ORIGINAL ARTICLE

Fully automatic model-based calcium segmentation and scoring in coronary CT angiography

Dov Eilot · Roman Goldenberg

Abstract

Purpose The paper presents new methods for automatic coronary calcium detection, segmentation and scoring in coronary CT angiography (cCTA) studies.
Methods Calcium detection and segmentation are performed by modeling image intensity profiles of coronary arteries. The scoring algorithm is based on a simulated unenhanced calcium score (CS) CT image, constructed by virtually removing the contrast media from cCTA. The methods are implemented as part of a fully automatic system for CS assessment from cCTA.
Results The system was tested in two independent clinical trials on 263 studies and demonstrated 0.95/0.91 correlation between the CS computed from cCTA and the standard Agatston score derived from unenhanced CS CT. The mean absolute percent difference (MAPD) of 36/39 % between the two scores lies within the error range of the standard CS CT (15–65 %).
Conclusions High diagnostic performance, combined with the benefits of the fully automatic solution, suggests that the proposed technique can be used to eliminate the need in a separate CS CT scan as part of the cCTA examination, thus reducing the radiation exposure and simplifying the procedure.

Keywords Calcium scoring · Computed tomography · Coronary CT angiography

D. Eilot · R. Goldenberg (✉)
Rcadia Medical Imaging, 157 Yafo Str., 35251 Haifa, Israel
e-mail: romang@rcadia.com

D. Eilot
e-mail: dov.eilot@gmail.com

Introduction

Coronary heart disease (CHD) is the narrowing or blockage of coronary arteries caused by accumulation of plaque on artery walls (coronary atherosclerosis). There are two noninvasive imaging tests clinically used today for the diagnosis of coronary atherosclerosis: calcium scoring (CS) and coronary CT angiography (cCTA).

Calcium scoring (CS) is used to estimate the amount of calcium in coronary arteries. Today, the CS is computed from unenhanced CT scan of the heart. The most commonly used technique for calcium quantification in CS CT scan is the Agatston score [1]. The score was initially designed for the electron beam computed tomography (EBCT) and defined for a study reconstructed according to the EBCT image specifications—non-overlapping 3-mm-thick slices.

The Agatston scoring is done in two steps: calcium segmentation and scoring. Calcium segmentation is performed by binarization of the original CS CT image using a fixed threshold of 130 HU and (manually/semi-automatically) choosing voxels that belong to coronary arteries. The score is then computed as the weighted sum of the areas (in mm^2) of 2D connected calcium components in axial slices. The weight assigned to each connected component is computed by the four-level step function (see Fig. 1) of the maximal image pixel intensity in that component.

Coronary CT angiography (cCTA) is a contrast-enhanced CT scan of the heart used to visualize the internal structure of coronary arteries. It is a common practice today to perform a CS study for patients referred to cCTA, prior to the cCTA exam. Until recently, high calcium score was a contraindication for cCTA, as CT scanners were unable to perform a diagnostic quality scan for heavily calcified coronaries. For the new generation of CT scanners that cope well with heavy calcification, such CS-based prediction of
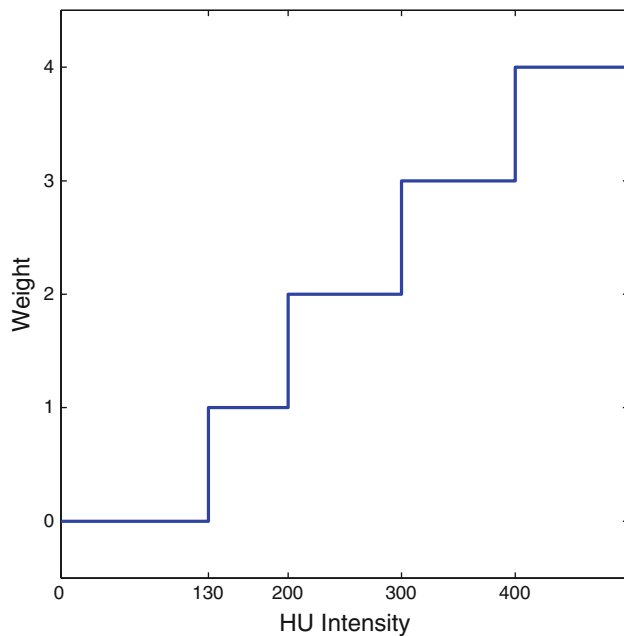
Fig. 1 Agatston weight function

expected diagnostic benefit of cCTA for a given patient is no longer required. Moreover, with the new dose-reduction techniques, the cCTA radiation dose was reduced from 10–15 to <1 mSv. Therefore, performing a CS scan, which adds another 0.8–2.5 mSv, prior to cCTA, effectively doubles the overall radiation dose [2].

Yet, since CS is commonly used by cardiologists for risk stratification, it is important to provide CS assessment for patients undergoing cCTA. This is especially true for patients in whom obstructive disease is ruled out by cCTA and calcium score is used to guide the aggressiveness of treatment for CHD prevention.

It was shown that calcium scoring can be accomplished directly from cCTA, without a separate CS scan. The feasibility of this approach has been demonstrated using manual [3] and semi-automatic methods.

Calcium segmentation in cCTA: the problem

Most reported semi-automatic techniques for calcium detection and segmentation in cCTA use thresholding—either fixed [2,4–6], or adaptive [7]. A learning-based approach for coronary calcium detection in cCTA (without segmentation and scoring) was also reported [8,9].

Using a fixed intensity threshold, as reported in [2,4,5], is problematic for the following reason: while biological tissues (blood, muscle, fat, etc.) have fixed and known CT attenuation levels, contrast material intensity depends on the used injection protocol and varies significantly from one study to another, from below 250 HU to above 600 HU. As calcium and contrast material intensities overlap [10], setting the threshold very high results in under-segmentation of calcium as reported in [4]. Setting the threshold too low results in calcium over-estimation (contrast is interpreted as calcium).

Bischoff et al. [7] suggested to use an adaptive, per-study threshold set to 150 % of the mean intravascular image intensity. This method is clearly superior to the fixed threshold approach, but still does not address another phenomenon often observed in cCTA studies: due to the partial volume effect, the maximal intensity of a small, eccentric, calcified lesion in cCTA could be lower than the maximal intensity of the contrast material in the artery cross section (see Fig. 2). Calcium pixels, that are closer to the vessel boundary, are influenced more by the lower intensity tissues outside the vessel and adjacent non-calcified plaque, than by contrast material. The lesion is still clearly visible, but cannot be segmented by thresholding [see the intensity profile in Fig. 2 (right)].

Calcium scoring in cCTA: the problem

As for the quantification of the calcium detected in cCTA, most reported methods performed the standard Agatston scoring on segmented calcium [3–5]. Bischoff et al. [7] amended the Agatston score by a calibration procedure to compensate for score underestimation in cCTA. Otton et al. [2] computed calcium score by multiplying the calcium volume by a factor empirically derived from a training set of studies.

Since the slice thickness in cCTA (0.60–0.75 mm) is much smaller than in CS CT (3 mm), the partial volume effect is less expressed, and calcium intensity in cCTA is significantly higher than in CS. The effect is amplified even more because
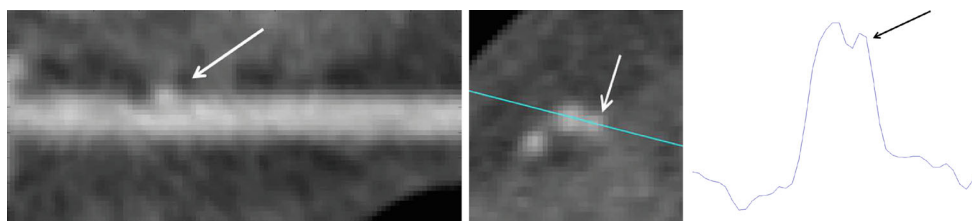

Fig. 2 Small, eccentric CA lesion: (*left*) longitudinal view; (*center*) axial view; (*right*) intensity profile along the cyan line in axial view. CA peak intensity is lower than the contrast peak intensity
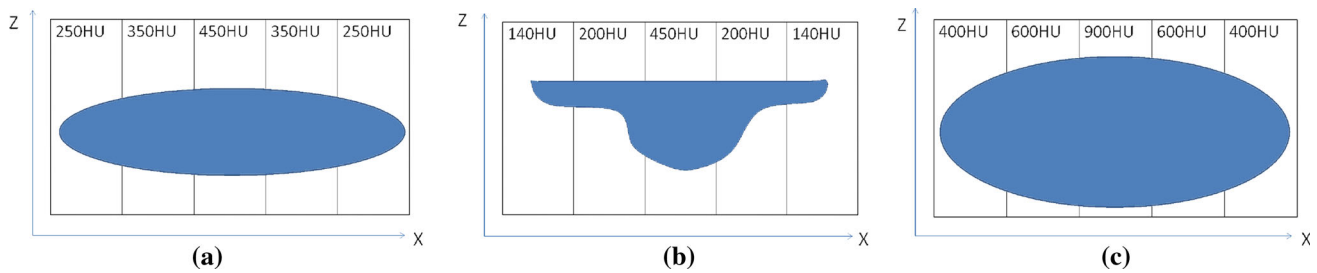
Fig. 3 Three CA lesions assigned the same Agatston score. All three occupy 5 pixels. Numbers at the *top* are image intensity values reflecting the amount of calcium in each pixel

calcium voxels in cCTA are averaged with high-intensity contrast material. As a result, the Agatston intensity-based weight factor produces different scores for cCTA studies. Moreover, since calcium intensity in cCTA is rarely below 400 HU, the Agatston weighing function (Fig. 1) is constantly in saturation, thus canceling the weighing effect.

Adapting the Agatston weight function to calcium intensity levels of cCTA (e.g., by moving the step positions) would not fully solve the problem. The Agatston score only weakly correlates with the real physical or geometric properties of calcium lesions. For example, the three lesions depicted in Fig. 3 receive the same score, despite obvious differences in size. Higher resolution cCTA that is more sensitive to differences in lesions geometry is likely to yield different scores for the three lesions. This observation suggests that the ability of score estimation methods based on regression analysis from quantities measured directly in cCTA, as reported in [2,7], to match the original Agatston score is inherently limited.

In what follows, we present new methods for calcium detection, segmentation and scoring in cCTA studies. The proposed calcium detection and segmentation method aims to overcome the intrinsic limitations of the threshold-based techniques and addresses the problems of varying contrast intensity in coronary arteries and low visibility of small calcified lesions in cCTA studies. The suggested scoring method is designed to deal with the nonlinear and non-geometric nature of the Agatston score to better match the standard score. The proposed methods are implemented as part of a fully automatic system for calcium score assessment in cCTA.

Methods

Similar to the standard CS procedure, we perform the automatic calcium scoring from cCTA in three steps:

– Coronary artery tree reconstruction,
– Detection and segmentation of calcified lesions in coronary arteries,
– Scoring.

Training data sets

Throughout the development, we used the following data sets for system training:

– T1—45 corresponding cCTA and CS studies collected from Carmel Medical Center, Haifa, Israel (25 patients) and Medical University of South Carolina, Charleston, SC (20 patients)—see Table 1 for details of used equipment and acquisition protocols. 100 corresponding calcium lesions were marked manually in cCTA and CS studies.
– T2—2,300 cCTA studies collected from 40 hospitals around the world. The set includes studies acquired on almost every existing type of CT scanners and using a wide variety of patient preparation, contrast injection, image acquisition and reconstruction protocols. These studies were used as a training set for our coronary stenosis detection system reported elsewhere [11]. 30 healthy (no visible coronary artery disease) coronary arteries of various sizes (diameters) were manually selected from the set.

None of the training studies were used for system testing and evaluation described in "Results" section.

Coronary artery tree reconstruction

For automatic coronary artery tree reconstruction and vessel segmentation, we use algorithms developed for the system for automatic stenosis detection in cCTA, previously reported in [11].

The automatic coronary tree reconstruction process includes segmentation of major anatomical structures (lungs, mediastinum, ascending aorta), localization of left and right coronary tree ostia, tracking tubular structures using DFS-based propagation in vessel enhanced image, vessel labeling, tree filtering by branch pruning, and centerline extraction. The resulting coronary artery tree graph is split into disjoint segments, and for each segment, vessel boundary is detected using tubular active surface model.

Table 1 Patient statistics and scanning parameters for CMC and MUSC studies

	CMC	MUSC
# patients	136	127
Age	59 ± 11	58 ± 11
Male/female	82/54	77/50
Scanner	256-slice, Brilliance iCT, Philips	128-slice, dual-source, Somatom definition flash, Siemens
CS CT: slice thickness/increment	2.5/2.5 mm	3/3 mm
CS CT: software	HeartBeat CS, Philips	Calcium scoring module of Aquarius iNtuition, TeraRecon
cCTA: contrast material	60–90 ml of Ultravist 370 mgI/ml (Bayer) at 6 ml/s	60–90 ml of Ultravist 370 mgI/ml (Bayer) at 6 ml/s
cCTA: acquisition	Step and shoot	Retrospective ECG-gating; prospective ECG triggering; prospectively ECG-triggered high-pitch spiral
cCTA: kV, mAs	120 kV, 300 mAs	100–120 kV, 320 mAs
cCTA: slice thickness/increment	0.67/0.33 mm	0.75/0.4 m
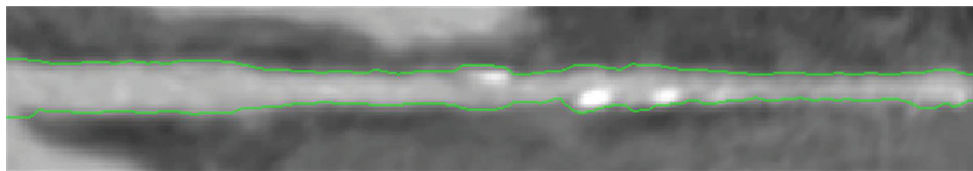
Fig. 4 Coronary artery straightened curved planar reformation (CPR) and external vessel boundary (*green*)

The calcium segmentation method presented below relies on an accurate artery centerline as input. To improve the centerline tracking accuracy, we use an iterative approach: after the vessel boundary is reconstructed, the centerline is adjusted to go through the center of vessel cross sections as defined by the boundary, and then the vessel is re-sampled along the new center line. Several iterations of boundary segmentation and center-line adjustment are performed till convergence.

Detailed descriptions of the used coronary tree reconstruction and vessel segmentation methods are available in [12] and [13], respectively. For an extensive review of other blood vessel extraction techniques, we refer the reader to [14]. A comprehensive review of various vessel segmentation methods can be found in [15].

The output of the tree reconstruction step is the coronary artery tree, represented by its centerlines and a list of disjoint coronary segments with detected external boundaries. Every coronary segment is re-sampled along its centerline using the straightened curved planar reformation (CPR) [16] (see Fig. 4). The external boundary is provided as a surface (or as a contour in every 2D vessel cross section) in the straightened CPR coordinate system (see Fig. 4).

In addition, the mean (μ_c) and the standard deviation (σ_c) of contrast material intensity levels inside the aorta are computed by the coronary tree reconstruction part of the system.

Calcium detection and segmentation

To deal with the problem of varying contrast intensity and low visibility of small calcified lesions (see "Calcium segmentation in cCTA: the problem" section), we propose a model-based segmentation approach.[1] The algorithm is based on fitting an adaptive intensity distribution model to vessel intensity profile for every cross section along the vessel. The model describes the intensity profile of the given vessel as it would look if it was healthy (with no calcium). High-intensity image deviations from the model prediction are interpreted as calcium lesions.

2D vessel modeling

Let $V = \{p_i | i = 1, \ldots, n_V\}$ be the set of pixels inside the vessel boundary in a given cross section of coronary artery, where n_V is the number of such pixels. Let us denote by $I(p_i)$ the intensity value of pixel p_i. We are looking for a model function $\hat{I} : V \rightarrow \mathbb{R}$ as:

$$\hat{I} = \arg\min_{I^*} \sum_{p_i \in V} w\left(I(p_i)\right) \left(I(p_i) - I^*(p_i)\right)^2, \qquad (1)$$

---

[1] The described method for calcium segmentation is patent pending.

Fig. 5 Intensity-based weight function w

Fig. 6 Paraboloid curvature constraint function $g(s) : 0 > a > g(s)$
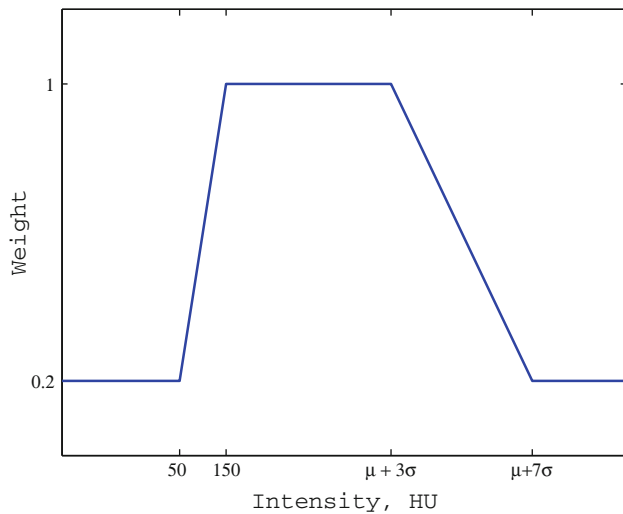
where I^* runs over all allowed model configurations and w is a weighing function preferring vessel pixels that are more likely to represent contrast material. The weighing is used to fit the model to the "healthy" part of the vessel (i.e., to the contrast filled lumen), while ignoring plaque areas.

Figure 5 shows the intensity-based weight function w used in Eq. 1, which reflects the following observations:

– Pixels above 150 HU and below $\mu_c + 3\sigma_c$ are likely to be contrast material,
– Pixels below 50 HU are likely to be non-calcified plaque, thrombus or other tissue outside vessel boundaries,
– Pixels above $\mu_c + 7\sigma_c$ are very unlikely to be contrast material.

The 150 and 50 HU thresholds were chosen by analyzing 2,300 cCTA studies from the T2 training set. The coronary tree reconstruction algorithm automatically segmented the ascending aorta and computed the contrast media mean intensity and standard deviation inside the aorta. 150 HU is the minimal mean intensity over the set. The average contrast intensity deviation in aorta was $\bar{\sigma}_{aorta} = 30$ HU, and the second threshold was chosen as $150 - 3\bar{\sigma}_{aorta} \approx 50$ HU. The weight function w is built by assigning a high weight to the first interval, low weight to the other two, and linearly connecting domains between the intervals (see Fig. 5).

To describe the vessel cross-section intensity profile, we use a parabolic model centered at the vessel centerline: $I^*(p_i) = ad^2(p_i) + b$, where $d(p_i)$ is the distance from pixel p_i to the vessel centerline (see [17,18] for related effort). Obviously, $a < 0$, since contrast intensity is higher than that of surrounding tissues. Also, since the central part of a large vessel is barely influenced by the partial volume effect, it is expected that the paraboloid curvature, defined by $|a|$, is larger for narrow vessels and smaller for wide ones.
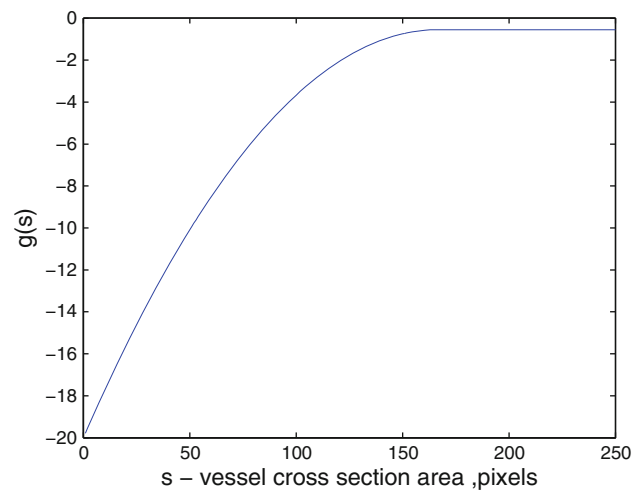
The presence of calcium, especially close to the centerline, despite our attempts to decrease its influence by assigning lower weights, can result in a wrong estimation of model parameters. Specifically, the optimizer tends to overestimate the factor $|a|$ (high curvature). To deal with this problem we impose a constraint on a in the form of $0 > a > g(s)$, where g is a negative, non-decreasing function of vessel cross-section area $s = |V|$. The function g (see Fig. 6) is found empirically, by analyzing parabolic models fitted to 30 healthy (no plaque) vessels of various sizes selected from the T2 training set and fitting a quadratic lower bound to the parabola parameter a. The optimal model parameters $(\hat{a}, \hat{b})$ are recovered by solving the constrained linear weighted least squares problem

$$(\hat{a}, \hat{b}) = \arg\min_{(a,b)} \sum_{p_i \in V} w\left(I(p_i)\right)\left(I(p_i) - ad^2(p_i) - b\right)^2 \tag{2}$$

$$\text{s.t. } 0 > a > g(|V|), \ b > 0$$

using a subspace trust-region technique based on the interior-reflective Newton method [19] implemented in *lsqlin* MATLAB function.

3D vessel modeling

The parabolic model is fitted to each vessel cross section independently to yield vectors $\hat{\mathbf{a}} = \{\hat{a}_j\}$ and $\hat{\mathbf{b}} = \{\hat{b}_j\}$, $j = 1, \ldots, L$, of paraboloid parameters, where L is the number of cross sections in the analyzed artery. In order to reduce the local model fitting error we smooth vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ using a bilateral-like filter defined as:

$$\tilde{a}_j = \frac{1}{W_j} \sum_{k \in N} D(j-k) A\left(\frac{s_j}{s_k}\right) C(k)\hat{a}_k, \tag{3}$$

where

- D is the spacial distance component,
- A is the vessel size difference component,
- s_j, s_k are vessel cross-section areas at cross sections j and k,
- C is the model fitting confidence level,
- N is the filter window around the cross section j,
- $W_j = \sum_{k \in N} D(j - k) A(s_j/s_k) C(k)$ is the normalization parameter.

The spacial distance component is implemented by a standard Gaussian filter

$$D(j - k) = e^{-\frac{(j-k)^2}{\sigma_d^2}},$$

with $\sigma_d = |N|/4$.

The A component, inspired by the bilateral filter approach [20], is used to reduce the influence of vessel cross sections of significantly different size than that of the cross section j:

$$A\left(\frac{s_j}{s_k}\right) = \left[\min\left(\frac{s_j}{s_k}, \frac{s_k}{s_j}\right)\right]^3.$$

The A filter component is especially important to discriminate between vessel part with normal tapering and segments with local positive or negative remodeling, and between different vessel branches when a bifurcation occurs inside the filter window N.

The confidence level C is used to prefer models based on more reliable data, i.e., pixels that are more likely to represent lumen filled with contrast material. The confidence level is computed as the ratio between the total pixels weight (see Fig. 5) and the cross-section area:

$$C(k) = \frac{\sum_{p_i \in V_k} w(p_i)}{|V_k|},$$

where $V_k = \{p_i\}$ is the set of pixels inside the vessel boundary in the cross section k.

The neighborhood N includes cross sections in a 6 mm long window around the cross section j. The resulting smoothed paraboloid parameters $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ are used to model cross-section intensity profile as $\tilde{I}(p_i) = \tilde{a}d^2(p_i) + \tilde{b}$ (see Fig. 7).

Special treatment is required for vessel bifurcations as the paraboloid model does not describe them correctly. Locations and directions of bifurcations are detected from branching points of the reconstructed coronary artery tree [11]. By analyzing the boundaries of bifurcating branches, we determine vessel segments and lumen sectors affected by bifurcations. In those segments, the parabolic intensity model is modified by assigning the parabolic peak value to all pixels in the lumen sector affected by the bifurcation (see Fig. 8).

Binarization

Given the vessel intensity model $\tilde{I}$ computed at the previous step, the calcium binary map M is generated by thresholding the difference between the actual image intensities and the values predicted by the model. The threshold is proportional

Fig. 7 Vessel cross-section intensity profile $I(p_i)$ (*left*). Paraboloid intensity model $\tilde{I}(p_i)$ (*center*). I and $\tilde{I}$ overlapped (*right*)
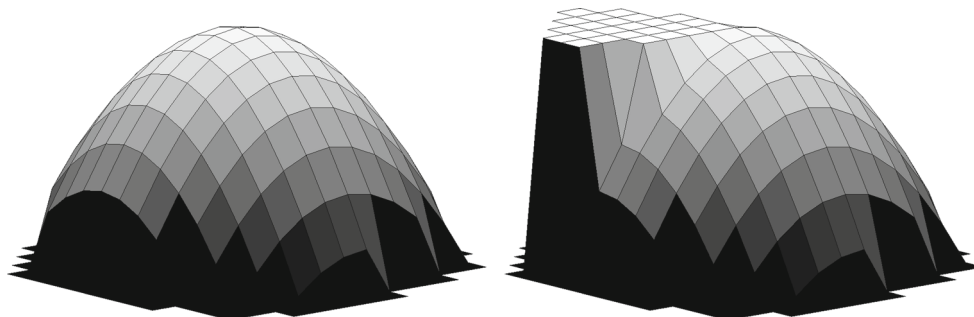

Fig. 8 *Left* parabolic intensity model. *Right* parabolic intensity profile modified to model a bifurcation
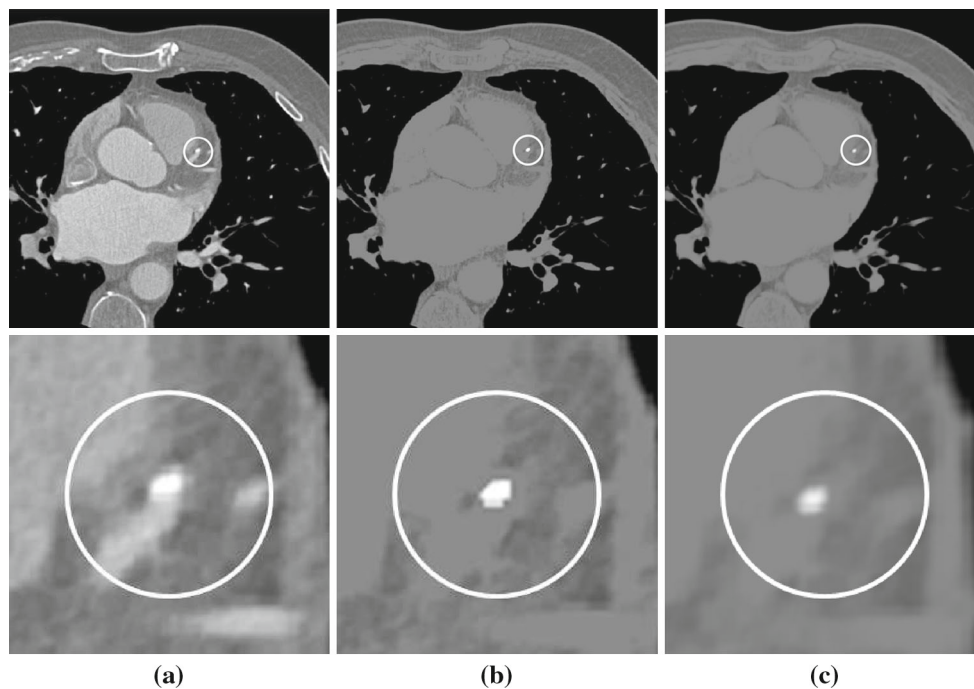
Fig. 9 Virtual CS image construction: an axial slice (*top*) and zoom-in into a calcified lesion in the left anterior descending (LAD) artery (*bottom*). **a** Original cCTA image; **b** after virtual contrast removal; **c** re-sampled to 3 mm slice thickness

to the image noise and set to $3\sigma_c$. In addition, to reduce the amount of false alarms, we require the pixel intensity to be higher than $\mu_c - \sigma_c$ to call it calcium. Formally, for every pixel p inside the artery:

$$M(p) = \begin{cases} 1 & \text{if } I(p) - \tilde{I}(p) > 3\sigma_c \wedge I(p) > \mu_c - \sigma_c, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

The binary map M, created for every coronary artery in its straightened CPR representation, is back-projected to the axial image volume to form the calcium binary mask B for the whole study.

Calcium scoring

The purpose of the method discussed in this section is, given the calcium binary mask B, to come up with a number as close as possible to the original Agatston score.

As suggested in "Calcium scoring in cCTA: the problem" section, due to the non-physical and non-geometric nature of the Agatston score, it is very unlikely that a scoring method based on measuring a geometric property of a lesion directly from cCTA can guarantee a good match to the original score. Instead, we propose to simulate the standard CS CT image formation process on the geometric model derived from cCTA and then estimate the Agatston score in a standard

way from the simulated image. The process hence includes two stages:[2]

1. Building a virtual CS CT image based on the cCTA,
2. Computing the calcium score using the standard Agatston scoring technique on the virtual CS CT study.

Building a virtual CS CT image is a two step process. First, a virtual contrast removal from coronary arteries is performed on a cCTA image. Second, the resulting image is re-sampled according to the standard CS image reconstruction specifications—non-overlapping 3 mm thick slices.

Virtual contrast removal is performed in the following way: voxels brighter than blood (>40 HU), that do not belong to coronary CA lesions (segmented in the previous step) are set to 40 HU (blood intensity)—see Fig. 9b. Note that, as a result of this transformation, the non-contrast, high-intensity voxels outside coronary arteries (e.g., bones, metal wires, non-coronary calcium lesions, etc.) are also set to 40 HU. This is not a problem, since scoring is performed only for coronary CA lesions and non-coronary voxels do not affect the score.

The resulting image is then re-sampled in Z direction to produce a standard CS image with 3 mm slice thickness and 3 mm inter-slice distance (Fig. 9c). Note that there is a degree of freedom in choosing the origin of the new 3 mm

---

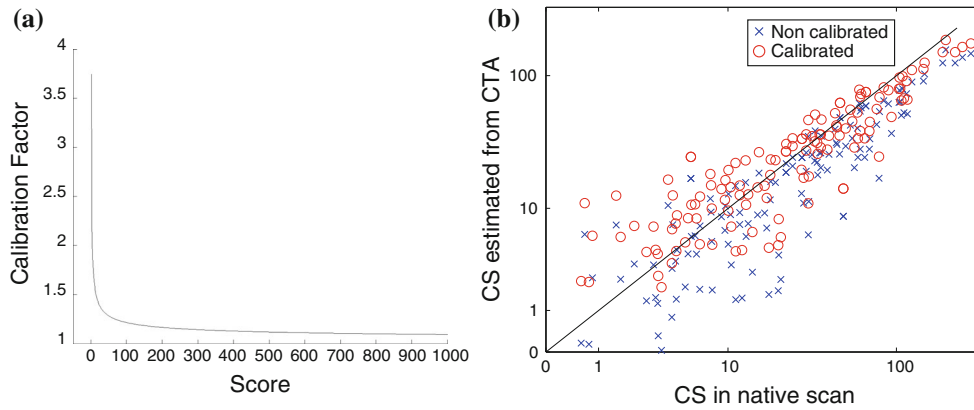[2] The described method for calcium scoring is patent pending.

Fig. 10 **a** Score calibration factor $\rho(S)$. **b** Uncalibrated and calibrated scores versus Agatston score from CS

sampling grid along the Z axis. Every chosen origin location defines a different, but valid virtual CS image, and hence a different score. A similar phenomenon was reported in [21] for conventional CS CT studies, where different scores were obtained by varying the reconstruction origin offset.

To improve the scoring robustness, we reconstruct several virtual CS images by choosing different origin offsets. Score is then computed for each image, and the average score is used as the final score. Here, we reconstruct $\lfloor z_{CS}/z_{cCTA} \rfloor$ different virtual CS images by offsetting the origin point in increments of z_{cCTA}. The re-sampling is performed by convolution with 3 mm long 1D rectangular filter along Z axis. The score S is then calculated by applying the standard Agatston scoring function.

Score calibration

The virtual CS image and hence the score strongly depend on the accuracy of calcium lesion segmentation in cCTA study. As mentioned earlier, image intensities of calcium and contrast material in cCTA studies overlap to some extent. Therefore, low-intensity calcium voxels can be erroneously recognized as contrast, thus resulting in under-segmentation of calcified lesions. Such low-intensity calcium voxels are often observed in small lesions and on the boundary of larger lesions. Obviously, the effect is more prominent for smaller lesions, where the ratio between the boundary and total lesion volume is larger.

In order to compensate for this calcium under-segmentation, we propose to multiply the initial score estimation by a lesion specific calibration factor. The overall study score is then computed as

$$S = \sum_i S_i \rho(S_i),$$

where S_i is the initial score of lesion i, $\rho(S_i)$ is the lesion i calibration factor and i runs over all calcium lesions detected in the study.

To get a rough idea about the form of the calibration function ρ, let us consider a simple example of a lesion that appears in cCTA as a sphere of radius R. Let us assume that, due to the partial volume effect, the outer layer of width Δ in every lesion is not recognized as calcium. Then the true radius of the lesion is $R + \Delta$ and the ratio between the true and observed lesion volumes is $(R + \Delta)^3/R^3$. Assuming the calcium score S is locally linear in lesion volume, i.e., $S \sim R^3$, $\rho(S) = 1 + 3\Delta S^{-\frac{1}{3}} + 3\Delta^2 S^{-\frac{2}{3}} + \Delta^3 S^{-1}$.

Motivated by this example, we approximate the calibration function $\rho(S)$ by a third degree polynomial of $S^{-\frac{1}{3}}$

$$\rho(S) = 1 + \alpha S^{-\frac{1}{3}} + \beta S^{-\frac{2}{3}} + \gamma S^{-1}. \tag{5}$$

In order to find the optimal parameters $[\alpha, \beta, \gamma]$, we used 100 corresponding calcium lesions marked in cCTA and CS studies of the T1 training set. For each lesion, we calculated the Agatston score S_i^{true} from CS, and the virtual calcium score S_i from cCTA. Parameters $[\alpha, \beta, \gamma]$ are selected as the solution to the least squares problem

$$[\alpha, \beta, \gamma] = \arg\min_{\alpha, \beta, \gamma} \sum_i \left(1 + \alpha S_i^{-\frac{1}{3}} + \beta S_i^{-\frac{2}{3}} + \gamma S_i^{-1} - \frac{S_i^{true}}{S_i}\right)^2, \tag{6}$$

$$\text{s.t. } \alpha > 0, \ \beta > 0, \ \gamma > 0.$$

The resulting calibration function is depicted in Fig. 10a. Figure 10b shows the calibration effect on 100 lesions. S_i before and after the calibration are plotted against S_i^{true}. The MAPD defined as:

$$APD = \begin{cases} 0, & \text{if } S_i^{true} + S_i = 0, \\ \frac{|S_i^{true} - S_i|}{(S_i^{true} + S_i)/2} \cdot 100\%, & \text{otherwise} \end{cases} \tag{7}$$

is 58 % before the calibration and 44 % after.

Results

Our system for fully automatic cCTA calcium scoring was implemented as a stand-alone PC application. As discussed in "Methods" section, the system includes three parts: coronary tree reconstruction, calcium detection and scoring. The coronary artery tree reconstruction part was tested in Rotterdam Coronary Artery Algorithm Evaluation Framework [22] and demonstrated some of the best results among competing fully automatic solutions [23] (COR Analyzer group).

Even though calcium detection is not the ultimate goal of the developed system, we tested the performance of the calcium lesion detection part separately. It was tested on a set of 15 corresponding CS CT and cCTA studies. Those studies are not part of the evaluation set used in the overall system performance trials described below. The size of the test set used in this experiment is relatively small due to a large amount of manual labor required to prepare the ground truth and to do the analysis—a per-lesion correspondence is to be established between the modalities, while the same lesion in CS can be split into several lesions in cCTA and vice versa.

Sixty-seven lesions detected in 15 CS studies were compared with those automatically detected in cCTA. Out of 67 lesions, the system correctly detected 63, missed 4 and generated 14 false alarms, yielding the sensitivity of 94 %. Three out of 4 missed lesions and all 14 false lesions were small: $1 \leq CS \leq 10$. Interestingly, 3 out of 14 false alarms were likely to be calcium lesions invisible in CS CT. For the small lesions subgroup, the system identified 17 out of 20 lesions (85 % sensitivity).

Figure 11 shows an example of a small calcified coronary lesion. The lesion peak intensity is lower than the maximal lumen intensity. Therefore, it cannot be correctly segmented, without producing false alarms, by constant thresholding. The lesion is correctly delineated using the proposed approach as the local difference between the image and the model is larger than $3\sigma_c$.

Testing the accuracy of calcium segmentation per se seems to be unfeasible, as no reliable ground truth is available, and the inter-observer variability in manual segmentation of calcium is too high.

The overall system performance for calcium score estimation was tested in two clinical double-blind trials conducted independently by Carmel Medical Center (CMC), Haifa, Israel, and Medical University of South Carolina (MUSC), Charleston, SC. Our system was installed in medical centers, and the trials were conducted by the hospital staff, without our participation. Full clinical reports on those trials were published in [24] and [25], respectively. Here, we summarize the two trials and provide a more technical insight into the achieved results.

The trials included 136 (CMC) and 127 (MUSC) consecutive patients who underwent both CS and cCTA procedures. Patients with stents, pacemaker leads and prior coronary bypass graft surgery were excluded. Patient statistics, used equipment and scanning protocols for the two trials are summarized in Table 1.

To eliminate any bias, the trials were conducted as double-blinded experiments. The standard Agatston score from CS CT was computed in a routine clinical way and not revealed to the person who ran the system. Vice versa, physicians, who performed the standard Agatston score assessment, did not know the score computed by the system. Trial organizers (hospital) received both results and performed the statistical analysis. The system was not trained and did not run on the test studies used in the trials prior to the experiment.

Agatston score (S_{CS}) was computed from CS studies by expert readers in the hospitals (routine medical reports were used), using commercially available semi-automatic/manual software packages (see Table 1).

Calcium score (S_{cCTA}) was computed from cCTA studies using our fully automatic system. The average processing time was 6 ± 1.3 min. 8 studies out of 127 (6.3 %) in MUSC were rejected by the system due to inability to automatically
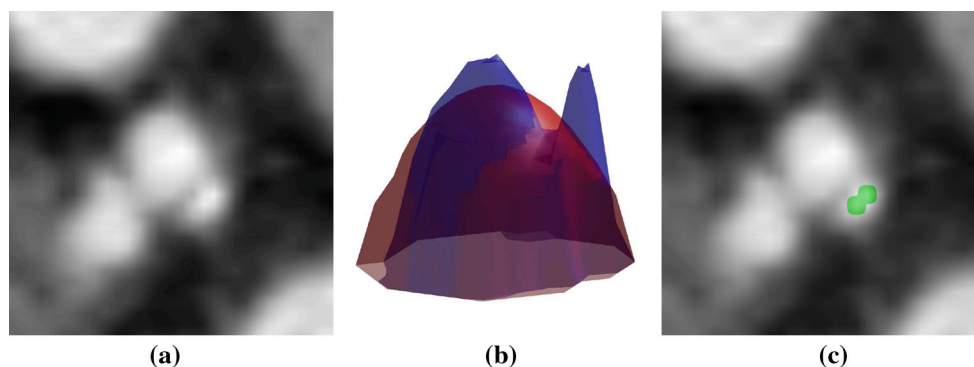

Fig. 11 **a** Coronary artery cross section. The intensity of the small, eccentric CA lesion is lower than the lumen peak intensity. **b** Lumen intensity profile (*blue*) overlapped with the parabolic intensity model (*red*). **c** Segmented lesion mask (*green*)
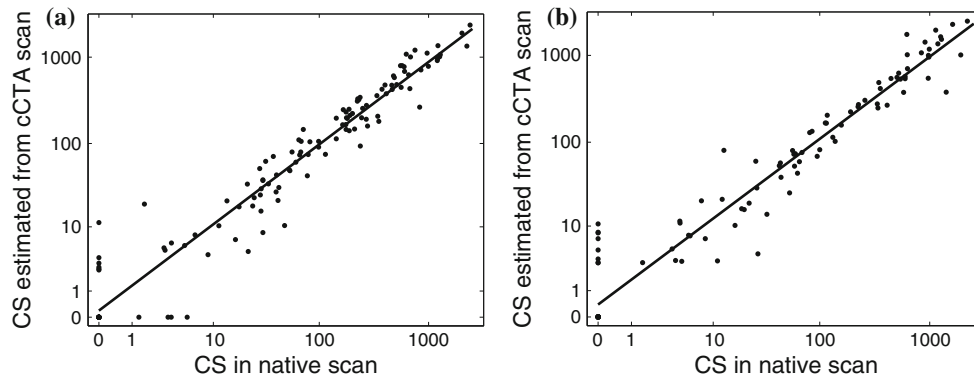
Fig. 12 S_{CS} versus S_{cCTA} scatter plots for CMC (**a**) and MUSC (**b**) studies
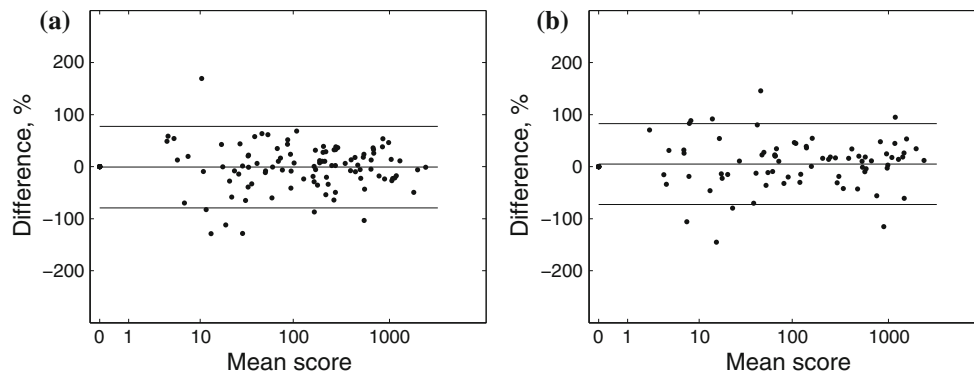

Fig. 13 Bland-Altman percent difference plots between S_{CS} and S_{cCTA} for CMC (**a**) and MUSC (**b**) studies

segment the coronary artery tree. No studies were rejected in CMC trial.

S_{CS} and S_{cCTA} correlated well (see Fig. 12), achieving Pearson correlation of 0.95 (CMC) and 0.91 (MUSC), $P < 0.0001$. The MAPD between S_{CS} and S_{cCTA} was 39 % in CMC trial and 36 % in MUSC.

Bland-Altman percent difference plots are presented in Fig. 13. The mean score is computed as MS = $(S_{cCTA} + S_{CS})/2$ and the relative difference as $(S_{cCTA} - S_{CS})/MS \cdot 100$ %. Horizontal lines in the plots are the bias (-1 % for CMC and 5 % for MUSC) and 95 % CI of limits of agreement ($\sigma = 40.1$ % for CMC and $\sigma = 39.6$ % for MUSC).

Table 2 presents confusion matrices of calcium score categorization into standard risk groups [26] according to S_{CS} and S_{cCTA} for the two trials.

Combining the results of the two evaluations, 211 out of 255 patients (82.7 %) were categorized into the same risk group by both S_{CS} and S_{cCTA}. 43 patients (16.9 %) were classified to an adjacent risk group, and only 1 patient (0.4 %) was classified to a non-adjacent risk group. There were 4 (1.6 %) false negative cases—patients with $S_{CS} > 0$ reported as having no coronary calcium according to cCTA ($S_{cCTA} = 0$)—all in the CMC group.

Table 2 CS risk categories confusion matrices for CMC (a) and MUSC (b) studies

CS					
cCTA	0	1–10	11–100	101–400	>400
(a)					
0	24	4	0	0	0
1–10	5	6	3	0	0
11–100	1	1	29	2	0
101–400	0	0	5	26	2
>400	0	0	0	2	26
(b)					
0	35	0	0	0	0
1–10	8	8	2	0	0
11–100	0	3	22	0	0
101–400	0	0	2	13	2
>400	0	0	0	2	22

Overall, the system performance was slightly better for MUSC studies than for CMC, both in terms of the MAPD—36 versus 39 %, and misclassification rate—15.9 versus 18.3 %. This can be possibly explained by the higher aver-

age noise level in CMC studies—$\bar{\sigma}_c = 42.6\,\mathrm{HU}$ versus $\bar{\sigma}_c = 30.8\,\mathrm{HU}$ in MUSC.

Discussion

Error analysis

Errors in calcium score estimation can be attributed to the two main factors—wrong calcium detection/segmentation and wrong scoring. The former can be caused by two major reasons:

- Centerline tracking and vessel boundary segmentation errors,
- Vessel intensity modelling and noise estimation errors.

Errors of the first type simply leave a calcium lesion outside the area analyzed by the algorithm. Figure 14 shows an example where the system failed to track an artery beyond a total occlusion. The calcium lesion in the untracked distal part of the vessel is undetected and hence not scored.

Errors in vessel intensity modeling are mainly caused by wrong contrast intensity assessment, or by inaccurate centerline detection. Figure 15 demonstrates an example of false alarm caused by wrong centerline position assessment.

Modeling errors can result in both misses and false alarms, while the binarization threshold (we used $3\sigma_c$) controls the position on the ROC curve. One possible way to improve the 2D intensity modeling is by using local estimations of con-
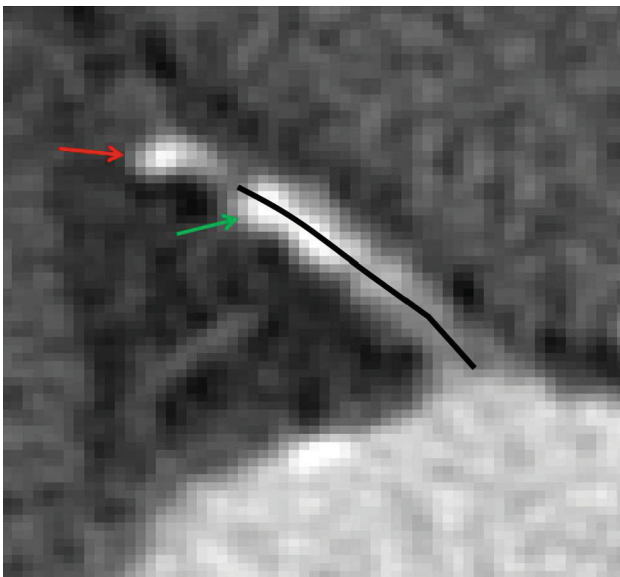

Fig. 14 Vessel tracking failure: tracking stopped due to a total occlusion. Calcium in the tracked proximal part of the vessel (*green arrow*) is scored, whereas the lesion in the distal part (*red arrow*) remains undetected

trast intensity and noise level for calcium segmentation. The aorta-based assessment used here may not be valid for every part of the coronary artery tree. Another source of error in the 3D intensity model is the excessive averaging of model parameters across bifurcations. This can be possibly addressed by explicit handling of bifurcation points in the bilateral filter (Eq. 3).

Scoring errors are due to the inability of the system to match the standard CS CT Agatston score even if all calcium visible in cCTA is perfectly detected and segmented.

Low calcium score studies are the major contributors to the absolute percent error. In general, small lesions are easier to misinterpret, especially for noisy studies (e.g., see Fig. 16). For studies with low total calcium score, the relative influence of small lesions is high. Therefore, the confidence of score assessment for low SNR studies with low calcium score is low. Automatically rejecting or warning about such studies can further improve the accuracy of the system. This is especially important because of the clinical significance of the distinction between zero and nonzero (small) CA score.

The model we used for score calibration is somewhat limited and based on spherical lesions. It can be possibly improved by taking into account other parameters, e.g., contrast intensity, vessel radius, lesion compactness, etc. In addition, the calibration can be made site specific by training on the site data to accommodate for differences in CT equipment, contrast injection, image acquisition and reconstruction protocols.

Another source of score assessment instability against small variations in image intensity is the Agatston score step weighing function (Fig. 1). A possible way to improve the robustness of the technique is by replacing it with a smooth function. Moreover, besides Agatston, there are other calcium scoring methods (e.g., calcium volume or calcium mass) that could be a better choice for assessment from cCTA in terms of accuracy and reproducibility. The main reason to use the Agatston score is a huge body of clinical research that has been performed to link it to the likelihood of cardiac events. Agatston score is the standard of care today, which was our motivation to use it when designing a clinically useful system.

Comparison with the prior art

Comparing with results published in prior art, our system performed better, or with a similar level of accuracy. Mühlenbruch et al. [5] and van der Bijl et al. [3] reported risk categories misclassification rate of 57 and 20–27 % respectively, compared with 16.9 % for our system. The method presented by Glodny et al. [4] demonstrated high correlation of 0.93 (Spearman) to the Agatston score, but the proposed score was not adapted to match the Agatston values, and hence, no error or risk group analysis is possible in this case.
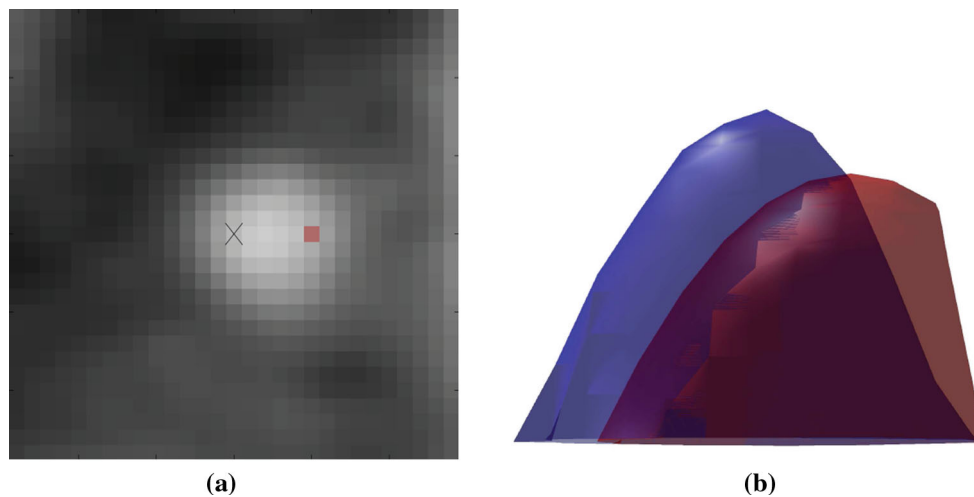
Fig. 15 **a** Vessel cross section. *X*—estimated center-line position; *red*—calcium false alarm. **b** Vessel intensity model (*blue*) and the actual intensity profile (*red*). Calcium is reported for pixels on the *right side*, where image intensity is higher than the model prediction
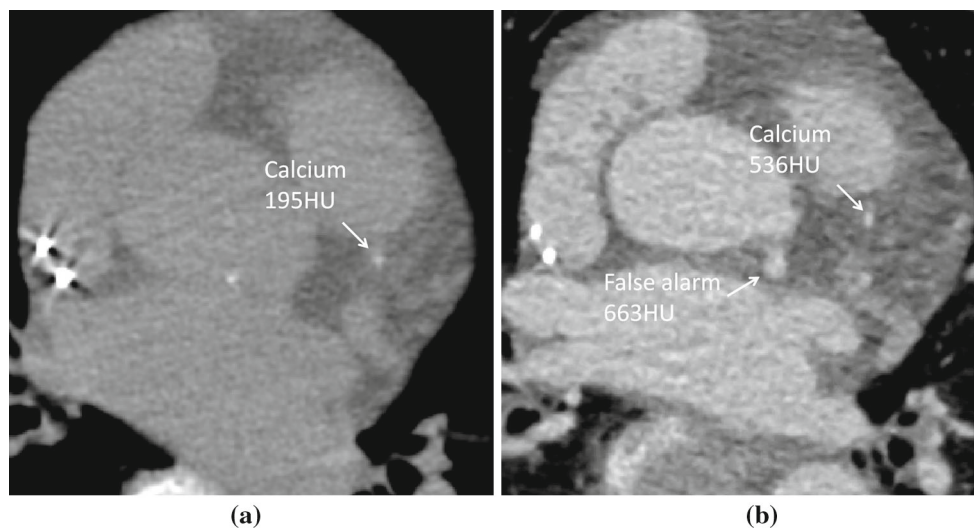

Fig. 16 Small CA lesion in CS study (**a**) and in low SNR cCTA (**b**) of the same patient. High-intensity noise pixel in cCTA (false alarm) is brighter than calcium. Notice that the same calcium lesion has higher intensity in cCTA, due to less expressed partial volume effect

Bischoff et al. [7] reported misclassification rate of 9 %, while combining the "1–10" and "11–100" Agatston risk categories into a single group. Applying the same analysis to our data collected in MUSC study (that used the same CT scanner as in [7]) yields a comparable misclassification rate of 11 %. The correlation of 0.95 between S_{CS} and S_{cCTA} reported in [7] is slightly better than ours in MUSC test (0.91) and equal to ours in CMC. Otton et al. [2] reported a very high correlation (0.99) to the Agatston score, but their method cannot be directly compared with ours, as manual lesion detection and segmentation was used there: lumen and vessel boundaries are drawn by automatic tool and then manually corrected; calcium is detected in the volume between the boundaries by thresholding.

It should be noted though that all prior reported methods mentioned above used a semi-automatic/manual selection of coronary lesions (candidates are detected and segmented automatically and then coronary lesions are manually selected), whereas our system performed the analysis in a fully automatic mode.

Comparison with the standard of care

In the previous section, we reported our system's diagnostic performance by treating the Agatston score derived from CS CT as the gold standard. The more practical question, though, is whether the accuracy of the proposed technique lies within the error range of the standard CS CT. It is

well known that many factors can significantly influence the Agatston score obtained from CS CT scan. Those include cardiac phase, reconstruction filter, field of view, reconstruction offset and others. Takahashi et al. [27] reported 32 % MAPD between scans that used different tube currents. Mao et al. [28] computed MAPD between consecutive CS CT scans reconstructed at different cardiac phases. The lowest MAPD (15 %) was reported for cardiac phase of 40 % and larger MAPD (23–25 %) for other cardiac phases (50, 60, 80 %). Small variation in starting position of CS CT scan reconstruction was found to be responsible for risk category misclassification of 9 % of the subjects [21].

Comparing between calcium score computed with CS CT and EBCT (for which the Agatston score was designed), reveals 65 % MAPD as reported by Stanford et al. [29]. Reinsch et al. [30] reported 14 % misclassification rate between EBCT and dual-source CT for 4 Agatston score risk categories (combined "1–99" category). Moreover, the calcium score computed from the EBCT itself also suffers from high inter-scan variability. Yoon et al. [31] reported 38 % MAPD between identical repeated EBCT scans.

The MAPD of 36–39 % and misclassification rates of 16–18 % (5 categories) and 11–15 % (combined "1–99" category) we measured for our system lie within the variability limits reported above.

Conclusions

In this paper, we presented new methods for automatic calcium detection, segmentation and scoring in cCTA studies. The proposed model-based approach for calcium lesion detection and segmentation deals with intrinsic calcium imaging limitations of cCTA that could not be addressed by thresholding techniques reported in prior art. The suggested scoring method uses a virtual CS CT image to match the standard Agatston score by coping with its highly nonlinear, non-physical and non-geometric nature.

The two methods were implemented as part of a fully automated system for calcium score estimation in cCTA studies. The system was tested in two independent clinical trials on 263 studies. The diagnostic performance of the system measured in these tests is better or comparable with the previously published results and lies within the error margins of the standard of care—Agatston score from CS CT.

High correlation between calcium scores derived by our system from cCTA and Agatston scores obtained using conventional CS CT, coupled with the benefits of our fully automatic solution, suggest that the proposed methods can be used to eliminate the need in a separate CS CT scan, thus reducing the radiation exposure for the patient and costs for the health-care system.

References

1. Agatston AS, Janowitz WR, Hildner FJ, Zusmer NR, Viamonte M, Detrano R (1990) Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol 15(4):827–832
2. Otton JM, Lønborg JT, Boshell D, Feneley M, Hayen A, Sammel N, Sesel K, Bester L, McCrohon J (2012) A method for coronary artery calcium scoring using contrast-enhanced computed tomography. J Cardiovasc Comput Tomogr 6(1):37–44
3. van der Bijl N, Joemai RM, Geleijns J, Bax JJ, Schuijf JD, de Roos A, Kroft LJ (2010) Assessment of Agatston coronary artery calcium score using contrast-enhanced CT coronary angiography. AJR Am J Roentgenol 195(6):1299–1305
4. Glodny B, Helmel B, Trieb T, Schenk C, Taferner B, Unterholzner V, Strasak A, Petersen J (2009) A method for calcium quantification by means of CT coronary angiography using 64-multidetector CT: very high correlation with Agatston and volume scores. Eur Radiol 19(7):1661–1668
5. Mühlenbruch G, Wildberger JE, Koos R, Das M, Flohr TG, Niethammer M, Weiss C, Gunther RW, Mahnken AH (2005) Coronary calcium scoring using 16-row multislice computed tomography: nonenhanced versus contrast-enhanced studies in vitro and in vivo. Invest Radiol 40(3):148–154
6. Hong C, Becker CR, Schoepf UJ, Ohnesorge B, Bruening R, Reiser MF (2002) Coronary artery calcium: absolute quantification in nonenhanced and contrast-enhanced multi-detector row CT studies. Radiology 223(2):474–480
7. Bischoff B, Kantert C, Meyer T, Hadamitzky M, Martinoff S, Schomig A, Hausleiter J (2012) Cardiovascular risk assessment based on the quantification of coronary calcium in contrast-enhanced coronary computed tomography angiography. Eur Heart J Cardiovasc Imaging 13(6):468–475
8. Tessmann M, Higuera FV, Fritz D, Scheuering M, Greiner G (2008) Learning-based detection of stenotic lesions in coronary CT data. In: Proceedings of the vision, modeling, and visualization conference. VMV'08, pp 189–198
9. Mittal S, Zheng Y, Georgescu B, Vega-Higuera F, Zhou SK, Meer P, Comaniciu D (2010) Fast automatic detection of calcified coronary lesions in 3d cardiac CT images. In: Proceedings of the first international conference on Machine learning in medical imaging. MLMI'10. Springer, Berlin, pp 1–9
10. Moselewski F, Ferencik M, Achenbach S, Abbara S, Cury RC, Booth SL, Jang IK, Brady TJ, Hoffmann U (2006) Threshold-dependent variability of coronary artery calcification measurements—implications for contrast-enhanced multi-detector row-computed tomography. Eur J Radiol 57(3):390–395
11. Goldenberg R, Eilot D, Begelman G, Walach E, Ben-Ishai E, Peled N (2012) Computer-aided simple triage (CAST) for coronary CT angiography (CCTA). Int J Comput Assist Radiol Surg 7(6):819–827
12. Begelman G, Goldenberg R, Levanon S, Ohayon S, Walach E (2011) Creating a blood vessel tree from imaging data, US Patent 7983459
13. Begelman G, Goldenberg R, Levanon S, Ohayon S, Walach E (2011) Method and system for automatic analysis of blood vessel structures to identify calcium or soft plaque pathologies. US Patent 7940977
14. Kirbas C, Quek F (2004) A review of vessel extraction techniques and algorithms. ACM Comput Surv 36(2):81–121
15. Lesage D, Angelini ED, Bloch I, Funka-Lea G (2009) A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. Med Image Anal 13(6):819–845

16. Kanitsar A, Fleischmann D, Wegenkittl R, Felkel P, Gröller ME (2002) CPR–curved planar reformation. In: IEEE visualization, pp 37–44

17. Krissian K, Malandain G, Ayache N, Vaillant R, Trousset Y (2000) Model-based detection of tubular structures in 3-D images. Comput Vis Image Underst 80(2):130–171

18. Worz S, Rohr K (2007) Segmentation and quantification of human vessels using a 3-D cylindrical intensity model. IEEE Trans Image Process 16(8):1994–2004

19. Coleman TF, Li Y (1996) A reflective Newton method for minimizing a quadratic function subject to bounds on some of the variables. SIAM J Optim 6(4):1040–1058

20. Tomasi C, Manduchi R (1998) Bilateral filtering for gray and color images. In: Proceedings of the sixth international conference on computer vision. ICCV'98. IEEE Computer Society, Washington, DC, pp 839–846

21. Rutten A, Išgum I, Prokop M (2008) Coronary calcification: effect of small variation of scan starting position on Agatston, volume, and mass scores. Radiology 246(1):90–98

22. Schaap M, Metz C, van Walsum T, van der Giessen A, Weustink A, Mollet N, Bauer C, Bogunović H, Castro C, Deng X, Dikici E, O'Donnell T, Frenay M, Friman O, Hernández Hoyos M, Kitslaar P, Krissian K, Kühnel C, Luengo-Oroz M, Orkisz M, Smedby Ö, Styner M, Szymczak A, Tek H, Wang C, Warfield S, Zhang Y, Zambal S, Krestin G, Niessen W (2009) Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. Med Image Anal 13(5):701–714

23. BIGR (2011) Rotterdam Coronary Artery Algorithm Evaluation Framework. http://coronary.bigr.nl/centerlines/results/results.php Online, Biomedical Imaging Group Rotterdam, Erasmus MC

24. Rubinshtein R, Eilot D, Halon D, Goldenberg R, Gaspar T, Lewis B, Peled N (2012) Automatic assessment of calcium score from contrast-enhanced 256-row coronary CT angiography. J Am Coll Cardiol 59(13s1):E1185

25. Ebersberger U, Eilot D, Goldenberg R, Lev A, Spears JR, Rowe GW, Gallagher NY, Halligan WT, Blanke P, Makowski MR, Krazinski AW, Silverman JR, Bamberg F, Leber AW, Hoffmann E, Schoepf UJ (2013) Fully automated derivation of coronary artery calcium scores and cardiovascular risk assessment from contrast medium-enhanced coronary CT angiography studies. Eur Radiol 23(3):650–657

26. Rumberger JA, Brundage BH, Rader DJ, Kondos G (1999) Electron beam computed tomographic coronary calcium scanning: a review and guidelines for use in asymptomatic persons. Mayo Clin Proc 74(3):243–252

27. Takahashi N, Bae KT (2003) Quantification of coronary artery calcium with multi-detector row CT: assessing interscan variability with different tube currents pilot study. Radiology 228(1):101–106

28. Mao S, Bakhsheshi H, Lu B, Liu SC, Oudiz RJ, Budoff MJ (2001) Effect of electrocardiogram triggering on reproducibility of coronary artery calcium scoring. Radiology 220(3):707–711

29. Stanford W, Thompson BH, Burns TL, Heery SD, Burr MC (2004) Coronary artery calcium quantification at multi-detector row helical CT versus electron-beam CT. Radiology 230(2):397–402

30. Reinsch N, Mahabadi A, Lehmann N, Möhlenkamp S, Hoefs C, Sievers B, Budde T, Seibel R, Jckel K, Erbel R (2012) Comparison of dual-source and electron-beam CT for the assessment of coronary artery calcium scoring. Br J Radiol 85(1015):e300–e306

31. Yoon HC, Goldin JG, Greaser LE, Sayre J, Fonarow GC (2000) Interscan variation in coronary artery calcium quantification in a large asymptomatic patient population. AJR Am J Roentgenol 174(3):803–809