon}, \quad (2)$$

$\forall i$, where ϵ is a small value to avoid division by zero, here set to 0.001, and $*g_{\sigma_1}$ denotes convolution with a Gaussian of scale σ_1 here set to 1 voxel. In this way we obtain a smoothed local estimate of the registration success. The weight image λ_i was inversely proportional to a value in the absolute difference image.

The probabilistic segmentation of the aorta S_p was determined by a weighted average of the transformed manual segmentations $S_i(\boldsymbol{u})$,

$$S_p(\boldsymbol{p}) = \frac{1}{\sum_{i=1}^{N} \lambda_i(\boldsymbol{p})} \sum_{i=1}^{N} \lambda_i(\boldsymbol{p}) S_i(\boldsymbol{u}_i(\boldsymbol{p})). \quad (3)$$

The resulting probabilistic segmentation was thresholded at 0.3 to obtain the final aortic segmentation result. The threshold was deliberately 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.

III.B. Candidate extraction

In clinical practice, calcifications are extracted by thresholding. Various threshold values can be applied, but most commonly a threshold of 130 HU is used. In this work, a thresholding at 130 HU was employed within the segmented aortic volume to extract candidate voxels that might represent aortic calcifications. To enable analysis of objects (potential aortic calcifications) connected voxels were clustered using three-dimensional connected component labeling with 26 connectivity. The algorithm used in this work is a straightforward three-dimensional extension of the algorithm for two-dimensional images described in Ref. 42 (Chap. 6.1, Algorithm 6.1.)

Connected components located inside the aortic volume were further considered and are referred to as candidate objects. In other words, all connected components which had at least 1 voxel contained in the binary segmentation were considered candidate objects for aortic calcifications. The extracted candidate objects not only represented calcifications in the wall of the aorta, but also noise, motion artifacts, artifacts coming from metal implants, pieces of spine, and various types of calcifications located in the vicinity of the aorta (e.g., in the aortic valve or in the coronary arteries) that were extracted due to a somewhat erroneous segmentation result.

Because of noise and limited image resolution, the descending aorta and the spine may appear connected in some situations. Therefore, at such a location, aortic calcification can appear spatially connected to the spine. To resolve this, we considered all candidate objects with a volume larger than 5000 mm^3. Calcification cannot be so large so these objects can safely be assumed to contain bony anatomy. All voxels that were part of such a candidate object and that were not contained in the aortic segmentation were removed from that candidate. This procedure may result in several remaining clusters of voxels (instead of a single connected component as previously extracted). Therefore, connected component labeling was repeated to obtain candidate objects for further analysis. An example of such a case is illustrated in Fig. 3.

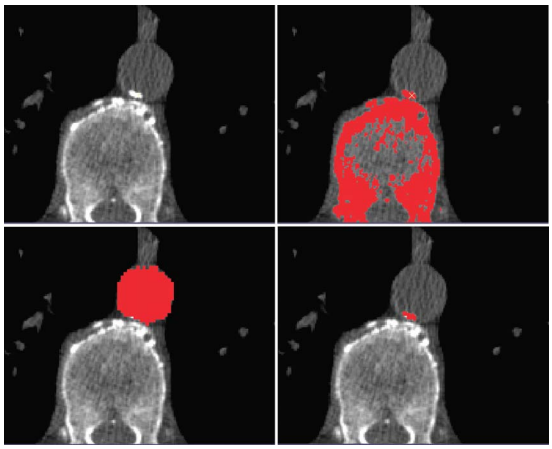

FIG. 3. An example of aortic calcification that appears to be connected to the spine: Original image (top, left). An overlay shows the candidate object exceeding the maximum allowed volume of 5000 mm^3 (top, right). The segmentation of the aorta (bottom, left). Final candidate object after removing parts outside the aorta segmentation (bottom, right).

III.C. 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 works[18,19] 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[43] through second order (L, L_x, L_y, L_z, L_{xx}, L_{yy}, L_{zz}, L_{xy}, L_{xz}, and L_{yz}) at scales σ=1, 2, 4, 8, 16 voxels calculated in the x, y, and z directions at the 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 segmented volume; the percentage of the candidate object's volume within the binary aortic segmentation; the value in a three-dimensional signed distance transform[42] of the aortic segmentation at the center of mass of the candidate object; the minimum value in the distance transform 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; and the minimum of the previous two values. This resulted in a total of 63 features, they are listed in Table I.

TABLE I. List of features describing candidate objects.

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

III.D. Classification

Statistical classifiers were used to separate the candidate objects. In our experiments, a 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 than a single classification step.

Before classification, all features were scaled to zero mean and unit variance to account for differences in the ranges of values for different features. In both stages sequential forward floating selection of features was employed to select the best features for the task.[44,45] This feature selection method 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 could be found. Classification performance was evaluated by accuracy over all candidate objects. The feature selection was performed on the training set only. Pilot experiments, using only the training set, were performed to select classifiers and find appropriate settings for parameters. It is not feasible to test all possible combinations of settings, but we experimented with the following. Three supervised classifiers were tested: Linear discriminant, quadratic discriminant, and k-nearest neighbor (kNN) classifier.[46] The best results were obtained with the kNN classifier. Therefore, in both classification stages a kNN classifier with Euclidean metric was employed. The performance of the kNN classifier was tested with the number of neighbors k ranging from 1 to 15 (odd values only) in both classification stages. In each of these experiments, a feature selection was performed. The maximum number of features to be selected was set to 3, 5, 15, 25, and 35. Once the best k and the maximum number of features were chosen for the first classification stage, an optimal threshold on the posterior probability was investigated. Because the goal was to discard only candidate objects which had a very high probability for being negative, tested values were 0.75, 0.85, and 0.95. The performance was evaluated based on the final classification (object-based) accuracy of the system. The settings that gave the best results were used for evaluation.

In the first classification stage, the number of neighbors' k was set to 10. The maximum number of features to be selected was set to 25. All objects with a posterior probability for a negative class larger than 0.85 were discarded. This means that at least nine out of the ten nearest neighbors from the training set of the sample at hand must have been negatives. Such candidate objects 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 11, and the maximum number of features to be selected was set to 25. The 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 aortic calcifications and others as noncalcifications.

III.E. 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.*[47] Our choice was to compute the volume of the detected calcifications implemented as described in Ref. 26.

III.F. Evaluation

Descriptive statistics was used to summarize the number of aortic calcifications and the corresponding volume scores. The 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 the 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 ("nonaortic 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 nonaortic 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 nonaortic 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.

IV. RESULTS

In 93 test scans the first observer selected 1036 aortic calcifications (positives). Distribution per scan was median =7; 25th percentile=3; 75th percentile=14; and range: 0–51. In terms of volume, these scans contained 89 750 mm^3 of aortic calcium (per scan median=308 mm^3; 25th percentile =62 mm^3; 75th percentile=1121 mm^3; and range: 0–9481 mm^3).

The automatic method extracted all these calcifications as candidate objects and 1239 other candidate objects (negatives). The distribution of the negatives per scan was median=15; 25th percentile=6; 75th percentile=28; and range of 0–312. In terms of volume, this corresponded to 36 940 mm^3 (per scan: median=54 mm^3; 25th percentile =10 mm^3; 75th percentile=363 mm^3; and range: 0–8 037 mm^3) of negatives.

By discarding large candidate objects, one object (on the average, 0.01 objects per scan) was eliminated from the test set without removing any calcifications in the aorta. This candidate object was a large piece of spine which occurred due to erroneous aortic segmentation. In total, 1238 (29 885 mm^3) negatives remained.

In the first classification stage 22 out of the maximum 25 features were selected. According to the ordering in Table I those were 6, 58, 3, 56, 15, 30, 4, 2, 5, 60, 10, 12, 63, 25, 7, 39, 61, 9, 40, 14, 55, and 34. In the second classification stage 23 out of the maximum 25 features were selected. Those were features numbered as 30, 1, 3, 58, 6, 10, 27, 2, 29, 34, 25, 60, 59, 61, 13, 48, 24, 21, 4, 12, 5, 63, and 40.

The computer system correctly identified 85.3% of all aortic calcifications and 97.9% of the calcified aortic plaque volume. For an average number of 9.5 (945 mm^3) true positive calcifications, an average of 1.3 (64 mm^3) false positive calcifications per scan were detected. Also, on the average there were 1.6 (20 mm^3) false negative objects.

The system assigned a zero score to 3/4 subjects (75%) where in both observers assigned a zero calcium score. Thus, the system assigned a positive score to 1/4 subjects (25%) where in both observers assigned a zero score. The two observers did not agree about zero scores in three subjects.

The detailed numbers for false positive and false negative errors are given in Table II. The main causes of false positives were noise and motion artifacts (62%), calcifications detected in the descending aorta below the level of the apex of the heart (15%), calcifications in the arteries branching out of the aortic arch (8%). Figure 4 shows examples of false positive errors, and Fig. 5 examples of false negative errors.

TABLE II. The number and volume of all candidate objects averaged over 93 test scans. The candidates are divided into 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 subcategories.

Category	Objects/scan	Volume (mm^3)/scan
True negatives	12.0	333
False positives		
Noise and artifacts (motion, swallowing)	0.8	12
Spine	0.05	8
Calcium in		
descending aorta below the heart	0.2	13
arteries above the arch	0.1	11
coronary arteries	0.06	9
mediastinum	0.05	6
aortic valve	0.04	5
Total	1.3	64
Total negatives	13.3	397
True positives	9.5	945
False negatives		
Descending aorta	0.8	6
Ascending aorta	0.5	12
Aortic arch	0.3	2
Total	1.6	20
Total positives (reference standard)	11.1	965

Figure 6 (a) correlates the volume scores assigned to each subject by the automatic system with that of the first observer. Figure 6 (b) correlates the scores of the first and second observers. Spearman rank correlation demonstrated excellent agreement with a correlation coefficient $\rho=0.960$ between automatic system and reference, and $\rho=0.961$ between the observers. Identical calcium scores for the computer system and the first observer were found in 25 (27%) scans. Identical calcium scores for the first and the second observers were obtained in 23 (25%) scans.

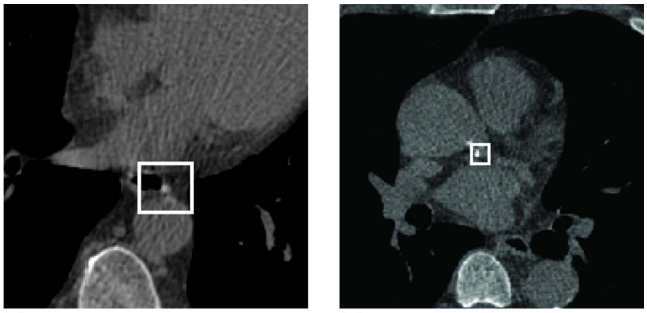

FIG. 4. Examples of false positive objects. Swallowing artifact around the esophagus identified as aortic calcium by the computer system (a). Left main coronary artery calcification selected as aortic calcification by the computer system (b). Both objects are blurred due to motion artifacts.

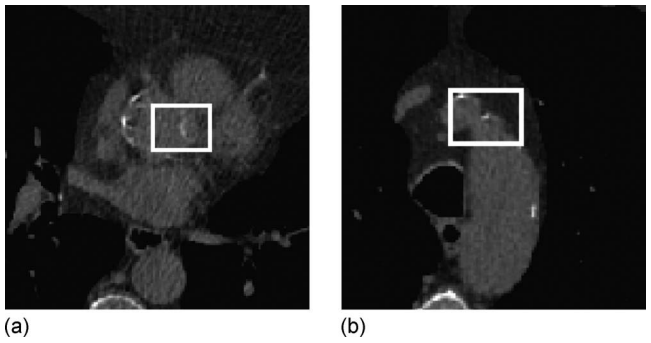

FIG. 5. Examples of false negative objects. Calcification in the ascending aorta (a). Because of the motion, this lesion seems positioned inside the aortic lumen. Calcification in the aortic arch at the point where an artery is branching off the aorta (b).

V. DISCUSSION

The presented system automatically segmented the aorta and correctly identified the vast majority of aortic calcifications. The results from the computer system correlated only slightly less with the results of a human reference standard (ρ=0.960) than two human observers with each other (ρ =0.961). Even for the presence or total absence of calcium in the aorta, the system performed well. There was only one subject to which both observers assigned a zero score, and the system assigned a positive score. The observers disagreed on a zero score in three cases.

The majority (62%) 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. Note that these errors could occur even with a slight oversegmentation of the aorta: The required overlap between the extracted connected component and the aortic segmentation was 1 voxel. Some of these errors are a question of judgment (e.g., calcification located in the aortic root or in the proximal coronary [see Fig. 4 (b)] or precise definition of the aortic contour (calcification located in the aortic arch or extending into a branching vessel). No general agreement exists where to place a distal cutoff point in the thoracic aorta. If a cardiovascular risk is evaluated across a population, calcium scores need to be determined in the same anatomical range. We have chosen the apex of the heart as the reference. By using a different reference point, e.g., the renal arteries, or the celiac artery, some of this variability might be reduced. However, in practice, this is not feasible because those arteries are not always included in the scan range of a thoracic 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 nonaortic 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 marked in Fig. 6 (a), where the disagreement between the automatic and reference calcium score was the largest, the results were dominated by (1) a calcification in mediastinum, (2) part of the spine adjacent to the descending aorta, and (3) a false positive in the left main coronary artery. 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. The disagreement on zero score scans between the observers was caused

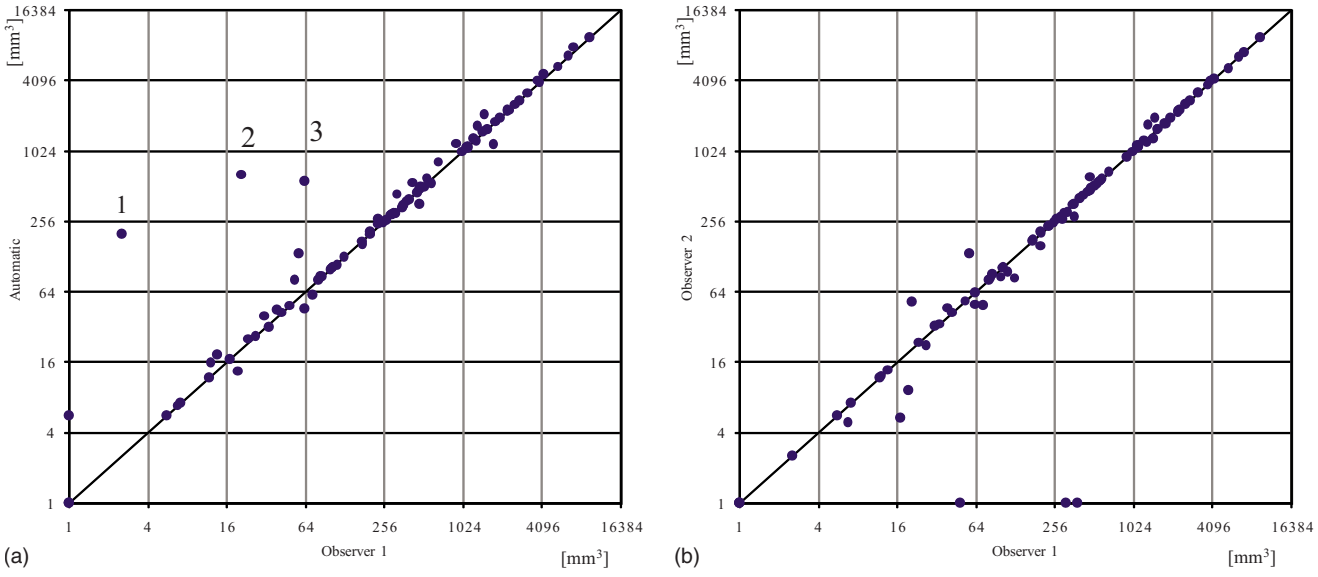

FIG. 6. Volume scores assigned to subjects by the first observer (x axis), vs (a) the score assigned by the computer system and (b) the second observer (y axis). Equal volumes are demonstrated by the solid line. The data is plotted with the logarithmic scale (x+1).

by calcifications in the ascending aorta and the aortic arch. Only in one case the appearance of the calcification in the ascending aorta was influenced by cardiac motion.

In some cases candidate objects appeared connected to the spine. They were separated into parts outside and parts inside the aorta. Only the latter was further analyzed. This separation was performed using the automatically obtained aortic segmentation. If the aortic segmentation was not very accurate, the candidate object might not completely overlap with its manual segmentation (reference standard). The difference in segmentation would reflect the volume score. We inspected these candidates in the training set and no major errors were found.

The system presented here bears some similarities with our method for the detection of coronary calcifications.[19] However, there are several important differences. First, the scans were acquired with different protocols. Images used in this work were acquired with low dose and without ECG synchronization. This means that the presented system had to deal with high noise level and artifacts originating from the cardiac motion. Second, the method presented in Ref. 19 analyzed candidates for coronary calcifications in the complete scan. The system presented here employs a precise multiatlas aortic segmentation which allowed for an analysis of the segmented volume only. Last, a dedicated set of position features was created using the results of the aortic segmentation.

Aortic calcifications have not widely been used as a cardiovascular risk marker, but a recent publication suggests that they are a strong independent 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 nongated 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.

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 Ref. 18. 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.

VI. CONCLUSION

A method for automatic detection and quantification of calcifications in low-dose, non-ECG-synchronized, noncontrast-enhanced CT scans of the chest has been presented. The system correctly detected 85.3% of all aortic calcifications with, on the average, 1.3 false positive calcifications per scan. Correlation with a human observer was very high (ρ=0.960). 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.

[a]Electronic mail: ivana@isi.uu.nl

[1]L. J. Shaw, P. Raggi, T. Q. Callister, and D. S. Berman, "Prognostic value of coronary artery calcium screening in asymptomatic smokers and nonsmokers," Eur. Heart J. **27**(8), 968–975 (2006).
[2]J. Shemesh, C. I. Henschke, A. Farooqi, R. Yip, D. F. Yankelevitz, D. Shaham, and O. S. Miettinen, "Frequency of coronary artery calcification on low-dose computed tomography screening for lung cancer," Clin. Imaging **30**(3), 181–185 (2006).
[3]R. Vliegenthart, M. Oudkerk, B. Song, D. A. M. van der Kuip, A. Hofman, and J. C. M. Witteman, "Coronary calcification detected by electron-beam computed tomography and myocardial infarction. The Rotterdam coronary calcification study," Eur. Heart J. **23**(20), 1596–1603 (2002).
[4]R. C. Detrano, M. Anderson, J. Nelson, N. D. Wong, J. J. Carr, M. McNitt-Gray, and D. E. Bild, "Coronary calcium measurements: Effect of CT scanner type and calcium measure on rescan reproducibility—MESA study," Radiology **236**(2), 477–484 (2005).
[5]A. Rutten, S. P. J. Krul, M. F. L. Meijs, A. M. de Vos, M.-J. M. Cramer, and M. Prokop, "Variability of coronary calcium scores throughout the cardiac cycle: Implications for the appropriate use of electrocardiogram dose modulation with retrospectively gated computed tomography," Invest. Radiol. **43**(3), 187–194 (2008).
[6]S. M. Kim, M. J. Chung, K. S. Lee, Y. H. Choe, C. A. Yi, and B.-K. Choe, "Coronary calcium screening using low-dose lung cancer screening: Effectiveness of MDCT with retrospective reconstruction," AJR, Am. J. Roentgenol. **190**(4), 917–922 (2008).
[7]J. Takasu, R. Katz, K. Nasir, J. J. Carr, N. Wong, R. Detrano, and M. J. Budoff, "Relationships of thoracic aortic wall calcification to cardiovascular risk factors: The multi-ethnic study of atherosclerosis (MESA)," Am. Heart J. **155**(4), 765–771 (2008).
[8]K. Nasir, R. Katz, J. Takasu, D. M. Shavelle, R. Detrano, J. A. Lima, R.

S. Blumenthal, K. O'Brien, and M. J. Budoff, "Ethnic differences between extra-coronary measures on cardiac computed tomography: Multi-ethnic study of atherosclerosis (MESA)," Atherosclerosis **198**(1), 104–114 (2008).
[9]Y. Itani, S. Watanabe, and Y. Masuda, "Relationship between aortic calcification and stroke in a mass screening program using a mobile helical computed tomography unit," Jpn. Circ. J. **70**(6), 733–736 (2006).
[10]C. R. Walsh, L. A. Cupples, D. Levy, D. P. Kiel, M. Hannan, P. W. F. Wilson, and C. J. O'Donnell, "Abdominal aortic calcific deposits are associated with increased risk for congestive heart failure: The Framingham heart study," Am. Heart J. **144**(4), 733–739 (2002).
[11]C. Iribarren, S. Sidney, B. Sternfeld, and W. S. Browner, "Calcification of the aortic arch: Risk factors and association with coronary heart disease, stroke, and peripheral vascular disease," JAMA, J. Am. Med. Assoc. **283**(21), 2810–2815 (2000).
[12]A. Eisen, A. Tenenbaum, N. Koren-Morag, D. Tanne, J. Shemesh, M. Imazio, E. Z. Fisman, M. Motro, E. Schwammenthal, and Y. Adler, "Calcification of the thoracic aorta as detected by spiral computed tomography among stable angina pectoris patients: Association with cardiovascular events and death," Circulation **118**(13), 1328–1334 (2008).
[13]Y. Z. Bagger, L. B. Tank, P. Alexandersen, G. Qin, and C. Christiansen, "Radiographic measure of aorta calcification is a site-specific predictor of bone loss and fracture risk at the hip," J. Intern Med. **259**(6), 598–605 (2006).
[14]F. E. de Leeuw, J. C. D. Groot, M. Oudkerk, J. C. Witteman, A. Hofman, J. van Gijn, and M. M. Breteler, "Aortic atherosclerosis at middle age predicts cerebral white matter lesions in the elderly," Stroke **31**(2), 425–429 (2000).
[15]R. Lee, N. Matsutani, A. C. Polimenakos, L. C. Levers, M. Lee, and R. G. Johnson, "Preoperative noncontrast chest computed tomography identifies potential aortic emboli," Ann. Thorac. Surg. **84**(1), 38–41 (2007) discussion 42.
[16]A. S. Agatston, W. R. Janowitz, F. J. Hildner, N. R. Zusmer, M. Viamonte, and R. Detrano, "Quantification of coronary artery calcium using ultrafast computed tomography," J. Am. Coll. Cardiol. **15**(4), 827–832 (1990).
[17]C. Hong, K. T. Bae, T. K. Pilgram, and F. Zhu, "Coronary artery calcium quantification at multidetector row CT: Influence of heart rate and measurement methods on interacquisition variability initial experience," Radiology **228**(1), 95–100 (2003).
[18]I. Isgum, B. van Ginneken, and M. Olree, "Automatic detection of calcifications in the aorta from CT scans of the abdomen," Acad. Radiol. **11**(3), 247–257 (2004).
[19]I. Isgum, A. Rutten, M. Prokop, and B. van Ginneken, "Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease," Med. Phys. **34**(4), 1450–1461 (2007).
[20]S. C. Saur, H. Alkadhi, L. Desbiolles, G. Székely, and P. C. Cattin, "Automatic detection of calcified coronary plaques in computed tomography data sets," *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science Vol. 5241 (Springer, Berlin, 2008), pp. 170–177.
[21]G. Brunner, U. Kurkure, D. R. Chittajallu, R. P. Yalamanchili, and I. A. Kakadiaris, "Toward unsupervised classification of calcified arterial lesions," *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science Vol. 5241 (Springer, Berlin, 2008), pp. 144–152.
[22]M. de Bruijne, "A pattern classification approach to aorta calcium scoring in radiographs," *Computer Vision for Biomedical Image Applications* (Springer, Berlin, 2005), Vol. 3765, pp. 170–175.
[23]C. A. van Iersel, H. J. de Koning, G. Draaisma, W. P. T. M. Mali, E. T. Scholten, K. Nackaerts, M. Prokop, J. D. F. Habbema, M. Oudkerk, and R. J. van Klaveren, "Risk-based selection from the general population in a screening trial: Selection criteria, recruitment and power for the Dutch-Belgian randomised lung cancer multi-slice CT screening trial (NELSON)," Int. J. Cancer **120**(4), 868–874 (2006).
[24]D. M. Xu, H. Gietema, H. de Koning, R. Vernhout, K. Nackaerts, M. Prokop, C. Weenink, J. Lammers, H. Groen, M. Oudkerk, and R. van Klaveren, "Nodule management protocol of the NELSON randomised lung cancer screening trial," Lung Cancer **54**(2), 177–184 (2006).
[25]J. J. Carr, J. C. Nelson, N. D. Wong, M. McNitt-Gray, Y. Arad, D. R. Jacobs, S. Sidney, D. E. Bild, O. D. Williams, and R. C. Detrano, "Calcified coronary artery plaque measurement with cardiac CT in population-based studies: Standardized protocol of multi-ethnic study of atherosclerosis (MESA) and coronary artery risk development in young adults (CARDIA) study," Radiology **234**(1), 35–43 (2005).
[26]B. Ohnesorge, T. Flohr, R. Fischbach, A. F. Kopp, A. Knez, S. Schröder, U. J. Schöpf, A. Crispin, E. Klotz, M. F. Reiser, and C. R. Becker, "Reproducibility of coronary calcium quantification in repeat examinations with retrospectively ECG-gated multisection spiral CT," Eur. Radiol. **12**(6), 1532–1540 (2002).
[27]U. Kurkure, O. C. Avila-Montes, and I. A. Kakadiaris, "Automated segmentation of thoracic aorta in non-contrast CT images," Proceedings of the fifth IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008 (ISBI 2008), 2008, pp. 29–32 (unpublished).
[28]M. Feuerstein, T. Kitasaka, and K. Mori, "Automated anatomical likelihood driven extraction and branching detection of aortic arch in 3-D chest CT," Second International Workshop on Pulmonary Image Analysis, 2009, pp. 49–60 (unpublished).
[29]T. Kitasaka, K. Mori, J. Hasegawa, J. Toriwaki, and K. Katada, "Automated extraction of aorta and pulmonary artery in mediastinum from 3D chest x-ray CT images without contrast medium," SPIE Medical Imaging 4684, 1496–1507 (SPIE, San Diego, 2002).
[30]P. Taeprasartsit and W. E. Higgins, "Method for extracting the aorta from 3D CT images," SPIE Medical Imaging 6512, pp. 65120J-1–65120J-16 (SPIE, San Diego, 2007).
[31]I. Isgum, M. Staring, A. Rutten, M. Prokop, M. Viergever, and B. van Ginneken, "Multi-atlas-based segmentation with local decision fusion—Application to cardiac and aortic segmentation in CT scans," IEEE Trans. Med. Imaging **28**, 1000–1010 (2009).
[32]L. G. Brown, "A survey of image registration techniques," ACM Comput. Surv. **24**, 325–376 (1992).
[33]J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," Med. Image Anal. **2**, 1–36 (1998).
[34]H. Lester and S. R. Arridge, "A survey of hierarchical non-linear medical image registration," Pattern Recogn. **32**, 129–149 (1999).
[35]D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, "Medical image registration," Phys. Med. Biol. **46**, R1–R45 (2001).
[36]A. Gholipour, N. Kehtarnavaz, R. Briggs, M. Devous, and K. Gopinath, "Brain functional localization: A survey of image registration techniques," IEEE Trans. Med. Imaging **26**(4), 427–451 (2007).
[37]P. Thévenaz and M. Unser, "Optimization of mutual information for multiresolution image registration," IEEE Trans. Image Process. **9**, 2083–2099 (2000).
[38]D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," IEEE Trans. Med. Imaging **18**(8), 712–721 (1999).
[39]D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," IEEE Trans. Med. Imaging **22**(1), 120–128 (2003).
[40]S. Klein, M. Staring, and J. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines," IEEE Trans. Image Process. **16**, 2879–2890 (2007).
[41]T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer Jr., "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," Neuroimage **21**(4), 1428–1442 (2004).
[42]M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed. (PWS, Pacific Grove, CA, 1999).
[43]L. M. J. Florack, *Image Structure* (Kluwer Academic, Dordrecht, 1997).
[44]A. K. Jain and D. Zongker, "Feature selection: Evaluation, application and small sample performance," IEEE Trans. Pattern Anal. Mach. Intell. **19**(2), 153–158 (1997).
[45]P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," Pattern Recogn. Lett. **15**(11), 1119–1125 (1994).
[46]R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. (Wiley, New York, 2001).
[47]R. W. Jayalath, S. H. Mangan, and J. Golledge, "Aortic calcification," Eur. J. Vasc. Endovasc Surg. **30**(5), 476–488 (2005).