

Automated Detection and Segmentation of large lesions in CT Colonography

S.E. Grigorescu^a, S.T. Nevo^a, M.H. Liedenbaum^b, R. Truyen^c, J. Stoker^b,
L.J. van Vliet^a, F.M. Vos^{a,b}

^a *Quantitative Imaging Group, Delft University of Technology, The Netherlands*

^b *Department of Radiology, Academic Medical Center, Amsterdam, The Netherlands*

^c *Clinical Science & Advanced Development, Healthcare Informatics, Philips Healthcare, The Netherlands*

Abstract

Computerized tomographic colonography (CTC) is a minimally invasive technique for the detection of colorectal polyps and carcinoma. Computer aided diagnosis (CAD) schemes are designed to help radiologists locating colorectal lesions in a efficient and accurate manner. Large lesions are often initially detected as multiple small objects due to which such lesions may be missed or misclassified by CAD systems. We propose a novel method for automated detection and segmentation of all large lesions, i.e. large polyps as well as carcinoma.

Our detection algorithm is incorporated in a classical CAD system. Candidate detection comprises pre-selection based on a local measure for protrusion and clustering based on geodesic distance. . The generated clusters are further segmented and analyzed. The segmentation algorithm is a thresholding operation with an adaptive threshold selected. The segmentation provides a size measurement that is used to compute the likelihood of a cluster to be a large lesion.

The large lesion detection algorithm was evaluated on data from 35 patients having 41 large lesions (19 of which malignant) confirmed by optical colonoscopy. At 5 FP/scan the classical system achieved a sensitivity of 78% while the system augmented with the large lesion detector achieved 83%. For malignant lesions, the performance at 5FP/scan was increased from 79% to 95%.

The good results on malignant lesions demonstrate that the proposed algorithm may provide relevant additional information for the clinical decision process.

Key words: CT colonography, computer aided detection, carcinomas, image segmentation, LH histogram

1. Introduction

Computerized tomographic colonography (CTC) is a minimally invasive technique for the detection of colorectal polyps and carcinoma. CTC involves acquisition and interpretation of original/reformatted 2D data and endoluminal 3D views. Although the data acquisition is fast and reliable, the interpretation of CTC data can be time consuming [1][2][3] and depends on human factors such as experience, fatigue, and perception [4][5]. This has led to a sustained effort to develop computer aided diagnosis (CAD) schemes that can help radiologists with locating colorectal lesions in CTC data in an efficient and accurate manner [3][6][7][8][9].

Large colorectal lesions are of high clinical importance, since lesion size relates to the risk of malignancy. Several previous articles address large polyps and so-called masses, with this later group not clearly defined. Zalis et al. [10] define them as lesions 30mm and larger, while Kiraly et al. [11] and Copel et al. [12] define them as lesions 20mm and larger. However, this is a conservative viewpoint with respect to the presence of malignancy (carcinoma) within the lesion and the chance of developing malignancy in the future (advanced adenoma). In daily practice, all lesions of 10 mm and larger are considered important and should be detected: 2.6% of all lesions equal or larger 10mm is malignant while the prevalence of advanced adenoma is 28% [13].

We therefore define large colorectal lesions as colorectal polyps or polypoid carcinoma having a diameter larger than 10mm (we will maintain this definition throughout the paper).

The performance of automated detection systems on large lesions is certainly promising (see, e.g. [14]). Still, the need for automatic detection and segmentation systems that are able to deal with a broad range of colorectal lesions was acknowledged previously [1][15]. Large lesions may have a rather flat surface shape (i.e. a large radius of curvature), which is not so easy to detect automatically due to its similarity to the physiological surface. Alternatively, large lesions can be initially detected as multiple smaller polyp candidates by automated systems, which may prevent a correct size measurement and/or a proper characterization of the candidate [2][3][16][17]. As a consequence,

these lesions may be either missed or misclassified. Such results are confusing and decrease the confidence of radiologists.

Näppi et al. [2] proposed a dedicated method for large lesion detection and segmentation that exploits the detection by existing CAD systems as sets of ‘polyps’. The algorithm selects CAD detections that lie in a neighborhood and fulfill a number of conditions. These detections are the input for a segmentation algorithm based on a level set approach. The final step of the algorithm is classification based on the computed features for each of the segmented regions. The reported results are promising, but the number of test cases is limited and the algorithm involves a large number of parameters making it sensitive to variations in the input data [18].

Luboldt et al. [16] proposed an algorithm for automated large lesion detection in contrast-enhanced CTC that searches for clusters of voxels fulfilling certain criteria in a region of a given thickness around the colon lumen. The approach was tested on five patients only. The results showed a high sensitivity at the expense of highly variable specificity rates. Luboldt et al. [16] use intravenous contrast-enhanced CTC data, an approach that is less appropriate in screening population due to potential side effects and costs [19][20][21].

The objective of this paper is to present a new method for automated detection and segmentation of large lesions, i.e. colorectal polyps or polypoid carcinoma larger than 10 mm in diameter. Such a lesion can be intraluminal, with a significant intraluminal component, such as lobulated, polypoid, or circumferential types of masses as well as nonintraluminal, associated with a mucosal wall-thickening type of growth pattern. It is assumed that the intraluminal component has CT density similar to that of muscular structures.

The detection step, inspired by the work of Näppi et al. [2], starts from intermediate results of a classical polyp CAD system. After a pre-selection step, we group together locations that satisfy certain proximity criteria. The result is a set of clusters of nearby locations that most likely correspond to large lesions. A key technical novelty of the algorithm is a segmentation step based on a simple thresholding operation in which an adaptive threshold is automatically selected. The threshold selection is done based on an LH histogram representation of the image data [23][24][25].

The major advantages of the proposed segmentation algorithm are its speed and simplicity. This study shows that the combination of large lesion detection and segmentation improves the performance of a classical CAD system.

2. Material and Methods

2.1. Experimental Data

The methods presented in this paper were *developed* using 21 patients from previous studies described in the work of Van Gelder et al. [26] and Pickhardt et al. [27]. The *evaluation* was done on separate (unseen) datasets containing patients from two different studies: a clinical population and a Fecal Occult Blood Test (FOBT) positive screening population. All patients were scanned in both prone and supine position. After CTC, patients underwent optical colonoscopy (OC) that served as the reference standard. Within a period of 3 years, all patients that harbored large lesions, i.e. large polyps (10mm and larger measured during OC) and/or carcinomas, were consecutively included. This rendered 35 patients in total having 41 large lesions, 19 of which were malignant. The lesion sizes measured during colonoscopy (using an opened biopsy forceps) or the pathology measurements, varied in the range 10-80mm, with a median value of 15mm. For three patients, scanning in either prone or supine position failed for technical reasons leaving 67 datasets. An expert labeled the large lesions in CT data based on the optical colonoscopy findings by indicating a point in 2D reformatted images. A CAD finding was considered true positive if it encompassed the indicated point. More details on the study data are given in the appendix.

2.2. Large lesion detection

A schematic view of the proposed large lesion detection scheme can be seen in Fig. 1. The input are suspicious locations on the colon surface detected using the deformation model described by Van Wijk et al. [28]. In the latter approach, bumps on the colon surface are flattened by solving a partial differential equation controlled by the second principal curvature. The distance between the original and the deformed surface defines the amount of *protrusion* of the bump. This quantity is used to select

potential polyp locations (Detection of Lesion Candidates). Apart from detection, a segmentation of the detected locations is also provided: the segmentation is given by the region that is brushed away during the deformation process. The segmentation is used to compute local features to characterize an object and to compute its likelihood of being a colorectal lesion.

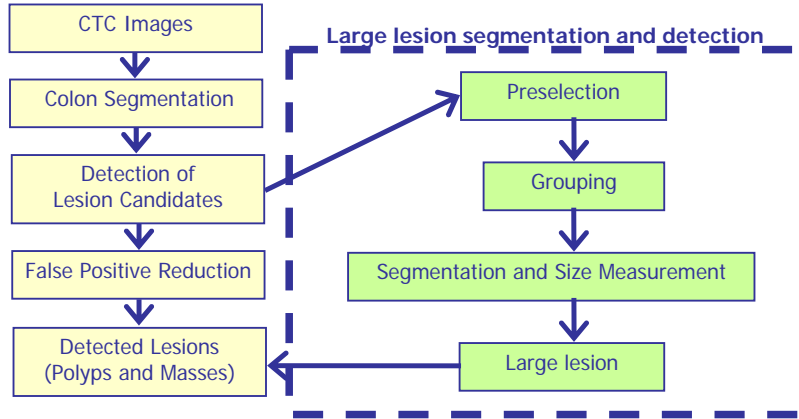


Fig. 1 A schematic view of the large lesion detector and its integration into a classical polyp detector.

The “preselection” step of the large lesion detector aims at discarding the most obvious faulty detections such as those generated by small food remnants, the enema tube, and small false positive detections emanating from image noise. This selection is done using information provided by the segmented region around the detected locations. We use the average intensity in the segmented regions for discarding detections of tagged food remnants. We regard as noisy detections those locations where the segmentation mask is too small or the protrusion is small. The parameters of this step were learned from the development set.

The second stage of the large lesion detection algorithm groups those locations that passed the preselection stage and lie close to each other. Two or more points are regarded as close to each other if the geodesic distance along the colon surface between them is smaller than a given threshold.

The final step of the grouping phase is to compute the likelihood P_C of the cluster of detections of being a large colorectal lesion. This is done by aggregating together the likelihood p_i associated with

each detection in the group: $P_C = \prod_{i=1:N} p_i$, where N is the number of detections grouped together.

At the end of this stage, each cluster of detections is considered as one hit whose likelihood of being a lesion is given by the computed P_C . These hits are put together with all other CAD detections (see Fig. 1) that were not included in the generated clusters. The likelihood of these latter detections remains unchanged.

2.3. *Large lesion segmentation*

The detection step described previously identifies the location of the large lesions without providing a proper segmentation. Consequently, additional steps are required for extracting the entire lesion volume, which is important for a correct size measurement and/or feature extraction for subsequent classification. One can reliably segment the lumen side of a lesion using thresholding with two fix values, one for air borders and one for tagged material borders. However, the identification of the internal, non-lumen side of a lesion is more complex. Large colorectal lesions usually have CT density similar to that of muscular structures. Most of the time the internal side of the lesion is surrounded by fat layers. Such layers have a slightly lower density than the tissue present in the lesions. In such cases, the internal border of a lesion can be seen and can be identified by simple thresholding. However, one cannot use one fixed threshold to find this border because the range of fat densities is not the same for all patients.

We propose a novel method for automatic selection of the threshold value that best separates a lesion from the surrounding fat. Since the lesion and muscular tissue have similar densities in CT data, this approach addresses the more general problem of separating fat and muscle.

Our method is based on the LH histogram representation of the local gray value data [23][25]. In such a representation, each voxel is mapped into LH space by looking at intensities values up and down the gradient direction at that point. The lowest and highest encountered values represent the L and H coordinates in the LH space, respectively. Consequently, points belonging to a material whose mean intensity is L form a blob in the LH space, centered at location (L, L) . Points that belong to the border region between two materials, having intensity L and H , are mapped onto a blob centered at location (L, H) , see Fig. 2(a)(b). The number of blobs in the LH space depends on the number of materials and

the number of borders in the original gray value volume. The size of the blobs depends on the noise level in the data.

Suppose, for simplicity, that two materials are to be segmented. Such a segmentation is equivalent to splitting the LH space in four quadrants (thick dashed lines, Fig. 2(b)) to ensure that each quadrant, but the lower right one, contains one blob. Consequently, finding the optimal threshold that best separates the two materials can be solved by finding a splitting of the LH space that best separates the three blobs. Such a splitting maximizes the distance between the centers of the clusters and can be found by successively trying a number of splittings and by looking at the distance between the centers of the clusters in each quadrant.

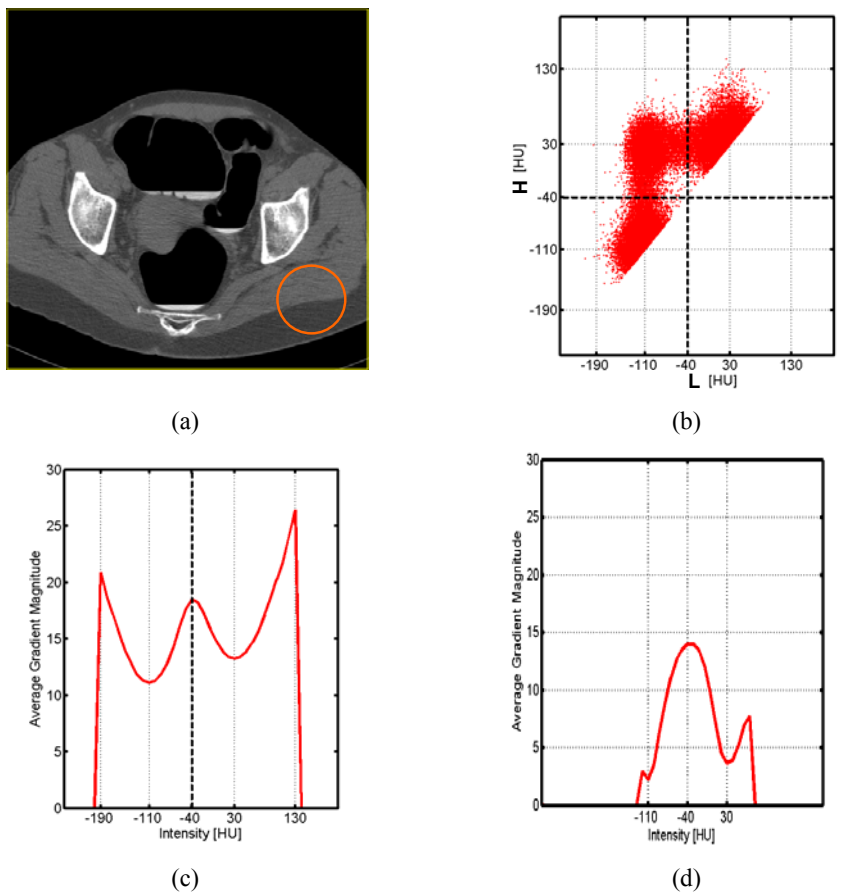


Fig. 2 The LH Histogram (b) and the average gradient magnitude as a function of threshold to select a border (c) in the subregion of (a). The average gradient magnitude for a filtered version of the subregion is shown in image (d).

For an arbitrary splitting, let us denote the centers of the clusters in the lower left, upper left, and upper right quadrant, respectively, by (\tilde{L}, \tilde{L}) , (\tilde{L}, \tilde{H}) , and (\tilde{H}, \tilde{H}) . \tilde{L} and \tilde{H} can be regarded as estimates of the intensities of the two material in the image. Note that L and H indicate the true intensities. The distances between these centers are proportional to the difference $(\tilde{H} - \tilde{L})$, which can be directly computed by only considering the cluster in the upper left quadrant (i.e. the border region), as follows:

$$\tilde{H} - \tilde{L} = \frac{\sum_{x \in \Omega} H\{P(x)\}}{N} - \frac{\sum_{x \in \Omega} L\{P(x)\}}{N} = \frac{\sum_{x \in \Omega} (\Delta s) |\nabla P(x)|}{N} = (\Delta s) \frac{\sum_{x \in \Omega} |\nabla P(x)|}{N} \quad (1)$$

in which $P(x)$ is the intensity at location x , Ω denotes the set of boundary voxels for the current partitioning; N is the number of boundary voxels; $H\{P(x)\}$ and $L\{P(x)\}$ yield the local maximum and local minimum, respectively, by following a path up and down the gradient direction at location x ; while Δs is the distance between the locations of $H\{P(x)\}$ and $L\{P(x)\}$. Eq.1 shows that the difference $(\tilde{H} - \tilde{L})$ is proportional to the average gradient magnitude along the borders of the object generated by the threshold used for generating the LH space splitting. Eq.1 assumes that Δs is approximately constant, a reasonable assumption when considering that Δs depends on the point spread function of the scanner. Specifically, the imaging of an edge by a CT scanner does not appear as a step function but as a blurred version of it [25]. Consequently, going from a low intensity to a high intensity requires covering a distance that is proportional to the amount of blurring, which in turn is determined by the point spread function of the scanner.

Based on the above considerations, one can find the threshold that best segments the encircled region in Fig. 2(a) by computing the average gradient magnitude for a number of successive thresholds and looking for a (local) maximum. For a fast implementation of this operation, we use the method proposed by Pekar et al. [29] leading to the result shown in Fig. 2(c).

The average gradient magnitude curve (Fig. 2(c) features a local maximum corresponding to the optimal splitting of the LH space (indicated by the thick dashed lines in the middle plot). We call this

local maximum “*the fat-muscle transition peak*”. This maximum is surrounded by two valleys corresponding to the “worst” splittings of the LH space – splitting that cut through the center of the blobs corresponding to the two materials. These valleys correspond to the low (L) and high (H) material intensities in the image. The local maxima on the sides of the plot of Fig. 2(c) are partly due to the presence of noise in the image. The higher the level of noise in the image the higher the amplitude of the side peaks. Additionally, the side peaks may emanate from transitions between air and tissue and between air and bone/tagged material present in the analyzed region.

The above technique can be used for discriminating fat densities from muscle densities in CT colonography data. Since our eventual goal is colorectal lesion segmentation, we restrict our analysis to a region of interest around a candidate lesion. More specifically, for the computation of the LH histogram, we use a cubic region of $80 \times 80 \times 80 \text{ mm}^3$ and, in this region, we select a tissue region 15mm thick around the colon lumen. For the selected region of interest, we compute the average gradient for successive thresholds in the aforementioned range. In order to identify the optimal threshold for fat muscle separation in the presence of the side peaks, we restrict the search for the location of a local maximum to the range $[-200;100]\text{HU}$, a typical range for fat and muscle densities [30].

We applied this approach to all detections in our development set. In 67% of the cases, we were able to automatically identify a fat-muscle peak. In the other 33% of the cases the combined influence of noise and other transitions suppressed the fat-muscle transition peak.

Suppression of the effect of noise can be done by smoothing the image prior to average gradient magnitude computation. A bilateral filter [31][32] does a weighted summation of the image intensities in the surrounding, in which the weights depend on the spatial and tonal difference between the considered points. The results shown in Fig. 2(d) are obtained with a spatial sigma of 2mm and a tonal one of 60 HU. Effectively, a bilateral filter attenuates the noise while preserving the edges.

On the cases in which the fat-muscle peak could not be identified, the proposed algorithm restarts the computation of the average gradient magnitude plot on the bilateral filtered data. As a result, the number of cases where a fat-muscle peak could be found increased from 67% to 77%. For the remaining cases, i.e. the one where the peak finding algorithm fails even after filtering, we use a

default threshold value. A visual inspection of these latter cases revealed that in most of these cases corresponded to regions in which a low amount of fat was present in the region of interest.

We base the choice of the default threshold value on the distribution of the threshold values computed for the cases where a fat-muscle peak *could* be identified in the average gradient magnitude plot. This distribution, shown in Fig. 3, suggests that the optimal threshold for fat muscle separation depends on the characteristics of the scanning protocol, scanner calibration, and patient preparation. Accordingly, we used as default values, the average of the estimated threshold values for the population to which the considered patient belonged.

It might be argued that the bilateral filter approach and/or a default value might be applied just as well as in those cases in which the threshold value was initially based on the LH histogram. This is not true. Fig. 4 shows the difference between the segmentation results with a threshold based on the fat-muscle peak location (middle column) and a default threshold derived from population statistics (right column).

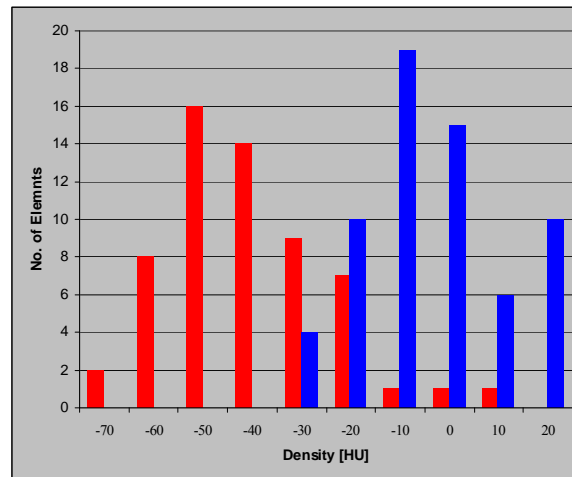


Fig. 3 Distribution of threshold values computed based on the fat-muscle peak location in the development data. The red distribution corresponds to the patients from the study described by Van Gelder et al. [26], while the blue one corresponds to the patients from the study of Pickhardt et al. [27].

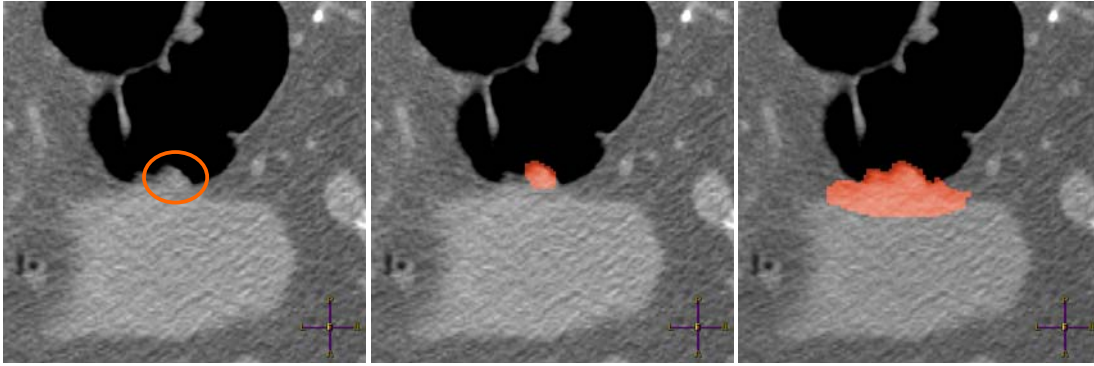


Fig. 4 Two lesions from the development set (left column) and the corresponding segmentation with a threshold based on the fat-muscle peak location (middle column) and a fixed threshold derived from population statistics (right column).

2.4. Post-Processing and Feature Computation for the Segmented Lesion

Once the threshold is determined, we apply it to the selected region of interest. The result is a segmentation of muscle-like structures. In order to eliminate ‘small’ connections between multiple muscle-like structures that might be present in the region of interest, we perform a morphological opening [33] followed by the identification of the connected component at the location of the initial detection.

Additionally, a few lesions are surrounded by tissues with similar attenuation to muscle-like tissue and no border can be observed. In order to deal with such cases, we model a lesion as a band like structure attached to the colon lumen. Consequently, any point in the lesion body should be connected to the lumen through a straight line that goes only through lesion region and not through non lesion tissue. Additionally, we eliminate those regions that lie deeper than 15mm (similar to the thickness used by Vining et al. [34] and Nappi et al. [2]). Such a clipping helps discarding muscular structures adjacent to the lesion (see Fig. 5). A side effect of it is that internal parts of large, invasive, lesions will be cut away. However, the extent of the lumen side of such lesions is not affected. In effect, the lesion is asserted to a somewhat compact object connected to the lumen. A fast way of doing this operation is by identifying the lumen side of the segmented region and propagate this border 15mm into the segmented region.



Fig. 5 An example from the development set [26] in which the boundary between the large lesion (pointed by the white arrow in the left image) and the surrounding muscular structures (indicated by the orange arrows) is not visible. The segmentation results are shown in orange. The second image shows the results in the absence of the post-processing step, while the third one shows the results after post-processing. In both cases a volume of interest around the lesion was used. The post processing step eliminates most of the abdominal muscle that was originally included in the segmentation.

Once the segmentation is available, we compute the diameter d of the segmented region as the largest Euclidean distance between two border points of the region. We further use this diameter in combination with P_C , the likelihood of a cluster of being a large lesion (see Section 2.2), to recalculate this likelihood P_C^{new} according to Eq. 3.

$$P_C^{new} = (P_C + d_r)/2, \text{ where } d_r = \begin{cases} d/50 & \text{if } d \leq 50 \\ 1 & \text{if } d > 50 \end{cases} \quad (3)$$

Roughly speaking, Eq. 3 implies that any detection with a diameter bigger than 50mm has a high likelihood of being a large lesion. Furthermore, if a cluster has a big likelihood, as computed during the grouping phase (see Section 2.2), and a big diameter than it is most likely a large lesion.

3. Results

3.1. Parameter Overview

The algorithm involves a number of parameters that were either determined experimentally based on the development data, or were based on literature evidence. The density range to identify soft tissue were selected based on evidence in the literature [9][17][29]. The maximum lesion thickness (used in the post-processing phase) was determined empirically, but is similar to values for colon wall analysis reported previously [2][34]. The threshold on the geodesic distance has been determined experimentally. This threshold prevents the chaining effect mentioned by Näppi et al. [2], while insuring that sufficient detection on a single lesion are grouped together. Small variations (under 20%) in the values do not affect the results.

3.2. Performance Evaluation

3.2.1. Detection

The performance of the detection step is assessed by the sensitivity and specificity of the algorithm for different threshold values for the likelihood of a detection to be a large lesion. This approach allows us to build a Free-Response Receiver Operating Characteristic (FROC) curve. We computed the sensitivity as the percentage of the available large lesions that were detected in either the prone or supine scans, i.e. per lesion sensitivity. The specificity was computed as the average number of false positives (FP) per scan. For a number of working points on the FROC curve, we also report the sensitivity range at 95% confidence interval. These intervals were computed using bootstrapping as follows: from the 35 patients, we selected randomly 50 samples (duplicates allowed) and measured the sensitivity for the selected number of FP. We repeated this procedure a number of times and obtained a series of sensitivity values. Such series provided us with the distribution of sensitivity levels for the selected FP rate and allowed us to estimate the range for the 95% confidence interval.

We use as input for the proposed large lesion detector intermediate results of the polyp detector described by Van Wijk et al. [28]. We will refer to latter detector as the classical CAD system. In Fig.

6 and Fig. 8, we show in blue the FROC curve for classical CAD system, which was obtained by varying the threshold on the likelihood of a detection of being a polyp.

The proposed detection/segmentation algorithms changes the list of hits provided by the classical CAD by grouping together a number of detections (as described in Section 2.2) and by computing a new value for the likelihood of each cluster of being a lesion (as described in Section 2.2 and Section 2.4). The hits that are not grouped together remain in the final list of detections with the likelihood computed by the classical CAD system. Consequently, one can regard the new system comprising the classical CAD system augmented with the information provided by the proposed detection/segmentation algorithms as an enhanced CAD system. The performance of this new system is shown in red in Fig. 6 and Fig. 8.

In a first analysis, we consider all large lesions available in the evaluation set (Fig. 6). At a rate of 8FP/scan and higher the classical and the enhanced CAD systems have the same sensitivity $83\% \pm 10\%$ (34 out of 41). The 7 false negatives (FN) comprised 3 sessile polyps of 10mm, a polyp on a fold of 15mm, 2 pedunculated polyps of 12mm and 15mm that were completely submerged, and a large obstructing carcinoma that was partially submerged. The polyp on the fold and one of the three sessile polyps were missed by the CTC observer in the first reading and could only be found retrospectively. The other two sessile polyps could not be found by the observer, not even retrospectively.

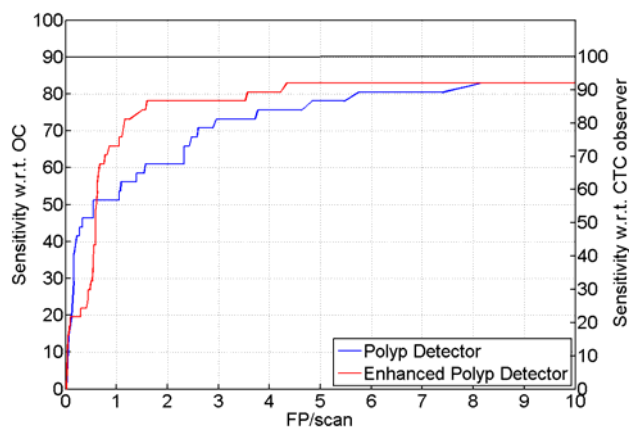


Fig. 6 FROC curve for the classical CAD system (in blue) and the enhanced CAD system (in red) for all large lesions in the test set. The right y-axis corresponds to sensitivity with regard to all large lesions except for the 4 cases that were missed by the CTC observer as first reader.

The large obstructing carcinoma missed by both detectors belonged to a patient for which only one scan was available. Furthermore, parts of the lesions were submerged, while those regions that were above fluid level were fairly flat (see the left image in Fig. 7).

The two detectors showed differences in the sensitivities when higher specificities were considered. At an average rate of 5FP/scan, the sensitivity of the enhanced CAD system stayed at $83\% \pm 11\%$, while that of the classical CAD system dropped at $78\% \pm 12\%$. For 2FP/scan, the enhanced CAD system achieved sensitivity of $78\% \pm 10\%$, while the sensitivity of the classical CAD system dropped to about $61\% \pm 14\%$. For the 2FP/scan rate, the enhanced CAD system discarded 2 carcinomas (see the middle and right images in Fig. 7). Both carcinomas caused obstruction and had a fairly flat surface.

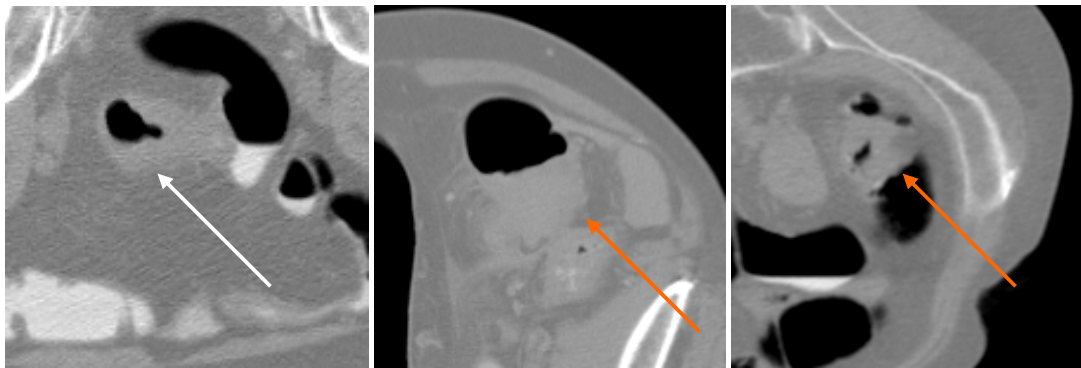


Fig. 7 The carcinomas missed by both CAD systems. The white arrow indicates the lesions missed at a rate of 8FP/scan, while the orange arrows point to the lesions missed by both detectors at a rate of 2FP/scan.

Subsequently, we made a comparison of the two CAD systems with regard to their ability to detect malignant lesions only (19 out of 41) and benign lesions only (22 out of 41). The enhanced CAD system performs better than the classical CAD system in detecting malignant lesions at an acceptable FP-rate, i.e. up to 5 FP/scan (see the plot in Fig. 8(a)). For 5FP/scan, the sensitivity of the enhanced CAD system is $95\% \pm 8\%$, while that of the classical CAD system is $79\% \pm 15\%$. For 2FP/scan, the sensitivity of the enhanced CAD system is $84\% \pm 10\%$, while that of the classical CAD system is $47\% \pm 20\%$. For benign lesions, the two CAD systems have similar performance (see the plot in Fig. 8(b)), i.e. $73\% \pm 20\%$ at 2 and 5 FP/scan. The benign lesions in our study are mainly polyps in the range of 10 - 20mm. This kind of lesions is still rather similar to the shapes which were used in the

design set of the classical CAD system [26] and accordingly were properly detected by it. Exceptions were the 4 cases that were missed by the CTC observer as the first reader and the 2 fully submerged polyps.

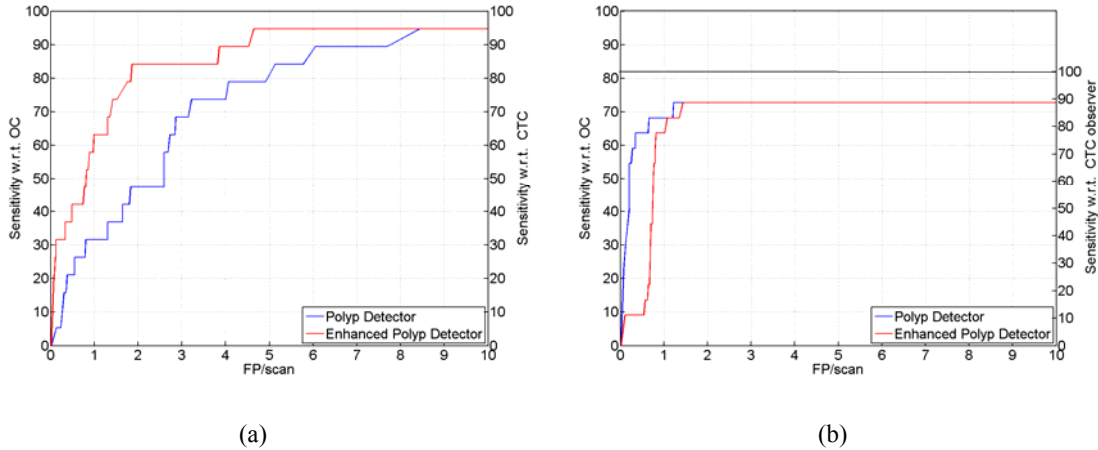


Fig. 8 Comparison of the classical CAD system (in blue) vs. the enhanced CAD system (in red) for malignant lesions (a) and for benign large lesions (b). The large lesions missed by both CAD systems are submerged lesions and lesions that were missed by the CTC observer as first reader.

The main causes of false positives were stool, enema tube, and ileocecal valve. One of the patients in Population 1 had not followed the prescribed diet and had a lot of residual feces in the large bowel. This patient accounted for 20% of the FP generated in the whole set of 35 patients.

The threshold selection based on the LH histogram was able to detect a fat-muscle peak in 70% of the detections. The bilateral filtering yielded another 12%, leaving the use of a default threshold value for 18% of the cases.

3.2.2. Segmentation

Next to the evaluation of the performance of the CAD system augmented with the information provided by the segmentation algorithm, we also looked separately at the quality of the segmentation results for the detected large lesions. A research fellow of the department of radiology (experience more than 300 colonoscopy verified CTC readings) graded the segmentation results on a five-point Likert scale. (1) 0-20% overlap; (2) 20-40% overlap; (3) 40-60% overlap; (4) 60-80% overlap; (5) 80-100% overlap. The marks on this scale corresponded to degree of perceived overlap with the actual

lesion volume: (1) 0-20% overlap; (2) 20-40% overlap; (3) 40-60% overlap; (4) 60-80% overlap; (5) 80-100% overlap. Effectively, this grading is similar to the measure used by Yao et al. [17] and Yao and Summers [35] except that in this reference a manual segmentation is performed, whereas we let the radiologist decide based on visual inspection. Such a grading penalizes both leakages and incomplete segmentations. The distribution of scores is given in Fig. 9.

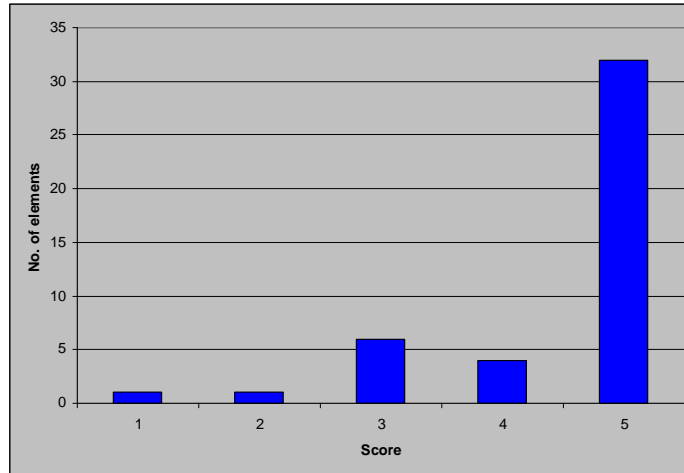
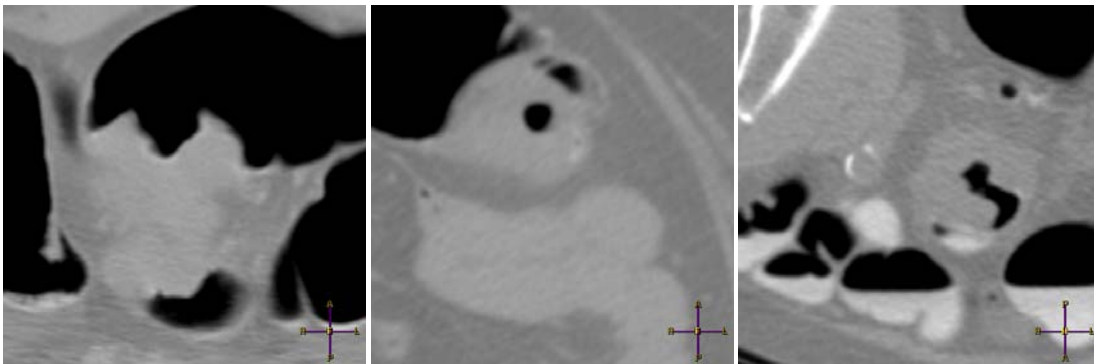


Fig. 9 The distribution of the scores given by the research fellow radiology. The marks on this scale corresponded to degree of perceived overlap with the actual lesion volume ranging from (1) 20% or less - (5) 80% or more overlap.

The segmentation results received the highest mark in 73% of all cases, with 65% of the malignant lesions falling into this category (e.g. Fig. 10), and 81% of the benign large lesions.



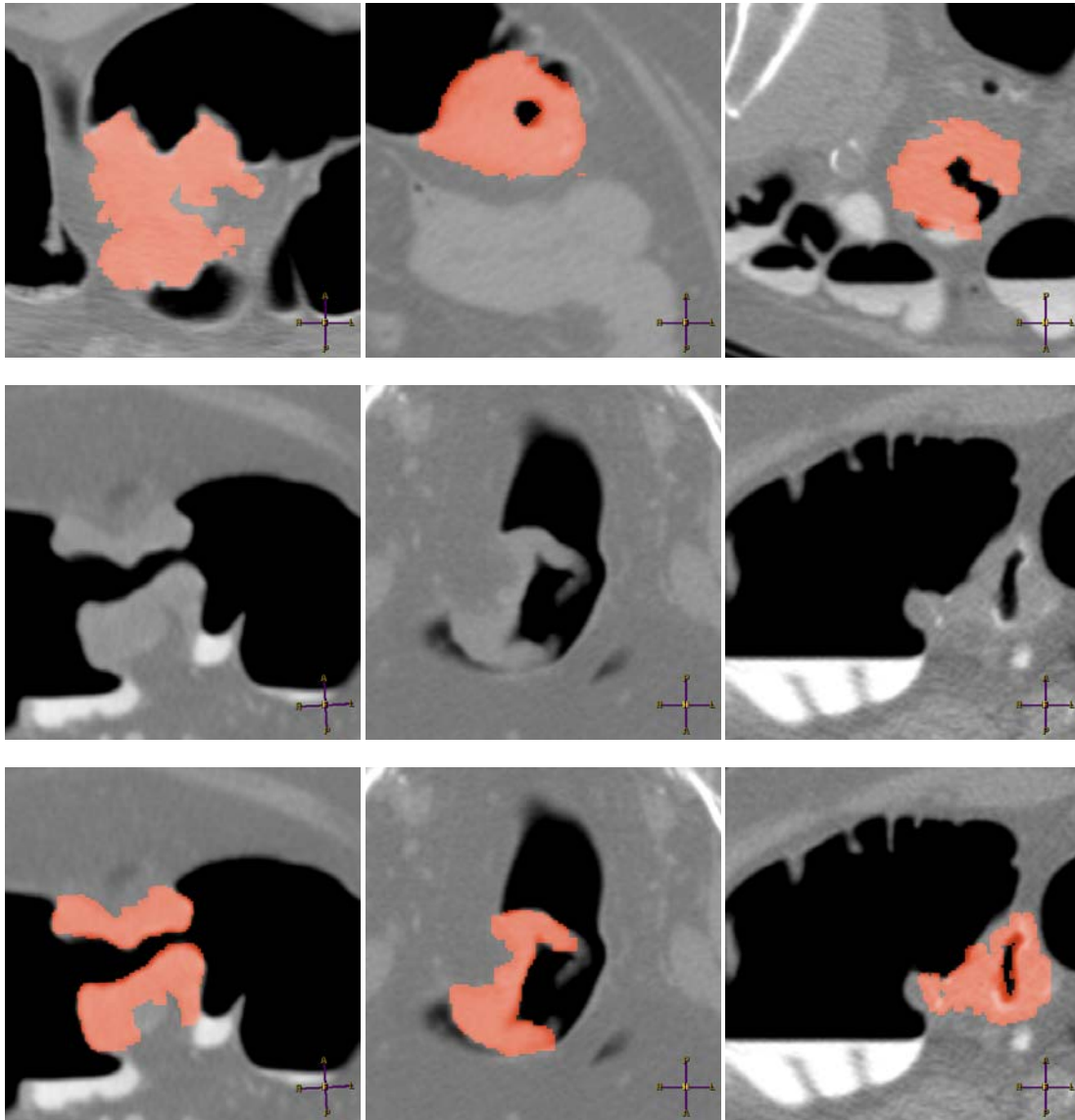


Fig. 10 Malignant lesions (first and third rows) from the test set and their corresponding segmentations (second and fourth rows, respectively). Notice that the lesion in the right image of the third column is adjacent to the ileocecal valve.

The main causes for failing segmentation were obstructing lesions and leakages into adjacent organs. Obstructing lesions were incompletely segmented since parts of the lesions extended further than 15mm from the lumen. The leakages into adjacent organs were due to the absence of a clear border demarcating the lesions. Apparently, the post-processing step only partially corrected for these leakages.

4. Discussion and Conclusions

We addressed the issue of automatic detection of large lesions and we showed that the performance of a classical polyp CAD system can be improved by adding two components: (i) detection and (ii) segmentation of large lesions.

The proposed algorithms are simple and fast. They use only a small number of parameters that were determined experimentally. On a PC workstation with an Intel Xeon processor running at 3.2 GHz, the detection step takes a few seconds per scan, depending on the number of detections that passed the preselection step. The segmentation algorithm takes on average 10s per lesion in case no bilateral filtering is needed in threshold selection. The cases involving bilateral filtering take on average 20s per lesion. Unlike segmentation methods involving iterative methods, like the ones proposed by Näppi et al. [2] and by Yao and Summers [35], the speed of our algorithm does not depend on the size of the segmented lesion. The mean total computation time per patient was 4 minutes for the classical polyp CAD system on the previous configuration.

The starting point of our large lesion detection system are the candidates generated by a classical polyp CAD system. A preselection step effectively aims to find bumps on the colon surface generated by large lesions. Subsequently, clusters are formed of the detections that pass the preselection. The generated clusters are further segmented and measured. The information provided by the grouping and segmentation steps is incorporated in a classical CAD system leading to a better detection performance, especially on malignant lesions.

The main source of FN of our large lesion detection algorithm were lesions that were either submerged or lack the bumpy appearance on the side facing the lumen. Next to cleansing, detection methods that can cope with non or minimal intraluminal protruding, flat lesions (see e.g. the work of Näppi, Frimmel et al. [2]) and/or obstructing lesions could prove useful. Other sources of FN were 10mm sessile lesions that were hard/impossible to find by a radiologist. The main sources of FPs were the ileocecal valve, the enema tube, and stool.

The large lesion segmentation algorithm assumes that there is a density difference between the lesion and the surrounding tissue. Such an assumption is not new, being used in one way or another in

previous work [2][17]. We asserted that lesion tissue has a muscle like density whereas the surrounding tissue shows a fat like density. Actually, it is a more general problem of finding the threshold that best separates fat and muscle regions. The results showed that there is a relation between the optimal fat-muscle separation threshold and the scan protocol and/or scanner calibration. Analyzing which aspects of the scanning protocol influence the value of the optimal threshold is a subject for future work. The proposed threshold selection method is robust against such variations. Many large lesions were successfully segmented (up to 73% of all cases). Not surprisingly, the algorithm failed to segment lesions that were surrounded by structures with muscle-like densities. In those cases, the border between a lesion and the surrounding structures cannot be identified in CT data. In this step, the only parameters determined empirically are the ones used in the bilateral filter. Another application of the proposed segmentation algorithm might be the segmentation of visceral fat [37]. In fact, the algorithm proposed by us is immediately applicable if only the proper region of interest is selected for LH histogram computation

Next to the role played in improving CAD performance, an accurate segmentation algorithm is important for correct size measurement, which is the decisive factor for further patient management. A recent study [38] showed that less experienced radiologists misclassified 29% of polyps larger than 10 mm as lesions smaller than 10 mm. The same study showed that lack of experience poses the risk of dismissing very large lesions (30mm) because only part of it is perceived as polyp. Our algorithm was able to properly segment 73% of the large lesions suggesting that it can provide a useful second opinion.

The good results on malignant lesions demonstrate that the proposed algorithm may provide relevant additional information for the clinical decision process.

Appendix: details of the study population

The evaluation was done on a separate (unseen) dataset containing 35 patients from two different studies:

- Population 1 (clinical patients): 22 symptomatic daily practice patients admitted for CTC from the Academic Medical Center, Amsterdam, The Netherlands were included. For bowel preparation the patients drank 4 liter of macrogol solution (KleanPrep, Helsinn Birex Pharmaceuticals Ltd, Dublin, Ireland) combined with 4*50 ml meglumine-ioxithalamate tagging material (Telebrix Gastro, 300 mg I/ml, Guerbet, Roissy, France) and 2 tablets of 5 mg bisacodyl starting one day before the examinations. The colon was distended by automatic insufflation of CO₂ (EZEM Protocol insufflator, Dordrecht, The Netherlands). The CT scans were performed onto two systems: a 4-slice CT scanner (Mx800, Philips Medical Systems, The Netherlands) was used for 13 patients and a 64-slice CT scanner (Brilliance, Philips Medical Systems, The Netherlands) was used for the other 9 patients. The scan parameters for the 4-slice scanner were: 120 kV, 4*2.5 mm collimation, 3.2 mm slice thickness, 1.25 mm pitch. The scan parameters for the 64-slice scanner were: 120 kV, 64*0.625 mm collimation, 0.9 mm slice thickness (3.0 mm after resampling), 0.984 mm pitch, 58 mAs for patients with an abdominal circumference <102.5 cm and 82 mAs for patients with an abdominal circumference ≥102.5 cm. For both scanners, a standard reconstruction filter ('C') was used.
- Population 2 (FOBT positive screening patients): 13 patients with a positive fecal occult blood test (FOBT) (increased risk) from the Academic Medical Center, Amsterdam, The Netherlands. For bowel preparation the patients drank 7*50 ml Telebrix tagging agent (meglumine ioxithalamate, 300 mg I/ml, Guerbet, Roissy, France) and had a non-fiber diet starting two days before the examinations. The colon was distended by automatic insufflation of CO₂ (EZEM Protocol insufflator, Dordrecht, The Netherlands). The CT scans were performed with a 64-slice CT scanner (Brilliance, Philips Medical Systems, The Netherlands) with scan parameters: 120 kV, 64*0.625 mm collimation, 0.9 mm slice thickness (3.0 mm after resampling), 1.2 mm pitch, 40 reference mAs with automatic current selection and dose modulation and a standard reconstruction filter ('C').

References

- [1] Chowdhury, T.A., Whelan, P.F., Ghita, O. A fully automatic CAD-ctc system based on curvature analysis for standard and low-dose CT data. *IEEE Transactions on Biomedical Engineering*, 55(3): 888-901, 2008.
- [2] Näppi, J., Frimmel, H., Dachman, A.H., Yoshida, H. Computerized detection of colorectal masses in CT colonography based on fuzzy merging and wall-thickening analysis. *Medical Physics* 31(4): 860-872, 2004.
- [3] Summers, R.M., Johnson, C.D., Pusanik, L.M., Malley, J.D., Youssef, A.M., Reed, J.E. Automated polyp detection at CT colonography: Feasibility assessment in a human population. *Radiology* 219(1): 51-59, 2001.
- [4] Bielen, D., Kiss, G. Computer-aided detection for CT colonography: Update 2007. *Abdominal Imaging* 32(5): 571-581, 2007.
- [5] Taylor, S., Halligan, S., Burling, D., Morley, S., Bassett, P., Atkin, W., Bartram, C. CT colonography: Effect of experience and training on reader performance. *European Radiology* 14(6): 1025-1033, 2004.
- [6] Gokturk, S.B., Tomasi, C., Acar, B., Beaulieu, C.F., Paik, D.S., Jeffrey Jr., R.B., Yee, J., Napel, S. A statistical 3-d pattern processing method for computer-aided detection of polyps in CT colonography. *IEEE Transactions on Medical Imaging* 20(12): 1251-1260, 2001.
- [7] Kiss, G., Van Cleynenbreugel, J., Marchal, G., Suetens, P. Computer aided detection in CT colonography, via spin images. *Lecture Notes in Computer Science, Proc. MICCAI 3217/2004*: 804-812, 2004.
- [8] Näppi, J., Yoshida, H. Feature-guided analysis for reduction of false positives in CAD of polyps for computed tomographic colonography. *Medical Physics* 30(7): 1592-1601, 2003.
- [9] Paik, D.S., Beaulieu, C.F., Rubin, G.D., Acar, B., Jeffrey Jr., R.B., Yee, J., Dey, J., Napel, S. Surface normal overlap: A computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT. *IEEE Transactions on Medical Imaging* 23(6): 661-675, 2004.
- [10] Zalis, M.E., Barish, M.A., Choi, J.R., Dachman, A.H., Fenlon, H.M., Ferrucci, J.T., Glick, S.N., Laghi, A., Macari, M., McFarland, E.G., Morrin, M.M., Pickhardt, P.J., Soto, J., Yee, J. CT colonography reporting and data system: A consensus proposal. *Radiology* 236(1): 3-9, 2005.
- [11] Copel, L., Sosna, J., Kruskal, J.B., Raptopoulos, V., Farrell, R.J., Morrin, M.M. CT colonography in 546 patients with incomplete colonoscopy. *Radiology* 244(2): 471-478, 2007.

- [12] Kiraly, A.P., Laks, S., Macari, M., Geiger, B., Bogoni, L., Novak, C.L. A fast method for colon polyp detection in high-resolution CT data. *International Congress Series, Proceedings CARS 2004* 1268: 983-988, 2004.
- [13] Lieberman D, Moravec M, Holub J, Michaels L, Eisen G. Polyp Size and Advanced Histology in Patients Undergoing Colonoscopy Screening: Implications for CT Colonography. *Gastroenterology* 135:1100–1105, 2008.
- [14] Summers, R.M., Yao, J., Pickhardt, P.J., Franaszek, M., Bitter, I., Brickman, D., Krishna, V., Choi, R. Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population. *Gastroenterology* 129(6): 1832-1844, 2005.
- [15] Yoshida, H., Näppi, J. CAD in CT colonography without and with oral contrast agents: Progress and challenges. *Computerized Medical Imaging and Graphics* 31(4-5): 267-284, 2007.
- [16] Luboldt, W., Tryon, C., Kroll, M., Toussaint, T.L., Holzer, K., Hoepffner, N., Vogl, T.J. Automated mass detection in contrast-enhanced CT colonography: An approach based on contrast and volume. *European Radiology* V15(2): 247-253, 2005.
- [17] Yao, J., Miller, M., Franaszek, M., Summers, R.M. Colonic polyp segmentation in CT colonography-based on fuzzy clustering and deformable models. *IEEE Transactions on Medical Imaging* 23(11): 1344-1352, 2004.
- [18] Yoshida, H., Dachman, A.H. Computer-aided diagnosis for CT colonography. *Seminars in Ultrasound, CT, and MRI* 25(5): 419-431, 2004.
- [19] Lauenstein, T., C. Mr colonography: Current status. *European Radiology* V16(7): 1519-1526, 2006.
- [20] Morcos, S., Thomsen, H. Adverse reactions to iodinated contrast media. *European Radiology* V11(7): 1267-1275, 2001.
- [21] Summers, R.M., Huang, A., Yao, J., Campbell, S.R., Dempsey, J.E., Dwyer, A.J., Franaszek, M., Brickman, D.S., Bitter, I., Petrick, N., Hara, A.K.. Assessment of polyp and mass histopathology by intravenous contrast-enhanced CT colonography. *Academic Radiology* 13(12): 1490-1495, 2006.
- [22] Pilkinton, D., Bitter, I., Summers, R.M., Campbell, S., Choi, J.R., Pickardt, P.J. The effect of edge-preserving image smoothing on automatic colonic polyp detection for CT colonography. *Proceedings of the SPIE*, 6143: 984-991, 2006.
- [23] Sereda, P., Bartroli, A.V., Serlie, I., Gerritsen, F.A. Visualization of boundaries in volumetric data sets using lh histograms. *IEEE Transactions on Visualization and Computer Graphics* 12(2): 208-218, 2006.

- [24] Serlie, I., Truyen, R., Florie, J., Post, F., Van Vliet, L.J., Vos, F. Computed cleansing for virtual colonoscopy using a three-material transition model. *Lecture Notes in Computer Science, Proc. MICCAI 2879/2003*: 175-183, 2003.
- [25] Serlie, I., Vos, F.M., Truyen, R., Post, F.H., Van Vliet, L.J. Classifying CT image data into material fractions by a scale and rotation invariant edge model. *IEEE Transactions on Image Processing* 16(12): 2891-2904, 2007.
- [26] Van Gelder, R.E., Nio, C.Y., Florie, J., Bartelsman, J.F., Snel, P., De Jager, S.W., Van Deventer, S.J., Laméris, J.S., Bossuyt, P.M.M., Stoker, J. Computed tomographic colonography compared with colonoscopy in patients at increased risk for colorectal cancer. *Gastroenterology* 127(1): 41-48, 2004.
- [27] Pickhardt, P.J., Choi, J.R., Hwang, I., Butler, J.A., Puckett, M.L., Hildebrandt, H.A., Wong, R.K., Nugent, P.A., Mysliwiec, P.A., Schindler, W.R. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med* 349(23): 2191-2200, 2003.
- [28] Van Wijk, C., Van Ravesteijn, V.F., Vos, F.M., Truyen, R., De Vries, A.H., Stoker, J., Van Vliet, L.J. Detection of protrusions in curved folded surfaces applied to automated polyp detection in CT colonography. *Lecture Notes in Computer Science, Proc. MICCAI 4191/2006*: 471-478, 2006.
- [29] Pekar, V., Wiemker, R., Hempel, D. Fast detection of meaningful isosurfaces for volume data visualization. *Proceedings Visualization VIS '01*: 223-230, 2001.
- [30] Masutani, Y., Yoshida, H., MacEneaney, P.M., Dachman, A.H. Automated segmentation of colonic walls for computerized detection of polyps in CT colonography. *Journal of Computer Assisted Tomography* 25(4): 629-638, 2001.
- [31] Pham, T.Q., Van Vliet, L.J. Separable bilateral filtering for fast video preprocessing. *IEEE Int. Conf. on Multimedia & Expo (ICME)*: 1-12, 2005.
- [32] Tomasi, C., Manduchi, R. Bilateral filtering for gray and color images. *IEEE Int. Conf. on Computer Vision*: 839-846, 1998.
- [33] Soille, P. *Morphological image analysis: Principles and applications* Springer-Verlag Telos, 1999.
- [34] Vining, D.J., Ge, Y., Ahn, D.K., Stelts, D.R. Virtual colonoscopy with computer-assisted polyp detection. *Computer-Aided Diagnosis in Medical Imaging, Amsterdam, The Netherlands*. 445-452, 1999.
- [35] Yao, J., Summers, R.M. Adaptive deformable model for colonic polyp segmentation and measurement on CT colonography. *Medical Physics* 34(5): 1655-1664, 2007.

- [36] Johnson, C.D., Toledano, A.Y., Herman, B.A., Dachman, A.H., Mcfarland, E.G., Barish, M.A., Brink, J.A., Ernst, R.D., Fletcher, J.G., Halvorsen, R.A., Hara, A.K., Hopper, K.D., Koehler, R.E., Lu, D.S.K., Macari, M., Maccarty, R.L., Miller, F.H., Morrin, M., Paulson, E.K., Yee, J., Zalis, M. Computerized tomographic colonography: Performance evaluation in a retrospective multicenter setting. *Gastroenterology* 125(3): 688-695, 2003.
- [37] Johnson, K.T., Harmsen, W.S., Limburg, P.J., Carston, M.J., Johnson, C.D. Visceral fat analysis at CT colonography. *Academic Radiology* 13(8): 963-968, 2006.
- [38] Burling, D., Halligan, S., Altman, D., Atkin, W., Bartram, C., Fenlon, H., Laghi, A., Stoker, J., Taylor, S., Frost, R., Dessey, G., De Villiers, M., Florie, J., Foley, S., Honeyfield, L., Iannaccone, R., Gallo, T., Kay, C., Lefere, P., Lowe, A., Mangiapane, F., Marrannes, J., Neri, E., Nieddu, G., Nicholson, D., O'Hare, A., Ori, S., Politi, B., Poulus, M., Regge, D., Renaut, L., Rudralingham, V., Signoretta, S., Vagli, P., Van der Hulst, V., Williams-Butt, J. Polyp measurement and size categorisation by CT colonography: Effect of observer experience in a multi-centre setting. *European Radiology* 16(8): 1737-1744, 2006.