2-1-2018

Evaluation of Motion Artifact Metrics for Coronary CT Angiography

Hongfeng Ma
Marquette University

Eric Gros
GE Healthcare

Aniko Szabo
Medical College of Wisconsin

Scott G. Baginski
Medical College of Wisconsin

Zachary R. Laste
Medical College of Wisconsin

See next page for additional authors

Accepted version. *Medical Physics*, Vol. 45, No. 2 (February 2018): 687-702. DOI.

Authors
Hongfeng Ma, Eric Gros, Aniko Szabo, Scott G. Baginski, Zachary R. Laste, Naveen M. Kulkarni, Darin Okerlund, and Tal Gilat Schmidt

Medical Physics, Vol. 45, No. 2 (February 2018): 687-702. DOI.

Evaluation of Motion Artifact Metrics for Coronary CT Angiography

Hongfeng Ma
Department of Biomedical Engineering at, Marquette University and Medical College of Wisconsin, Milwaukee, WI
Eric Gros
GE Healthcare, Waukesha, WI,
Aniko Szabo
Division of Biostatistics, Medical College of Wisconsin, Milwaukee, WI
Scott G. Baginski
Department of Radiology, Medical College of Wisconsin, Milwaukee, WI
Zachary R. Laste
Department of Radiology, Medical College of Wisconsin, Milwaukee, WI
Naveen M. Kulkarni
Department of Radiology, Medical College of Wisconsin, Milwaukee, WI
Darin Okerlund
GE Healthcare, Waukesha, WI
Taly G. Schmidt

Department of Biomedical Engineering at, Marquette University and Medical College of Wisconsin, Milwaukee, WI

Abstract

Purpose

This study quantified the performance of coronary artery motion artifact metrics relative to human observer ratings. Motion artifact metrics have been used as part of motion correction and best-phase selection algorithms for Coronary Computed Tomography Angiography (CCTA). However, the lack of ground truth makes it difficult to validate how well the metrics quantify the level of motion artifact. This study investigated five motion artifact metrics, including two novel metrics, using a dynamic phantom, clinical CCTA images, and an observer study that provided ground-truth motion artifact scores from a series of pairwise comparisons.

Method

Five motion artifact metrics were calculated for the coronary artery regions on both phantom and clinical CCTA images: positivity, entropy, normalized circularity, Fold Overlap Ratio (FOR), and Low-Intensity Region Score (LIRS). CT images were acquired of a dynamic cardiac phantom that simulated cardiac motion and contained six iodine-filled vessels of varying diameter and with regions of soft plaque and calcifications. Scans were repeated with different gantry start angles. Images were reconstructed at five phases of the motion cycle. Clinical images were acquired from 14 CCTA exams with patient heart rates ranging from 52 to 82 bpm. The vessel and shading artifacts were manually segmented by three readers and combined to create ground-truth artifact regions. Motion artifact levels were also assessed by readers using a pairwise comparison method to establish a ground-truth reader score. The Kendall's Tau coefficients were calculated to evaluate the statistical agreement in ranking between the motion artifacts metrics and reader scores. Linear regression between the reader scores and the metrics was also performed.

Results

On phantom images, the Kendall's Tau coefficients of the five motion artifact metrics were 0.50 (normalized circularity), 0.35 (entropy), 0.82 (positivity), 0.77 (FOR), 0.77(LIRS), where higher Kendall's Tau signifies higher agreement. The FOR, LIRS, and transformed positivity (the fourth root of the positivity) were further evaluated in the study of clinical images. The Kendall's Tau coefficients of the selected metrics were 0.59 (FOR), 0.53 (LIRS), and 0.21 (Transformed positivity). In the study of clinical data, a Motion Artifact Score, defined as the product of FOR and LIRS metrics, further improved agreement with reader scores, with a Kendall's Tau coefficient of 0.65.

Conclusion

The metrics of FOR, LIRS, and the product of the two metrics provided the highest agreement in motion artifact ranking when compared to the readers, and the highest linear correlation to the reader scores. The validated motion artifact metrics may be useful for developing and evaluating methods to reduce motion in Coronary Computed Tomography Angiography (CCTA) images.

1 Introduction

Coronary CT Angiography (CCTA) has shown benefit for detecting and diagnosing coronary artery disease.[1-3] Numerous improvements have been made to CCTA acquisition techniques to increase the temporal resolution and to reduce the effects of vessel motion, including retrospective and prospective ECG gating,[4] dual-source acquisition,[7] and single-beat wide-cone beam imaging.[5, 6] Algorithms have also been introduced to reduce the effects of vessel motion after acquisition, including motion compensated reconstruction algorithms, motion correction algorithms, and automatic determination of the lowest-motion phase for viewing.[8-14] While these efforts have reduced the effects of motion, residual vessel motion artifacts such as blur, deformation, and

shading may still be present. As reported by a recent study, motion artifacts caused by small levels of coronary artery motion can challenge the visual grading of stenosis, despite the high temporal resolution of the CT acquisition.[22]

Quantifying the level of motion artifact is key for developing and evaluating acquisition and algorithmic methods to reduce the effects of motion. Algorithms have been proposed that use motion metrics to optimize CCTA reconstruction to reduce motion artifacts.[11, 12, 14] Several algorithms have also been proposed to determine the lowest-motion phase, with the algorithms generally representing the level of motion through a metric and then selecting the phase that optimizes this metric.[8-10, 23] Some algorithms use metrics of motion, such as the correlation coefficient between images at adjacent phases, to determine the lowest-motion phase.[8, 9, 14] Other algorithms use metrics of vessel image quality or motion artifact severity to identify the lowest-motion phase, such as metrics of entropy, positivity, and circularity.[10, 11]

Although motion artifact metrics have been widely used to improve CCTA images, it has been challenging to validate how well the metrics represent the level of motion artifact. The lack of ground-truth motion data is one challenge for validating motion artifact metrics. Another challenge is the complexity of motion artifacts, which depend on the interrelated combination of both patient and CT scanner factors. Patient factors include heart rate, artery diameter and orientation, and the presence of pathologies such as calcifications and soft plaques. Gantry speed, start angle, contrast level, and spatial resolution are examples of scan factors that affect motion artifacts.

This study presents a procedure for evaluating continuously valued metrics that quantify the level of vessel motion artifact against reader scores. Five motion artifact metrics, both previously proposed metrics and novel metrics, were evaluated using dynamic phantom data and clinical CCTA images. This study focused on metrics of through-plane vessel quality, as the fastest coronary artery segments are when the vessels are in the through-plane orientation[17] The motion artifact metrics were calculated on the images and then compared with ground-truth scores that were obtained by reader studies using a multiple pairwise comparison method.

This paper first describes the typical features of coronary artery motion artifacts (Section **2.A**) followed by a description of the investigated motion artifact metrics (Section **2.B**). The phantom (Section **2.C**), clinical images (Section **2.D**) and observer study methods (Section **2.E**) are then described, with results presented in Section **3**.

2 Methodologies and materials

2.A. Motion artifacts

Motion artifacts are caused by inconsistencies in projection data due to vessel motion. The appearance of motion artifacts depends on a complex relationship between patient and acquisition factors. The artifacts depend on the vessel motion velocity relative to the projection direction, and thus depend on heart rate, heart rate variability, gating, gantry speed, and gantry angle. Artifact size and intensity generally increase with vessel size and intensity, respectively. Because of the complexity of motion artifacts, it is difficult to quantify these motion artifact effects individually. Therefore, previous studies have proposed motion artifact metrics that evaluate the overall vessel image quality.[8-13]

Despite the complexity of motion artifacts factors, motion artifacts present as vessel deformation and low-intensity shading artifacts. Motion artifacts can be classified into different patterns of vessel deformation, as shown in Fig. 1:

- **Crescent**: The vessel appears with a crescent shape. The orientation of the crescent is determined by both the CT gantry start angle and the direction of vessel movement. Figure 1(b) shows that low-intensity shading artifacts are present in addition to the vessel deformation.

- **Tails and horns**: The vessel has a distinguishable core and one or more high intensity tails [see Fig. 1(c)]. When the vessel displacement is small, the tails are short which look like horns [Fig. 1(d)]. Dark shading is observed between the tails/horns.

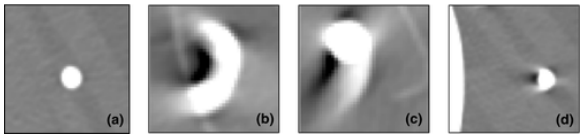

Figure 1 Motion artifacts patterns. (a) Artifact free image (b) Crescent, (c) Tails, (d) Horns. The images are regions of interest (ROIs) extracted from the phantom study described in Section 2.C.

2.B. Evaluated motion artifact metrics

This section describes the motion artifact metrics investigated in this work. Each metric is calculated on a region of interest (ROI) extracted around the vessel, which includes the vessel, shading artifacts and the vessel background. The methods used to extract the vessel ROIs for the phantom and clinical data study are described in Sections **2.C** and **2.D**.

Motion artifact metrics in previous studies were developed for quantifying the relative level of motion artifacts within a single exam. For example, metrics were developed for finding the lowest-motion phase within one exam or optimizing the reconstruction parameters for a specific exam to reduce motion artifacts. In these previous applications, the metrics need only to represent the relative change in motion artifact level for a particular patient and vessel. The performance of these metrics is unknown for absolute quantification, i.e., consistently measuring motion artifact levels across patients and vessels. The metric values may vary with factors such as vessel size, contrast, and the appearance of calcifications and soft plaques. These factors may be relatively constant within one exam but vary across patients. Absolute quantification of motion artifact level may be important for developing and comparing motion reduction techniques across different patients. In addition, metrics validated to measure the absolute level of motion artifact level will also be beneficial for relative evaluation of motion artifact level.

This study investigated three existing metrics (positivity, entropy, and normalized circularity) for the task of absolute motion artifact quantification, as well as two additional novel metrics: Fold Overlap Ratio (FOR) that quantifies vessel symmetry and Low-Intensity Region Score (LIRS) that quantifies the intensity and area of dark shading regions. Unlike previously proposed motion artifact metrics, the FOR and LIRS metrics were designed for absolute quantification of motion artifact level. For example, FOR measures vessel symmetry independent of vessel diameter and intensity. LIRS is designed as a function of values that are relative to the vessel intensity and size. Sections **2.B.1** through **2.B.5**provide details about the investigated motion artifact metrics.

2.B.1. Positivity

A metric of positivity was previously proposed to guide motion correction and reconstruction.**11** The positivity metric is designed to penalize outlier pixels with low-intensity values. Positivity is defined as

$$L_{pos} = \sum_{h_j \in \mathrm{ROI}} \left(\begin{cases} 0, & h_j > T \\ (h_j - T)^2, & otherwise \end{cases} \right),$$

(1) $$L_{pos} = \sum_{h_j \epsilon ROI} \left(\begin{cases} 0, & h_j < T \\ (h_j - T)^2, & otherwise \end{cases} \right),$$

where h_j is the intensity of the jth pixel of the vessel ROI. Shading artifacts are assumed to have lower intensity than the myocardium. In previous work, the myocardium intensity was calculated as the mean value of the pixels surrounding the coronary artery. The threshold *T* was defined as the myocardium intensity minus the standard deviation of the myocardium to identify the shading artifacts while reducing sensitivity to noise. The range of positivity is [0, ∞).

2.B.2. Normalized circularity

A metric of circularity was previously proposed to quantify motion artifacts,[14] as through-plane vessels appear as circles when static and deform with motion. The circularity metric is calculated on a binary image representing the segmented vessel region (Segmentation procedure described in Section **2.C**). Circularity is defined as

$$L_{circ} = \frac{p^2}{4\pi A},$$

(2) $$L_{circ} = \frac{p^2}{4\pi A} \;;$$

where *A* and *p* are area and perimeter of the segmented binary vessel, respectively. The circularity of a perfect circle is equal to one, with noncircular shapes having circularity greater than one. Since *A* and *p* are measured on a pixelized image, the circularity value may be less than one in some cases due to discretization errors. As in previous work, the circularity values were transformed to have a range of zero to one, with a value of zero indicating high deformation and a value of one indicating a perfect circle. The transformation function, (Eq. 3) assumes that vessels with $L_{circ} > 2$ represent high deformation and are assigned a metric value of zero. The transformed circularity is called normalized circularity in this study.

$$L_{circ\_n} = \begin{cases} 1 - |L_{circ} - 1| & L_{circ} \le 2 \\ 0 & L_{circ} > 2 \end{cases}$$

(3) $$L_{circ\_n} = \begin{cases} 1 - |L_{circ} - 1| & L_{circ} \le 2 \\ 0 & L_{circ} < 2 \end{cases}$$

2.B.3. Entropy

A metric of entropy was previously proposed for quantifying motion artifacts.[15] Entropy is given by

$$L_{ent} = -\sum_{h \in \mathrm{ROI}} p(h) \ln p(h),$$

(4) $$L_{ent} = \sum_{h \epsilon ROI} p(h) \ln p(h),$$

Where *h* is the intensity of a pixel in a vessel ROI. *p*(*h*) is the probability of the pixels having intensity *h*. As in previous motion artifact metric work, *p*(*h*) was estimated using a Parzen-window technique,[16]

$$p(h) = \frac{1}{\mathrm{N}} \sum_{h_j \in \mathrm{ROI}} \mathrm{R}(h - h_j),$$

$$(5)\ p(h) = \frac{1}{N} \sum_{h_j \epsilon ROI} R(h - h_j)$$

where N is the number of pixels in the ROI, and $R(x)$ is a Gaussian kernel.

The entropy metric is relatively small when the intensities within the extracted ROI are concentrated within a narrow range. The entropy metric increases as the intensity values become more widely distributed. The entropy range is from zero to one with zero representing an ROI with uniform intensity.

Motion causes vessel deformation and low-intensity shading. These artifacts expand the distribution of image intensities, causing higher entropy values. One advantage of the entropy method is that it does not require segmentation of the vessel or shading artifact regions.

2.B.4. Fold overlap ratio

The Fold Overlap Ratio (FOR) is a metric of symmetry proposed to measure vessel deformation. It is calculated from a binary image of the segmented vessel region. The segmentation method will be described in Section **2.C**. The binary image is folded along an axis passing through the segmented vessel centroid. The segmented binary vessel pixels are then divided into two subsets, V_1 and V_2, by the axis, where V_1 represents the region that was held stationary while V_2 is the region that was folded over the axis. FOR is defined as the ratio of the number of pixels in the intersection of V_1 and V_2 to the number of pixels in the union of V_1 and V_2

$$L_{FOR} = \frac{\| V_1 \cap V_2 \|_0}{\| V_1 \cup V_2 \|_0},$$

$$(6)\ L_{FOR} = \frac{\|V_1 \cap V_2\|_0}{\|V_1 \cap V_2\|_0},$$

A static through-plane vessel is circular. Therefore, the FOR of a static through-plane vessel is close to one. A deformed vessel may have a high FOR when folded across some axes. We selected two orthogonal axes for folding, vertical and horizontal, with the smaller FOR selected to represent the FOR of the vessel,

$$L_{FOR\_V} = \min\{L_{FOR\_ver}, L_{FOR\_hor}\},$$

$$(7)\ L_{FOR\_V} = \min\{L_{FOR\_ver}, L_{FOR\_hor}\},$$

where L_{FOR_ver} and L_{FOR_hor} are the FOR values obtained by folding across the vertical and horizontal axes, respectively. The calculation of the FOR metric is illustrated in the flowchart of Fig. 2.

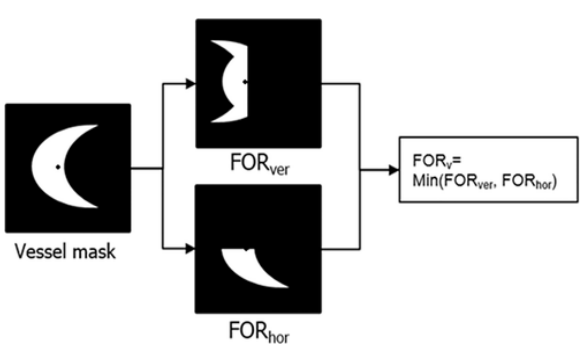

Figure 2 Vessel FOR calculation flowchart. The dark point in the binary images represents the vessel region centroid.

2.B.5. Low-Intensity Region Score

Low-intensity motion shading artifacts depend on vessel size, contrast, and motion, as well as scan conditions and the reconstruction algorithm. As the severity of the motion artifact increases, the low-intensity shading generally becomes larger in area and more negative in intensity. However, in addition to the effects of motion, the size of the shading artifact increases with vessel size, while the shading intensity decreases with increasing vessel contrast. The Positivity metric described in Section **2.B.1**penalizes the size and intensity of the artifact region, without considering the vessel properties. Therefore, the positivity metric may overpenalize large and bright vessels, while being less sensitive to motion artifacts in vessels that are small or have less contrast.

To overcome these potential issues with the positivity metric, we instead propose a Low-Intensity Region Score (LIRS). The Low-Intensity Region *Intensity Score* (LIR-*IS*) quantifies the low-intensity shading relative to the background intensity. LIR-*IS* is expressed as:

$$L_{LIR\_IS} = \frac{\bar{I}_{LIR} + 1024}{\bar{I}_{background} + 1024},$$

(8) $$L_{LIR\_IS} = \frac{\bar{I}_{LIR} + 1024}{\bar{I}_{background} + 1024},$$

where $\bar{I}_{LIR}$ is the mean intensity in the dark shading region and $\bar{I}_{background}$ denotes the mean intensity of the vessel background (i.e., the myocardium in clinical images), with both intensities offset by 1024 to ensure a positive relative fraction. This metric requires a method to identify the low-intensity shading regions, similar to the positivity metric. Since the intensity within the dark shading region is always lower than background tissue, the range of the LIR-*IS* metric is (0, 1], where zero indicates severe artifact. If the image has no identified dark shading artifacts, the LIR-*IS* is set equal to one.

The Low-Intensity Region *Area Score* (LIR-*AS*) is defined as

$$L_{LIR\_AS} = 1 - \frac{A_{LIR}}{A_{ves}},$$

(9) $$L_{LIR\_AS} = 1 - \frac{A_{LIR}}{A_{VES}},$$

where A_{LIR} is the total area of all segmented low-intensity shading regions in the vessel ROI, and $A_{ves}$is the area of the segmented vessel region. The LIR-*AS* metric quantifies the low-intensity shading artifact size relative to the vessel size, so that the metric can be sensitive to artifacts in small vessels without penalizing larger vessels. Since the low-intensity shading region is usually smaller than the vessel region, LIR-*AS* ranges from zero to one, with one indicating a region without dark shading artifact.

The LIRS metric is defined as the average of the LIR-*IS* and the LIR-*AS* metrics. The range of the LIRS metric is (0, 1], with zero corresponding to severe artifact and one corresponding to no artifact.

$$L_{LIRS} = \frac{L_{LIR\_IS} + L_{LIR\_AS}}{2}$$

(10) $$L_{LIRS} = \frac{L_{LIR\_IS} + L_{LIR\_AS}}{2}$$

2.C. Phantom data

The motion artifact metrics were first evaluated using an experimental dynamic cardiac phantom. A phantom was used in this study because it provided a test case with known vessel diameters and contrast levels and with known motion phases against which to test the metrics.

The cardiac phantom modeled the cardiac chambers and coronary arteries (Fig. 3). Six artery models were added to the phantom, each filled with iodine-based contrast. The vessel diameters were 2 mm for vessel 1 (V1), 3 mm for vessels 2 and 3 (V2 and V3), 4 mm for vessels 4 and 5 (V4 and V5), and 5 mm for vessel 6 (V6). The vessels contained varying levels of iodinated contrast and contained calcifications and soft plaques in some slices. Phantom motion, simulating a 60 bpm cardiac cycle, was controlled by a dynamic platform (Quantitative Standard Pulsating Phantom, Fuyo Corporation, Tokyo, Japan).

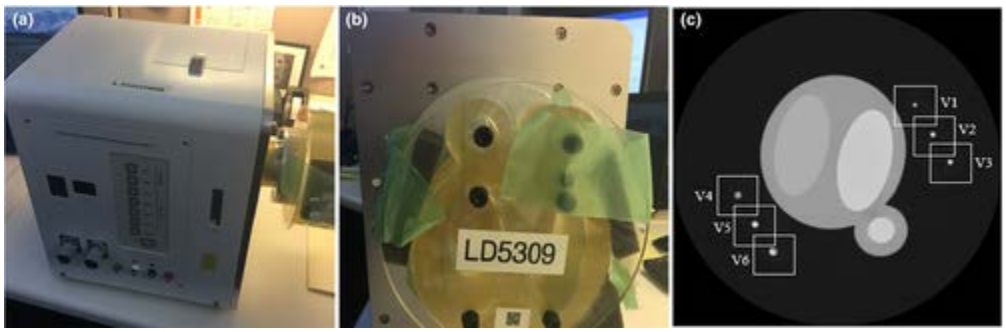

Figure 3 (a) and (b) The dynamic cardiac phantom shown in two orientations and (c) CT image of the static phantom depicting the vessels and the extracted ROIs. The diameters of V1 through V6 are 2, 3, 3, 4, 4, 5 mm. [Color figure can be viewed at wileyonlinelibrary.com]

CT images of the phantom were collected using a wide-cone beam axial scan (256 slice, 16-cm detector coverage) at 120 kVp tube voltage, 600 mA tube current, and 0.35 s gantry rotation (Revolution CT, GE Healthcare). A variety of motion artifacts were generated by repeating the scan using eleven different gantry start angles ranging from 37 to 333 degrees. For each gantry start angle, images were reconstructed at 55%, 60%, 65%, 70%, and 75% of simulated R-R interval, representing vessel velocities of 65, 53, 33, 17, and 10 mm/s. All images were reconstructed by Filtered Back Projection (FBP) with 0.4883 × 0.4883 mm pixel dimensions and 0.625 mm slice thickness. No motion correction algorithms were applied to the data.

Regions of interest (ROIs) of size 25 × 25 mm were extracted around each vessel in each reconstructed image for further analysis (Fig. 3). The selected ROI size ensured that the deformed vessel and its shading artifacts were included in the ROI. For vessel V1, ROIs that contained the myocardium were manually processed to exclude those pixels from further analysis.

The circularity, FOR, and LIRS metrics require binary images of the deformed vessel and/or low-intensity shading regions. In the phantom images, the vessel intensity is greater than 500 HU, while the background intensity is approximately −460 HU. A threshold of −200 HU was selected as the threshold T_{ves} to segment the bright vessel regions, with all pixels greater than T_{ves}considered to be part of the deformed vessel. A threshold, T_{LIR}, of −600 HU was selected for segmenting the low-intensity shading regions, with any pixel with value less than T_{LIR} identified as shading artifact. To investigate the sensitivity of the FOR and LIRS metrics to these threshold parameters in the phantom images, the metrics were also calculated with the T_{ves} threshold varied between −250 and −150 HU in increments of 5 HU, and with T_{LIR}varied between −650 and −550 HU in increments of 5 HU. The change in metric value with the change in threshold was evaluated across the ROIs.

The LIRS and positivity metrics require the mean vessel background value, which in clinical images would be equal to the mean myocardium intensity. For the phantom images, the mean vessel background value was

calculated as the average of all pixels in the extracted ROI not in the segmented vessel and shading regions. Part of the myocardium is visible in the extracted V1 vessel ROI. To prevent this region from biasing the observer or metric results, the myocardium pixel values for this vessel were set to background levels prior to vessel segmentation. Segmentation of the vessel and low-intensity shading regions was performed on all extracted vessel ROIs. Figure 4 displays example segmented vessel and low-intensity shading regions for all motion phases

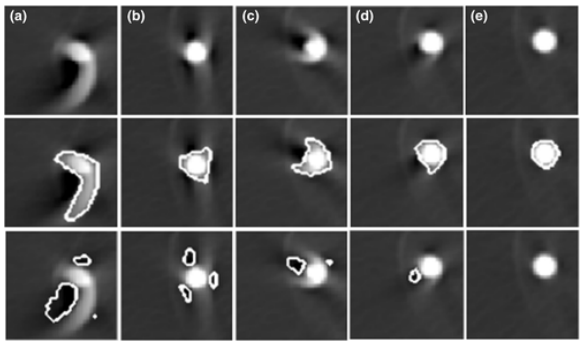

Figure 4 Vessel and shading region segmentation results for phantom images. The top row displays the original images. The center row displays the vessel segmentation results. The bottom row displays the shading artifact segmentation results. (a) phase 55%, (b) phase 60%, (c) phase 65%, (d) phase 70%, and (e) phase 75% of the simulated R-R cycle.

2.D. Clinical data

The phantom vessel ROIs, which contain only vessel and artifact structures, are relatively simple compared to clinical images. The simplified phantom images provide a useful initial evaluation of the motion artifact metrics for the ideal case where the artifacts are easy to distinguish through thresholding. After identifying candidate metrics using the phantom data, this study then evaluated the metrics on clinical images to investigate effectiveness for absolute motion artifact quantification across different patients and varying conditions such as heart rate, vessel size, noise levels, and contrast level.

Fourteen previously acquired CCTA exams were used for this study. The exams were collected at 100 and 120 kVp tube voltage, depending on patient size (Revolution CT, GE Healthcare). Tube current was modulated through automatic exposure control for each patient. Gantry speed was 0.35 s per rotation, with randomly varying gantry start angle. All images were collected by axial scan mode, as is the protocol for the wide-cone beam acquisition on the investigated scanner.

The patient heart rates ranged from 52 to 82 bpm. Twenty phases were reconstructed across the 14 exams by filtered backprojection, with reconstructed phases ranging from 43% to 82% of the R-R interval. The images were reconstructed with 17 to 26 cm Field of View (FOV) and 0.625 mm slice thickness. No motion correction algorithms were applied to the data.

For each dataset, slices containing the through-plane Right Coronary Artery (RCA) were manually identified. Since the purpose of this study is to investigate the performance of the motion artifact metrics independent of the vessel and shading segmentation algorithms, the RCA ROIs were extracted manually. The vessel and shading regions required by the normalized circularity, FOR and LIRS metrics were also segmented manually by three expert readers. The ground-truth vessel and shading region segmentations were obtained by combining the reader segmentations using the Simultaneous Truth And Performance Level Estimation (STAPLE) method.[21] The myocardium region, which is required by the LIRS and positivity metrics, was defined as pixels in the ROI that were not in the segmented vessel, shading artifact, and lung regions.

2.E. Ground truth and metric performance evaluation

2.E.1. Observer study

A ground-truth motion artifact level is required for each vessel ROI to assess the effectiveness of the proposed metrics. Because of the complex combination of factors that cause motion artifacts, we hypothesize that velocity is insufficient for providing a ground-truth artifact level. For example, Fig. 5 demonstrates the inconsistency between vessel velocity and artifact level, which will be further quantified in the phantom study.

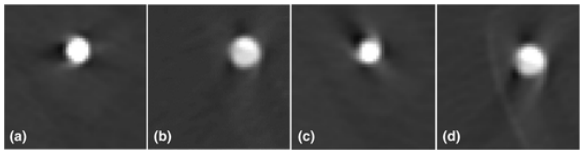

Figure 5 Example images demonstrating the inconsistency between vessel velocity and motion artifact. Velocities of the ROIs are (a) 10 mm/s, (b) 17 mm/s, (c) 17 mm/s, and (d) 33 mm/s. The motion artifact in image (a) is more severe than image (b) despite the slower velocity. Images (b) and (c) have the same velocity but different levels of artifact. The vessel in image (d) has the highest velocity, but with moderate artifacts.

This study performed human observer studies to provide a ground-truth artifact score. Likert scale and pairwise comparison are two tools that are commonly used for subjective image quality assessment. A previous study[18] demonstrated that pairwise comparison yielded more accurate reader assessment than the Likert scale. This study used a pairwise comparison reader study to obtain ground-truth motion artifact scores against which to evaluate the continuously valued motion artifact metrics.

Two separate observer studies were performed for phantom data and clinical data. Forty vessel ROIs were selected randomly for each observer study. For the phantom study, the selected ROIs spanned the range of acquired vessel diameters, motion phases, slices, and gantry start angles, as can be seen in the labels in Fig. 6. Figure 7 displays the clinical ROIs, with the patient heart rate and motion phase labeled above each image.

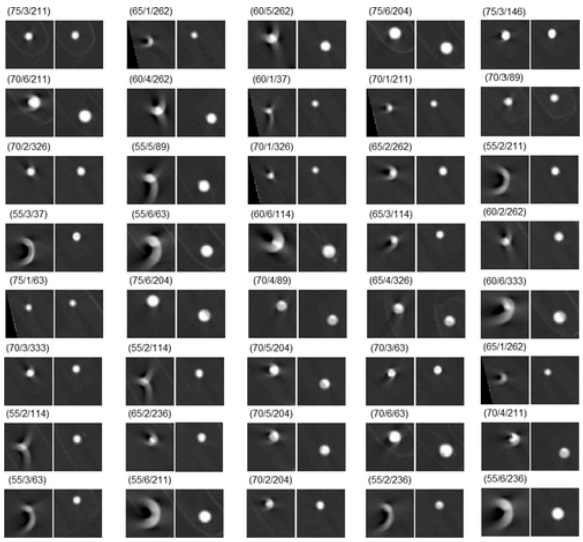

Figure 6 Each pair of images represents (left) a phantom vessel ROI randomly selected for the observer study and (right) the image of the same vessel segment during a static scan. The label above each selected image states the motion phase percentage/diameter (mm)/and the gantry start angle of the selected vessel ROI.

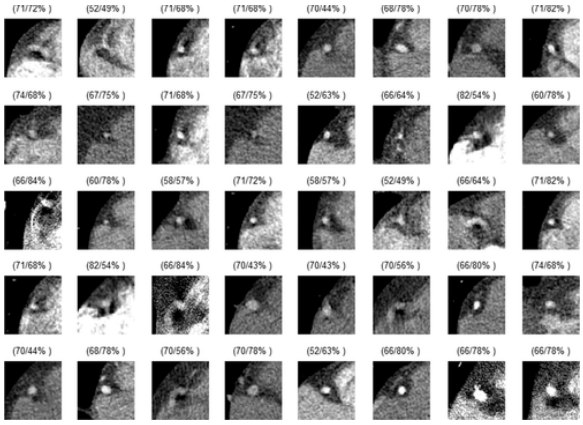

Figure 7 Each image represents a clinical vessel ROI randomly selected for the observer study. The label above each image states the patient's heart rate/motion phase percentage of the selected vessel ROI.

The observer studies were performed using the same monitor settings for all readers. Three readers with experience in CCTA imaging were sequentially and individually shown all 780 pairs of the 40 ROI images. For the phantom data, the readers were trained physicists and engineers in the field of cardiac imaging. For the clinical data, the readers were radiologists specializing in cardiothoracic (S. G. B. and Z. R. L.) or body (N. M. K) imaging. All images were presented to the readers at the same window level and window width. The ROI images were magnified by a factor of three for display, with this magnification factor held constant throughout the observer study. For each pairwise comparison, the readers were asked to select the image with the least motion artifacts. The readers could also indicate a “tie” if they could not distinguish a difference in image quality between the two images. The order of the presented image pairs was randomized across the three readers. The readers were blinded to image information such as the phase, gantry start angle, or vessel diameter, blinded to the other reader scores, and were blinded to the motion artifact metric values.

At the beginning of each reader evaluation, all image ROIs were initialized with a score of zero. For each pairwise comparison, the score of the image selected by the reader was incremented by one, while the score of the unselected image was decremented by one. If the reader selected a tie, no score was added or removed from either of the images. At the end of the evaluation, the score for each image and reader represented the number of times that image was selected as having better image quality minus the number of times the image was selected as having lower image quality. An image's final score was the sum of the three scores obtained from the three independent reader evaluations, with higher scores representing lower motion artifact severity.

2.E.2. Ranking agreement between metrics and reader scores (Kendall's Tau Coefficient)

The motion artifact metrics described in Section **2.B** were calculated for each vessel ROI used in the observer studies. The Kendall's Tau coefficient was calculated to quantify the ranking agreement between a motion artifact metric and the ground-truth reader score. Kendall's Tau coefficient is a statistic to measure ordinal association, or ranking relationship, between two measured quantities.[19] It quantifies the similarity of the orderings when ranked by two quantities, which in this study were one of the calculated metrics (X) and the aggregate reader score (Y).

The aggregate reader score increases with decreasing motion artifact. For each pair of images (p_1, p_2), we defined the signed difference of the reader scores for two images as

$$\Delta_p^Y = \begin{cases} 1, & if Y_{p1} > Y_{p2} \\ 0, & if Y_{p1} = Y_{p2} \\ -1, & if Y_{p1} < Y_{p2} \end{cases}$$

$$(11)\ \Delta_p^Y = \begin{cases} 1, & if\ Y_{p1} > Y_{p2} \\ 0, & if\ Y_{p1} = Y_{p2} \\ -1, & if\ Y_{p1} < Y_{p2} \end{cases}$$

Using the definition in Eq. 11, $\Delta_p^Y = 1$ when image p_1 has fewer artifacts than image p_2.

For metrics that increase with decreasing motion artifact (normalized circularity, FOR, LIRS), Δ_p^X was similarly defined as the sign of the difference of the metric value, i.e., X_{p1} and X_{p2}

$$\Delta_p^X = \begin{cases} 1, & if\ X_{p1} > X_{p2} \\ 0, & if\ X_{p1} = X_{p2} \\ -1, & if\ X_{p1} < X_{p2} \end{cases}$$

$$(12)\ \Delta_p^X = \begin{cases} 1, & if\ X_{p1} > X_{p2} \\ 0, & if\ X_{p1} = X_{p2} \\ -1, & if\ X_{p1} < X_{p2} \end{cases}$$

For metrics that decrease with decreasing motion artifact (entropy, positivity), Δ_p^X was calculated according to the following expression, so that Δ_p^X = 1 when image p_1 has fewer artifacts than image p_2.

$$\Delta_p^X = \begin{cases} 1, & if\ X_{p1} < X_{p2} \\ 0, & if\ X_{p1} = X_{p2} \\ -1, & if\ X_{p1} > X_{p2} \end{cases}$$

$$(13)\ \Delta_p^X = \begin{cases} 1, & if\ X_{p1} < X_{p2} \\ 0, & if\ X_{p1} = X_{p2} \\ -1, & if\ X_{p1} > X_{p2} \end{cases}$$

In this study, the 40 ROIs were compared with 780 comparisons, with each ROI compared to all other ROIs. Δ_p^X, Δ_p^Y was calculated for each of the comparisons. A pair $p1p2$ was concordant if the X and Y scores agreed on the ranking of the two ROIs, i.e., $\Delta_p^X, \Delta_p^Y > 0$. Then, for L image pairs, the Kendall's Tau coefficient was the difference of the fraction of concordant and discordant pairs. Let Δ_p^X, Δ_p^Y be the indicator of concordance. C_p = 1 indicates are concordant. C_p = −1 means the pair of images are discordant. Kendall's Tau coefficient was then calculated as:

$$\tau = \frac{\sum_{p=1}^{L} C_p}{L}$$

$$(14)\ \tau = \frac{\sum_{p=1}^{L} C_p}{L}$$

The range of τ is from −1 (100% negative association) to 1 (100% positive association).

We used a bootstrap method to evaluate the confidence intervals of the estimated Kendall's Tau coefficients, using the following procedure:[20]

- *M* images were randomly selected from *N* images without replacement (i.e., for each bootstrap trial, some images may be selected more than once, while others may not be selected). In this study, *M* = 40, *N* = 40.
- The Kendall's Tau, τ_b, coefficient was calculated for this resampled data, as described in Eqs. 11-13. The duplicates contributed as tied observations.
- Steps (1) – (2) were repeated *n* times to obtain τ_b, $b = 1,2,3,\ldots,n$. This study used n = 1000.
- The bootstrap standard error was estimated as:

$$SE_{boot}(\tau) = \sqrt{\frac{\sum_{i=1}^{n}(\tau_i - \bar{\tau})^2}{n-1}},$$

(15) $$SE_{boot}(\tau) = \sqrt{\frac{\sum_{i=1}^{n}(\tau_i - \bar{\tau})^2}{n-1}},$$

where $\bar{\tau}$ was the average τ_b, b = 1,2,3,…,n.

2.E.3. Ranking agreement between two metrics

The bootstrap method described in Section **2.E.2** was also used to compare agreement between two metrics. In each iteration of the bootstrap process, both the Kendall's Tau coefficient τ_{b1} of X1 with Y and τ_{b2} of X2 with Y were calculated. The standard error of (τ_{b1}–τ_{b2}) was estimated as

$$SE_{boot}(\tau_1 - \tau_2) = \sqrt{\frac{\sum_{i=1}^{n}((\tau_{b1} - \tau_{b2}) - (\bar{\tau}_1 - \bar{\tau}_2))}{n-1}},$$

(16) $$SE_{boot}\ (\tau_1 - \tau_2) = \sqrt{\frac{\sum_{i=1}^{n}((\tau_{b1} - \tau_{b2}) - (\bar{\tau}_1 - \bar{\tau}_2))}{n-1}},$$

The null hypothesis that two metrics have equivalent ranking agreement to the ground-truth reader score was tested with the asymptotically normal distributed test statistic

$$Z = \frac{\tau_1 - \tau_2}{SE\,(\tau_1 - \tau_2)} \sim N\,(0, 1)$$

$$Z = \frac{\tau_1 - \tau_2}{SE(\tau_1 - \tau_2)} \sim N(0,1)$$

2.E.4. Linear correlation between metrics and observers

For some applications, it may be important for the motion artifact metrics to correlate linearly to the reader scores, in addition to having good ranking agreement. A linear relationship means that the metric and reader scores agree in how they quantify the relative difference between motion artifact levels. Linear regression was performed for each investigated metric against the ground-truth reader scores to evaluate whether the metrics correlated linearly with the reader scores

3 Results

3.A. Phantom study results

Figure 8 shows the results of the phantom data observer study, with the ground-truth aggregate reader score displayed for each ROI. Images with higher score generally demonstrated less artifacts, such that the images in Fig. 8 and are displayed from low to high artifact level.

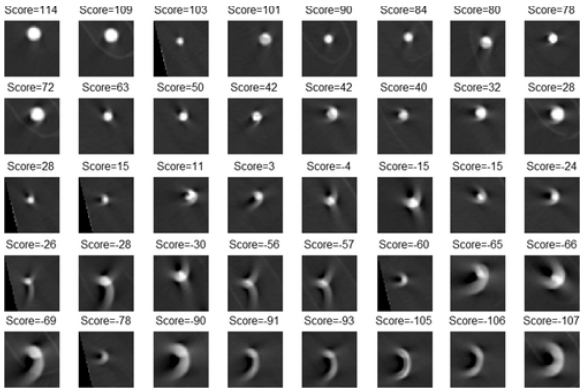

Figure 8 The 40 phantom images used in the reader study displayed in descending order of reader score. The number displayed above each image is the ground-truth aggregate reader score, which was calculated as the sum of the three reader scores.

Figure 9 displays the scatter plots of the aggregate reader score for the phantom vessel ROIs against the vessel velocity and diameter. Generally, the reader score decreased with increasing vessel velocity, signifying more artifact at higher velocities, as seen in Fig. 9(a). However, the plot demonstrates overlap in reader scores across different velocities, suggesting that velocity is not a unique indicator of motion artifact level. Vessel diameter had a negligible effect on reader score, as shown in Fig. 9(b).

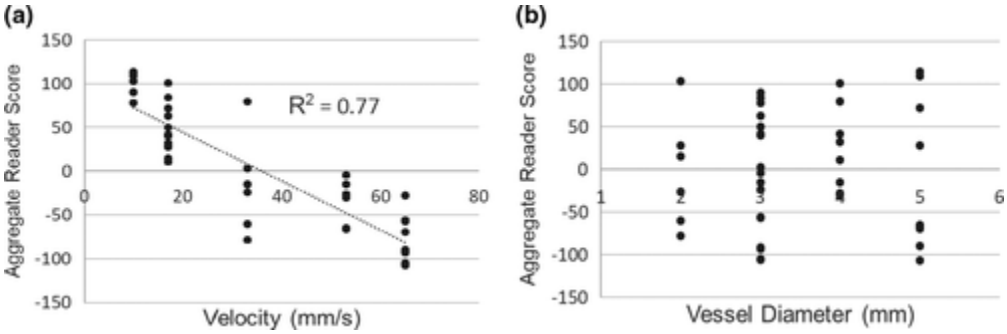

Figure 9 Scatter plots displaying the relationship between the ground truth and the (a) vessel velocity, and (b) diameter.

As described in Section 2.C, the vessel and low-intensity regions were segmented by thresholding, with thresholds T_{LIR} and T_{ves} set to −200 and −600 HU, respectively. An additional investigation was performed to investigate the sensitivity of the metrics to these threshold settings. Figure 10(a) plots the change in FOR, ΔFOR, for a range of threshold settings. ΔFOR was calculated as the difference between the FOR metric calculated at a particular threshold setting to the FOR metric calculated at the reference −200 HU threshold. The FOR metric increased with increasing T_{ves}values, suggesting that a smaller segmented vessel region results in a lower estimation of the motion artifact level. However, these changes were relatively small, with a total change of ~0.07 across the 100 HU threshold range, compared to the total FOR range of 0 to 1. Figure 10(b) similarly plots the change in LIRS metric, ΔLIRS, for a range of settings. As the LIRS threshold T_{LIR} is increased, more pixels are identified as shading artifact, causing an increased estimation of the level of motion artifacts. While the LIRS metric was more sensitive to the threshold setting than the FOR metric, the change in LIRS was relatively robust (±0.1) within a 50 HU range around the selected T_{LIR} threshold level of −600 HU.

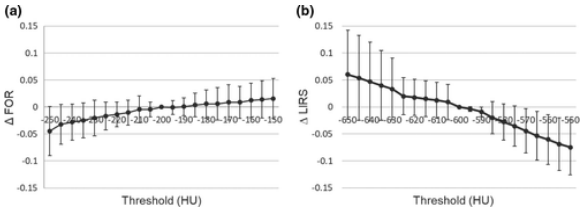

Figure 10 (a) ΔFOR, the average change in FOR metric, is plotted for a range of T_{ves} settings and (b) ΔLIRS, the average change in LIRS metric, is plotted for a range of T_{LIRS} settings. Both plots display the average metric change across the 40 phantom ROIs, with the error bars representing the standard deviation in the change of metric value. For reference, both the FOR and LIRS metrics range from 0 to 1, with increasing metric value corresponding to increasing image quality.

Figure 11 plots the Kendall's Tau coefficients and standard error for all investigated metrics relative to the ground-truth reader score. The Kendall's Tau coefficients of the metrics were: 0.35 (entropy), 0.50 (normalized circularity), 0.82 (positivity), 0.77 (FOR), and 0.77 (LIRS). For comparison, the Kendall's Tau coefficients between the different pairs of readers were 0.81, 0.84, and 0.87.

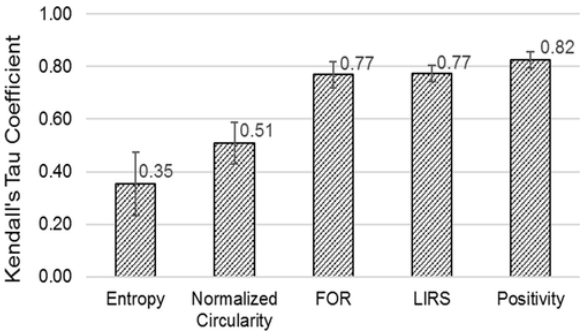

Figure 11 Ranking agreement between the investigated motion artifact metrics and the ground-truth reader score for the phantom study. The error bars represent the standard error.

The FOR, LIRS, and positivity metrics were found to have statistically significantly higher Kendall's Tau agreement than the entropy and normalized circularity metrics ($P < 0.05$). There was no statistically significant difference between the Kendall's Tau values of the FOR, LIRS, and positivity metrics ($P > 0.22$), suggesting that agreement with readers in ranking order is statistically equivalent for these metrics.

Figure 12 displays scatter plots of the investigated metrics plotted against the ground-truth reader score. The results of the linear regression are also displayed on each plot.

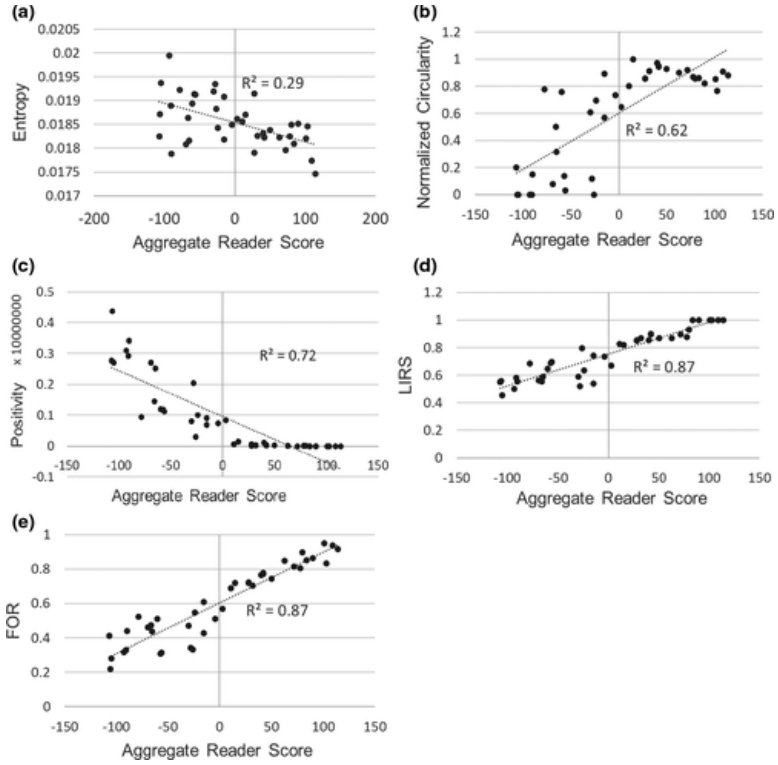

Figure 12 Scatter plots displaying the relationship between the ground-truth aggregate reader score and the investigated metrics (a) entropy, (b) normalized circularity, (c) positivity, (d) LIRS, and (e) FOR. The results of the linear regression are also displayed on each plot.

Because positivity has good ranking relationship with the ground truth but poor linear correlation, a transformed positivity metric (TPOS) was also investigated, where the transformed metric is the fourth root of the positivity metric. Results are shown in Fig. 13.

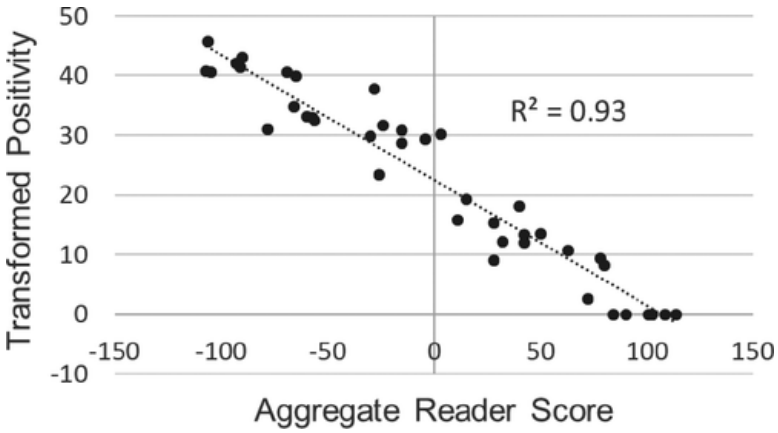

Figure 13 Scatter plot displaying the relationship between the transformed positivity metric and the ground-truth aggregate reader score.

The FOR, LIRS, and transformed positivity metrics were selected for further study on clinical data because they demonstrated both high ranking agreement and linear correlation to the ground-truth reader score for the relatively easier task of quantifying artifact level in the phantom images.

As an example of metric performance, Fig. 14 displays the 40 ROIs sorted by descending values of the FOR metric. While the metric rankings in Fig. 14 are not identical to the reader ranking shown in Fig. 8, both images demonstrate similar ranking trends. The images with higher FOR have vessels that appear more circular with less

shading artifacts. Images with lower FOR generally contain higher motion artifacts, i.e., longer vessel tails and more shading regions.

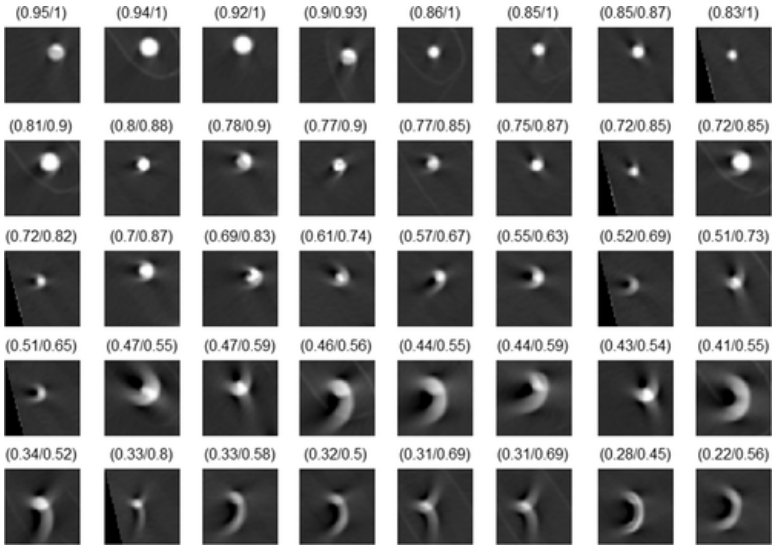

Figure 14 The 40 images used in the reader study displayed in descending order of the FOR metric. The values of the FOR/LIRS metrics are displayed above each image.

3.B. Clinical study results

Figure 15 shows the results of the observer study on clinical data with the ground-truth aggregate reader score displayed on each ROI. The images with higher score demonstrated less artifacts. Similar to Fig. 8, the images are displayed in decreasing order of reader score, i.e., displayed from low to high artifact level.

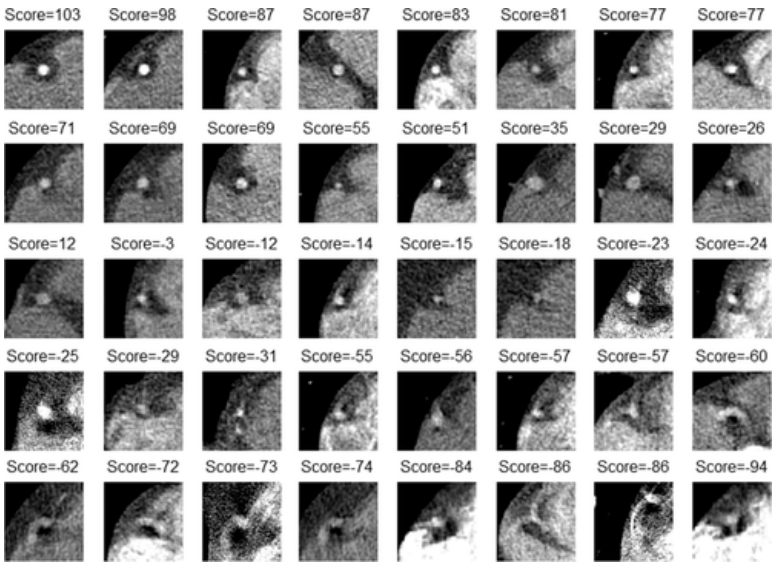

Figure 15 The 40 clinical images used in the reader study displayed in descending order of aggregate reader score. The number displayed above each image is the ground-truth aggregate reader score, which was calculated as the sum of the three reader scores.

Figure 16 plots the Kendall's Tau coefficients representing the agreement between the selected metrics and the ground-truth reader scores for the clinical ROIs. The Kendall's Tau coefficients for the phantom study are also plotted for comparison. The coefficients of the selected metrics were: 0.21 (TPOS), 0.59 (FOR), and 0.53 (LIRS). For comparison, the Kendall's Tau coefficients between the different pairs of readers were 0.65, 0.68, and 0.74.

Figure 17 displays scatter plots of the metrics plotted against the ground-truth reader score. The results of the linear regression are also displayed on each plot.

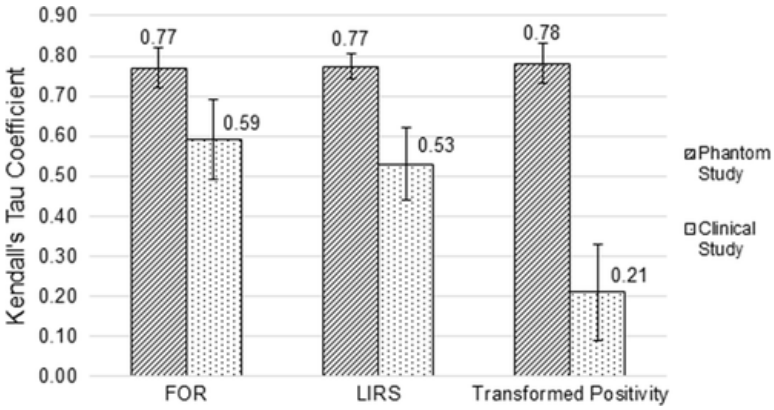

Figure 16 Ranking agreement on clinical images between the selected metrics and the ground-truth scores plotted with standard error. The Kendall's Tau coefficients for the phantom study are also plotted for comparison.

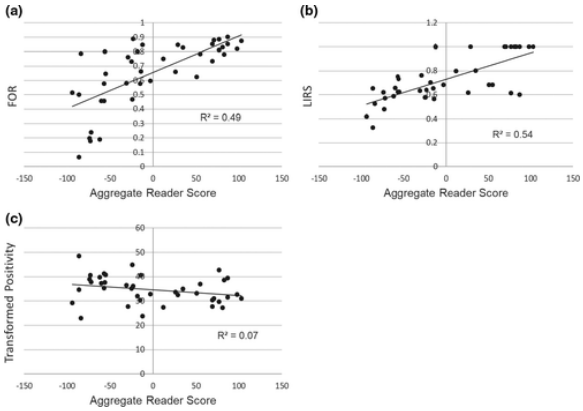

Figure 17 Scatter plots displaying the relationship between the ground-truth aggregate reader score and the single metrics (a) FOR, (b) LIRS, and (c) transformed positivity. The results of the linear regression are also displayed on each plot.

Of the selected metrics, transformed positivity showed weak agreement to the ground-truth reader scores (Kendall's Tau = 0.21, R^2 = 0.07). FOR (Kendall's Tau = 0.59) and LIRS (Kendall's Tau = 0.53) demonstrated statistical significantly higher Kendall's Tau than the transformed positivity metric ($P < 0.05$).

As shown in Fig. 16, the agreement between the metrics and the reader scores was lower for clinical images than phantom data. While transformed positivity demonstrated good performance on phantom data, both the Kendall's Tau coefficient and linear correlation were weaker when applied to clinical data. One potential advantage of positivity is that the metric uses a simple thresholding step to identify regions of low-intensity shading. In phantom data, this thresholding step was successful in identifying regions of low-intensity shading. In the clinical images, the thresholding step erroneously identified some low-intensity pixels from the lung and myocardium as artifact, leading to the low agreement to reader scores.

The FOR and LIRS metrics, which were designed for absolute artifact quantification, demonstrated both good ranking relationship (Kendall's Tau of 0.59 for FOR metric, 0.53 for LIRS metric) and linearity (R^2 of 0.49 for FOR, 0.54 for LIRS) to the ground-truth scores on clinical images. These metrics evaluate complementary motion artifact features. In the phantom images, vessel deformation and shading regions were typically jointly visible. For clinical images, the vessels deformation and low-intensity shading may not always be jointly visible, as can be seen in some of the ROIs in Fig. 7. Therefore, a combination of these two metrics may be beneficial as an overall measure of motion artifact level. A compound metric, Motion Artifact Score (MAS), was defined as the

product of the FOR and LIRS metrics. The Kendall's Tau coefficient of MAS to the ground-truth score was found to be 0.65, which is higher than the individual metric coefficients (0.59 for FOR, 0.53 for LIRS), although this improvement was not statistically significant ($P > 0.25$). Figure 18displays a scatter plot of the MAS metric against the ground-truth reader score. The linear correlation was also higher for the compound MAS metric than for the individual metrics with an $R^2$of 0.64, compared to 0.49 for the FOR metric and 0.54 for LIRS.
Figure 19 displays the 40 ROIs sorted by descending values of the MAS metric which demonstrate similar ranking trends as the reader results in Fig. 15.

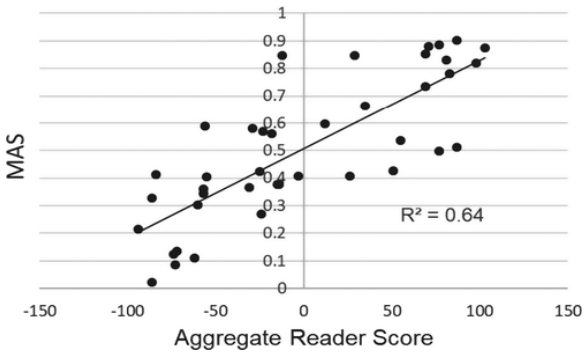

Figure 18 Scatter plot displaying the relationship between the MAS and the ground-truth aggregate reader score.

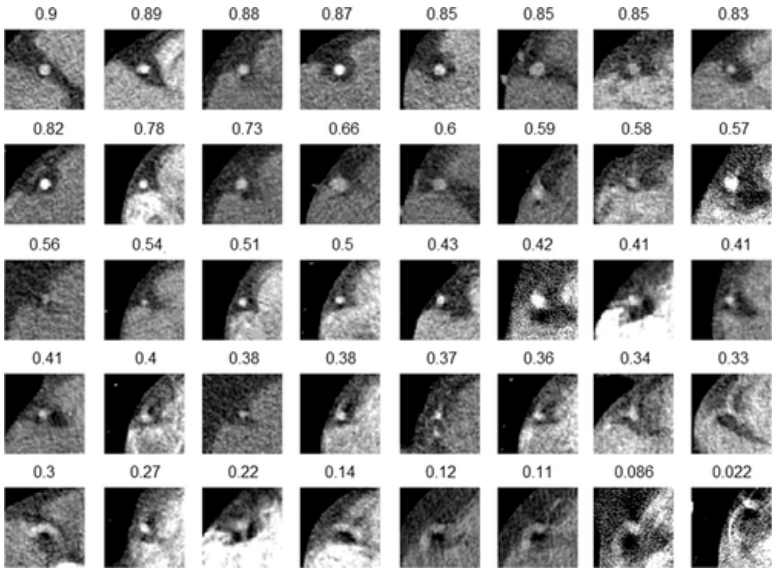

Figure 19 The clinical images used in the reader study displayed in descending order of the MAS metric, with the MAS value displayed above each image.

4 Discussion

This study evaluated continuously valued motion artifact metrics against ground-truth reader scores for both phantom and clinical data. The results of Fig. 5 demonstrate that while motion artifacts generally increase with vessel velocity, velocity does not consistently represent motion artifact severity. Velocity is only one of several factors that affect motion artifacts. For example, the direction of vessel motion relative to the projection direction affects the motion artifact severity.

Kendall's Tau is a measure of ranking agreement, quantifying how often a metric and the readers agree that one image is better than the other. For some applications, such as selecting the lowest-motion phase for display, ranking agreement may be the most important property of a metric. For other applications, such as optimizing motion compensation algorithms, it may be beneficial to have metrics that correlate linearly with the reader

scores. Of the individual metrics investigated in the phantom study, the proposed metrics of FOR and LIRS, along with the previously used metric of positivity, demonstrated the highest ranking agreement with the reader scores (Kendall's Tau >0.75), with statistically similar performance for these three metrics. The positivity metric performance on clinical images was worse than on phantom images due to errors in identifying low-intensity artifacts by the thresholding step. The FOR and LIRS metrics demonstrated good ranking agreement with the reader scores (Kendall's Tau >0.53) and linearity (R^2 > 0.49) on clinical images. One limitation of this study is that different readers evaluated the phantom and clinical images, making it difficult to directly compare the results of the two studies.

A compound metric of Motion Artifact Score, defined as the product of the FOR and LIRS metrics, demonstrated better performance than the individual FOR and LIRS metrics for clinical data (Kendall's Tau = 0.65, R^2 > 0.63). The ranking agreement between the MAS metric and the readers (Kendall's Tau = 0.65) was similar to ranking agreement between readers (average Kendall's Tau = 0.69).

The FOR and LIRS metrics require segmentation of the vessel region. The LIRS metric requires segmentation of the dark shading artifacts, while the positivity metric requires a threshold to identify dark shading artifacts. All metrics require identification of the vessel ROI. In the phantom study, the regions were segmented by simple thresholding. Results demonstrated that the metrics were systematically affected by the threshold settings, but that the changes in FOR or LIRS metrics were relatively small (<0.1) for a 100 HU range of threshold settings. In the study of clinical images, the vessel ROIs, vessel region, and dark shading regions were manually segmented, so that the motion artifact metrics could be evaluated independently of segmentation approaches. Automated identification of vessel regions and automated segmentation will be challenging for clinical images due to noise, contrast dynamics, anatomical structure, and artifacts due to metal and beam hardening. The results of the phantom study suggest that the metrics are sensitive to variability in the segmented regions, therefore robust segmentation algorithms will be needed. Future work is required to develop and validate these segmentation algorithms.

The metrics of vessel deformation used in this study, normalized circularity and FOR, assume that the arteries are circular, requiring images in which the vessels are through-plane in the transverse slices. Algorithms have been previously proposed to identify through-plane vessel regions from the acquired 3D cardiac volume.[10], [11] The FOR metric of vessel symmetry could potentially be modified for in-plane vessels, by folding the vessel region across the main vessel axis.

One limitation of this study is that the metrics were evaluated only for images acquired by axial scanning using wide-cone-beam acquisition on a single scanner model and with filtered backprojection reconstruction. The metrics were not evaluated in the presence of helical artifacts. However, with the advent of scanners with 160 mm of coverage, the entire heart can be captured in a single axial scan. This eliminates the need for helical cardiac scanning and the subsequent risk of helical artifacts. As seen in Figures 6 and 7, the images used for validation contained a wide range of motion artifact presentations, so as to provide a variety of artifacts for validation despite the limitation of evaluation on one scanner. The clinical study used images reconstructed from varying heart rates and anatomical configurations, as well as different pixel spacing, contrast and noise levels to mitigate this limitation.

The motion artifact metrics validated in this work may be useful for comparing cardiac CT protocols, as well as for developing, validating, and comparing motion correction algorithms.[13] The metrics may also be useful for optimization-based reconstruction algorithms that compensate for motion.[11], [12] The metrics may also improve algorithms that select the lowest-motion phase for display.[8]-[10] Future work is needed to investigate the performance of the developed metrics for such clinical applications that require evaluation of motion artifact severity.

5 Conclusions

This study evaluated coronary artery motion artifact metrics using observer studies on phantom images and clinical images. The metrics of Low-Intensity Region Score (LIRS), Fold Overlap Ratio (FOR) and their product, MAS, resulted in the highest agreement in motion artifact ranking when compared to the readers. The FOR, LIRS, and MAS metrics also demonstrated the highest linear correlation to the reader scores. The validated motion artifact metrics may be useful for developing and validating algorithms to reduce motion in Coronary Computed Tomography Angiography (CCTA) images.

Acknowledgment

This study was funded in part by a research grant from GE Healthcare. Biostatistical services were supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Number UL1TR001436. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The authors thank Ali Aleiou, Parag Khobragade, and Wesley Richerson for help with phantom and segmentation observer studies.

Conflicts of interest

Hongfeng Ma received research funding from GE Healthcare during the course of this project. Eric Gros and Darin Okerlund are employees of GE Healthcare. Taly Gilat Schmidt receives research funding from GE Healthcare.

References

1Mowatt G, Cook JA, Hillis GS, et al. 64-Slice computed tomography angiography in the diagnosis and assessment of coronary artery disease: systematic review and meta-analysis. *Heart*. 2008; **94**: 1386– 1393.

2Plank F, Friedrich G, Dichtl W, et al. The diagnostic and prognostic value of coronary CT angiography in asymptomatic high-risk patients: a cohort study. *Open Heart*. 2014; **1**: e000096.

3Gaemperli O, Valenta I, Schepis T, et al. Coronary 64-slice CT angiography predicts outcome in patients with known or suspected coronary artery disease. *Eur Radiol*. 2008; **18**: 1162– 1173.

4Desjardins B, Kazerooni EA. ECG-gated cardiac CT. *Am J Roentgenol*.2004; **182**: 993– 1010.

5Elmquist R. The Impact of 320 Detector Rows: Aquilion ONE in the Pediatric Setting| imagingBiz electronic journal for leaders in medical imaging services. 2009.

6Latif M, Sanchez F, Sayegh K, et al. Initial evaluation of coronary ct angiography image quality on the revolution ct system. 2016.

7Johnson TR, Nikolaou K, Wintersperger BJ, et al. Dual-source CT cardiac imaging: initial experience. *Eur Radiol*. 2006; **16**: 1409– 1415.

8Wick CA, McClellan JH, Arepalli CD, et al. Characterization of cardiac quiescence from retrospective cardiac computed tomography using a correlation-based phase-to-phase deviation measure. *Med Phys*. 2015; **42**: 983– 993.

9Manzke R, Köhler T, Nielsen T, Hawkes D, Grass M. Automatic phase determination for retrospectively gated cardiac CT. *Med Phys*.2004; **31**: 3345– 3362.

10Stassi D, Dutta S, Ma H, et al. Automated selection of the optimal cardiac phase for single-beat coronary CT angiography reconstruction. *Med Phys*. 2016; **43**: 324– 335.

11Rohkohl C, Bruder H, Stierstorfer K, Flohr T. Improving best-phase image quality in cardiac CT by motion correction with MAM optimization. *Med Phys*. 2013; **40**: 031901.

12Iatrou M, Pack JD, Bhagalia R, Beque D, Seamans J. 2010.Coronary artery motion estimation and compensation: a feasibility study. In IEEE Nuclear Science Symposuim & Medical Imaging Conference.
13Tang Q, Cammin J, Srivastava S, Taguchi K. A fully four-dimensional, iterative motion estimation and compensation method for cardiac CT. *Med Phys*. 2012; **39**: 4291– 4305.
14Kim S, Chang Y, Ra JB. Cardiac motion correction based on partial angle reconstructed images in x-ray CT. *Med Phys*. 2015; **42**: 2560– 2571.
15Kyriakou Y, Lapp RM, Hillebrand L, Ertel D, Kalender WA.Simultaneous misalignment correction for approximate circular cone-beam computed tomography. *Phys Med Biol*. 2008; **53**: 6267.
16Parzen E. On estimation of a probability density function and mode. *Ann Math Stat*. 1962; **33**: 1065– 1076.
17Husmann L, Leschka S, Desbiolles L, et al. Coronary artery motion and cardiac phases: dependency on heart rate—implications for CT image reconstruction 1. *Radiology*. 2007; **245**: 567– 576.
18Phelps AS, Naeger DM, Courtier JL, et al. Pairwise comparison versus Likert scale for biomedical image assessment. *Am J Roentgenol*. 2015; **204**: 8– 14.
19Kendall MG. A new measure of rank correlation. *Biometrika*.1938; **30**: 81– 93.
20Hollander M, Wolfe DA, Chicken E. *Nonparametric Statistical Methods*. Hoboken, NJ: John Wiley & Sons; 2013.
21Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging*. 2004; **23**: 903– 921.
22Contijoch F, Stayman JW, Mcveigh ER. The impact of small motion on the visualization of coronary vessels and lesions in cardiac CT: a simulation study. *Med Phys*. 2017; **44**: 3512– 3524.
23Vembar M, Garcia MJ, Heuscher DJ, et al. A dynamic approach to identifying desired physiological phases for cardiac imaging using multislice spiral CT. *Med Phys*. 2003; **30**: 1683– 1693.