Detection of coronary calcifications from computed tomography scans for automated risk assessment of coronary artery disease

Ivana Išgum
Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

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

Bram van Ginneken
Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

(Received 30 July 2006; revised 26 January 2007; accepted for publication 29 January 2007; published 23 March 2007)

A fully automated method for coronary calcification detection from non-contrast-enhanced, ECG-gated multi-slice computed tomography (CT) data is presented. Candidates for coronary calcifications are extracted by thresholding and component labeling. These candidates include coronary calcifications, calcifications in the aorta and in the heart, and other high-density structures such as noise and bone. A dedicated set of 64 features is calculated for each candidate object. They characterize the object's spatial position relative to the heart and the aorta, for which an automatic segmentation scheme was developed, its size and shape, and its appearance, which is described by a set of approximated Gaussian derivatives for which an efficient computational scheme is presented. Three classification strategies were designed. The first one tested direct classification without feature selection. The second approach also utilized direct classification, but with feature selection. Finally, the third scheme employed two-stage classification. In a computationally inexpensive first stage, the most easily recognizable false positives were discarded. The second stage discriminated between more difficult to separate coronary calcium and other candidates. Performance of linear, quadratic, nearest neighbor, and support vector machine classifiers was compared. The method was tested on 76 scans containing 275 calcifications in the coronary arteries and 335 calcifications in the heart and aorta. The best performance was obtained employing a two-stage classification system with a *k*-nearest neighbor (*k*-NN) classifier and a feature selection scheme. The method detected 73.8% of coronary calcifications at the expense of on average 0.1 false positives per scan. A calcium score was computed for each scan and subjects were assigned one of four risk categories based on this score. The method assigned the correct risk category to 93.4% of all scans.
[DOI: 10.1118/1.2710548]

Key words: coronary calcium scoring, coronary calcifications, computer-aided diagnosis, cardiac CT

I. INTRODUCTION

The coronary arteries supply the heart muscle with oxygenated blood. In the case of luminal narrowing of the coronaries through atherosclerotic processes, oxygen supply to the heart can become insufficient and myocardial infarction occurs. This is often the first sign of coronary atherosclerosis.[1,2] Conventional screening tests for cardiovascular disease, such as blood pressure and cholesterol level measurements, are only moderately sensitive for identifying subjects at risk.[1,3] Computed tomography (CT) calcium scoring can increase this sensitivity. For calcium scoring a noncontrast enhanced CT scan of the heart is obtained. On this scan calcifications in the coronary arteries can be detected and quantified.[4] The amount of calcium can be expressed in terms of Agatston, volume, or mass score,[5,6] and these scores estimate the risk for future cardiac morbidity and/or mortality.[7]

Atherosclerotic plaques can also be visualized by x ray, magnetic resonance imaging, and ultrasound.[8,9] Recently, a feasibility study was performed for coronary calcification quantification by dual-energy imaging in combination with a flat panel detector.[10] The advantage of such a system would be lower effective dose to a patient and lower costs, but maybe less accurate calcium quantification. In this work we focus on calcium scoring with CT only.

Several vendors of commercial workstations for medical image analysis offer tools for coronary calcium scoring, but these tools all require manual selection of the calcifications. Typically, all candidates for calcifications are extracted by thresholding and component labeling. From these candidates, each calcification within one of the coronary arteries has to be identified by a human operator. After this has been done, a calcium score for each coronary artery as well as total coronary calcium score can be easily computed by the software. Another tool for calcium scoring,[11] which was used in some studies,[12] required manual tracking of coronary vessel centerlines. Once the course of each coronary artery had

been defined, the software automatically detected objects above the desired threshold close to the vessel centerline and computed a calcium score. Calcium scoring is thus a time-consuming task that needs to be performed by a trained human operator. As a consequence, calcium scoring is not performed routinely on all CT scans in which the heart is imaged. Automated tools for identification and quantification of coronary calcifications are desirable to reduce evaluation time and input from expert human observers. This would be especially advantageous in large-scale screening or follow-up studies that make use of calcium scoring. To our knowledge, no work on automated coronary calcium scoring has been published. The purpose of this work is to present a system that automatically assigns a risk category to a subject based on completely automated coronary calcification detection and scoring.

Coronary calcium can be located in any of the three main coronary arteries and their subbranches: left main (LM), left anterior descending (LAD), left circumflex (LCX), right coronary artery (RCA), and posterior descending (PDA). Furthermore, calcium is frequently found in the ascending and the descending aorta, and in the heart valves. In addition, other high-density objects such as bone and metal implants are often located inside or nearby the heart. Finally, noise is present in the images and often exceeds the threshold for calcification extraction. This means that automated coronary calcium detection not only requires the ability to separate calcified and noncalcified lesions, but also to separate coronary calcifications and all other high-density structures, among which are also noncoronary calcifications. Examples of coronary and noncoronary calcifications are shown in Fig. 1. To illustrate the difficulty of the task, Fig. 1(b) shows a calcification in the LCX that resembles calcium in mitral valve shown in Fig. 1(c). In such cases, a radiologist cannot differentiate between the two only by a local visual inspection. Typically, a human observer will track the artery through several slices, although often partially invisible, to establish if the candidate is located within a coronary. Another difficulty is caused by heart motion, which blurs calcium loci, as also shown in Fig. 1. In this case the object itself and its surroundings change properties such as size, shape, and intensity.

Our method does not require a segmentation of the coronary arteries. Segmentation of the coronary arteries would in principle be beneficial for identifying coronary artery calcifications, but several practical reasons speak against it. Calcium scoring is performed on non-contrast-enhanced scans. The lack of contrast material makes contrast between coronary arteries and surrounding tissue low or nonexistent. Thus, accurate automatic segmentation of the coronaries is extremely difficult. In addition, motion artifacts may occur despite ECG synchronization techniques. This can blur the coronaries up to a degree that they are no longer visible. Also, the relatively low dose used for scanning leads to noise, especially in the basal portions of the heart caused by high absorption of radiation by organs below the diaphragm.

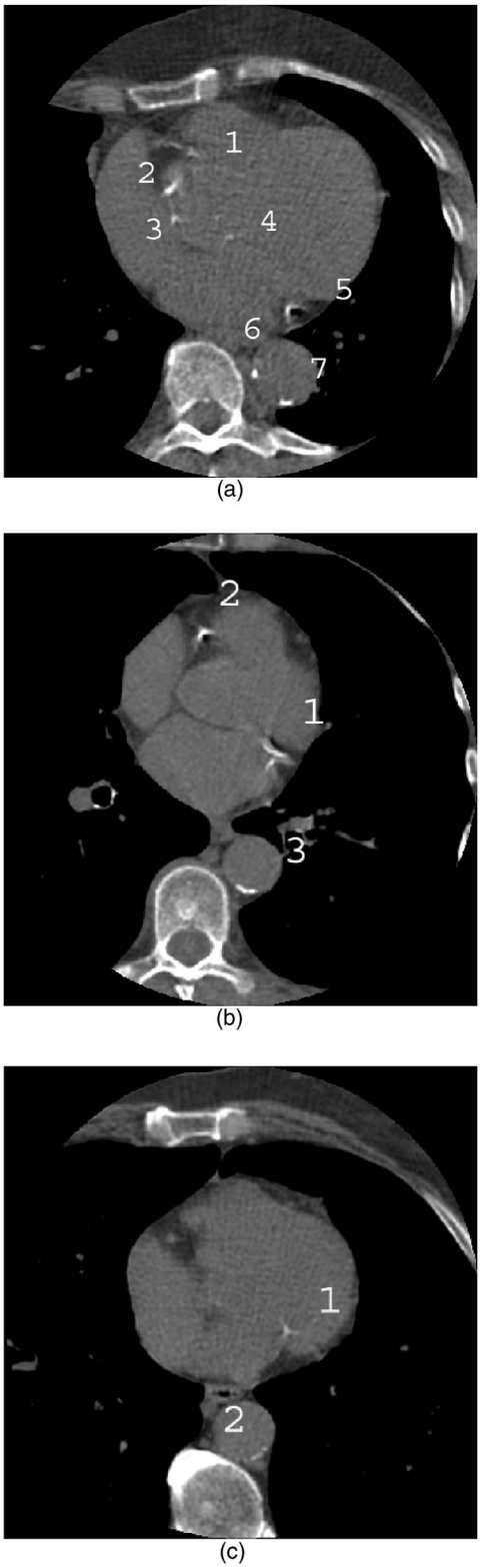

FIG. 1. (a) Calcification in (1) RCA located at the point where the artery is branching off the ascending aorta; (2), (3), and (4) calcium in the ascending aorta; (5) calcium in LCX; and (6) and (7) calcifications in the descending aorta. (b) (1) calcification in the LCX; (2) calcification in the RCA; and (3) calcium in the descending aorta. (c) (1) calcified mitral valve annulus and (2) calcification in the descending aorta. Note that calcification (1) in the (b) image resembles mitral valve calcification (1) in the (c) image. Also, note that in these figures calcifications in the heart or coronaries are blurred due to heart motion.

As a result, it is even harder to differentiate arteries from surrounding tissue. We, therefore, opted for an approach without segmentation of coronary arteries.

The method we present first extracts objects possibly representing coronary calcifications by thresholding and component labeling. Subsequently, each object is represented by features describing its size, shape, location, and appearance. From that moment, the detection of coronary calcifications is a problem in the realm of pattern recognition. We present three classification systems. In the first two systems a direct classification between coronary calcifications and other objects is attempted, with and without a feature selection scheme. The third classification system consisted of two stages where in the first stage easily recognizable noncoronary calcifications were discarded, and in the next stage more complex separation between coronary calcification and the remaining objects was performed. Experiments have been performed with a range of classifiers. Best results were obtained with *k*-nearest neighbor classifiers (*k*-NN) and support vector machines (SVM), and those are reported.

Preliminary, simpler versions of this system, trained and tested with much smaller data sets, have been presented in Refs. 13 and 14.

The paper is organized as follows: Section II describes the data. Section III gives a detailed overview of the method. Next, results are presented. Section V provides a discussion and we end with a conclusion in Sec. VI.

II. MATERIALS

In this study 304 cardiac non-contrast-enhanced CT scans were used. They were obtained as part of a study investigating the association between age at menopause and risk of cardiovascular disease. The scans were acquired on a Philips CT scanner (Mx8000IDT, Philips Medical Systems, The Netherlands) with 16×1.5 mm collimation, and a pixel size of 0.43×0.43 mm^2. They were prospectively triggered at 70% of the R-R interval (mid-diastole), the most motionless phase in scans acquired with our scanner. Peak voltage of 120 kVp and tube current between 40 and 70 mAs depending on the subject weight were used. Data were reconstructed to 512×512 matrices. The vertical range included approximately the region from the level of the tracheal bifurcation to the base of the heart. Coronary calcium scoring is usually performed on cardiac CT scans with a slice thickness of 3 mm.[4] Therefore, all scans were resampled to 3 mm thick slices by averaging pixel intensities of two neighboring slices. This resulted in a voxel size of $0.43\times0.43\times3$ mm^3.

The reference standard was set manually by a trained observer with 1 year experience (AR) who identified a point in each calcified lesion in the heart. A three-dimensional (3D) region growing was used to connect all neighboring voxels above a threshold value of 130 HU, the commonly used value for calcification extraction.[4] Additionally, each lesion was given a label depending on its location. The labels were coronary calcification and noncoronary calcification. Noncoronary calcifications included calcium in the aorta, aortic and mitral valves, and myocardium.

TABLE I. The complete set of features to characterize candidate coronary calcification objects.

Feature	Description
1	volume
2,3	λ_1/λ_3, λ_2/λ_3
4–6	x-, y- and z-coordinate in the image coordinate system
7–9	x-, y- and z-coordinate in the heart coordinate system
10	distance to the heart border
11	heart area ratio
12	distance to the aorta border
13	average object intensity
14	maximum object intensity
15–19	L for scales $\sigma=1,2,4,8,16$
20–24	L_x for scales $\sigma=1,2,4,8,16$
25–29	L_y for scales $\sigma=1,2,4,8,16$
30–34	L_z for scales $\sigma=1,2,4,8,16$
35–39	L_{xx} for scales $\sigma=1,2,4,8,16$
40–44	L_{xy} for scales $\sigma=1,2,4,8,16$
45–49	L_{xz} for scales $\sigma=1,2,4,8,16$
50–54	L_{yy} for scales $\sigma=1,2,4,8,16$
55–59	L_{yz} for scales $\sigma=1,2,4,8,16$
60–64	L_{zz} for scales $\sigma=1,2,4,8,16$

The data set was randomly divided in a training set of 228 scans and a test set of 76 scans. The training set contained 490 coronary calcifications and 1961 noncoronary calcifications. The test set contained 275 coronary calcifications and 335 noncoronary calcifications. The test set was not used in any way during development and training of the system.

III. METHODS

A. Candidate objects extraction

Thresholding is a standard way of identifying candidates for vascular calcification.[4] Scans were thresholded at 130 HU, the same threshold value used for setting the reference standard. Three-dimensional component labeling was used to connect neighboring voxels and obtain candidate objects. All objects bigger than 2500 voxels (1387 mm^3) were discarded as coronary calcifications are not expected to have such a large volume.

B. Features

Properties that may distinguish coronary calcifications from other objects include their size, shape, position, intensity values, and variations of intensities within the object and their surroundings. We designed a set of features that captures these properties. The complete list of features is given in Table I.

1. Size

Coronary calcifications typically have a smaller volume than bony structures such as the spine and ribs. Small objects most often represent noise. Therefore, an object's volume expressed in number of voxels was used as a size feature.

2. Shape

Although coronary calcifications usually cannot be distinguished from other candidates from their shape alone, some of the negatives clearly have a shape that discriminates them from positives. Examples are calcifications in mitral valves, which are often platelike, unlike coronary calcifications, which are elongated or spherical. Therefore, principal component analysis was applied to the set of voxel coordinates. This yields three eigenvectors λ_1, λ_2, and λ_3, with $\lambda_1 \leq \lambda_2 \leq \lambda_3$ for each candidate.[15] They represent the three principal axes of the objects. Two features were computed from the eigenvalues: λ_1/λ_3 and λ_2/λ_3. These features discriminate between platelike and elongated structures.

3. Spatial features

A first set of three spatial features was defined as the (x,y,z) location of the object's center of gravity in absolute image coordinates.

A scanning protocol for cardiac CT typically requires that the complete heart is contained in the field of view. As a result, the set of structures included in the field of view may vary substantially between scans depending on subject size and anatomy. In some scans the complete rib cage is visible, while in others no ribs are contained. Similarly, some scans include the complete spine and sternum, while in others those structures are partially or completely absent. In the vertical direction similar variations may occur. Because of these variations, the absolute position of an object in the scan is not always a reliable indication of its location within the body. For this reason, knowledge about the location of the heart in a scan can provide valuable features for detecting coronary calcium, because coronary calcifications are located inside the heart. Similarly, location of the aorta in a scan can be important to exclude calcifications there. Therefore, an algorithm was developed for automatic delineation of these organs. This algorithm is described in Appendix A.

The segmentation of the heart was used to describe the object's location relative to it. A coordinate system was defined by determining the smallest box around the heart. The lower-left-frontal corner of the box represents the position (0, 0, 0) and the upper-right-back corner the position (1, 1, 1). The location of the center of gravity of the object in this coordinate system yielded three additional spatial features.

Coronary arteries branch off the ascending aorta towards the heart edge and stretch along the heart surface. Therefore, the smallest Euclidean distance of an object to the border of the heart was determined. The distance was multiplied with minus one for objects inside the heart.

A position feature less sensitive to slight errors in the segmentation was computed as the ratio of two areas inside the heart mask in axial slices. The areas were defined as two parts of the heart divided by the line through the object's center of gravity and perpendicular to the line having the shortest distance between the object and the heart border. If the object was outside of the heart, this feature was set to zero.

The segmented heart volume often contains calcifications in the aorta. To distinguish these from coronary calcium, the segmentation of the aorta was used to determine the distance of an object to the border of the aorta. This distance was used as a feature. For objects inside the aorta, the distance was multiplied by minus one.

This leads to a total of nine spatial features.

4. Appearance features

The first two appearance features were average and maximum intensity value inside the candidate. They are usually lower for calcifications than for bony structures and higher than for noise.

The next group of gray value features was based on image derivatives.[16] These are computed by convolving the data with derivative of the Gaussian kernel. The standard deviation or scale of this kernel is a free parameter. The use of multiple scales has been shown to be beneficial in many applications (the scale space paradigm). We used scales of 1, 2, 4, 8, and 16 voxels and Gaussian derivatives through second order denoted as $(L, L_x, L_y, L_z, L_{xx}, L_{xy}, L_{xz}, L_{yy}, L_{yz}, L_{zz})$. This set of derivatives encodes information about the appearance of objects and their surroundings. This combination of 10 derivatives and 5 scales gave 50 appearance features. They were determined at the maximum intensity point of each candidate object.

The exact computation of Gaussian derivatives for the large scans in our database is very expensive both in terms of computation time and required memory. Therefore, we used approximations of Gaussian derivatives that can be very efficiently computed at a limited set of positions. A detailed description how to calculate these approximate Gaussian derivatives is given in Appendix B.

C. Classification

The positive class in the data set was coronary calcium. Negative samples were all other candidate objects. These negative samples include noncoronary calcifications and other high density objects in the scan, such as bones and noise. Three classification approaches were applied. In the first system a direct classification between positives and negatives without feature selection was employed. The second scheme utilized also a direct classification, but included a feature selection. Finally, in the third scheme a two-stage classification with feature selection was used. First, all objects with a probability for being coronary calcification above a certain threshold are sent to the next stage. In the second stage, another classifier is used to perform complex separation between the remaining candidates.

The proper choice for a particular classifier is very important in pattern recognition problems. It is difficult to estimate which classifier will perform the best on a given data set. Thus, it is desirable to test the performance of several classifiers with the data at hand.[17,18] We have experimented with linear discriminant (LDC), quadratic discriminant (QDC), support vector machines (SVMs) and k-nearest neighbor (k-NN) classifiers. Particular design choices (number of fea-

tures to be selected, classifier, value for k) were made on the basis of pilot experiments on the training data.

As both k-NN and SVM classifiers are sensitive to feature scaling, prior to feature selection and classification, all features were scaled to zero mean and unit variance.

1. Direct classification without feature selection

As a reference system we performed direct classification with a k-NN classifier, without any feature selection. The optimal choice for the number of neighbors k in a (k-NN) classifier was determined to be one.

2. Direct classification with feature selection

With respect to the number of positive samples in the training set, the feature set contained many features. It is not likely that all features are equally useful in coronary calcification detection. Therefore, feature selection was employed to find a subset of features with good performance. Sequential floating forward feature selection (SFFS)[19] was used. This is a wrapper-based algorithm that tests the performance of a specific classifier using different feature sets. In this way, the set of features is optimized for the classifier at hand, in this case a k-NN classifier. SFFS has shown good performance on practical problems compared to other feature selection techniques.[19] The algorithm is based on a "plus 1, take away r" strategy. At every 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. Feature selection was performed on the training set with Euclidean metric and a leave-one-object-out methodology. Number of neighbors k was set to one. Classification performance was evaluated in terms of accuracy. Subsequently, classifications with LCD, QDC, and 1-NN classifiers were compared.

3. Two-stage classification with feature selection

A SVM classifier can be used to construct nonlinear decision boundaries by a proper choice of its kernel. We used a radial basis kernel (a Gaussian function), which is preferable to other kernels because it has only a single free parameter γ. In addition, the penalty parameter C also needs to be determined. The performance of an SVM classifier depends strongly on a proper choice of these parameters. Following Ref. 20 we used a fivefold cross-validation grid-search on the training data to determine the optimal values for C and γ. The drawback of this procedure is a long training time, and therefore the number of samples in the training set had to be reduced. We used a first classification stage with a k-NN classifier with feature selection to remove candidates that are negative with a high probability. Feature selection was set to select at most 20 features and k was set to 10. It was performed on the training set with Euclidean metric and a leave-one-object-out methodology. Objects that had a posterior probability for being negative bigger than 0.75 were discarded. All remaining objects were selected for further analysis. In the second stage, performance of a SVM classifier was tested.

The same experiment was repeated with a k-NN classifier employed in both stages. The settings for the first stage remained unchanged. In the second stage SFFS was set to select at most 15 features and a 1-NN classifier was employed. Again, feature selection was performed on the training set with Euclidean metric and a leave-one-object-out methodology. Note that feature selection was not employed with SVM. The necessary parameter optimization makes training times for feature selection prohibitively long, and SVMs have been shown to be able to achieve good results without feature selection.

D. Risk category determination

Calcium scores of coronary calcifications detected by the method were computed for each scan. Subjects were assigned to one of four risk categories for coronary artery disease. The categories were based on the Agatston score (AS) and were indicating a risk group:[21]

(1) low risk: $0 \leq AS \leq 10$
(2) intermediate risk: $10 < AS \leq 100$
(3) high risk: $100 < AS \leq 400$
(4) very high risk: $AS > 400$

The Agatston score of an object was computed as in Refs. 22 and 23, because this particular algorithm is used most often in clinical practice. The maximum intensity of an object in each transversal slice I_{max} was found. That value determined the weight factor w as follows:

$$\text{If } 130 \text{ HU} \leq I_{max} < 200 \text{ HU}, \quad \text{then } w = 1.$$

$$\text{If } 200 \text{ HU} \leq I_{max} < 300 \text{ HU}, \quad \text{then } w = 2.$$

$$\text{If } 300 \text{ HU} \leq I_{max} < 400 \text{ HU}, \quad \text{then } w = 3.$$

$$\text{If } 400 \text{ HU} \leq I_{max}, \quad \text{then } w = 4.$$

The Agatston score for a calcification is given by

$$AS_{obj} = \sum_{i=1}^{n} A_i * w_i, \tag{1}$$

where A_i is an area of the object in the slice i, and n is the number of slices in which the object is present.

Total Agatston score is the sum of Agatston scores of all calcifications in the scan:

$$AS = \sum_{obj=1}^{m} AS_{obj} \tag{2}$$

with m the number of calcifications in a scan.

IV. RESULTS

Originally, the training set contained 228 scans. For computation of spatial features, heart and aorta segmentations

TABLE II. Results of classification presented in number of objects. Columns indicate results of the (1) direct classification system without feature selection (no SFFS); (2) direct classification with feature selection (SFFS); (3) two-stage classification and SVM (SVM); and (4) two-stage classification with feature selection and *k*-NN (two-stage *k*-NN). Rows list number of true negatives (TN), false positives (FP), true positives (TP), and false negatives (FN).

	No SFFS	SFFS	SVM	Two-stage *k*-NN
TN	34 381	34 397	34 405	34 399
FP	29	13	5	11
TP	134	196	182	203
FN	141	79	93	72

were needed. There was 1 scan in which heart segmentation failed and 27 scans where either the heart or the aorta segmentation algorithm did not give completely satisfactory results. As this would result in some erroneous spatial features for certain objects, these scans were removed from the training set. This resulted in 200 training scans and 76 test scans.

After candidate objects were extracted, the training set contained 151 720 objects: 439 coronary calcifications, 1721 noncoronary calcifications, and 149 560 negatives. These negatives included noise, bone, and metal implants. The test set contained 35 198 candidates: 275 coronary calcifications, 335 noncoronary calcifications, and 34 588 negatives.

Discarding objects bigger than 2500 voxels did not remove any coronary calcifications from both the training and the test sets. However, in the training set 6 calcifications in the heart and 1293 negatives were discarded, and from the test set 513 negatives were removed.

Performance of LDC and QDC was substantially worse than that of the two nonlinear classifiers, SVM and *k*-NN, so only the results of the latter two will be reported.

A. Direct classification without feature selection

The experiment was performed using a *k*-NN classifier, where *k* was set to 1. On average the method made 2.2 errors per scan: 0.4 false positives and 1.8 false negatives.

Misclassified coronary calcifications were located in all coronary arteries. Among false positives there were five calcifications in the aorta and one calcified aortic valve. The remaining 23 false positives were noncalcifications. Results in terms of objects are listed in Table II.

Fifty-seven (75.0%) subjects were assigned to the correct risk category. These results are listed in Table III.

B. Direct classification with feature selection

The feature selection scheme selected ten features. Among them were four position features and six appearance features. The features were selected in the following order (cf. numbering in Table I): 9, 18, 8, 28, 27, 19, 22, 32, 4, 11. Note that neither shape nor size features were among the selected ones.

The *k*-NN classifier made on average 1.2 mistakes per scan: 0.2 false positive and 1 false negative object. Results per object are listed in Table II.

TABLE III. Risk categories of coronary artery disease assigned to the subjects by a direct classification with a 1-NN classifier without feature selection. Columns indicate the category assigned to a subject by the reference standard (ref), and rows indicate the category assigned by the automated method (sys).

	1 (ref)	2 (ref)	3 (ref)	4 (ref)
1 (sys)	33	7	1	0
2 (sys)	1	10	4	1
3 (sys)	0	0	4	5
4 (sys)	0	0	0	10

Among the 13 false positives, 11 were noise in the heart (five times noise in the left ventricle, close or on the border with the lung, twice noise in the ascending aorta, once noise in the right ventricle, once in the left atrium, and two times noise right next to the airway), one time it was an artifact in the heart caused by the heart movement (a high density objects next to the calcification in the right coronary artery), and once it was noise outside of the heart (stomach). All but one of the false positives were small in size. Note that all false positives were noncalcifications. False negatives were located in all coronary arteries and varied in size.

A total of 71 (93.4%) subjects were assigned to the correct risk category. The results of a category assignment are listed in Table IV. In all five cases where the category was not correctly assigned, risk was underestimated by one category. Incorrect category assignment was always caused by a big calcification in LAD.

C. Two-stage classification with feature selection

After the first classification stage 218 out of 275 (79.3%) coronary calcifications were correctly detected. Out of 34 410 negative candidates, 42 remained (0.1%).

1. SVM

The optimal parameters for the SVM were determined to be $C=16$ and $\gamma=0.0039$.

After both stages, the method made on average 1.3 errors: 0.1 false positive, and 1.2 false negative errors.

All false positives were noncalcifications. They consisted of four objects representing noise and one object represent-

TABLE IV. Risk categories of coronary artery disease assigned to the subjects by a direct classification with a *k*-NN classifier and with feature selection. Columns indicate the category assigned to a subject by the reference standard (ref), and rows indicate the category assigned by the automated system (sys).

	1 (ref)	2 (ref)	3 (ref)	4 (ref)
1 (sys)	34	1	0	0
2 (sys)	0	16	1	0
3 (sys)	0	0	8	3
4 (sys)	0	0	0	13

TABLE V. Risk categories of coronary artery disease assigned to the subjects by a two-stage classification with SVM. Columns indicate the category assigned to a subject by the reference standard (ref), and rows indicate the category assigned by the automated method (sys).

	1 (ref)	2 (ref)	3 (ref)	4 (ref)
1 (sys)	34	2	0	1
2 (sys)	0	15	4	0
3 (sys)	0	0	5	6
4 (sys)	0	0	0	9

ing bone. The noise was located twice in the left ventricle, once around the sternum and once next to the LAD. The bone was part of the sternum.

Sixty-three out of the 76 (82.9%) subjects were assigned the correct risk category. In the case of incorrect category assignment, the risk was always underestimated.

Out of the 13 scans where category was incorrectly determined, one subject was assigned more than one category off. Results are listed in Table V.

2. k-NN

The second stage experiment was repeated with a *k*-NN classifier and SFFS. Fourteen features, six spatial and eight appearance features, were selected in the following order (cf. numbering in Table I): 8, 58, 9, 7, 20, 48, 12, 5, 34, 18, 54, 10, 28, 14. Size and shape features were not selected.

On average, after both classification stages, 1.1 errors per scan were made: 0.1 false positive error and 1 false negative error. Results of both classification stages are listed in Table II.

False positives were four times noise around base of the heart, once small calcification in the ascending aorta, two times a piece of bone around the sternum, and four times noise in the heart (twice near the LAD, once next to RCA, once inside the left ventricle). False negatives were located in all coronary arteries.

A total of 71 out of 76 (93.4%) subjects were assigned the correct risk category. In all five scans where category was not correctly determined, it was underestimated by one. The results of the category assignments for the experiment with the *k*-NN classifier are listed in Table VI.

TABLE VI. Risk categories of coronary artery disease assigned to the subjects by a *k*-NN classification with feature selections in two stages. Columns indicate the category assigned to a subject by the reference standard (ref), and rows indicate the category assigned by the automated system (sys).

	1 (ref)	2 (ref)	3 (ref)	4 (ref)
1 (sys)	34	3	0	0
2 (sys)	0	14	1	0
3 (sys)	0	0	8	1
4 (sys)	0	0	0	15

V. DISCUSSION

The best results were obtained in a two-stage classification with *k*-NN classifier and feature selection. The method correctly detected 73.8% coronary calcifications at the expense of an average of 0.1 false positive objects per scan. Furthermore, subjects were assigned to the correct risk category with high accuracy: 93.4% of subjects were assigned correctly and no scan was more than one category off.

In all experiment setups, the best results were obtained with complex nonlinear classifiers *k*-NN and SVM. Their performance was always superior compared to LDC and QDC.

For the good performance of *k*-NN classifier, feature selection was essential. Although classification in two stages was slightly better in terms of objects, results of risk category assignment were comparable. In general the differences between direct and two-stage classification were small. The reason why two-stage classification with *k*-NN classifier and feature selection performed slightly better than one-stage classification might be the fact that in the first stage outlier positives and obvious negatives were discarded, and, subsequently, a new classifier with different features was trained for the remaining, more difficult data.

SVM can only be used in two stages, because determining the optimal parameters C and γ with the large number of samples in our training data would require vast amounts of computation time. Thus, first another classifier had to be employed to reduce number of samples to be classified by SVM.

Our experiments have shown that feature selection was important for good performance of the system. Spatial features derived from the segmentation of the heart and the aorta were selected. This indicates that both heart and aorta segmentation are important for the performance of the system. However, both algorithms occasionally did not give a satisfactory result, and the heart segmentation failed completely in one case. Therefore, it is important to investigate the possibility of improving the precision of the segmentation algorithm.

In direct classification with feature selection the most important features were those derived from the object's position relative to the heart. This is not surprising given that coronary calcifications have a typical position relative to the heart border. Coronary arteries branch off the ascending aorta and stretch towards and along the border of the heart. Appearance features were also among the selected features. They describe the texture of the object and its surrounding depending on the object's size and scale. Appearance features at scales 4, 8, and 16 pixels seemed more important than those of smaller scales. In two-stage classification again spatial and appearance features were selected. Again, features describing the position relative to the heart were among the first selected. Distance to the border of the aorta was also a selected feature. We would expect this feature to be important because it can differentiate calcifications in the wall of the aorta from coronary calcifications. Appearance features at scale of 8 pixels were the most often selected. It is, however, very

difficult to know in such a high dimensional space why a certain feature is more important than another. Both feature selection procedures have not selected size and shape features. Besides the fact that those are only three features in the set of 64 features, there might be other reasons why those were not important. Coronary calcifications should either have elongated shape as stretching along the vessel wall or be small spherelike objects. However, artifacts caused by the heart movement blur the calcifications and, as a consequence, they have any arbitrary shape. The size of the coronary calcifications is in a certain range of values, but the size of many other candidates is similar.

If the method would be used in an interactive setting where a human operator would have to correct the errors of the automated method, in the order of 1.1 corrections per scan would have to be made. We believe that in such a setting, the required evaluation time would be reduced compared to the currently used manual scoring systems. In this scenario, it might be useful to investigate if, instead of hard classification, a soft classification would be advantageous. That way it might be possible to present to the observer all objects about which the system is not sure.

The performance level of the automated system is still below that of human observers. Both intra- and interobserver agreement have been reported to be excellent,[12] which indicates that coronary calcium scoring is a relatively easy task for an experienced human operator. However, it has been recently reported in a large study that the average relative rescan difference for Agatston score is 20.1%.[24] This is a large difference that is caused by movement of the coronaries and by noise. If we translate this figure to our test set, this means that if our 76 subjects were rescanned, 4 of them can be expected to be assigned to a neighboring risk category. Our best system determined the correct risk category for 71 of the 76 subjects, and in the remaining 5 subjects category was underestimated by one. This performance is thus comparable to the variation due to rescanning.

The main aim of the method is to assign a subject automatically to the correct risk category for coronary artery disease. This means that the accuracy of the calcium score is more important than the sensitivity and specificity of the method in terms of classification of objects as the calcium score is more influenced by larger lesions than by small ones. The classification, however, was performed object based. It is clear from our results that the main source of error in risk category assignment is misclassification of a, typically single, large false negative object. Figure 2 shows examples of a large false negative and a large false positive object. In every training and test set there are inevitably some big outlier objects. Those examples are difficult for classification, because only very few examples are provided, with properties different from other samples of the same class. To improve the performance of our method, the system should be modified or extended to focus on correct classification of large candidate objects. Features specifically designed to accurately describe properties of large objects could be developed. For example, the current appearance features were calculated at the point of the maximum intensity of the candidate. This is a suitable choice for small and medium sized objects, but might not be the best choice for big objects. Additionally, in our data set large objects are vastly outnumbered by the small ones. Therefore, performance might be improved by a separate detection and classification system for large and small candidates. A much larger training database, especially with many more big calcifications, would be required to do this.

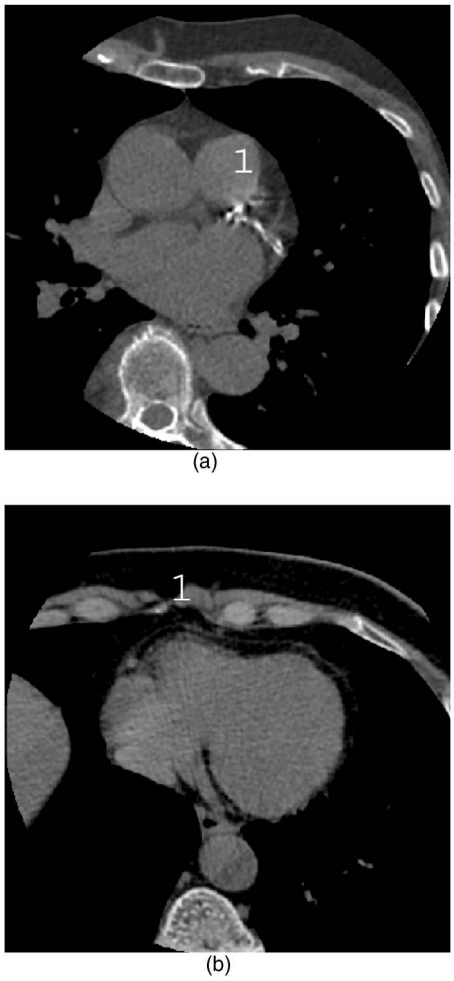

FIG. 2. (a) Misclassified calcium in left coronary artery. Calcification is substantially blurred due to the heart motion. (b) Misclassified bone around the sternum, which was the largest false positive.

VI. CONCLUSION

The first fully automated method for automated risk assessment of coronary artery disease based on CT calcium scoring has been presented. The method has been tested on noncontrast, ECG-gated CT scans of the heart.

A dedicated set of features was designed to describe coronary calcifications. Extensive classification experiments were carried out with three classification strategies. The best result was obtained in two-stage classification strategy with k-NN classifier and feature selection. Of the coronary calcifications, 73.8% were correctly detected at the expense of 0.1

false positive objects per scan. In 93.4% of the cases, the category was correctly assigned.

APPENDIX A: SEGMENTATION OF THE HEART AND THE AORTA

We developed a single algorithm that was used for both the heart and the aorta segmentation tasks, albeit with slight differences. The algorithm can be divided in two stages.

1. Detection of the starting point

First a starting point for the heart or the aorta segmentation was detected. The aortic arch was not included in the volume of the scans, so the ascending and the descending aorta had to be detected as two separate structures. Because the heart and the ascending and descending aorta have a circular shape in axial slices, a two-dimensional (2D) Hough transform was used to find a circle that fits the target structure. The search for the circle indicating the heart or the aorta started in the top axial slice, which, according to the protocol, should contain the top of the heart. Because gray values inside the target structures are those of blood, muscle, and fat, the Hough transform was performed in images double thresholded at values $t1$ and $t2$.

Circles were accepted if their score, the number of edge points voting for the circle, exceeded s_{min}. Moreover, because the heart and aorta are not perfectly circular, several circles might be identifying the same structure. Therefore, overlapping circles were grouped, and only the best one in the group was considered for further detection.

2. Organ edge delineation

The second part of the algorithm used the detected circle to find the border of the structures. For this purpose, a slice based ray shooting algorithm constrained by a statistical shape model was employed. First, the original image was processed to obtain a more detailed heart edge map than the one used for circle detection. A gradient magnitude image computed with Gaussian derivatives at a scale of 2 pixels was used.

Next, ray shooting was performed in the same slice where the circle was detected. The center of the detected circle was used as the origin. Rays were cast at angular distance of α deg. The search along the ray was limited to range of pixels between $[r_1, r_2]$, where these parameters were expressed in multiples of r, the distance of the circle pixel from the origin. Along the ray, the pixel with the maximum value in the gradient magnitude image was considered to be the edge point.

A 2D statistical shape model of the heart was fit. This model was constructed from previously manually segmented n_{ss} organ borders from n_{ps} different subjects. The model was a point distribution model[25] in which the positions are concatenated in a vector. A training set of vectors was aligned by translation, scaling, and rotation, and principal component analysis was applied, to explain 99% of the model variance. This resulted in 24 modes of deformation for the heart and 19 for the aorta. Fitting the model to a set of points is performed by computing a translation, scaling, and rotation parameter, projection on the principal components, and truncating the projections to at most three standard deviations.[25]

The ray shooting procedure and shape model fitting were iteratively applied n_{it} number of times.

For all further slices, the same process was repeated, but, as starting point, the edge found in the previous slice was taken. The origin for ray shooting was determined as the center of mass of the segmented area.

The complete algorithm was performed for each of the detected circles. The circle that yielded a segmentation in most of the transversal slices was considered to be the heart edge. In the case of the aorta, ascending and descending, two circles that yielded a segmentation in most of the slices were taken as its segmentation. The parameter settings used in this algorithm are listed in Table VII.

TABLE VII. Parameter settings for the heart and aorta extraction algorithms.

Parameters	Heart extraction	Aorta extraction
t_1 [HU]	−300	−300
t_2 [HU]	−150	0
s_{min}	50	100
α (deg)	4	4
r_1 (mm)	0.86	$0.344r$
r_2 (mm)	$0.215r$	$0.516r$
n_{ss}	91	230
n_{ps}	11	3
n_{it}	3	2
n_s	⋯	6
a_b (%)	⋯	10
b_v (HU)	⋯	130

APPENDIX B: BOX DERIVATIVES

Spatial derivatives are often computed by convolving an image with a Gaussian kernel. For the efficiency reasons we approximated the Gaussian kernels with boxes. In Refs. 26 and 27 this approach was used for filtering 2D images, and here we present a 3D implementation. Figure 3 illustrates a box approximation of a Gaussian kernel for zero, first, and second order (including cross-derivatives) kernels. This type of approximation can be extended to any order of derivative, with higher order derivatives requiring more boxes. It is also possible to refine the approximation by using more boxes, e.g., approximate the Gaussian with three boxes. The height, position, and extent of the boxes depend on the order of derivative and the scale σ of the Gaussian to be approximated. They were determined by minimizing the squared sum of the differences between the true Gaussian kernel and

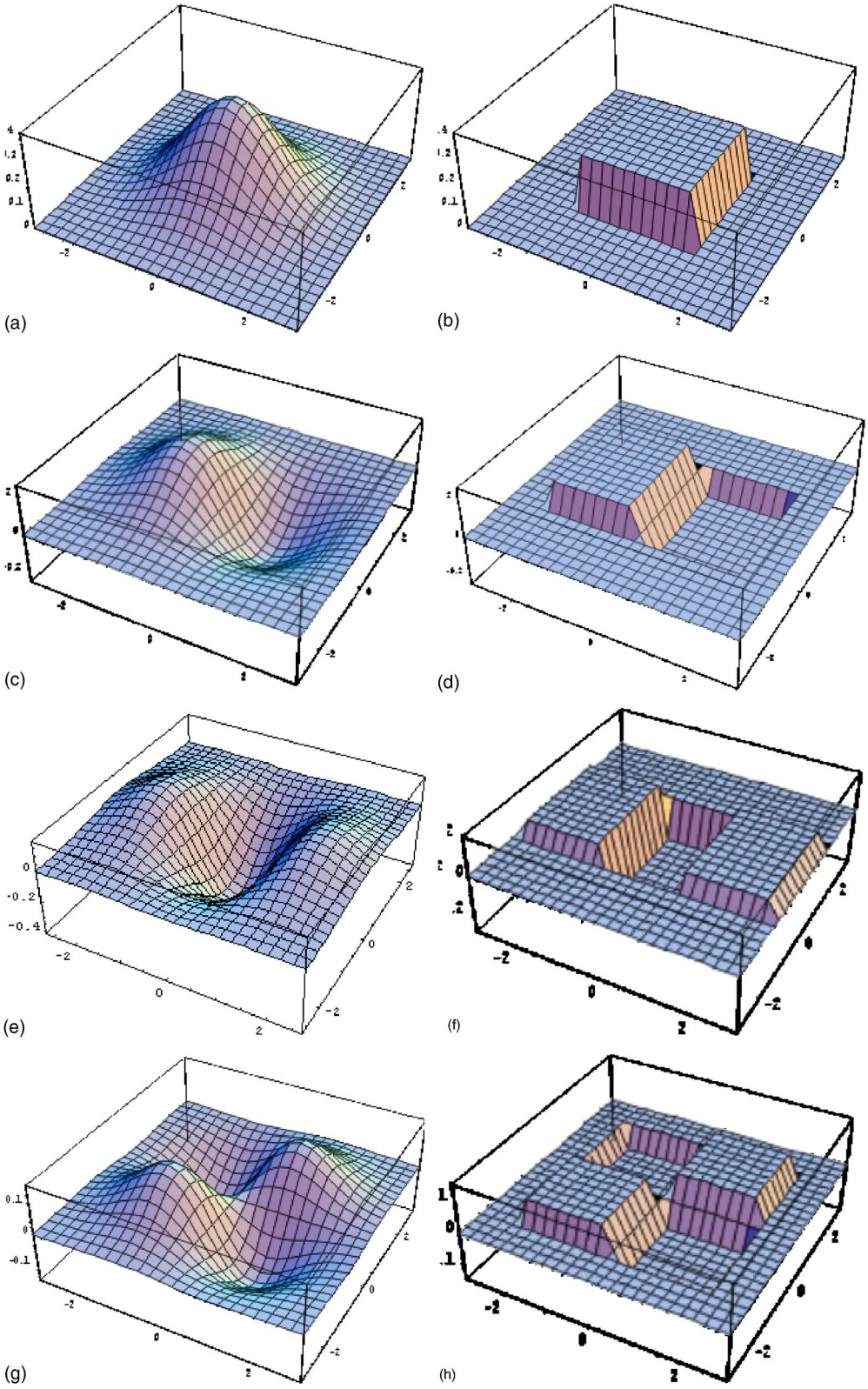

FIG. 3. (a) A Gaussian function and (b) a box approximating it. (c) An example of the first order derivative and (d) two boxes approximating the function. (e)–(h) Examples for the second order derivatives. The height of the box is equal to the weight factor, and the size of the box depends on the approximated Gaussian scale.

its box approximation. Scans in our data set were not isotropic and therefore a correction for the box size in z-direction was applied.

Sums of the intensity values in the boxes can be calculated using an integral image, as in Ref. 26. In a 3D case the integral image was computed as follows:

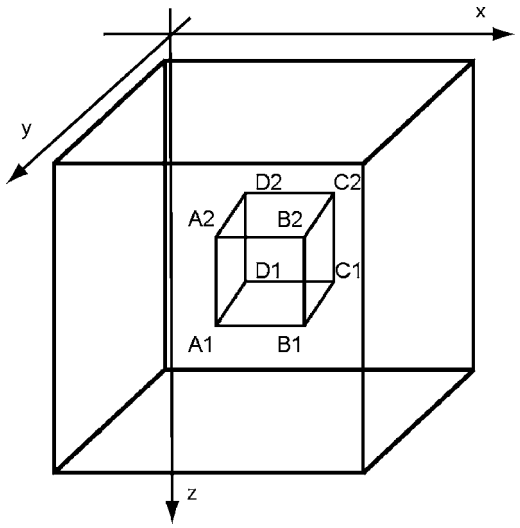

FIG. 4. Sum of the values in the box can be computed according to formula using integral volume.

$$ii(a,b,c)=\sum_{x\le a}\sum_{y\le b}\sum_{z\le c} im(x,y,z), \tag{B1}$$

where *im* is the original volume and *ii* is the integral volume. This was calculated according to

$$t(x,y,z)=t(x-1,y,z)+im(x,y,z), \tag{B2}$$

$$ii(x,y,z)=t(x,y,z)+ii(x,y-1,z)+ii(x,y,z-1)-ii(x,y-1,z-1), \tag{B3}$$

where *im* is the original volume; *t* is a temporary image containing cumulative sums with $t(-1,y,z)=0, t(x,-1,z)=0$, and $t(x,y,-1)=0$; and *ii* is the integral volume. Computing the sum of voxels in any box using the integral volume requires just eight addition operations:

$$s=ii(B1)-ii(A1)-ii(C1)+ii(D1)-ii(B2)+ii(A2)+ii(C2)-ii(D2), \tag{B4}$$

where *s* is the sum of the voxel values in the box and $A1$, $A2$, $B1$, $B2$, $C1$, $C2$, $D1$, $D2$ are its vertices, as illustrated in Fig. 4. Thus the total cost of computing a zeroth order Gaussian approximation L thus requires eight operations [single box, Fig. 3(a) and 3(b)]; computing a first order derivative approximation (L_x,L_y,L_z) requires 16 operations [two boxes, Fig. 3(c) and 3(d)]; computing a second order derivative approximation (L_{xx},L_{yy},L_{zz}) requires 24 operations (three boxes, Fig. 3(e) and 3(f)); and, finally, a mixed order derivative approximation (L_{xy},L_{xz},L_{yz}) requires 32 operations (four boxes, Fig. 3(g) and 3(h)). These numbers of operations are independent of the scale σ of the Gaussian, require only a single data volume in memory (the integral image), and can be computed at any limited set of points without having to process a neighborhood (unlike, for example, a separable implementation of Gaussian filtering or recursive filtering).

[1]G. T. Kondos, J. A. Hoff, A. Sevrukov, M. L. Daviglus, D. B. Garside, S. S. Devries, E. V. Chomka, and K. Liu, "Electron beam tomography coronary artery calcium and cardiac events, a 37-month follow-up of 5635 initially asymptomatic low- to intermediate-risk adults," Circulation **107**(20), 2571–2576 (2003).
[2]M. J. LaMonte, S. J. FitzGerald, T. S. Church, C. E. Barlow, N. B. Radford, B. D. Levine, J. J. Pippin, L. W. Gibbons, S. N. Blair, and M. Z. Nichaman, "Coronary artery calcium score and coronary heart disease events in a large cohort of asymptomatic men and women," Am. J. Epidemiol. **162**(5), 421–429 (2005).
[3]P. Greenland, L. LaBree, S. P. Azen, T. M. Doherty, and R. C. Detrano, "Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals," J. Am. Med. Assoc. **291**(2), 210–215 (2004).
[4]J. Shemesh, R. Evron, N. Koren-Morag, S. Apter, J. Rozenman, D. Shaham, Y. Itzchak, and M. Motro, "Coronary artery calcium measurement with multi-detector row CT and low radiation dose: Comparison between 55 and 165 mAs," Radiology **236**(3), 810–814 (2005).
[5]A. S. Agatston, W. R. Janowitz, G. Kaplan, J. Gasso, F. Hildner, and M. Viamonte, Jr., "Ultrafast computed tomography-detected coronary calcium reflects the angiography extent of coronary arterial atherosclerosis," Am. J. Cardiol. **74**(12), 1272–1274 (1994).
[6]T. Q. Callister, B. Cooil, S. P. Raya, N. J. Lippolis, D. J. Russo, and P. Raggi, "Coronary artery disease: Improved reproducibility of calcium scoring with an electron-beam CT volumetric method," Radiology **208**(3), 807–814 (1998).
[7]Y. Arad, K. J. Goodman, M. Roth, D. Newstein, and A. D. Guerci, "Coronary calcification, coronary disease risk factors, C-reactive protein, and atherosclerotic cardiovascular disease events: The St. Francis heart study," J. Am. Coll. Cardiol. **46**(1), 158–165 (2005).
[8]Z. A. Fayad, V. Fuster, K. Nikolaou, and C. Becker, "Computed tomography and magnetic resonance imaging for noninvasive coronary angiography and plaque imaging: Current and potential future concepts," Circulation **106**, 2026–2034 (2002).
[9]G. S. Mintz, J. J. Popma, A. D. Pichard, K. M. Kent, L. F. Satler, Y. C. Chuang, C. J. Ditrano, and M. B. Leon, "Patterns of calcification in coronary artery disease: A statistical analysis of intravascular ultrasound and coronary angiography in 1155 lesions," Circulation **91**(7), 1959–1965 (1995).
[10]T. Xu, J. Ducote, J. Wong, and S. Molloi, "Feasibility of real time dual-energy imaging based on a flat panel detector for coronary artery calcium quantification," Med. Phys. **33** 1612–1622 (2006).
[11]S. Yaghoubi, W. Tang, S. Wang, J. Reed, J. Hsiai, R. Detrano, and B. Brundage, "Offline assessment of atherosclerotic coronary calcium from electron beam tomograms," Am. J. Card. Imaging **9**(4), 231–236 (1995).
[12]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).
[13]I. Isgum, B. van Ginneken, and M. Prokop, "A pattern recognition approach to automated coronary calcium scoring, in 17th International Conference on Pattern Recognition (ICPR'04), Vol. 3, 2004, pp. 746–749.
[14]I. Isgum, B. van Ginneken, A. Rutten, and M. Prokop, "Automated coronary calcification detection and scoring, in 4th International Symposium on Image and Signal Processing and Analysis, edited by S. Loncaric, H. Babic, and M. Bellanger, Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia, 2005, pp. 127–132.
[15]M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed. (PWS, Pacific Grove, CA, 1999).
[16]L. M. J. Florack, *Image Structure* (Kluwer Academic, Dordrecht, The Netherlands, 1997).
[17]A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," IEEE Trans. Pattern Anal. Mach. Intell. **22**(1), 4–37 (2000).
[18]N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods* (Cambridge University Press, Cambridge, 2000).
[19]A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," IEEE Trans. Pattern Anal. Mach. Intell. **19**(2), 153–158 (1997).
[20]C.-C. Chang and C.-J. Lin, LIBSVM: a library for support vector ma

chines, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm(2001).
[21]J. A. Rumberger, B. H. Brundage, D. J. Rader, and G. Kondos, "Electron beam computed tomographic coronary calcium scanning: A review and guidelines for use in asymptomatic persons," Mayo Clin. Proc. **74**, 243–252 (1999).
[22]B. Ohnesorge, T. Flohr, R. Fischbach, A. F. Kopp, A. Knez, S. Schroder, U. J. Schopf, 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).
[23]A. Agatston, W. Janowitz, F. Hildner, N. Zusmer, J. Viamonte, and M. R. Detrano "Quantification of coronary artery calcium using ultrafast computed tomography," J. Am. Coll. Cardiol. **15**(4), 827–832 (1990).
[24]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).
[25]T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham, "Active shape models—Their training and application," Comput. Vis. Image Underst. **61**(1), 38–59 (1995).
[26]P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Conference on Computer Vision and Pattern Recognition (IEEE Computer Society, Los Alamitos, CA, 2001), Vol. 01, pp. 511–518.
[27]P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," Int. J. Comput. Vis. **63**, 153–161 (2005).