Automatic stent detection in intravascular OCT images using bagged decision trees

Hong Lu,¹ Madhusudhana Gargesha,¹ Zhao Wang,¹ Daniel Chamie,² Guilherme F. Attizzani,² Tomoaki Kanaya,² Soumya Ray,³ Marco A. Costa,² Andrew M. Rollins,¹ Hiram G. Bezerra,² and David L. Wilson¹,4,*

¹Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, 44106, USA
²Cardiovascular Imaging Core Laboratory, Harrington Heart & Vascular Institute, University Hospitals Case Medical Center, Cleveland, OH, 44106, USA
³Department of Electrical Engineering & Computer Science, Case Western Reserve University Cleveland, OH, 44106, USA
⁴Department of Radiology, Case Western Reserve University, Cleveland, OH, 44106, USA
*david.wilson@case.edu

Abstract: Intravascular optical coherence tomography (iOCT) is being used to assess viability of new coronary artery stent designs. We developed a highly automated method for detecting stent struts and measuring tissue coverage. We trained a bagged decision trees classifier to classify candidate struts using features extracted from the images. With 12 best features identified by forward selection, recall (precision) were 90%–94% (85% – 90%). Including struts deemed insufficiently bright for manual analysis, precision improved to 94%. Strut detection statistics approached variability of manual analysis. Differences between manual and automatic area measurements were 0.12 ± 0.20 mm² and 0.11 ± 0.20 mm² for stent and tissue areas, respectively. With proposed algorithms, analyst time per stent should significantly reduce from the 6–16 hours now required.

© 2012 Optical Society of America

OCIS codes: (100.0100) Image processing; (110.4500) Optical coherence tomography.

References and links

1. M. A. Beijk, M. Klomp, N. J. Verouden, N. van Geloven, K. T. Koch, J. P. Henriques, J. Baan, M. M. Vis, E. Scheunhage, J. J. Pick, J. G. Tijssen, and R. J. de Winter, “Genous endothelial progenitor cell capturing stent vs. the Taxus Liberté stent in patients with de novo coronary lesions with a high-risk of coronary restenosis: a randomized, single-centre, pilot study,” Eur. Heart J. 31(9), 1055–1064 (2010).
2. J. F. Granada, S. Inami, M. S. Aboodi, A. Tellez, K. Milewski, D. Wallace-Bradley, S. Parker, S. Rowland, G. Nakazawa, M. Vorpahl, F. D. Kolodgie, G. L. Kaluza, M. B. Leon, and R. Virmani, “Development of a novel prohealing stent designed to deliver sirolimus from a biodegradable abluminal matrix,” Circ Cardiovasc Interv 3(3), 257–266 (2010).
3. H. G. Bezerra, M. A. Costa, G. Guagliumi, A. M. Rollins, and D. I. Simon, “Intracoronary optical coherence tomography: a comprehensive review clinical and research applications,” JACC Cardiovasc. Interv. 2(11), 1035–1046 (2009).
4. G. Guagliumi, G. Musumeci, V. Sirbu, H. G. Bezerra, N. Suzuki, L. Fiocea, A. Matiashvili, N. Lortkipanidze, A. Trivisonno, O. Valsecchi, G. Biondi-Zoccai, and M. A. Costa; ODESSA Trial Investigators, “Optical coherence tomography assessment of in vivo vascular response after implantation of overlapping bare-metal and drug-eluting stents,” JACC Cardiovasc. Interv. 3(5), 531–539 (2010).
5. G. Guagliumi, V. Sirbu, G. Musumeci, H. G. Bezerra, A. Aprile, H. Kyono, L. Fiocea, A. Matiashvili, N. Lortkipanidze, A. Vassileva, J. J. Popma, D. J. Allocco, K. D. Dawkins, O. Valsecchi, and M. A. Costa, “Strut coverage and vessel response to a new-generation paclitaxel-eluting stent with an ultrathin biodegradable abluminal polymer: Optical Coherence Tomography Drug-Eluting Stent Investigation (OCTDESI),” Circ Cardiovasc Interv 3(4), 367–375 (2010).
6. H. Kyono, G. Guagliumi, V. Sirbu, N. Rosenthal, S. Tahara, G. Musumeci, A. Trivisonno, H. G. Bezerra, and M. A. Costa, “Optical coherence tomography (OCT) strut-level analysis of drug-eluting stents (DES) in human coronary bifurcations,” EuroIntervention 6(1), 69–77 (2010).
7. S. Tahara, H. G. Bezerra, V. Sirbu, H. Kyono, G. Musumeci, N. Rosenthal, G. Guagliumi, and M. A. Costa, “Angiographic, IVUS and OCT evaluation of the long-term impact of coronary disease severity at the site of...
overlapping drug-eluting and bare metal stents: a substudy of the ODESSA trial,” Heart 96(19), 1574–1578 (2010).

8. S. Tahara, D. Chamić, M. Baihars, C. Alraies, and M. Costa, “Optical coherence tomography endpoints in stent clinical investigations: strut coverage,” Int. J. Cardiovasc. Imaging 27(2), 271–287 (2011).

9. G. Guagliumi, V. Sirbu, H. Bezerra, G. Biondi-Zoccai, L. Fiocca, G. Musumeci, A. Matiashvili, N. Lortkipanidze, S. Tahara, O. Valsecchi, and M. Costa, “Strut coverage and vessel wall response to zotarolimus-eluting and bare-metal stents implanted in patients with ST-segment elevation myocardial infarction: the OCTAMI (Optical Coherence Tomography in Acute Myocardial Infarction) Study,” JACC Cardiovasc. Interv. 3(6), 680–687 (2010).

10. G. Guagliumi, M. A. Costa, V. Sirbu, G. Musumeci, H. G. Bezerra, N. Suzuki, A. Matiashvili, N. Lortkipanidze, L. Mihailescu, A. Trivisonno, O. Valsecchi, G. S. Mintz, O. Dressler, H. Parise, A. Maehara, E. Cristea, A. J. Lansky, R. Mehran, and G. W. Stone, “Strut coverage and late malapposition with paclitaxel-eluting stents compared with bare metal stents in acute myocardial infarction: optical coherence tomography substudy of the Harmonizing Outcomes with Revascularization and Stents in Acute Myocardial Infarction (HORIZONS-AMI) Trial,” Circulation 123(3), 274–281 (2011).

11. G. T. Bonnema, K. O. Cardinal, S. K. Williams, and J. K. Barton, “An automatic algorithm for detecting stent endothelialization from volumetric optical coherence tomography datasets,” Phys. Med. Biol. 53(12), 3083–3098 (2008).

12. C. Xu, J. M. Schmitt, T. Akasaka, T. Kubo, and K. Huang, “Automatic detection of strut struts with thick neo-intimal growth in intravascular optical coherence tomography image sequences,” Phys. Med. Biol. 56(20), 6665–6675 (2011).

13. S. Gurmeric, G. G. Isguder, S. Carlier, and G. Unal, “A new 3-D automated computational method to evaluate in-stent neo-intimal hyperplasia in in-vivo intravascular optical coherence tomography pullbacks,” Med Image Comput Comput Assist Interv 12(Pt 2), 776–785 (2009).

14. G. J. Ughi, T. Adriaenssens, K. Onsea, P. Kayaert, C. Dubois, P. Sinnaeve, M. Coosemans, W. Desmet, and J. D’hooge, “Automatic segmentation of in-vivo intra-coronary optical coherence tomography images to assess stent strut apposition and coverage,” Int. J. Cardiovasc. Imaging 28(2), 229–241 (2012).

15. C. Kauffmann, P. Motreff, and L. Sarry, “In vivo supervised analysis of stent reendothelialization from optical coherence tomography,” IEEE Trans. Med. Imaging 29(3), 807–818 (2010).

16. A. Wang, J. Eggemont, N. Dekker, H. M. Garcia-Garcia, R. Pawar, J. H. C. Reiber, and J. Dijkstra, “Automatic stent strut detection in intravascular optical coherence tomographic pullback runs,” Int. J. Cardiovasc. Imaging. (2012).

17. N. Bruining, K. Sihan, J. Ligthart, P. Cummins, S. De Winter, and E. Regar, “Automated three-dimensional detection of intracoronary stent struts in optical coherence tomography images,” J. Am. Coll. Cardiol. 58(20), B181 (2011).

18. S. Tsantis, C. G. Kagadis, K. Katsanos, D. Karnabatidis, G. Bourantas, and G. C. Nikiforidis, “Automatic vessel lumen segmentation and strut strut detection in intravascular optical coherence tomography,” Med. Phys. 39(1), 503–513 (2012).

19. Z. Wang, H. Kyono, H. G. Bezerra, H. Wang, M. Gargesha, C. Alraies, C. Xu, J. M. Schmitt, D. L. Wilson, M. A. Costa, and A. M. Rollins, “Semiautomatic segmentation and quantification of calcified plaques in intracoronary optical coherence tomography images,” J. Biomed. Opt. 15(6), 061711 (2010).

20. D. Chamić, Z. Wang, H. Bezerra, A. M. Rollins, and M. A. Costa, “Optical coherence tomography and fibrous cap characterization,” Curr Cardiovasc Imaging Rep 4(4), 276–283 (2011).

21. Z. Wang, D. Chamić, H. Bezerra, D. L. Wilson, M. Costa, and A. M. Rollins, “Three-dimensional volumetric quantification of fibrous caps using intravascular optical coherence tomography,” Proc. SPIE 8213, 8213-34 (2012).

22. Z. Wang, D. Chamić, H. G. Bezerra, H. Yamamoto, J. Kanovsky, D. L. Wilson, M. A. Costa, and A. M. Rollins, “Volumetric quantification of fibrous caps using intravascular optical coherence tomography,” Biomed. Opt. Express 3(6), 1413–1426 (2012).

23. Pierre Soille, “Geodesic transformations” in Morphological Image Analysis: Principles and Applications, 2nd ed. (Springer-Verlag, 2004).

24. L. Breiman, “Bagging predictors,” Mach. Learn. 24(2), 123–140 (1996).

25. P. Bühlmann and B. Yu, “Analyzing bagging,” Ann. Stat. 30(4), 927–961 (2002).

26. N. Y. Graham, “Smoothing with periodic cubic splines,” Bell Syst. Tech. J. 62, 101–110 (1983).

27. E. T. Y. Lee, “Choosing nodes in parametric curve interpolation,” Comput. Aided Des. 21(6), 363–370 (1989).

28. L. Rokach and O. Maimon, “Feature selection” in Data Mining with Decision trees: Theory and Application (World Scientific, 2008).

1. Introduction

Cardiovascular disease is the leading cause of death worldwide. Stent implantation by means of percutaneous coronary intervention is the most common coronary revascularization procedure, and approximately 2 million people worldwide receive stent implantation each year. To minimize rates of restenosis, there is a high prevalence of drug eluting stent usage.
worldwide. However, safety concerns exist, particularly the risk of late stent thrombosis, a rare clinical condition that raises great concern due to its high associated morbidity and mortality. Pathological studies have suggested that the absence of stent strut coverage due to delayed vascular healing is a potential surrogate metric for risk of stent thrombosis. New stent designs aim to aid appropriate vascular healing. For example, new stents from OrbusNeich Medical Technologies include anti-CD34 antibodies to aid capture of EPCs [1] or anti-CD34 antibodies combined with anti-proliferative abluminal sirolimus elution [2].

To optimize device designs and drive the field forward, sensitive, in vivo assessments are needed for serial preclinical studies and for clinical evaluations. Intravascular Optical Coherence Tomography (iOCT) is the only imaging modality with the resolution and contrast necessary to enable accurate measurements of luminal architecture and neointima stent coverage [3]. Strut tissue coverage as assessed by iOCT has become an important surrogate biomarker of stent viability [4–10]. The Cardiovascular Imaging Core Lab in the Harrington Heart & Vascular Institute, University Hospitals Case Medical Center, Cleveland, Ohio, hereafter called Core Lab, has provided iOCT image analysis service to >20 international trials of stent devices. A well trained “image analyst” typically takes 6-16 hours to analyze manually a single stent pullback, containing 100–200 frames over the length of the stent with about 9 struts per frame, limiting the size and number of studies that can be performed. In addition to stent device trials, there is a need to provide analysis for treatment decisions. For example, with fast software, it would be possible to present the number and location of malapposed struts in 3D, providing instant feedback on the potential need for a second intervention. iOCT could also play a role in drug management; e.g. if a stent is fully covered, then anti-platelet therapy might be minimized. Alternatively, high numbers of uncovered struts have been related to stent thrombosis, and a patient under this condition may require a prolonged anti-platelet regime. Especially, for these live-time clinical applications to become a reality, fast, reproducible stent analysis will be required. Clearly, highly automated image analysis software will be key for full realization of iOCT stent imaging.

There are early reports of software for analysis of iOCT stent images. The most obvious feature of metallic stent struts is a bright reflection followed by a dark shadow, and strut detection approaches incorporate this observation and more. In most algorithms, authors devised image processing schemes and manually optimized parameters [11–16]. For example, Bonnema et al. used thresholds for strut reflection, shadow darkness, and “concentrated energy” along single A-lines to detect stent struts [11]. Xu et al. employed an improved ridge detector using a steerable filter to detect struts with thick tissue coverage [12]. Gurmeric et al. used angular intensity distribution of the image to identify shadows and detected the brightest pixels in the shadow regions [13]. Ughi et al. applied thresholds for peak intensity, shadow intensity and speed of intensity rising and falling to define a strut [14]. Kauffmann et al. combined morphological, gradient and symmetry operators together with active contour models to detect struts [15]. Wang et al. detected the brightest pixel along each A-line and clustered these candidate pixels using edges identified by Prewitt compass filters [16]. There are two reports of feature extraction and classification methods. Bruining et al. used a basic set of features (mean, maximum, sum of values above mean) of each A-line and performed feature-based classification using a k-nearest neighbor classifier [17]. Tsantis et al. detected struts using wavelet features and probabilistic neural networks [18]. To determine stent cross sectional areas for analysis of tissue coverage, splines [13] and ellipsoids [12] have been used to connect detected struts. Together, these reports encourage the further development of much needed automated analysis methods. There were limitations in these early reports. Some with particularly good results used ex vivo imaging of tissue engineered vessels or in vivo imaging of femoral artery [11,18], rather than coronary arteries with probably worse image quality and more heterogeneous implantation. Some features were based on single A-lines, rather than capturing the 2D nature of a strut; some analyzed a limited number of cases; error analysis did not always identify false positive and false negative stent strut detection; very few compared
software errors to variation among analysts; many did not use some obvious information such as “stents tend to be near the lumen;” none clearly distinguished between bright and non-bright struts; and some methods were not clearly extensible. We have addressed these issues in our study. A comprehensive comparison of processing results is given in the Discussion.

Our group has been making detailed manual measurements in the Core Lab and developing highly automated software for analyzing iOCT images, especially for assessing vulnerable plaques [19–22] and stents. Here, focusing on stents, we developed and evaluated performance of an algorithm using machine learning classification to detect stent struts and a new contour identification method to measure tissue coverage area in stents. Bagged decision trees classifier was used because it is less sensitive to noise as compared to standard decision tree, giving improved accuracy and stability. A machine learning, classification approach should be ideal for strut detection. It allows one to extract multiple, physically meaningful features and train the classifier on 1000s of manually detected struts. In this way, we should avoid bias that appears with the use of manually developed image processing heuristics, necessarily considering many fewer cases. We calculated intensity statistics of the candidate strut region and shadow region separately and used both sets of features to detect strut locations. Other researchers used classification to simply find the A-line containing a strut. Features from both the strut and shadow were selected early in the feature selecting process, indicating their importance. Although overtraining can be an issue in machine learning, we applied standard methods to remove redundant features and limit over training. A new method was proposed for extracting the stent contour and tissue coverage area. We used periodic cubic spline to reconstruct stent contour, which allows local flexibility while maintaining a nearly circular contour. One potential criterion for accepting a software solution is that its performance should lie within the variability among human observers. With this in mind, multiple manual analyses were done on the same iOCT pullbacks allowing us to compare detection performance of software against inter-analyst variation. We also compared our new contouring method to stent and tissue coverage areas measured by analysts.

2. Materials

Images were collected by a Fourier-Domain OCT (FD-OCT) system (C7-XR™ OCT Intravascular Imaging System, St. Jude Medical, St. Paul, Minnesota). The system was equipped with a tunable laser light source sweeping from 1250 nm to 1370 nm, providing 15-μm resolution along the A-line. Pullback speed was 20 mm/sec over a distance of 54.2 mm, and the interval between frames was 200 μm, giving 271 total frames. Stents were imaged over 100 to 200 frames, depending upon the length of the stent. All iOCT images were acquired from the database in the Core Lab. We analyzed 508 iOCT images and 4392 struts taken from 12 pullbacks. Of these, 6 were baseline cases taken immediately after stenting and 6 were follow-up studies occurring at 2-18 months following implantation. Each polar-coordinate \((r, \theta)\) image consisted of 504 A-lines, 972 pixels along the A-line, and 16 bits of gray-scale data. These data were log transformed to a floating point data type for automatic image analysis.

3. Image analysis algorithms

We developed algorithms for detecting stent struts and for measuring the area of tissue covering the stent. In iOCT images, stent struts often give a bright reflection with a shadow behind it. In other cases, reflected light from the strut is not detected, mostly due to the orientation of the strut wire, and only the shadow is evident. We call these bright, analyzable struts and non-bright struts, respectively. These definitions are consistent with manual analysis in the Core Lab. In the case of a bright, analyzable strut, the front surface of the strut will occur near the brightest point in the reflection, allowing one to accurately assess tissue thickness covering a single strut. With non-bright struts, since there is some ambiguity as to...
the location of the strut, they are not used to measure strut-level tissue coverage in the Core Lab. Below, we describe our method for detecting bright analyzeable struts.

3.1. Detect bright, analyzeable stent struts

Our iOCT stent strut detection algorithm consists of multiple steps: (1) detect the expanded lumen boundary; (2) detect A-lines containing a shadow; (3) detect bright spots; logical AND of steps 1-3 giving candidate struts; (4) compute features from candidate struts; (5) classify candidate struts as either struts or else, using a bagged decision trees classifier trained on a large data set; and (6) eliminate extra hits using a simple rule. All processing is done on polar coordinate \((r, \theta)\) iOCT images. This view is geometrically transformed to create the anatomical \((x, y)\) view for visualization. Figure 1 gives an overview of steps.

![Fig. 1. Classification-based stent strut detection algorithm.](image)

We apply Steps 1-3 to obtain a large number of purposely “over called” candidate struts. To determine the lumen, we use a dynamic programming method described previously by our group [19]. Briefly, in polar \((r, \theta)\) coordinates, we detect edges along \(r\) and then use dynamic programming to find the lumen contour having the highest cumulative edge strength from top to bottom along \(\theta\). The guide wire gives a very bright reflection and very dark shadow which obscures the lumen boundary and any stent materials behind it. We determine A-lines corresponding to the guide wire shadow and make all pixels zero. This effectively makes the guide wire a “don’t care” region which is easily bridged by the dynamic programming method for obtaining the lumen contour.

Figure 2 shows the process of shadow detection. An intensity at each angle is calculated by summing a predetermined number, \(SL\), of pixels after the lumen border along each A-line in the \((r, \theta)\) image. We then detected the extended minima [23] of this 1D intensity profile to...
Fig. 2. Shadow detection. (A) Intensity profile obtained by summing a predetermined number of pixels, \( SL \), after the lumen border along A-lines in (C). (B) Intensity minima indicative of a shadow are shown (red solid line). The dashed curve is the H-minima transformation, which suppresses minima having a depth \(<TD1\). Intensity minima obtained in (B) are used to generate the shadow mask in (D), where white bands indicate A-lines containing detected shadows. Note that even very thin shadows in the input image (C) are accurately detected. Parameters are given in the text.

determine A-lines having a shadow. There is a threshold parameter \( TD1 \) for this operation, which is the difference between the negative peaks and their neighborhood.

We next identify bright spots which might correspond to reflections from struts, using a morphological extended maxima detection algorithm [23]. Briefly, regional maxima detection is performed on the H-maxima transformation (Fig. 3B) of the image to detect extended maxima (Fig. 3C). A single parameter, \( TD2 \), is the threshold for gray-scale difference between the bright spots and their neighborhood. To eliminate some irrelevant bright spots,

Fig. 3. Bright spot detection and expanded lumen boundary ROI. (A) Input \((r, \theta)\) image. (B) Image after h-maxima transformation with all the maxima whose depth is lower than \( TD2 \) suppressed, leaving a smoother image suitable for regional maxima detection. (C) Overlay of raw image, ROI mask (region between green lines), and extended maxima (in red) detected by taking the regional maxima of (B). (D) Candidate struts left after the logical AND of ROI mask, shadow mask, and extended maxima.
we identify a region of interest (ROI) of width \( W_{ps} \) centered on the lumen boundary, where struts should occur (Fig. 3C). A logical AND of a ROI mask, bright spots, and the shadow mask gives candidate struts in an image (Fig. 3D). Particularly in this case with residual blood there are a large number of candidate struts to be further processed and classified.

In Step 4, we compute image features for identification of bright analyzable struts. As shown in Table 1, features are divided into 3 categories: intensities and shape of the bright spot, intensities of the shadow region, and combinations from both regions. The shadow region is a rectangular region following the bright spot, with its center lying along the same line as the centroid of the bright spot. The width of the shadow is determined by the shadow detection algorithm. Shadow length, \( SL \), is a parameter determined by observation. Table 1 lists all features, and bolded ones are those remaining following the feature selection process described later.

Intuitively satisfying features are inspired from manual observations on a large number of candidate struts. For example, since the strut reflection spot is usually bright and the shadow is usually dark, intensity statistics (features 1-5, 8-12) should be very discriminative. Solidity (feature 6) is area/convex area (convex area is the smallest convex polygon that can contain the region), which should be different for strut and non-strut considering their different shapes of bright spot. A strut reflection spot is only a few pixels, so the area of the bright spot (feature 7) is a useful feature. Percentage of dark area (feature 13) is the percentage of dark pixels with values below a predetermined threshold of \( DTh \) in the shadow region. The mean of the dark area (feature 14) is the gray-scale mean of these dark pixels. Struts have a higher percentage of dark area and lower mean of dark area, than non-strut bright spots. Residual blood in the lumen is not a source of error, since it usually doesn’t cause a detectable shadow. However, some “candidate struts” were due to reflections from residual blood in front of a real strut, where the real strut behind it will give a large value for the maximum intensity of the shadow region. In this case, the difference between the two maximum intensities (feature 15) is more informative than the maximum of the bright spot or shadow regions alone. Slope of the intensity profile (feature 16) is the change in intensity from the brightest pixel of the bright spot to the 30th dark pixel in the shadow region. The 30th dark pixel is chosen to avoid dark pixels in the noise. A bright spot from strut reflection followed by shadow should have a steeper slope and higher percentage of dark pixels along the slope (feature 17) than non-strut bright spot without shadow behind it.

Table 1. Features for classification

| Strut Region Features | Shadow Region Features | Combination Features |
|-----------------------|-----------------------|---------------------|
| 1. Maximum Intensity  | 8. Maximum Intensity  | 15. Difference between Two Maxima |
| 2. Minimum Intensity  | 9. Minimum Intensity  | 16. Slope of Intensity Profile |
| 3. Mean Intensity     | 10. Mean Intensity    | 17. Percentage of Dark Pixels along Slope |
| 4. Median Intensity   | 11. Median Intensity  |                     |
| 5. Intensity Variance | 12. Intensity Variance|                     |
| 6. Solidity           | 13. Percentage of Dark Area |                     |
| 7. Area               | 14. Mean of Dark Area |                     |

Features in bold face were the ones finally chosen to use in the algorithm. See Experimental Methods for a description of the feature selection process and Results for more details.

In Step 5, we use bagged decision trees to classify candidate struts as analyzable bright struts or else. This popular classification technique is reported to be less sensitive to noise than standard single decision tree, giving improved accuracy of classification. A bagged decision trees classifier creates bootstrapped replicas of the training data set and separate decision trees are trained on each replica to create an ensemble. Bagging reduces the variance of noisy predictions from data and improves the stability of the classifier [24,25]. For a binary decision, a majority vote is taken from the output of the single decision trees. If \( \geq 50\% \) of trees
vote that the candidate is a bright strut, it is marked accordingly. Feature selection, and training and validation experiments are described later. The optimal number of decision trees for our experiment was set to be 20 by a trial-and-error process with consideration to detection performance statistics described later.

Following classification, we further process results to eliminate “multiple hits” (Step 6). Occasionally, extra bright spots are found along the shadow of a strut. Typically, this occurs when there is a bright reflection alongside the shadow, due to a reflecting object such as a foam cell or calcification, or from extra reflection echoes. In these instances, extra bright spots were eliminated by keeping only the brightest group of pixels, leaving the final detected struts.

3.2. Determination of stent and tissue coverage areas

To obtain tissue coverage area, we compute the difference in areas bounded by the stent and lumen contours. The stent contour is obtained by smoothly connecting detected struts with a nearly circular curve. A 2D periodic cubic spline curve satisfactorily estimates stent shape [26,27]. Nevertheless, problems occur when too few struts are detected in a frame. When there is no strut detected in half of the frame, we omit the frame. When there is no strut in a quarter of the frame, we add an “interpolation point” by linearly interpolating the distances to lumen of the detected struts in the neighborhood. We refine lumen segmentation by masking out all A-lines containing detected struts in the raw image and reapplying the dynamic programming algorithm with these A-lines having zero intensity value.

4. Experimental methods

4.1. Manual “ground truth”

Manually analyzed ground truth image data were obtained from expert analysts in the Cardiac Imaging Core Lab. In a subset of data, three analysts, blinded to each other, used manual segmentation tools in Amira (www.visageimaging.com) to annotate stent struts from six iOCT pullback image sequences - three baseline and three follow-up cases. These 6 cases were used to validate automatic detection and to analyze inter-analyst variability, so as to provide a benchmark for computer detection accuracy. Inter-analyst variability was assessed as follows. There was agreement in marking a strut if the 2D Euclidean distance was less than a tolerance, Tol-1 = 95 µm. Results were insensitive to this parameter. We created three groupings of the 3 analysts, with each analyst taking turn as the “gold standard.” We then determined agreement between each of the other two with the gold standard. In this way, we obtained 6 measurements of the number of true positives (TPs), false positives (FPs), and false negatives (FNs). It is understood that groupings are not independent; i.e., precision (recall) obtained by comparing Analyst-1 to Analyst-2 equals recall (precision) of Analyst-2 to Analyst-1. Additional cases were annotated by single analysts using the image analysis software integrated in the OCT imaging system C7-XR. In this latter case, we also compared automated versus manually-determined tissue coverage areas.

In cases having 3 analysts, we faced a conundrum when trying to determine a gold standard for comparison to our software, because there was imperfect agreement of analysts. The major source of variability among analysts is that they have different thresholds for labeling a strut as bright versus non-bright. We considered using a majority vote to ascertain detected bright struts, but instead determined to use the most experienced reader, Analyst-1, as the gold standard. This analyst labeled more stent struts to be “bright and analyzable” than the other two, thereby nudging classification software towards more aggressive labeling of bright struts. We trained and validated our classification software against bright struts. However, because many FPs have shadows with a small bright spot, we also compared results against the aggregate of bright and non-bright struts.
4.2. Classifier training and validation

We applied two different training/validation paradigms. First, we applied a 5-fold cross validation across all “pooled” images from either baseline or follow-up cases. (Separating baseline and follow-up gave superior results to those obtained with combined image data, mostly because there is no tissue coverage in baseline cases.) Second, to further ensure generality, we did a leave-one-stent-out cross validation. To form positive and negative examples for training, we identified bright analyzable struts from “ground truth” data and called spatially overlapping candidate struts positive examples. Candidate struts not identified in the gold standard data were deemed negative examples. For both 5-fold validation on pooled data and leave-one-stent out, we computed detection statistics listed below. Note that we used precision which is more meaningful than specificity which is sensitive to the number of TNs. In this problem we could consider TNs to include all pixels not containing a strut, a rather meaningless number.

\[
\text{Recall} = \frac{TP}{TP + FN}, \quad (1) \\
\text{Precision} = \frac{TP}{TP + FP}, \quad (2) \\
F = 2 \times \left( \frac{PR \times RC}{PR + PC} \right). \quad (3)
\]

Recall (RC), or sensitivity, is a measure of the percentage of correctly detected struts of all the true, manually annotated struts. Precision (PR) is the percentage of correctly detected struts of all the predicted struts. The F score combines precision and recall, giving a scalar value to optimize during feature selection. Statistics were computed for each validation set in turn, giving for example, 5 validation results for the 5-fold cross validation. Means and standard deviations are reported.

4.3. Feature selection

Overtraining is a well-known problem in machine learning. Especially with a large number of features and limited data, it is possible for the classifier to discriminate the training data very well but not generalize to validation data, actually leading to degradation in validation statistics. We performed a forward feature selection technique to find the most discriminative features [28]. Briefly, we began by testing 17 feature subsets, each containing one feature and found the subset with the best performance on validation data. We then evaluated 16 feature subsets each consisting of 2 features: the best feature found in the first step and one of the remaining 16 features. The best feature subset containing 2 features was kept. This process continued with feature subsets containing 3, 4, and more features. The process ended when the performance stopped improving or started to degrade. Since this process was time consuming, we used one stent data set. Classification performance varied little among different stents, indicating that feature selection was not biased to the data set used for feature selection.

5. Results

5.1. Parameter settings

We performed experiments to decide the few parameters in the algorithm. In candidate strut detection, we set $TD1$ and $TD2$ very low and set $W_{ps}$ very large to purposely overcall and ensure a recall higher than 95%. These parameter settings resulted in a large number of false positives which were removed in the classification step. Final results were not sensitive to $TD1$, $TD2$, and $W_{ps}$ values over a large range. Feature extraction included two manually adjusted parameters, $SL$ and $DTh$. Again, these were chosen with consideration to many randomly chosen image frames. $DTh = 0.5$ is the noise level for OCT data ranging 0-255, and $SL = 0.5$ mm, after which the shadow gets uniform and no new statistic information can be
obtained. These observed basic properties apply to any iOCT images we have seen. In general, results were insensitive to parameters. Since stent image volumes were randomly selected from the Core Lab database containing studies from multiple sites around the world, very probably, future parameter adjustment will be unnecessary unless arteries, stents, or instruments change significantly.

5.2. Feature selection

Result of forward feature selection is shown in Fig. 4. Precision, recall, and F score improve dramatically as the first few features are added and continue to improve gradually with the addition of more features. We chose to use the first 12 features, which gave a good trade-off between precision and recall. In addition, we found that these 12 features gave good results in the leave-one-stent-out experiment. The 12 features chosen are: maximum intensity, mean intensity, median intensity, solidity and area of bright spot, mean intensity, intensity variance, percentage of dark area and mean of dark area of shadow region; difference between maximum intensity of bright spot and maximum intensity of shadow, slope of intensity profile, percentage of dark pixels along slope. Table 1 lists all of the features as well as the final 12.

![Fig. 4](image)

**Fig. 4.** Change of algorithm performance as the number of features increases. 12 features give the best trade-off between precision and recall as marked by red cross on each curve.

5.3. Strut detection / classification validation

Figure 5 shows results from processing Steps 1 through 6. Detection of candidate struts in Steps 1-3 gave struts and numerous FPs (Fig. 5B). Almost all FPs were removed in Step 5 (Fig. 5C). Finally, elimination of extra hits (Step 6) gave a result which perfectly agreed with annotation of an analyst (Fig. 5D).

In Fig. 6, we give precision and recall of major steps for baseline and follow-up cases, for both 5-fold cross-validation and leave-one-stent-out. Following detection of candidate struts (Steps 1-3), most struts were detected, giving a recall of ≈96%, but precision was low because of the large number of FPs. After classification and removal of extra struts (Steps 5 and 6), precision increased significantly, with little change in recall. As discussed earlier, validation against all struts (analyzable bright struts and non-bright struts) eliminated some FPs to obtain “actual precision,” PRa. Comparing PR and PRa, we found a slight improvement at baseline and a larger improvement in the follow-up cases, where there were more dim struts with covering tissue. Performance metrics were similar for 5-fold cross-validation and leave-one-stent-out, suggesting generalizability.
Fig. 5. Detection of bright, analyzable struts. (A) Input image from a baseline case. (B) Detection of "candidate struts" following Steps 1-3, including many FPs. (C) Struts following classification (Steps 4-5). (D) Final result after elimination of extra hits (Step 6), eliminating FPs at 3 o'clock and 7 o'clock.

Fig. 6. Stent strut detection evaluation. The first and second groups of bars in each panel are from pooling fivefold cross validation (FFCV) and leave-one-stent-out (LOSO), respectively. Between candidate and classification steps, FPs are removed and precision increase significantly. There is little effect on recall. The two validation strategies gave similar results, indicative of generalizability.
Figure 7 shows some non-bright struts and response of our detection software. When comparing automatic detection in Fig. 7A to analyst-determined bright analyzable struts in Fig. 7B, there were 3 FPs (3, 7, and 12 o’clock) and 1 FN (6 o’clock). The FN was obtained because the strut was very close to the guide wire, which was masked out in automated processing. In Fig. 7D and 7E, magnified images show that there were bright spots at two of the “FP” strut locations, bringing to question whether these instances should have been labeled as FPs. We call these struts ambiguous.

Table 2 summarizes detection statistics for both our software and analysts. Both training/validation methods (5-fold cross-validation on pooled data and leave-one-stent-out) gave similar results. This was not the case when more features were added, indicative of over-training in leave-one-stent-out. Note that leave-one-stent-out is a more strenuous test because in any one stent pullback, images tended to be similar as compared to pooling images across all stents. Using each analyst as ground truth, we obtained 6 sets of TPs, FPs, and FNs for inter-analyst variability. Collecting results across these sets, we got RC = PR = 96% ± 2% and PRa = 97% ± 1% for baseline cases, and RC = PR = 92% ± 4% and PRa = 96% ± 2% for follow-up cases. Standard deviations were measures of spread and were obtained differently for software detection and analysts. Inter-analyst variability approached the “error” of the algorithm. The range of difference in recall and precision was 0-5% when comparing automatic detection performance and inter-analyst variability of “analysts 2 VS 1” and “analysts 3 VS 1.”

### Table 2. Stent strut detection evaluation statistics

| Baseline cases | Follow-up cases |
|----------------|-----------------|
| Pooling        | Analysts        | Pooling        | Analysts        |
| FFCV           | LOSO            | 2 VS 1         | 3 VS 1          |
| Recall         | 90% ± 2%        | 90% ± 3%       | 94%             | 95%             | 94% ± 1%        | 94% ± 1%       | 93%             | 97%             |
| Precision      | 90% ± 1%        | 91% ± 3%       | 96%             | 99%             | 81% ± 3%        | 85% ± 6%       | 97%             | 98%             |
| Actual Precision | 94% ± 1%       | 94% ± 3%       | 98%             | 99%             | 94% ± 1%        | 94% ± 2%       | 97%             | 98%             |

* FFCV: five-fold cross validation. LOSO: Leave-one-stent out cross validation.

### 5.4. Tissue coverage

Tissue coverage area was estimated by subtracting the area of the lumen from the area of the stent. Construction of stent contours is illustrated in Fig. 8. When a sufficient number of struts were detected, a smooth contour was automatically constructed by fitting a periodic cubic
spline (Fig. 8A). When a smaller number of struts were detected, the contour was inaccurate (Fig. 8B). However, after automatically adding “interpolation points” described previously, a good estimate of the stent contour was obtained (Fig. 8C). Lumen contour segmentations are shown in Fig. 9, where we show the lumen before (Fig. 9A) and after refinement with struts excluded (Fig. 9B). With refinements, dynamic programming nicely fills the missing gaps and gives a smooth representation of the lumen (Fig. 9B).

![Fig. 8. Stent contour formation. (A) With a large number of struts (green), a good stent contour is obtained (red). (B) With an insufficient number of detected struts, the stent contour can be in error. (C) The contour from B is corrected by automatically adding “interpolation points” (blue) between detected struts.](image)

![Fig. 9. Lumen border segmentation. (A) Initial lumen border from dynamic programming, which includes errors due to struts on the lumen surface. (B) Lumen border refined by masking out A-lines containing struts prior to dynamic programming.](image)

We compared automated stent and tissue coverage area measurements against manual assessments in Bland-Altman plots (Figs. 10 and 11, respectively). We used 191 image frames from follow-up cases containing at least 1 strut in each half of a frame.

![Fig. 10. Bland-Altman plot of stent area measurement.](image)
Fig. 11. Bland-Altman plot of tissue coverage area measurement.

Figures show similar differences as areas increase and no obvious bias towards high or low values. Collapsing data across all areas, differences were $0.12 \pm 0.20 \text{ mm}^2$ and $0.11 \pm 0.20 \text{ mm}^2$ for stent and tissue areas, respectively. The coefficients of variation ($\sigma$/mean) were 3% and 14% for stent and tissue coverage areas, respectively.

5.5. 3D visualization of detected struts

Figure 12 shows a 3D visualization of detected stent struts for a 9-month follow-up case. The 3D reconstruction was obtained in Amira from 40 2D frames of iOCT images. The length of the section is 8mm. The 3D reconstruction clearly shows the geometrical features of Xience® (Abbott Vascular) stent: parallel zigzag (in-phase) rings connected with horizontal struts. Note that this segmented stent was obtained in an entirely automated fashion.

Fig. 12. 3D visualization of stent. Vessel wall is in red and detected struts are in white. 3D reconstruction shows the characteristic pattern of Xience stent (arrows).

6. Discussion

Our 2D stent strut detection software compares favorably to accuracy reported in the literature and to variability of trained analysts. We tested against a case mix seen by the Core Lab: coronary artery stent studies acquired from sites around the world with equal numbers of baseline and follow-up. When validating against bright, analyzable struts, we obtain $90 \pm 3\%$
and 94 ± 1% recall and 90 ± 3% and 85 ± 6% precision for baseline and follow-up cases, respectively. As described later, several FPs have image evidence of a strut, and if we consider “non-bright struts” marked by analysts, actual precision, PRa, values improve to 94 ± 3% and 94 ± 2%, respectively. Bonnema et al. reported a sensitivity (recall) of 93%, specificity of 99%, and precision of 95% on ex vivo, tissue engineered blood vessels [11], a much more controlled setting than clinical imaging; Bruining et al. reported “success rate,” the fraction of detected stents not requiring manual editing, of 77% on baseline cases and 50% on follow-up cases with a relatively large data set (4024 frames) [17]; Tsantis et al. reported sensitivity of 90 - 92% and specificity of 95 - 97% on femoral artery stent images, which have elongated stent struts and less motion than in coronary images [18]; Gurmeric et al. compared the total number of software detected struts to that for manual analysis without consideration to individual FPs or FNs and obtained an accuracy of 91% ± 11% [13]; Kauffmann, Motreff, and Sarry achieved a “detection” rate of 35.4 - 73.4% in vivo and up to 84.4% with in vitro images [15]; and Wang et al. achieved a sensitivity of 94% and a FP rate of 4% [16]. For improved comparison of methods, one should use similar or identical cases obtained with similar IOCT systems. In addition, comparisons will be improved if reports are given in terms of standardized statistics, not always reported in above studies. For medical imaging studies quantitatively analyzed by experts, an excellent criterion for software acceptance is that results should approach inter-analyst variability. Comparing strut detection against our 3 analysts, we found that the difference in precision and recall were within 5%, and as argued in the next paragraph, many “errors” were relatively unimportant. Given qualifications of above studies, we believe that our method is at least as accurate as any method reported in the literature. It is also clear that our analysis methods are more complete than those previously reported and that they could provide a model for future studies.

Upon closer examination, many “errors” recorded by software either may not be errors or are relatively unimportant. Sometimes FPs occurred at locations where there was evidence of a strut. In Fig. 7, we show cases where software detected a strut not marked as a bright, analyzable strut by analysts. On close examination, we found image evidence of a shadow and a bright spot in two detected struts, even though they were recorded as FPs. Arguably, the software gave “true” results. When we added non-bright struts found by analysts, precision improved to 94% because the number of FPs reduced. Of errors remaining, we obtained a FP fraction, FPF = FP/ (TP + FP) of 6%, which can be further divided. FP detection of a malapposed strut would be serious because such instances are ill advised for stents [10]. We were concerned that residual blood would introduce FP malapposed struts, but none were evident in 508 images analyzed. In two stents with relatively large amounts of residual blood, detection statistics were not compromised. Most FPs occurred at bright spots in the tissue, often near a true strut shadow, and contributed 4.5% of the 6% total. This type of error could lead to inaccurate tissue thickness assessments and should be removed by manual correction. The other 1.5% FPs originated from extra reflection echoes, seam artifacts, catheter sheath, etc. About 2% of the 10% FNs occurred at sites near the guide wire shadow (Fig. 7A). These missed struts should not introduce significant bias, as the location of the guide wire is random. In general, we are relatively unconcerned by FNs because our software will enable all images to be analyzed rather than every third one, as most often done manually [8]. Hence, our software will find almost 3-times as many struts to analyze, allowing us to “miss” some.

Properly applied, machine learning classification is well suited to the problem of stent strut analysis. A machine learning approach allows us to consider 1000s of manually detected struts to optimize our algorithm, a process impossible with manually optimized algorithms. A great advantage is that one can add features such as 3D features, mimicking the analyst’s ability to look forward and backward alone frames when identifying struts and stent contours. Bagged decision trees worked well. It was chosen because it is relatively robust to noise and has better accuracy than a single decision tree. Even though we had significant training data, we saw evidence of overtraining when using all 17 features. The forward feature selection
approach successfully allowed us to remove redundant features and limit over training. We used crisp classification of a strut, corresponding to a majority vote (>50%) across the trees of the algorithm. An attractive option is to use probabilistic classifiers as done recently by Tsantis et al. [18], allowing one to sweep out a receiver operating curve (ROC) and select a threshold which trades off FN versus FP errors.

Automated measurements of stent and tissue coverage area are promising. Differences between automated and manual measurements are 0.12 ± 0.20 mm² for stent area and 0.11 ± 0.20 mm² for tissue coverage area, values similar to those reported by Gurmeric et al. [13]. Coefficients of variation (σ/mean) of measurement difference are 3% and 14%, respectively. However, Bland-Altman plots reveal some outlier frames corresponding to significant differences between automatic and manual contours. The difference is understood from the method for creating manual contours. In addition to bright and non-bright struts, occasionally an analyst will add “interpolation points,” which can depend upon frames before and after the current frame, a process not captured in our 2D method. Automated results are easily edited if needed, and we believe that automated results can be improved using a 3D method.

Our algorithms should greatly reduce analysis time as compared to the fully manual method currently used. An analyst would simply identify start and end frames, and software would proceed automatically. Unattended strut detection and area measurements using software without speed optimization is about 15 minutes for 100 frames. Although automated results are promising, we would advocate analyst review of every frame. Regions with branching would be excluded because tissue coverage and strut malapposition make little sense at a bifurcation, and other analyses might be used. Visual review and editing should be relatively quick. FP struts will be removed by a simple click, and, as argued previously, FNs are relatively unimportant, unless they negatively affect the stent contour. Assuming that all downstream analyses (percentage of covered, uncovered and malpositioned struts, NIH thickness, etc.) can be automated within a comprehensive program, analyst time per stent should be very much reduced from the 6-16 hours now required. For manually analyzed studies, intra and inter-analyst variability limits the statistical power of comparisons between stent designs. Repeatable analysis with standardized software should reduce variability and improve power.

7. Conclusion

In conclusion, our strut detection and tissue coverage area measurement algorithms are quite promising and should greatly speed analysis when incorporated within a comprehensive software package. It is believed that in the future it should be possible to analyze new stent designs using iOCT quickly, cheaply, and robustly with off line analysis. More challenging, but perhaps not insurmountable, will be live-time analysis of stents for clinical decision making, where careful review might not be an option.

Acknowledgments

The project described was supported by the National Heart, Lung, and Blood Institute through NIH R21 HL108263 and by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1RR024989. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. H. L. was partially supported by the Chinese Government Scholarship. Z. W. was partially supported by the American Heart Association predoctoral fellowship (#11PRE7320034).