Improving initial polyp candidate extraction for CT colonography

This article has been downloaded from IOPscience. Please scroll down to see the full text article.


(http://iopscience.iop.org/0031-9155/55/7/019)

The Table of Contents and more related content is available

Download details:
IP Address: 129.49.137.134
The article was downloaded on 19/03/2010 at 15:13

Please note that terms and conditions apply.
Improving initial polyp candidate extraction for CT colonography

Hongbin Zhu, Yi Fan, Hongbing Lu and Zhengrong Liang

1 Department of Radiology, State University of New York, Stony Brook, NY 11794, USA
2 Department of Biomedical Engineering, Fourth Military Medical University, Xi’an, Shaanxi 710032, People’s Republic of China
E-mail: jzl@mil.sunysb.edu

Received 23 August 2009, in final form 4 February 2010
Published 19 March 2010
Online at stacks.iop.org/PMB/55/2087

Abstract
Reducing the number of false positives (FPs) as much as possible is a challenging task for computer-aided detection (CAD) of colonic polyps. As part of a typical CAD pipeline, an accurate and robust process for segmenting initial polyp candidates (IPCs) will significantly benefit the successive FP reduction procedures, such as feature-based classification of false and true positives (TPs). In this study, we introduce an improved scheme for segmenting IPCs. It consists of two main components. One is geodesic distance-based merging, which merges suspicious patches (SPs) for IPCs. Based on the merged SPs, another component, called convex dilation, grows each SP beyond the inner surface of the colon wall to form a volume of interest (VOI) for that IPC, so that the inner border of the VOI beyond the colon inner surface could be segmented as convex, as expected. The IPC segmentation strategy was evaluated using a database of 50 patient studies, which include 100 scans at supine and prone positions with 84 polyps and masses sized from 6 to 35 mm. The presented IPC segmentation strategy (or VOI extraction method) demonstrated improvements, in terms of having no undesirably merged true polyp and providing more helpful mean and variance of the image intensities rooted from the extracted VOI for classification of the TPs and FPs, over two other VOI extraction methods (i.e. the conventional method of Nappi and Yoshida (2003 Med. Phys. 30 1592–601) and our previous method (Zhu et al 2009 Cancer Manag. Res. 1 1–13). At a by-polyp sensitivity of 0.90, these three methods generated the FP rate (number of FPs per scan) of 4.78 (new method), 6.37 (Nappi) and 7.01 (Zhu) respectively.

(Some figures in this article are in colour only in the electronic version)
1. Introduction

According to up-to-date statistics from the American Cancer Society (ACS 2008), colorectal cancer ranks as the third most common cause of both cancer deaths and new cancer cases in 2008 for both men and women in the United States. Fortunately, early detection and removal of colonic polyps prior to their malignant transformation can effectively decrease the incidence of colon cancer (Eddy 1990). As a potential minimally invasive screening technique, computed tomographic colonography (CTC) or CT-based virtual colonoscopy (VC) has shown several advantages over traditional optical colonoscopy (OC) (Pickhardt et al. 2003). To improve the efficiency of CTC in detecting polyps, computer-aided detection (CAD) of polyps has shown potential of being a second reader assisting physicians for finding polyps in the colon (Summers et al. 2005).

Generally, starting from a segmented colon wall, most of the current CAD schemes detect initial polyp candidates (IPCs) represented by a group of voxels, i.e. a volume of interest (VOI), based on which the CAD schemes conduct feature analysis for the purpose of reducing false positives (FPs) (Bielen and Kiss 2007). To generate the IPCs, several techniques have been reported. First, voxels on the colon wall with a possibility of being part of a polyp candidate were labeled or filtered, based on various heuristic rules like curvature (Summers et al. 2000, Yao et al. 2004), shape index and curvedness (Yoshida and Nappi 2001, Nappi and Yoshida 2003, Suzuki et al. 2008), HT (Hough transformation) score (Acar et al. 2002) and surface normal overlap (Paik et al. 2004) in a certain value range. These voxels were then clustered based on the spatial adjacency to form suspicious patches (SPs). Since multiple SPs might be generated for one colonic object, SPs which are close but not connected to each other were merged if the Euclidean distance (ED) between them was less than a predefined threshold (Yoshida and Nappi 2001, Acar et al. 2002, Nappi and Yoshida 2003, Wang et al. 2005, 2008, Suzuki et al. 2006, 2008, Zhu et al. 2009). Finally, the VOIs of IPCs were generated with different methods. For example, in the studies (Summers et al. 2000, Paik et al. 2004, Bhotika et al. 2006), the SPs were directly utilized as the VOIs. Alternatively, a cubic or rectangular sub-volumes centered at the centroid of such unmerged (Yao et al. 2004) or merged (Acar et al. 2002, Suzuki et al. 2006, 2008) SPs were assumed as the VOIs of IPCs for the purpose of including more voxels of the polyp candidates. To reduce the redundant voxels in the above sub-volumes, non-rectangular volume areas created by sphere fitting (Kiss et al. 2002), three-dimensional (3D) surface fitting (Chowdhury et al. 2006) and ellipsoid fitting (Wang et al. 2005) were explored. However, some lumen voxels might still be included in such fitted volume area. Loyal to the truth that the contrast at the SP–lumen interface (outer border) is conspicuous enough to be detected reliably, several researchers have been focusing on constructing the interface between the polyp candidates and the normal tissue (inner border) for an adequate VOI of IPC. For example, some researchers grew the SPs toward the tissue area with region growing techniques, like conditional morphological dilation (Nappi and Yoshida 2003), while others believed that the contrast at the interface of polyp candidates and normal tissue was strong enough for them to find the inner border by the use of edge-detectors like (i) Harr transform-based edge finder (Wang et al. 2005, Zhu et al. 2009), (ii) combination of canny edge detection and Radon transform (Jerebko et al. 2003) and (iii) deformable models like the curvature-driven level set evolution (van Wijk et al. 2006) and the active contour model (Yao et al. 2004).

For adequate VOI extraction, correct merging of SPs and accurate segmentation of the inner border would be critical but still problematic. The merging criteria employed in the above methods are all based on the ED. However, the ED cannot serve as the criteria alone to determine whether multiple SPs belong to one object or not since ED ignores the height
Improving initial polyp candidate extraction for CT colonography

Figure 1. The outline of a CAD scheme to evaluate the two strategies presented.

information of SPs. Therefore, we introduce a geodesic distance (GD)-based merging (GDM) strategy in this study to address this problem. Additionally, in CTC images, the image contrast between a polyp and its surrounding normal tissue could be too small to be detected by those edge detectors (Jerebko et al 2003, Yao et al 2004, Wang et al 2005, van Wijk et al 2006, Zhu et al 2009). On the other hand, the morphological dilation method (Nappi and Yoshida 2003), though it did not use any edge detector, might generate the concave inner border (Zhu et al 2009), while the inner border is often expected to be flat or convex (Haggitt et al 1985). Therefore, we introduce a convex dilation (CD) strategy to ensure that a flat and/or convex inner border will be generated for the final VOI of each IPC, and we hope such VOI can benefit the successive FP reduction procedures. The presented GDM and CD strategies are adapted to our previous CAD pipeline (Zhu et al 2009) and then evaluated using a CTC database.

The remainder of this paper is organized as follows. Section 2 introduces the two strategies for VOIs of IPCs. Section 3 reports the evaluation results. Several conclusions are drawn in section 4 through some discussions.

2. Method

To facilitate the presentation of the two strategies, the whole CAD scheme is outlined in figure 1. At the beginning, voxels are labeled and clustered based on spatial adjacency to form the SPs (the details are given, e.g., in Zhu et al (2009)). These SPs are merged with the GDM strategy and then grow toward the inner border with the CD strategy. To evaluate the strategies, several widely used features are extracted from the generated VOIs of the IPCs and fed into the well-known support vector machine (SVM) for FP reduction.

2.1. Geodesic distance-based merging (GDM) strategy

Frequently, multiple SPs can be generated for one polyp candidate, especially for those lobulated ones. Figure 2 shows an example of three SPs being generated for a 35 mm lobulated tubulovillous adenoma. Therefore, SPs located within a merging distance \( L_m \), which was the ED like 12.5 mm in Yoshida and Nappi (2001), Nappi and Yoshida (2003), Suzuki et al (2008), Wang et al (2008) and Zhu et al (2009) and 10 mm in Acar et al (2002), were merged together as one. However, two or more unrelated SPs may be mistakenly merged. Figure 3(a) shows an example of stool standing closely to a polyp in the same colon segment, while figures 3(b)–(d) show another example of stool locating closely to a polyp in different segments of the colon. Each of the two stools shown in figure 3 is merged to the related polyp to form two IPCs if \( L_m \) is set to be 12.5 mm, which will undoubtedly impair the performance of the following FP reduction.

The reason for such undesirable merging lies in that the ED between the two objects just reflects their smallest spatial distance, but ignores the height information of the objects standing out of their surroundings. Such conjecture can be illustrated by the 2D pictures in figure 4, where both the EDs from \( \text{P1} \) to \( \text{P2} \) and from \( \text{P3} \) to \( \text{P4} \) are smaller than the merging distance \( L_m \) of 12.5 mm, and therefore, \( \text{P1} \) and \( \text{P3} \) shall be merged with \( \text{P2} \) and \( \text{P4} \),
Figure 2. The endoluminal or endoscopic view of a 35 mm lobulated tubulovillous adenoma near the rectum of a 54 year old female. The three arrows indicate three separated SPs, respectively, which are detected for this single polyp due to the three lobules.

Figure 3. (a) The endoluminal view of a 9 mm polyp and stool which are close to each other. (b) A 6 mm polyp and stool in the 2D axial slice. Obviously, they are located in different segments of the colon. (c) The endoluminal view of the two colonic objects in (b), where the view angle is set mainly for the polyp. (d) Similar to (c), but at another view angle for the stool.
Improving initial polyp candidate extraction for CT colonography

Figure 4. The two dark circles represent two major objects, on which there are two bumps (mimicking polypoid-like objects) in light gray, respectively. Parts of the bumps encompassed by the closed curves are the peak and neck areas of the objects, denoted by $P_1$–$P_4$, which represent the SPs detected by an analysis of each voxel on the borders of the major objects.

$\text{LED} < \text{LM}_G$ \
$\text{LED} < \text{LM}_G$ \
$\text{LED} < \text{LM}_G$ \
$\text{LED} < \text{LM}_G$

respectively, according to the previous works. By inspecting these two pictures in figure 4, we would conclude that the ED from $P_3$ to $P_4$ is similar to that from $P_1$ to $P_2$; however, their relative larger heights make them much more likely to be different objects ($P_3$ and $P_4$) standing on the underlying major object (the dark circle in figure 4) than other two objects ($P_1$ and $P_2$). The SPs $P_1$ and $P_2$ appear more likely to be lobules of their underlying major object (the dark circle) and may be merged, but $P_3$ and $P_4$ should not be merged.

Based on the above analysis, we note that the height of the SPs should also be considered in the merging process. The GD between any two points on a manifold (the colon wall in this study) is the length of the shortest path on the manifold connecting them, i.e. the geodesic path (Kimmel and Sethian 1998). In figure 4, the geodesic paths from $P_1/P_3$ to $P_2/P_4$ are the gray and dark curves rather than the straight lines (i.e. EDs) connecting them, whose lengths (i.e. GDs) reflect the height information of the SPs according to the underlying major object. Therefore, we introduce a GDM strategy to improve the SP merging process. Figure 5 shows a flow chart for the GDM strategy, where $\text{LED}$ and $\text{GD}$ denote the ED and GD respectively between any two SPs. Notation $\gamma$ represents a constant threshold. The first condition inherits the criteria of the previous works (Yoshida and Nappi 2001, Wang et al. 2008, Zhu et al. 2009). The second criteria are introduced here for the rationale that if the ratio of the GD to ED between any two SPs is small enough (the values of GD and ED are similar), the height of the patches would be relatively small and they should be conceptually considered as lobules of one object; otherwise, they would be different objects and should not be merged.

The GD can be computed accurately using the Dijkstra method (Dijkstra 1959) if taking the manifold as a graph (Kimmel and Sethian 1998). In this study, for simplicity, the GD is approximated by the number of steps (denoted with $n_D$) of a geodesic dilation, which starts from one SP and ends at another with the dilation process limited only on the colon wall. After applying the geodesic dilation with $n_D$ steps, the generated geodesic path would be a sequence of $n_D+1$ voxels. As shown in figure 6, two extreme cases might happen: (1) all the successive voxel pairs on the path are six-connected (the solid arrow) and (2) all the successive voxel pairs on the path are diagonally connected (the dashed arrow). These two cases result in the smallest and largest possible lengths of the geodesic path with fixed $n_D$ respectively.
Therefore, an accurate GD would satisfy \( n_D \cdot \delta \leq L_{GD} \leq \sqrt{3} \cdot n_D \cdot \delta \), where \( \delta \) represents the resolution of the CTC image. In practice, the GD was simply approximated as the average of the two extreme cases

\[
L_{GD} = \begin{cases} 
\sqrt{3+1} \cdot \frac{n_D \cdot \delta}{2} & \text{if } n_D < \kappa \\
\infty & \text{otherwise},
\end{cases}
\]

(1)

where \( \kappa \) is a predefined constant used to stop the geodesic dilation for the case that the GD is extremely large, such as the stool and the polyp shown in figure 3(b). This upper bound is used to cover all the possibilities when applying the second criteria in figure 5, where the constant \( \gamma \) reflects the relative height of the SPs standing on the underlying major object for determining whether they are conceptually two objects or just parts of a single object.

2.2. The convex dilation (CD) strategy

Based on the polyp pathology in general (Haggitt et al. 1985), Wang et al. (2005) introduced the Harr transform-based fitting method to extract an elliptical inner border to satisfy the expectation that the inner border should be convex or flat. As mentioned in section 1, the small contrast might fail the Harr transform-based edge finder (figure 11(d)); thus in this study, we introduce a convex dilation (CD) strategy to extract the VOI based on the conditional morphological dilation method (Nappi and Yoshida 2003). For illustration, we start from a brief introduction of the method in Nappi and Yoshida (2003). An initial VOI is extracted by dilating for a few steps from a merged SP. The set of voxels \( \{p_i\} \) forming the initial VOI is denoted by \( V_0 \). Then the dilation process continues by a manner as depicted by \( V_{n+1} = V_n \cup N(V_n) \), where \( V_n \) represents the set of voxels at the step \( n \) of the dilation and \( N(V_n) \) represents the set of voxels in the soft tissues (excluding those in lumen) adjacent but not belonging to \( V_n \), i.e.

\[
N(V_n) = \{ p_i \mid p_i \in V_n^T \text{ and } p_i \notin V_n \},
\]

(2)

where \( V_n^T \) represents all the neighbors of \( V_n \) in the tissue region. Such an iterative procedure would stop at a predefined maximum step \( n_M \) if a minimum growth rate could not be detected at any intermediate step \( n_m < n_M \).

As mentioned by Zhu et al. (2009), the inner borders of the VOIs generated by the above dilation process would not be convex when the minimum growth rate could not be detected for sessile and flat lesions (figure 11(c)). In this paper, instead of using the Harr transformation-based edge finder (Wang et al. 2005, Zhu et al. 2009), we add an additional constraint to the aforementioned dilation process to address the problem, where \( N(V_n) \) is rewritten as

\[
N(V_n) = \{ p_i \mid p_i \in V_n^T \text{ and } p_i \notin V_n \text{ and } \zeta(p_i) > \alpha \},
\]

(3)
Improving initial polyp candidate extraction for CT colonography

Figure 7. Illustration of the CD strategy in the 2D case. Following the dilation direction (denoted by the arrow), the dilation process proceeds from the upper left area (pixels belong to \( V_n \)) to the lower gray area (pixels in the tissue area outside \( V_n \)). Pixels on the interface of these two areas represent the neighbors of \( V_n \).

where \( \alpha \) is a constant threshold, and \( \zeta(p_i) \) is defined as

\[
\zeta(p_i) = \frac{\sum_{|p_j - p_i| < r} H(p_j)}{\sum_{|p_j - p_i| < r} [1 - H(p_j)]},
\]

where \( H(p_j) \) is a step function

\[
H(p_j) = \begin{cases} 
1 & p_j \in V_n \\
0 & p_j \notin V_n.
\end{cases}
\]

The underlying rationale of the above equations can be illustrated by a 2D case as shown in Figure 7. The upper left area represents the current \( V_n \). Considering pixel \( p_1 \) during the iteration at step \( n+1 \), the numerator and denominator on the right-hand side of equation (4) actually measure the areas of the upper left and lower (including the interface) regions, respectively, in the circle with the radius \( r \). Therefore, \( \zeta(p_1) \) measures the area ratio of the two regions. Straightforwardly, if such a ratio is larger than 1, the upper left area is locally concave at pixel \( p_1 \). If it is less than or equal to 1 (i.e. the case at \( p_2 \)), the area is locally convex (at \( p_2 \)). As a result, if the parameter \( \alpha \) in equation (3) is set to be less than 1, the dilation process can make sure that the upper left area will grow into the lower area at concave pixels (such as \( p_1 \)), but not at convex pixels (such as \( p_2 \)). The dilation process (equation (3)) will stop automatically when \( \zeta(p_i) \) is less than \( \alpha \) at all pixels on the inner border. Therefore, the final dilated region will be convex everywhere.

Two parameters are involved in the above equations, i.e. the constant threshold \( \alpha \) and the radius \( r \). The first parameter determines the final convexity, which denotes how convex the final inner border would be with a smaller value resulting in a more convex inner border. The second one reflects the degree of spatial locality, which depends on the size of the object under consideration. Straightforwardly, a larger radius should be used for a larger object. Theoretically, the curvedness represents the size of the object (Dorai and Jain 1997). However, it is hard to get the meaningful curvedness on the colon wall (Nappi et al. 2001). Fortunately, the size of the object can be roughly reflected by the volume of the merged SP since a larger SP will be generated for a larger lesion candidate; hence, a proper value can be assigned to \( r \) accordingly (to be described in section 3).
3. Results

3.1. CTC database

The two strategies presented were adapted to our previous CAD pipeline (Zhu et al 2009), as shown in figure 1, for evaluation. The pipeline was applied to a CTC database including 50 patient studies collected with IRB approval. The patients were aged from 50 to 80 years. Each patient was scanned at both supine and prone positions by 4- and 8-MDCT scanners (Light Speed Ultra, GE Medical Systems, Milwaukee, WI), resulting in 100 CT scans. The scanning protocol included mAs modulation in the range of 120–216 mA with kVp of 120–140 values, 1.25–2.5 mm collimation and reconstruction interval of 1 mm. The slice thickness of the CTC images ranged from 0.96 to 1.25 mm, and the in-plane pixel size from 0.53 to 0.76 mm. In the database, a total of 84 clinically significant polyps and masses (larger than 30 mm), sized in the range of 6–35 mm, were confirmed by both optical and virtual colonoscopies. The size distribution of these polyps is shown in figure 8, and the numbers of sessile, pedunculated and flat polyps were 37, 40 and 7 respectively. In this evaluation, the supine and prone scans of each patient study were considered as different datasets, and the presented strategies were evaluated in terms of by-polyp detection sensitivity and the FP rate denoting the number of FPs per dataset or CT scan. Therefore, there were 168 polyp views corresponding to true polyps and masses in 100 CT scans in the CTC database. For short, we just use polyp instead of polyp view hereafter to indicate all the polyp views exhibited in the supine and prone scans, and the terms ‘sensitivity’ and ‘FP rate’ in the following text mean the by-polyp sensitivity and the number of FPs per dataset.

As the reference, the center coordinates of polyps or masses were determined by two radiologists. To label a CAD detection as a TP or FP, we employed a two-step procedure. In the first step, an initial label was automatically put on each IPC. In this step, an IPC was considered as a TP candidate if its ED to the reference location of a true polyp was less than $L_{TP}$. Within $L_{TP}$ to the reference location, if there was no TP candidate, the true polyp was actually missed by the CAD algorithm; otherwise, the one with the smallest ED would be marked as TP, and all others were FPs. All CAD detections farther than $L_{TP}$ to all of the reference locations were directly taken as FPs. In the experiments, $L_{TP}$ was set to be 20 mm to include the large masses. We noted that wrong labeling might occur if an FP was the
Improving initial polyp candidate extraction for CT colonography

closest IPC to a true polyp within 20 mm. Therefore, in the second step, two radiologists double checked the labels of the IPCs near each true polyp, and wrong labelings were rectified manually. As a result, a 7 mm flat polyp was missed (false negative), and all other true polyps were successfully detected and labeled as TPs.

3.2. Evaluation of the strategies

The initial suspicious patches (SPs) were detected by the use of the initial detector in our previous CAD pipeline (Zhu et al 2009). The merging of the initial SPs was then performed by the above-presented GDM strategy for an IPC. For each merged SP, its VOI was segmented by the presented CD strategy. Features were then extracted from the VOIs to differentiate TPs and FPs from all the IPCs. The differentiation task was accomplished by the use of the SVM classifier. The gain by the presented GDM and CD strategies was shown in comparison with our previous IPC detection methods.

3.2.1. The geodesic distance-based merging (GDM) strategy. In the GDM strategy, we extended $L_m$ to be 20 mm in order to cover all possible polyp sizes. The predefined constant $\kappa$ was determined for lesions as large as 40 mm (any lesion larger than 40 mm is expected to be detected easily without CAD help). The approximated GD should be larger than this key size $\kappa$. Therefore, $\kappa$ was evaluated as $\lceil \frac{80}{[(\sqrt{3}+1)\delta]} \rceil + 1$, where $\lceil x \rceil$ denotes the smallest integer which is larger than $x$.

In the evaluation, any case in which the polyp lobules (as shown in figure 2) were not merged was referred to as an unmerged polyp (UMP), i.e. a failure. On the other hand, any case in which neighboring polyps or non-polyp structures were merged with a polyp (as shown in figure 3) was referred to as a wrongly merged polyp (WMP), another kind of failure. In the context of VOI extraction for polyp detection, we are only interested in the correct segmentation of true polyps and we do not care about those unmerged or wrongly merged non-polyp structures. Therefore, we will focus on the detection of the true polyps and the reduction of FPs. Under this condition, all the generated initial SPs by the initial detector in Wang et al (2008) and Zhu et al (2009) were input into the GDM module shown in figure 1 with various $\gamma$ values, see figure 5. The incidences of UMPs and WMPs were calculated by per polyp. Multiple unmerged detections of one polyp were counted as one case of UMP. However, if $n$ polyps were involved in a wrongly merging scenario, $n$ cases of WMP would be counted. For example, if one polyp was merged with a stool, it would be counted as one WMP. If two polyps were merged with a stool, it would be counted as two WMPs.

Figure 9 plots the incidences of UMPs and WMPs according to different $\gamma$ by experiments. The worst cases were 18 UMPs (with the incidence of 10.7%) and 27 WMPs (with the incidence of 16.1%) for the wide range value of $\gamma$. When $\gamma$ is in the range [1.20, 1.24], there is no failure. For comparison, we removed the second criteria shown in figure 5, so that the above-presented GDM strategy could be reduced to exactly the same as that in the previous works with different $L_m$ (Yoshida and Nappi 2001, Nappi and Yoshida 2003, Wang et al 2008, Zhu et al 2009). Figure 10 plots the incidences of UMPs and WMPs according to various $L_m$. When $L_m = 11.5$ mm, 22 UMPs and 21 WMPs were generated with the incidences to be 13.1% and 12.5% respectively.

3.2.2. The convex dilation strategy. As discussed in section 2.2, the threshold $\alpha$ determines the final convexity of the inner border of a VOI. However, there is no theory or clinical reports about how convex the actual inner border of the polyp would be. In this study, we selected the value of $\alpha$ to be 0.9 by visually inspecting all the polyps in the CTC database. As for the
other parameter, i.e. the radius $r$, its value can be roughly estimated according to the volume of each merged SP. By experiments, we found that $r = 5$ was adequate for merged SPs with a volume smaller than 40 mm$^3$, which indicates a polyp of intermediate size (6–9 mm), while a larger value $r = 9$ worked well for larger polyps.

For comparison, two previous VOI extraction methods were implemented: (1) the conditional morphological dilation method (Nappi and Yoshida 2003) and (2) the Harr transform-based inner border extraction method (Zhu et al 2009). In the former, the diameter
Improving initial polyp candidate extraction for CT colonography

**Figure 11.** An experimental result of the CD strategy. (a) A 12 mm sessile polyp in an axial slice, where the voxels highlighted on the tip indicate the SP. (b) The generated initial VOI $V_0$ in the slice (including the SP area). (c) The result of the previous method in Nappi and Yoshida (2003). (d) The result of the previous method in Zhu *et al* (2009), where the intermediate curve growing into the normal tissue represents the inner border found by the Harr transformation-based edge finder. (e) The result of the presented CD strategy with $r = 9$ and $\alpha = 0.9$.

of the structuring element was set to be one voxel as in Nappi and Yoshida (2003), while the maximum number of dilation steps was constrained by 5.5 mm instead of 7.5 mm in the previous work. Through visual inspection on all polyps in the CTC database, we found that the smaller number of dilation steps excluded more redundant voxels. As for the latter method, it had no control parameters and was directly applied to all the IPCs.

Together with the CD strategy, there were three VOI extraction methods involved in the comparison. Figure 11 shows an example of VOI extraction results for a 12 mm sessile polyp. From figure 11(a), it was difficult to visually figure out a meaningful inner border of the polyp. Figure 11(b) shows the dilated initial VOI, $V_0$ (referred to in section 2.2), and figure 11(c) is the result of the original method (Nappi and Yoshida 2003) with the concave inner border. Figure 11(d) shows the result by using the Harr transformation-based edge finder (Zhu *et al* 2009). Obviously, the VOI was over-estimated, and the internal information of the VOI, e.g., the image intensity distribution, might be overwhelmed by the large amount of normal tissues. The CD strategy gave a more reasonable result as shown in figure 11(e) with a convex inner border as expected (Haggitt *et al* 1985).

Based on the VOIs generated by the three methods for the IPCs from the corresponding merged SPs, two texture features, the mean and variance of the CT intensities, were extracted since they were directly derived from the CT intensity distribution in the VOI and their outcomes would reflect the relative performance of different VOI extraction strategies. Therefore, there were three IPC sample pools from the 100 CT scans with two features per sample. For each sample pool, the 100 CT scans were randomly split into two equal-sized groups (50 scans in each group), and the feature vectors of the IPCs in the two groups formed the training and testing sets, respectively. The training set of each sample pool was trained by the SVM classifier with radial basis function (RBF) kernel, and the testing set was then tested for classification. The performances of the three VOI extraction procedures were measured by the free-response receiver operating characteristic (fROC) curves based on the testing results, where the vertical axis reflects the sensitivity and the horizontal axis represents the FP rate. The fROC curves were parameterized by the same decision thresholds for the three methods.

The standard SVM (Chang and Lin 2001) was geared for our specific CAD application (Zhu *et al* 2009) by (1) modifying the training goal to minimize the number of FPs at the specified detection sensitivity level of TPs, instead of maximizing the overall sensitivity, and (2) putting an optimal weight on TPs. The reason for these two modifications is that the training set of CAD of colon polyps is often heavily imbalanced with much more FPs than TPs, and the bias toward the FP class should be removed. In the SVM training process, a tenfold
leave-one-out cross-validation process was employed. Three parameters were involved: the cost function value, the RBF kernel parameter and the weight on TPs. The optimal values of them were determined with a 3D grid search method (actually a roughly exhausting process). More details are available in our previous work (Zhu et al. 2009).

So far, the biases in the training process were properly considered by the two modifications above and the tenfold cross-validation. However, even with the cross-validation in the training process, severely biased testing result was reported in previous works (Sundaram et al. 2008) when the training and testing sets were selected with bias. Therefore, we repeated the random grouping process for 20 times to overcome the bias of case selection when determining the training and testing sets. As a result, there were 20 fROC curves for each VOI extraction method. The fROC curves were averaged in terms of by-polyp sensitivities sampled at 0.05, 0.10, . . . , 0.95, as shown in figure 12, where the FP rates at the sampling points were interpolated using the cubic spline technique. In the experiment, extrapolation was used when averaging the FP rate at 95% sensitivity, since 95% was larger than the largest sensitivity levels in several trials. When all the largest sensitivities were greater than 90%, the error bars were put at the three operating points at a 90% sensitivity level, as shown in figure 12, representing the confidence intervals of the FP rate at the 95% confidence level. The confidence intervals were retrieved based on the 20 sampled FP rates at a 90% sensitivity level through using the bootstrap method (Davison and Hinkley 1997) with 10 000 repetitions. Table 1 lists the FP rates of the three methods at the three operating points, and table 2 lists the relative $p$-values through using the $z$-test with the samples after bootstrapping.

4. Discussion and conclusion

It can be seen in figure 9 that the incidence of UMPs decreases but that of WMPs increases with the threshold $\gamma$. This is because $\gamma$ denotes the ratio of the GD and ED between any

![Figure 12. The fROC curves of the testing procedures. The three curves, averaged from the 20 trials at sampling sensitivities of 0.05, 0.10, . . . , 0.95, depict the performance of the three VOI extraction methods using the two features. The three error bars denote the confidence intervals (95% confidence level) for the FP rate at the detection sensitivity of 90%.

2098 H Zhu et al

![Image 156x547 to 407x745]
Table 1. Performance of the three VOI extraction methods. Numbers in parentheses are the 95% confidence intervals.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FP rate 6.37 (6.07–6.87)</td>
<td>7.01 (6.56–7.36)</td>
<td>4.78 (4.51–5.01)</td>
</tr>
</tbody>
</table>

Table 2. p-Values of the three operating points, I, II and III of the methods in Nappi and Yoshida (2003), Zhu et al (2009) and the CD strategy respectively.

<table>
<thead>
<tr>
<th>I–II</th>
<th>II–III</th>
<th>III–I</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-Value</td>
<td>0.1105</td>
<td>3.16 × 10⁻⁵</td>
</tr>
</tbody>
</table>

two SPs and indicates how high the SPs stand on the colon wall. According to the flowchart shown in figure 5, a smaller \( \gamma \) discourages the merging process even though the SPs stand low on the colon wall and leads to more cases of unmerged lobules of a polyp. In contrast, a larger \( \gamma \) encourages the merging processes although the SPs stand high on the colon wall and leads to higher incidence of WMPs. In figure 9, the zero incidence of both UMPs and WMPs was achieved if \( \gamma \) was set to be about 1.22. However, the previous ED-based merging method with any \( L_{m} \) would always generate undesirable merging cases as shown in figure 10, e.g. the by-polyp incidences of UMPs and WMPs were 13.1% and 12.5% when \( L_{m} = 11.5 \text{ mm} \) in the experiments. Even more WMPs/UMPs were generated with larger/smaller \( L_{m} \).

The fROC curves in figure 12 show the averaged performance over 20 runs of the classifier on the features extracted from the VOIs by the three different methods. It is seen that the two curves of the methods of Nappi and Yoshida (2003) and Zhu et al (2009) lie closely and even overlap near the 70% sensitivity. The curve of the presented CD strategy obviously stays upon both of the other two curves. Qualitatively, the CD strategy for VOI extraction outperforms the other two methods. More quantitative conclusion should be investigated through quantitative analysis like JAFROC analysis (Chakraboty and Berbaum 2004).

As listed in table 1, at a detection sensitivity of 90%, the corresponding FP rates are 6.37, 7.01 and 4.78 respectively for the methods of Nappi and Yoshida (2003), Zhu et al (2009) and the presented CD strategy. The performance of the previous two methods does not differ significantly (p-value = 0.1105) from each other, but does differ from that of the CD strategy with the p-values (<0.001) given in table 2. Thus, it would be concluded that the two features from the VOIs generated by the CD strategy perform the best by yielding the lowest FP rate at the detection sensitivity of 90%.

It is noted that the method of Nappi and Yoshida (2003) performs slightly better than the method of Zhu et al (2009) according to figure 12 and table 1. This can possibly be explained as follows. The very low image contrast between polyp and normal tissue renders a challenge for the Harr transform-based edge finder, and the generated VOIs may include more normal tissue voxels (see figure 11(d)). These redundant voxels may dilute the characterizing capability of the two features, i.e. the mean and variance of the CT intensities, extracted from the VOI. However, it is unknown whether the method of Nappi and Yoshida (2003) would outperform the method of Zhu et al (2009) if other features are involved. We also noted that the control parameters of the three methods, e.g., the \( \alpha \) and \( r \) in this study, and the maximum number of dilation steps in Nappi and Yoshida (2003), were chosen by visually inspecting the entire database. These parameter values might not be optimal for extracting features with best performance, and potential bias might exist in the resulted fROC curves as shown in figure 12.
Performance according to various parameter values should be evaluated with fROC curves to get the optimal parameters.

The evaluation was based on a CTC database including 100 scans with 84 polyps and masses sized in the range of 6–35 mm (as shown in figure 8) and shaped variously (as detailed in section 3.1). The number of polyps is not small for evaluation, as compared to that (averaged about 32) surveyed in Sundaram et al (2008). However, as mentioned in section 3.2, there are several parameters dependent on the size of polyps, such as the merging distance $L_{mm}$, the constant $\kappa$ used in the GDM strategy and radius $r$ used in the CD strategy. In this study, the first two parameters were determined by the size of the largest lesion, i.e. 40 mm. Therefore, lobules of lesions larger than 40 mm might not be merged. The last parameter, i.e. the radius $r$, was designed for lesions of 6 mm and larger, and it should be refined for polyps smaller than 6 mm. Another parameter, the ratio $\gamma$ between GD and ED for two close patches in the GDM strategy, determines the relative height with which the two patches should be merged or not. We note that $\gamma \approx 1.22$ is the optimal value for the current database leading to zero incidence of UMP and WMP. However, it might be sub-optimal if applied to a different database. Some cases of UMP/WMP might appear if it is an under/overestimation of the true optimal value. Therefore, the parameters involved in the two techniques, the GDM and CD strategies, could be further refined with a larger polyp database.

Two radiologists were invited to double check the results of the first step in the TP/FP labeling process, where the TP/FP labels were automatically put on each IPC. Manual corrections followed if wrong labeling occurred. Such double check provides most confident labeling for the successive evaluation process of the CD strategy. However, it makes the labeling process not being fully automatic. In our future work, we will explore other criteria, like the overlap percentage of the detection to the ground truth, to achieve a fully automatic labeling process with confident outcome.

The evaluation experiments for the two technical components of this work, i.e. GDM and CD, were conducted on all polyps equal to or larger than 6 mm, and the results showed the advantages of the two components against the corresponding existing methods. However, such evaluation is limited to identify the performance of these two components in a single polyp group of all sizes. A more meaningful evaluation would be on multiple polyp groups, e.g., one of size range 6–9 mm versus another of size 10 mm and larger, since such two polyp groups have different clinical significance. This multiple group experimental design would be more necessary for evaluation of a whole CAD pipeline and is under progress (Zhu et al 2010).

In conclusion, compared to the ED-based merging method (Wang et al 2008, Zhu et al 2009), the GDM strategy can effectively reduce the incidence of the undesirable merging cases, like UMPs and WMPs, and even achieve zero incidence at certain circumstance. The CD strategy can ensure the convexity of the inner border of the IPCs to meet the expectation (Haggitt et al 1985). By using two simple, while VOI-related, features, the experimental results showed that the resulted VOIs, qualitatively at least, outperform those with concave inner borders (Nappi and Yoshida 2003) or generated by the intensity contrast-based edge finder (Zhu et al 2009). In our future work, other features rooting from the VOI will be utilized to validate the improvement of the presented strategies with a larger polyp database.

Acknowledgments

This work was partially supported by NIH grant #CA082402 and #CA120917 of the National Cancer Institute. Dr Lu was supported by the National Nature Science Foundation of China under grant no 60772020. The authors would like to acknowledge the use of the Viatronix
V3D-Colon Module, the helpful discussion with Matthew Barish, MD, Perry Pickhardt, MD, and Robert Richards, MD, on the convexity of polyp model, with John Chen, PhD, and Guanxiang Zhang, PhD, on the statistics in the evaluation process and the assistance from Hongyu Lu, PhD, and Chaijie Duan, PhD, on data processing. The authors thank the anonymous reviewers for their helpful comments.

References


American Cancer Society 2008 Cancer Facts and Figures 2008 (Atlanta, GA: American Cancer Society)


Chakraborty D and Berbaum K 2004 Observer studies involving detection and localization: modeling, analysis, and validation Med. Phys. 31 1331–30


Davison AC and Hinkley D V 1997 Bootstrap Methods and Their Application (Cambridge: Cambridge University Press)

Dijkstra E W 1959 A note on two problems in connection with graphs Numer. Math. 1 269–71


Eddy D M 1990 Screening for colorectal cancer Ann. Intern. Med. 113 373–84

Haggitt R C, Glotzbach R E, Soffer E E and Wruble L D 1985 Prognostic factors in colorectal carcinomas arising in adenomas: implications from lesions removed by endoscopic polypectomy Gastroenterology 89 328–36


Nappi J, Dachman A, MacEneaney P and Yoshida H 2001 Computer-aided detection of polyps in CT colonography: evaluation of volumetric features in differentiating polyps from false positives Int. Congress Ser. 1230 676–81


