

Accurate Polyp Segmentation for 3D CT Colonography Using Multi-Staged Probabilistic Binary Learning and Compositional Model*

Le Lu^{1,2} Adrian Barbu¹ Matthias Wolf² Jianming Liang² Marcos Salganicoff² Dorin Comaniciu¹

¹Integrated Data Systems Dept. Siemens Corporate Research, Princeton, NJ 08540

²Computer Aided Diagnosis Group, Siemens Medical Solutions USA, Malvern, PA 19355

Abstract

Accurate and automatic colonic polyp segmentation and measurement in Computed Tomography (CT) has significant importance for 3D polyp detection, classification, and more generally computer aided diagnosis of colon cancers. In this paper, we propose a three-staged probabilistic binary classification approach for automatically segmenting polyp voxels from their surrounding tissues in CT. Our system integrates low-, and mid-level information for discriminative learning under local polar coordinates which align on the 3D colon surface around detected polyp. More importantly, our supervised learning system has flexible modeling capacity, which offers a principled means of encoding semantic, clinical expert annotations of colonic polyp tissue identification and segmentation. The learning generality to unseen data is bounded by boosting [12, 11] and stacked generality [14]. Extensive experimental results on polyp segmentation performance evaluation and robustness testing with disturbances (using both training data and unseen data) are provided to validate our presented approach. The reliability of polyp segmentation and measurement has been largely increased to 98.2% (ie. errors $\leq 3\text{mm}$), compared with other state of art work [4, 15] of about 75% \sim 80%.

1 Introduction

Colon cancer is the number two cause of cancer death for both men and women combined, but it is one of the most preventable of cancers because doctors can identify and remove the pre-cancerous growth known as a polyp. Virtual colonoscopy (3D Computed Tomography Colonography, CTC) is emerging as a powerful screening tool because of its non-invasiveness, low cost and high sensitivity. “Flying through” the colon in 3D CTC is intuitive, but the physician needs to manually adjust the navigation speed and change the angle in order to see a polyp clearly. For example, a polyp may be hidden behind a colonic fold and thus could be missed during the physician’s visual inspection. Therefore research in computer aided detection (CAD) of polyps

in CTC has been very active and several such CAD systems [10, 4, 9, 16] have been proposed. Once detected either manually or automatically, the polyp must be measured and classified. This analysis is required as part of physician’s report. Studies show that there is a large variability in physicians’ measurements. Therefore an accurate, consistent and automated polyp measurement tool is highly demanded.

Polyp segmentation is defined as extracting and isolating a polyp from the colon wall at a given location. In addition to its significant value for polyp measurement in clinic practice, a polyp segmentation tool is also of great importance in computer aided detection of polyps. After a polyp candidate is extracted and segmented (or enhanced by our trained class-conditional probabilities), a number of features including intensity, texture and shape can be more precisely computed by incorporating trainable semantic knowledge and effectively used in the automated polyp detection, classification and diagnosis process [10, 4, 9, 16] to improve robustness. Furthermore, polyp segmentation is also essential in creating ground truth database for training, improving, testing and validating CAD systems. To address all of these needs, this paper presents a novel approach for accurate polyp segmentation using multi-staged probabilistic binary learning and compositional model.

Polyp segmentation is a very challenging task because polyps are abnormal growths from the colon wall and the “expected” segmentations are often a semantic, perceptual boundary with low imaging contrast support. Furthermore, polyps contain multiple shape categories (sessile, pedunculated, flat, etc.) with tremendously large 3D shape/appearance variation (refer to figure 1). All the existing methods for polyp segmentation in the literature can be regarded as unsupervised segmentation or clustering in 3D data volumes. The methods include fuzzy clustering [15], deformable model [15] or snakes [5], variational level-set method [7], or heuristic surface curvature constraints [4, 15]. As reported, these unsupervised approaches work well for up to 70% polyps due to the unclear polyp/nonpolyp boundary, large within-class polyp appearance/shape variations, or their limited heuristic shape assumptions [4, 15].

*Thanks Dr. Sarang Lakare, Siemens IKM-CKS, for useful discussion.

To overcome the aforementioned limitations, we leverage robust statistical learning of encoded expert knowledge annotated in a colon polyp database and a novel compositional representation by sampling and modeling a polyp surface using a collection of ordered 1D radial axes (namely polar coordinates). We will demonstrate that our polyp segmentation system can successfully measure various types (sessile, pedunculated, flat) of polyps (size ranging from 1.8mm to 30mm), with 20% ~ 25% higher reliability and accuracy compared with previous work [4, 15]. This superior performance is attributed to our two novel contributions:

- Our multistaged probabilistic learning framework, which decomposes a complex learning task as a sequence of better trainable sub-tasks. A local-to-global scaled 3D data evidences are gradually integrated with this learning process to achieve robustness.
- Our compositional model, which makes statistical learning practically feasible when applying to a highly complex 3D computer vision problem, by “dimension reduction¹” to avoid the well-known “curse of dimensionality” (i.e., the number of training samples verse the high dimensionality of 3D polyp surfaces).

The rest of the paper is organized as follows. After an overview in section 2, we detail our hierarchical multistaged binary learning representation and compositional framework for polyp segmentation in section 3. Extensive experimental results and comparison with previous work [15, 4] are given at section 4, where the robustness aspect of polyp tip detection and our hierarchical information stacking are also addressed. Finally we conclude the paper and discuss future work.

2 Overview

Motivation & Task: Polyp segmentation is to extract a polyp from a given initial position either clicked by a physician or provided as output by a CAD software using a closed 1D curve boundary in 3D CTC. In addition to the high workload and limited reading time, some physicians read the case in 2D and some in 3D, leading to a large variation in the initial detections. The state-of-art automated polyp detection systems [4, 17, 10, 15, 16] generate many false alarms and predict the polyp locations with limited accuracy. Our experiments with ~ 300 polyps in our database show that the detected polyp positions provided by a CAD software may have the error/inconsistency range of $3 \sim 12$ voxel distance against the labeled annotations. This large variability in the initial positions reduces the polyp measurement reliability significantly as reported in [17, 15]. Therefore, to compensate the above inaccuracy/inconsistency and improve the overall system’s performance, our robust and learning based polyp segmentation system is proposed. First we find the

¹The essential dimensionality of a 3D shape profile is much higher than (a collection of) 1D curves.

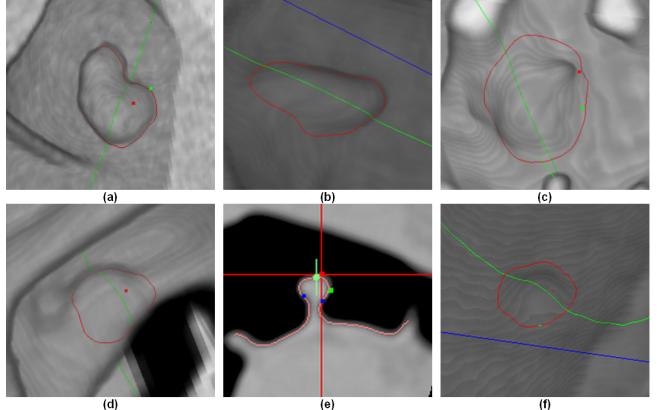


Figure 1. Illustrative examples of polyp segmentation in Computed Tomography Colonography (CTC) for sessile polyps (a,b), mass (c), polyp on a fold (d), pedunculated polyp (e) and flat polyp (f). (a,b,c,d,f) are 3D colon volume renderings and the red-colored curves presents the segmentation result of our system. (e) shows in 2D where blue dots are the detected boundary points for interior/exterior polyp.

polyp tip as a voxel on the polyp surface by a 3D point detector, followed by clustering and centering. A polar coordinates system is then constructed using the detected tip as its origin with a set of evenly sampled radial axes (1D curves on colon surface). The polyp boundary learning is performed along these 1D sampling axes using two-layered stacking learning [14] using a 3D box detector and 1D curve parsing detector [1] to effectively determine which portion of the curve is polyp and which portion is nonpolyp. Finally the binary probabilistic boundary decisions on all 1D curves/axes are assembled jointly based on their axis connectivity in the polar coordinates to form a complete polyp segmentation in 3D.

Features: Our hierarchical, multistaged probabilistic boosting learning framework [11, 12] relies on a large pool of features, which are extracted from colon surfaces under local-to-global scales. We have developed an efficient approach for computing steerable features in 3D under arbitrary spacial configuration[19], while, in this paper, we design two specific steerable patterns for our polyp segmentation task as shown in figure 2. In (a), we show an axis-based pattern for polyp tip detection. This sampling pattern contains three sampling axes as the gradient directions computed in any given voxel v ’s neighborhood under three scales, centered at v . Along each axis, nine grids are evenly sampled and 71 intensity, gradient and curvature based local features are computed for each grid. This process is repeated for half and quarter downsampled CT volumes. Altogether we have $M = 81 = 3 \times 9 \times 3$ grid nodes which bring $71 \times 81 = 5751$ features. For the box-based pattern in (b), there are $7 \times 7 \times 5$ sampling grids under each volume scale where total 52185 features are computed. This is used for polyp interior/exterior material/texture classification. In polyp boundary detection, 440 probabilistic curve

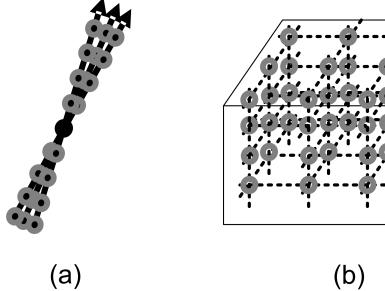


Figure 2. Steerable sampling grid pattern for (a) 3D point detector and (b) 3D box detector.

parsing based features [1] are used for feature selection. In our three learning steps, the polyp-class probabilistic evidences are collected and grouped from a voxel-level detector (for polyp tip voxels), to a box-level detector (for boxes of voxels of polyp interior materials), to a curve-parsing detector using stacked probabilities [14, 18] generated by the box detector. This local-to-global, class-conditional information integrating framework is very robust and bring excellent accuracy on our polyp segmentation task.

3 Algorithm

Our hierarchical, multistaged learning of polyp segmentation method in a compositional framework includes four steps as described below.

3.1 Polyp Tip Detection using 3D Point Detector

The polyp tip is defined as a surface colon voxel inside the polyp region, and reasonably close to the center of the region² in this paper. Our polyp tip detection process is composed of two steps: classification; clustering and centering, as described below. Note that we only work in the domain of surface voxels (ie. voxels with non-zero gradients). The constructed gradient surface is shown in figure 3 (b), generated by a version of Canny edge detection in 3D.

Classification: From the labeled polyp tip position τ for each polyp in our training set, we select its neighboring surface voxels ($\{v\} : dist(\tau, v) < \ell_1$) with Euclidean distances less than a predefined metric ℓ_1 as the training samples of positive class. Other surface Voxels ($\{v\} : dist(\tau, v) > \ell_2$) with relative Euclidean distances larger than ℓ_2 are collected as the negative class. We enforce that $\ell_1 < \ell_2$ so that the trained classifier is forced to distinguish central polyp voxels from non-polyp voxels. After this, we solve this binary classification problem using a trained 3D point detector (ie. probabilistic boosting tree³ classifier [11] with axis-based

²The exact geometric center is often ambiguous because of polyp's irregular 3D shape. Therefore our annotation is usually an approximate geometric center of the polyp region. This labeling process also have intrinsic, moderate inter-reader variability. As discussed later, our compositional polyp shape representation compensates this variability well to obtain stable, robust and accurate segmentation results.

³Probabilistic boosting tree (PBT) is a powerful two-class and multi-class discriminative learning framework [11]. PBT can be generally considered as a generalization of adaboost learning inside the decision tree structure. In practice PBT often offers good performances for complex clas-

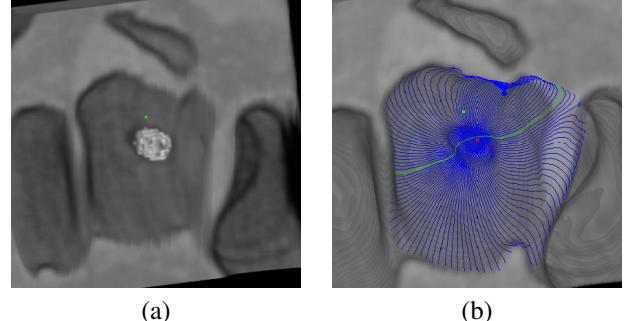


Figure 4. (a) **Left:** all detected voxels in \mathbb{S} shown in white, and the green dot as the center of the volume, (b) **Right:** the red dot as the final polyp tip detected after classification, clustering and centering, and polar coordinates (in blue and green) fitting. As shown in this example, the final polyp tip (red dot) can be successfully detected up to 7.8 voxel distance away from the initial input (green dot).

steerable features in figure 2 (a)). We denote the classifier trained in this step as PBT_1 . After PBT_1 is learned, each surface voxel candidate v gets a positive-class probability value $prob(v)$ (from 0 to 1). If $prob(v) > T_1$, this voxel is classified as a polyp tip voxel. An example of polyp tip voxel detection is illustrated in figure 3 (c).

Clustering & Centering: From all detected polyp tip voxels ($\mathbb{S}\{v\} = \{v\} : prob(v) > T_1$) in each volume, we find one particular voxel as the output polyp tip by clustering and centering. We T_1 is determined based on the PBT learning performance as shown in figure 9 (a). Connected component analysis (CCA) [2] is first applied to partition the set $\mathbb{S}\{v\}$ into a list of n clusters $C_1\{v\}, C_2\{v\}, \dots, C_n\{v\}$. For each of the clusters, we obtain its overall fitness of being positive polyp tip class using $P_i = \sum_{v \in C_i\{v\}} \{prob(v)\}$, $i = 1, 2, \dots, n$ then we simply select the cluster with the maximal fitness: $C_j\{v\}$ where $P_j \geq P_i$, $i = 1, 2, \dots, n$ and discard all others. Optionally, a 3D multivariate normal distribution based spatial prior $\mathcal{G}(v(x, y, z) | \mu, \Sigma)$ can be integrated into P_i to reflect the confidence of the initial polyp position input μ from a CAD software or manual click. Details are omitted due to space limit. Finally we calculate the geometric mean $\bar{\mu}(x, y, z)$ of the winning cluster $C_j\{v(x, y, z)\}$ and map it onto the colon surface by finding \bar{v} where $dist(\bar{v}, \bar{\mu}) \leq dist(v, \bar{\mu})$ and $v \in \mathbb{S}\{v\}$.

3.2 Polyp Mesh Surface Construction and Polar Coordinates Fitting

The Marching Cubes algorithm [6] is a widely used method of constructing a polygonal mesh of an isosurface from a 3D field of scalar values or voxels (ie. 3D volume) in both medical and geological scans. An illustrative example of an original volume and its isosurface extracted using Marching Cubes algorithm is shown in figure 3 (a,d). For details, refer to [6]. On the extracted mesh based polyp surface, we identify the polyp tip τ as the closest vertex on the

sification problems, by using boosting classifiers with the embedded tree-clustering scheme, in a divide-and-conquer manner. For more detail, refer to [11].

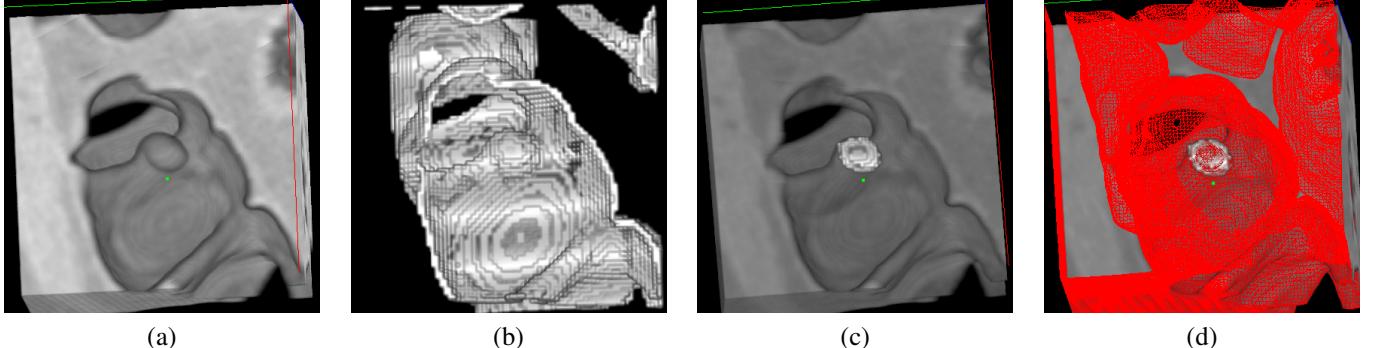


Figure 3. (a) An example of CTC volume containing a typical polyp where the green dot shows the output position from a CAD software, i.e. the center of the volume. (b) The voxel-based gradient surface generated from (a) using 3D Canny edge detection. (c) The detected polyp tip voxels highlighted with whiter intensity. (d) The mesh-based CTC colon surface constructed by using "Marching Cubes" Algorithm [6].

mesh from \bar{v} (ie. \bar{v} is transferred as τ). Then a local polar coordinate system is constructed using τ as its origin. Given the normal vector n_t at τ , we first define the other two axes which are orthogonal to n_t . In theory, the configuration of these two axes is not unique (ie. it can be any two axes in n_t 's orthonormal plane P_t passing τ which are mutually orthogonal as well.), but we fix this configuration by enforcing one axis (Y) passing through the intersection point of P_t and a predefined line vector, such as $[1, 0, z]$ where z is an arbitrary value. Denote n_t as axis X , another constrained axis as Y , and Z is computed as their cross-product. The cutting planes P_n are defined as planes passing through n_t or X , and another line segment Y' in P_t which evenly rotates from Y under a clockwise manner. The axes of final constructed polar coordinates are actually the intersected curves between P_n and the polyp mesh surface extracted using Marching Cubes algorithm [6]. The coordinates on each axis are defined upon their geodesic distances from τ (the origin) on the corresponding cutting curve or axis. The total number of axes is set as I , and they are spatially ordered in a loop. Axis i is adjacent to axes $mod(i - 1, I)$ and $mod(i + 1, I)$. An example is illustrated in figure 4.

The intention of fitting 2D polar coordinates (an spatially ordered collection of 1D curve axis) to represent the 3D polyp surface is due to its flexibility and convenient formulation to parametrization. Compared with Conformal mapping [13], no user identified 3D-2D correspondences are needed and there is no constraint on 3D surface with disk topology. More importantly, our representation of using a collection of 1D curves as sampling coordinates to model the 3D polyp surface shape, decreases the intrinsic dimensionality of the problem (ie. 1D versus 3D modeling). As shown later, this decomposable formulation makes it feasible to construct or model a large variety of 3D shapes using an assembly of simpler 1D curves which are more learnable and flexible.

3.3 Polyp Interior/Exterior Detection using Structural 3D Box Detector

From the annotated polyp boundaries $B_i, i = 1, 2, \dots, I$ on polar coordinates, we can label all voxels along each

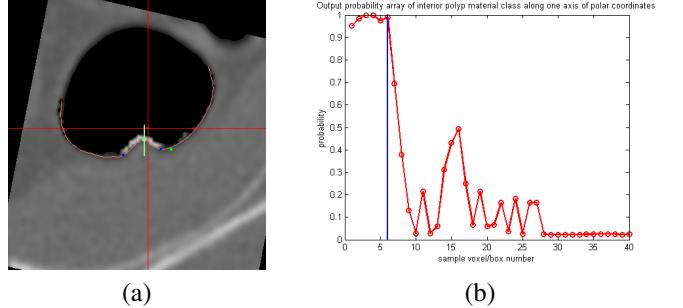


Figure 5. (a) **Left:** The polyp interior class-conditional probability responses along two examples of polar coordinates axes inside a cutting plane. Higher probabilities are shown as surface voxels with whiter intensities; vice versa. The green vertical line segment is the normal vector at the origin of fitted polar coordinates. The orange curves represent the polar coordinates axes. The intersection of orange curves and green line segment is the origin of polar coordinates. The two blue dots show the polyp interior/exterior boundaries from the expert annotations on two axes respectively. (b) **Right:** The plot of polyp interior class-conditional probability responses on the right axis originating from the origin of polar coordinates. The blue bar represents the location of the polyp boundary from annotation.

axis as positive class (ie. voxels lying inside the boundary ($\mathbb{S}_p\{v_{ij}\} = \{v_{ij}\} : (PC(v_{ij}) \leq B_i - \varepsilon)$), or negative class (ie. voxels lying outside the boundary ($\mathbb{S}_n\{v_{ij}\} = \{v_{ij}\} : (PC(v_{ij}) \geq B_i + \varepsilon)$)). $PC(v_{ij})$ is the polar coordinates value for any voxel v_{ij} on the i th axis; B_i is the polar coordinates of the labeled boundary point on the same axis; and ε is a small value to keep the distance margin between \mathbb{S}_p and \mathbb{S}_n for learnability (as discussed in section 3.1). Again this binary learning problem is solved by training a Probabilistic Boosting Tree based classifiers [11] as PBT_2 with 3D box based steerable features. Boxes are placed centering at each voxel and aligned with its local gradient surface. Compared with the 3D point detector PBT_1 (polyp tip), the resulted box detector PBT_2 contains more structural information within a larger neighboring spatial span for better discrimination of [11]. The feature number of PBT_2 for feature selection is 52185 versus 5751 in PBT_1 . Refer to figures 5 and 6 (a) for illustrative examples of polyp interior/exterior detection.

3.4 Polyp Surface Boundary Detection using Boosted 1D Curve Parsing and Stacking

We treat the final step of polyp boundary segmentation as a statistical curve parsing problem [1] of finding the polyp/nonpolyp breaking points on 1D curves. Given the output probabilities from polyp interior/exterior detection (as shown in figure 5 (a,b)), another layer of boundary boosting learning is applied. This is equivalent to learn a new classifier in the embedded, semantic space learned by another classifier (as a wrapper) in the previous step. The theory of stacked generality [14, 18] also provides its mathematical background. To learn the true polyp boundary of 1D vectors of noisy values (5 (b)), we label the coordinates along each axis as positive class (ie. coordinates within a small distance threshold of the annotated boundary ($\mathbb{S}_+ \{v_{ij}\} = \{v_{ij}\} : (B_i - \varepsilon_1) \leq PC(v_{ij}) \leq B_i + \varepsilon_1\}$), or negative class (ie. coordinates outside a small distance threshold of the annotated boundary ($\mathbb{S}_- \{v_{ij}\} = \{v_{ij}\} : (PC(v_{ij}) \geq (B_i + \varepsilon_2)) \mid (PC(v_{ij}) \leq B_i - \varepsilon_2)\}$)). $\varepsilon_2 > \varepsilon_1$ is normally set to keep the distance margin ($\varepsilon_2 - \varepsilon_1 > 0$) between \mathbb{S}_+ and \mathbb{S}_- for reducing learning ambiguity.

From \mathbb{S}_+ and \mathbb{S}_- , we first evaluate v_{ij} 's probability value ρ_{ij} using PBT₂. Thus we obtain an array of $\{\rho_{ij}\}$ on each axis i as $\mathbf{A}(\rho_{ij}) = (\rho_{i1}, \rho_{i2}, \rho_{i3}, \dots, \rho_{ij}, \dots)$. Then the following types of features are computed according to each v_{ij} or ρ_{ij} from $\mathbf{A}_f(\rho_{ij}) = \{\rho_{i1}, \rho_{i2}, \rho_{i3}, \dots, \rho_{ij}\}$ and $\mathbf{A}_b(\rho_{ij}) = \{\rho_{i,j+1}, \rho_{i,j+2}, \rho_{i,j+3}, \dots\}$ respectively, for two-class PBT learning:

- 1) *probability count features*: The count of φ s above a threshold θ in array \mathbf{A}_f or \mathbf{A}_b ;
- 2) *probability gap features*: The size vector of the largest gaps of the θ -thresholded probability array for \mathbf{A}_f or \mathbf{A}_b , sorted in decreasing order.

$\mathbf{A}_f(\rho_{ij})$ represents ρ_{ij} 's forward array spatially from origin to v_{ij} , and $\mathbf{A}_b(\rho_{ij})$ represents ρ_{ij} 's backward array spatially from $v_{i,j+1}$ to end. The θ -thresholded probability array means replacing every element φ in \mathbf{A}_f or \mathbf{A}_b with 0 (if $\varphi < \theta$) or 1 (if $\varphi \geq \theta$); the gap size is the count of continuous 0s. We set $\theta = i \times 0.05$, $i = 0, 1, \dots, 19$ and the size of gap vectors to be 20 for each θ . If there are less than 20 existing gaps, 0 will be used to fill up the 20-element vector. Therefore total $440 = (20 + 20 \times 20) \times 2$ features are computed as the feature selection pool for PBT learning. The resulted classifiers is denoted as PBT₃.

With PBT₃, we can compute each point v_{ij} on the polar coordinates a new probability value ρ_{ij} by evaluating its forward and backward probability arrays $\mathbf{A}_f(\rho_{ij}), \mathbf{A}_b(\rho_{ij})$ output by PBT₂. To determine a unique boundary point on any axis i , the simplest way to select the v_{ik_i} with its $\rho_{ik_i} = \max\{\rho_{i1}, \rho_{i2}, \rho_{i3}, \dots, \rho_{ij}, \dots\}$. Assuming that the selection of each v_{ik_i} is solely based on $\{\rho_{i1}, \rho_{i2}, \rho_{i3}, \dots, \rho_{ij}, \dots\}$ and independent with the formation on other axes, the final polyp segmentation boundary will probably be rugged by sequentially connecting boundary points $v_{1k_1}, v_{2k_2}, \dots$, on all axes

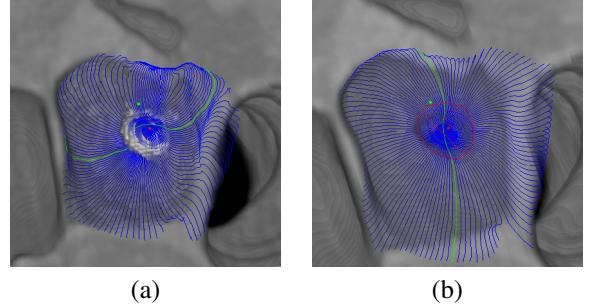


Figure 6. Two-layered polyp segmentation based on stacked generality [14]. (a) The polyp interior class probability response map on the polyp surface (higher probabilities shown as whiter intensities and vice versa), (b) Final polyp segmentation boundary contours (in red) overlaid on the local polar coordinates. Additionally two red line segments represent the two main axes for polyp dimensioning.

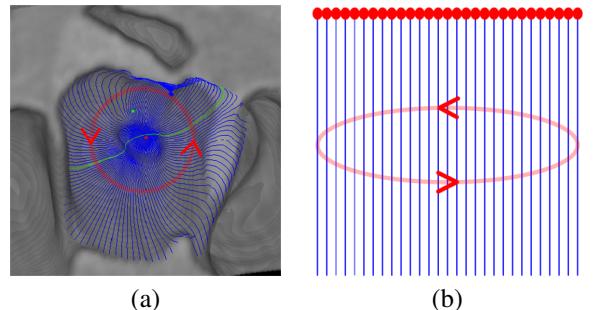


Figure 7. (a) The loopy structure topology of the axes of polar coordinates overlayed on polyp surface, (b) The equivalent looped sequential trellis or looped 2D graph topology constructed.

of the polar coordinates. For smoothness, these boundary decisions can be assembled using Gaussian blurring over their polar coordinates values $PC(v_{1k_1}), PC(v_{2k_2}), \dots$, according to neighborhoods of successive axes. Alternatively, Viterbi-like dynamic programming algorithm [3] can also be employed on this sequential smoothing problem with a looped trellis structure (as interpreted in figure 7). Loopy belief propagation [8] is another applicable solution by treating polar coordinates as a general looped 2D graph. Gaussian blurring is indeed a spatially homogeneous smoothing filter, while Viterbi and loopy belief propagation are nonhomogeneous functions by building pairwise content-sensitive potential functions. All three approaches function as assemblers of propagating 1D axis segmentations across neighborhoods of polar coordinates.

At the end, we can enumerate the set of positive polyp surface voxels (with their associated probability ρ_s) inside the polyp boundary segmentation for each polyp. The remaining polyp dimensioning process is straightforward: 1) finding a pair of positive voxels as v^{11}, v^{12} with the maximal Euclidean distance to build the first polyp axis; 2) using their distance $dist(v^{11}, v^{12})$ as the estimated polyp size; 3) then computing other pairs of voxels $v^{2'}, v^{2''}$ which is orthogonal to v^{11}, v^{12} (ie. $(v^{2'} - v^{2''}) \odot (v^{11} - v^{12}) = 0$); 4) finally selecting the pair of voxels v^{21}, v^{22} with the maximal Euclidean distance as the second polyp axis, from the

filtered voxles in 3). Optionally the polyp dimension on the third orthonormal axis can be computed in a similar manner.

4 Experiments

We present both quantitative and qualitative results on polyp segmentation and dimensioning measurement of our method and its comparison with the state-of-art work [4]⁴. To present quantitative results, we define two metrics of the error measure between the detection result and the annotation: the difference of polyp size in both millimeter and relative percentage %; and the overlapping ratio Υ between the sets of segmented and annotated positive polyp surface voxels. We train and test the dataset of 274 polyps acquired from 10 different sites on Siemens and GE scanners, while [4] is tested on another dataset of 154 polyps from the same source that we have no access. The parameter settings used at different training stages, are listed as follows. The positive/negative training margins of PBT_1 are $\ell_1 = 4$ voxel and $\ell_2 = 6$ voxel; and the threshold of tip detection is $T_1 = 0.5$. The training distance margin of PBT_2 is $\varepsilon = 2$ voxel. $\varepsilon_1 = 2$ and $\varepsilon_2 = 4$ are set as the positive/negative training margins of PBT_3 . The axis number of polar coordinates I is 120 (ie. the colon polyp surface is sampled every 3°). The training Receiver Operating Characteristic (ROC) curves in three PBT learning stages are illustrated in figure 9 (a). There are 5732 positives and 376730 negatives in training PBT_1 , 1203829 positives and 2949828 negatives in PBT_2 , and 32880 positives and 1146319 negatives in PBT_3 . All boosting training procedure takes about 20 hours to finish on a P4-3.2G PC. The runtime of polyp segmentation using our system is within $2 \sim 4$ seconds. This segmentation/measurement process can be easily made to be parallel for all polyp candidates in a whole colon CT scan.

We first explore the differences between our segmented polyp size measurements and the labeled sizes. Our compositional framework of assembling the 2D or 3D polyp segmentation boundary from 1D decisions on the polar coordinates axes is, intrinsically insensitive to partial under-segmentations. On the other side, the polyp axis pairs of v^{11}, v^{12} and v^{21}, v^{22} are constrained to have high ρ_s (> 0.75) during polyp dimensioning. This strategy prevents the polyp size measurement inaccuracy caused by less likely, but possible peaky over-segmented boundaries with weak ρ_s . Using [4], there are 80.5% polyps with size measurement error $E \leq 3mm$ and 72.1% polyps with $E \leq 2mm$ in the dataset of 154 instances (size ranging from $1.8mm$ to $28mm$). Our cross-validation results are 98% polyps with $E \leq 3mm$ and 92% polyps with $E \leq 2mm$ for both the training set (221 polyps) and the testing set (53 polyps). Random splitting is used to construct the training and testing set from total 274 polyps. Based on our approach, the mean error is $0.82mm$ or 7.39% , and the standard deviation of errors

⁴We do not have access to another recent work [15] which has slightly worse performance than [4] in literature.

is $0.95mm$ or 7.09% , respective to the absolute or relative error metric separately. Using [4], the results are $2.00mm$ or 27.6% for mean, and $2.97mm$ or 26.7% for standard deviation. This comparison proves our statistically significant performance improvement over previous work [4, 15], in terms of the polyp dimensioning accuracy. We further evaluate our approach based on the segmented/annotated surface area overlapping ratio Υ that is computed using the size (ie. voxel number) of the joint set of segmented polyp voxels and annotated polyp voxels divided by the size of the set of annotated polyp voxels. There are only 2.9% polyps with $\Upsilon \leq 60\%$, and 86.9% polyps with $\Upsilon \geq 75\%$, and 65.8% with $\Upsilon \geq 90\%$ in our approach. Refer figures 8, and 9 for more details.

In section 3.4, we discussed three polyp boundary binding techniques with smoothness. Gaussian blurring is the default method which participates to generate the above experimental results. Due to the space limit, we only report the qualitative comparison results: loopy belief propagation [8] gains slightly in segmentation accuracy, but dynamic programming [3] has slightly worse performance than Gaussian blurring. We leave the further investigation as future work.

4.1 Robustness of Polyp Tip Detection and Disturbance Testing

It is often known that the initial polyp location input is not accurate enough to precisely locate the polyp when using a CAD software, and has large inter-reader variation from clicks of different clinical experts. Our compositional assembling of 1D curves/axis to form 2D/3D polyp shape representation offers the excellent and comprehensive modeling capacity for a large variety of 3D polyp shapes (learned in PBT_3) using a limited quantity of 1D curve bases (learned in PBT_2). This is tested using random disturbance as described below. $DS-I$ denotes disturbances on the searching origin (normally, set as the center of CTC subvolume containing the polyp for investigation) for polyp tip detection of PBT_1 ; $DS-II$ stands for directly using the disturbed clinician's clicks to initialize the polyp segmentation of PBT_2 and PBT_3 by skipping PBT_1 . Two illustrative cases of $DS-II$ are demonstrated in figure 10. For example, on the right, the range of our estimated polyp size is $[8.54mm, 9.12mm]$ and the labeled value is $8.7mm$. The resulted multiple polyp boundaries (in red) lie consistently within a close band relative to the annotated boundary (in purple). This variation range is comparable or smaller than the inter-reader variability reported in [10, 16, 17]. Furthermore, the numerical performance evaluation details on $DS-I$ and $DS-II$ are described in table 1. On the other hand, the required manual initialization is critical and can be a sensitive factor for the final performance of fuzzy clustering [15], deformable model [15, 5], or level-set based approaches [7]. Our system is automatic, unless required of starting from a specified user-click. In summary, our compositional representation of learning complex 3D polyp surface shape model is well validated with its robustness against extensive $DS-I$ and $DS-II$ disturbances.

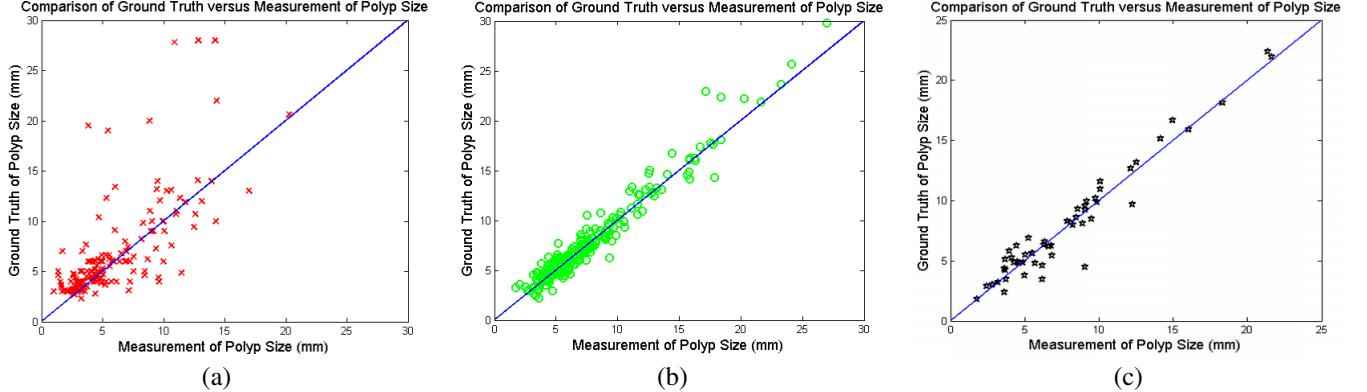


Figure 8. Comparison on the polyp size measurement errors of our supervised, data-driven learning approach and the heuristic based method [4]: (a) **Left:** the error plots of ground truth verse measurement polyp sizes of 154 examples (ranging from 1.8mm to 28mm) using [4], (b) **Center:** the training error plots of ground truth verse measurement polyp sizes of 221 examples (ranging from 1.8mm to 30mm) in our system, (c) **Right:** the testing error plots of ground truth verse measurement polyp sizes of 53 examples (ranging from 1.8mm to 22mm) in our system. Notice that only 11 out of 274 total available polyp examples have sizes $\leq 17\text{mm}$. Particularly, we randomly chose 8 polyps in the training set and other 3 instances for testing. Therefore the generality of our learning system is highly promising for the full polyp size range that we explore in this paper.

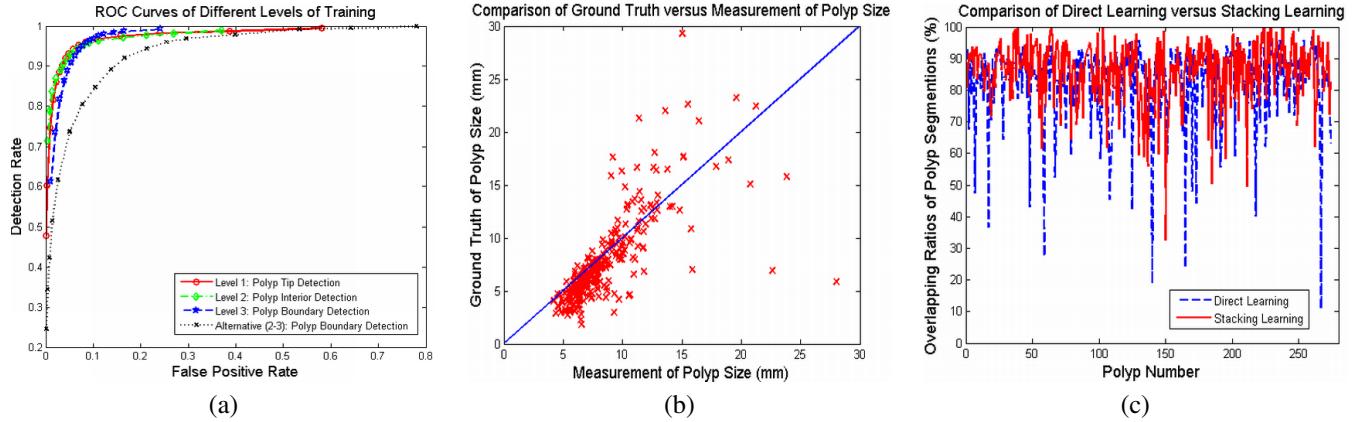


Figure 9. (a) **Left:** Receiver operating characteristic curves of different levels of training in our multistaged polyp segmentation system. Levels 1-2-3 represent the binary PBT based learning processes for polyp tip detection, polyp interior detection and polyp boundary detection respectively. Alternative (2-3) refers direct polyp boundary learning after tip detection, boosted using 208740 geometric features sampled along each polar coordinates axis from the origin to any scanning surface voxel. It is implemented to compare with the combination of level 2 and 3 in our current system. By leveraging from stacked generality [14], the training task of levels 2 and 3 is easier and better performed than the directly geometric learning (2-3). (b) **Center:** Performance on the polyp size measurement errors of direct geometry learning method. Based on 274 polyp examples (size ranging from 1.8mm to 28mm), there are 86.86% polyps with size measurement error $\leq 3\text{mm}$ and 79.93% polyps with size measurement error $\leq 2\text{mm}$ for direct learning. These results improve the performance of [4] at 80.5% and 72.1% respectively, but are not as good as of our two level stacking learning approach with 98% and 92%. The mean error is 1.54mm and the standard deviation of errors is 2.22mm. Again it is better than the values of 2.00mm, 2.97mm in [4], but worse than 0.82mm, 0.95mm in our stacking learning. (c) **Right:** Comparison of overlapping ratios of polyp segmentation: direct learning (in green) versus stacking learning (in red). From the results of all 274 polyp examples, our stacking learning achieves $\text{mean}(\Upsilon) = 86.53\%$ and $\text{std}(\Upsilon) = 10.68\%$; direct learning obtains $\text{mean}(\Upsilon) = 81.58\%$, $\text{std}(\Upsilon) = 13.51\%$. There are 0.36% versus 4.38% polyp measurements with $\Upsilon \leq 50\%$, and 86.86% versus 79.20% polyp measurements with $\Upsilon \geq 75\%$ for stacking learning and direct geometry learning respectively.

4.2 Comparison of Direct Geometric Learning Versus Stacking Learning [14]

We also implement a more straight way of learning 1D polyp shapes from the geometrical features on the curve (ie. the axis of polar coordinates). In our implementation, each voxel v_{ij} gets its characteristic geometry feature set by concatenating all 3D box steerable features of four points which are evenly sampled along its axis i originating from the coordinates origin t to v_{ij} . The resulting feature number is

$208740 = 52185 \times 4$ for PBT boosting training. Compared with the feature number 440 in PBT_3 , 208740 is too large and more prone to overfitting given the same amount of training samples. This type of learning structure is denoted as *direct geometric learning* and our two-layered architecture of PBT_2 and PBT_3 is referred as *stacking learning* [14, 18]. Based on the results in figures 8 (b,c), against 9 (a,b,c), our two-layered stacking learning approach outperforms the one-layered direct geometric learning in polyp segmentation,

| | No-DS | DS-I(20%) | DS-I(40%) | DS-I(60%) | DS-II(15%) | DS-II(20%) | DS-II(25%) | DS-II(30%) |
|----------|--------|-----------|-----------|-----------|------------|------------|------------|------------|
| Mean(mm) | 0.8207 | 0.8177 | 0.8157 | 0.8582 | 0.816 | 0.7997 | 0.9603 | 0.9583 |
| Std (mm) | 0.9502 | 0.9349 | 0.9512 | 0.9649 | 1.06 | 0.9267 | 1.1738 | 1.1346 |
| ROS (%) | 98.0 | 97.5 | 96.67 | 97.5 | 95.0 | 95.83 | 94.17 | 93.3 |
| Mean(%) | 7.39 | 7.41 | 7.19 | 7.71 | 7.88 | 7.46 | 8.43 | 8.52 |
| Std (%) | 7.09 | 7.23 | 7.47 | 6.96 | 7.12 | 9.14 | 7.63 | 9.73 |

Table 1. Robustness testing results of polyp size measurement over two types of disturbances (DS-I, DS-II) on polyp tip detection (in stage I), compared with our original result with no disturbance (No-DS on the leftest column). The first two rows are the means and standard deviations of absolute measurement errors in millimeters (mm); the third row shows the ratio of success (ROS) which is defined as the percentage of the polyps with absolute error $\leq 3\text{mm}$ out of all polyps; the forth and fifth rows are the means and standard deviations of relative measurement errors in percentage (%) (divided by their labeled polyp sizes respectively). All percentages (%) following DS-I or DS-II mean the disturbance magnitudes relative to individual polyp sizes. For DS-II, by assuming that the polyp tip approximately locates in the center of polyp surface, and the polyp size is a diameter-like measurement, the valid, maximal possible DS-II is roughly 50%. Otherwise the disturbed polyp tip input will lie outside of the polyp region. From above, the first type of disturbance DS-I does not bring errors with statistical differences than No-DS, which proves the stability of our polyp tip detection process in section 3.1. DS-II is considered as a more severe type of disturbance for polyp size measurement accuracy. The performance DS-II degrades gracefully on increasing levels of noise, which demonstrates the robustness of our compositional polyp shape representation.

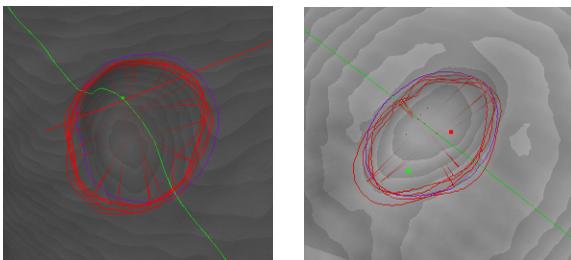


Figure 10. Two examples of our robust and consistent polyp segmentation/measurement performance under multiple, different user clicks as hard-coded polyp tip inputs. The resulted multiple polyp boundary curves are drawn in red, and the annotated boundary curve is shown in purple.

with statistical significance. However the implemented data-driven boosting nature of direct geometric learning still helps to improve over the heuristic approaches in literature [4, 15].

5 Conclusion

In this paper, we have proposed a hierarchical, multi-staged probabilistic binary classification approach for automated segmentation of colonic polyp surfaces from other, attached colon tissues in 3D CTC. Our system integrates low-, and mid-level information for discriminative learning under the local polar coordinates which are built to align on the 3D colon surface around polyp. More importantly, our flexibly architected, supervised probabilistic learning system offers a principled means of encoding a large capacity of semantic, clinical expert annotations on colonic polyp surface tissue identification and segmentation (for example, figure 1). Results on extensive performance evaluations using both training and unseen data validate our significant improvement over the previous state-of-art work [4, 15]. This has very important clinical impacts of offering more accurate polyp segmentation/measurement, detection and classification solutions for computer-aided diagnosis systems [10, 4, 9, 16].

References

- [1] A. Barbu, et al., Hierarchical Learning of Curves Application to Guidewire Localization in Fluoroscopy, *CVPR* 2007.
- [2] R. Duda, P. Hart and D. Stork, *Pattern Classification*, Wiley Interscience, 2002.
- [3] F. Jelinek, *Statistical Methods for Speech Recognition*, MIT press, 1998.
- [4] A. Jerebko, S. Lakare, P. Cathier, S. Periaswamy, L. Bogoni: Symmetric Curvature Patterns for Colonic Polyp Detection. *MICCAI* (2) 2006: 169-176.
- [5] M. Kass, A. Witkin and D. Terzopoulos, Snakes: active contour models, *Int. J. Comput. Vis.*, pp.321-331, 1998.
- [6] W. Lorensen and H. Cline, Marching Cubes: A High Resolution 3D Surface Construction Algorithm, *SIGGRAPH*, 1987.
- [7] M. Lynch, T. Chowdhury, O. Ghita and P. Whelan, Determining Candidate Polyp Morphology from CT Colonography using a Level-Set Method, *EMBE*, 2005.
- [8] K. Murphy, Y. Weiss, M. Jordan, Loopy Belief Propagation for Approximate Inference: An Empirical Study, *UAI*, 1999.
- [9] J. Nappi and H. Yoshida, Automated detection of polyps with CT colonography: evaluation of volumetric features for reduction of false-positive findings, *Academic Radiol.*, 9(4): 386397, 2002.
- [10] R. Summers, A. Jerebko, M. Franaszek, J. Malley, and C. Johnson, Colonic polyps: Complementary role of computer-aided detection in ct colonography, *Radiology*, 225:391399, 2002.
- [11] Z. Tu. Probabilistic boosting-tree: Learning discriminative methods for classification, recognition, and clustering. In Proc. Int'l Conf. Computer Vision, 2005.
- [12] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. IEEE Conf. Computer Vision and Pattern Recognition, 2001.
- [13] S. Wang, Y. Wang, M. Jin, X. Gu and D. Samaras, Conformal Geometry and Its Applications on 3D Shape Matching, Recognition and Stitching, *IEEE Trans. PAMI*, 2007.
- [14] D. H. Wolpert, Stacked generalization, *Neural Networks* 5(2): 241-259, 1992.
- [15] J. Yao, M. Miller, M. Franaszek and R. Summers, Colonic polyp segmentation in CT Colonography-based on fuzzy clustering and deformable models, *IEEE Trans on Medical Imaging*, 23(11):1344-1352, 2004.
- [16] H. Yoshida et al., Computerized detection of colonic polyps at CT colonography on the basis of volumetric features: pilot study, *Radiology*, 222(2)327336, 2002.
- [17] H. Yoshida and A. H. Dachman, CAD techniques, challenges, and controversies in computed tomographic colonography, *Abdominal Imaging*, Springer, 30(1):26-41, 2005.
- [18] S. Zheng, Z. Tu and A. Yuille, Detecting object boundaries using low-, mid-, and high-level information, *CVPR*, 2007.
- [19] Y. Zheng, et al., Fast Automatic Heart Chamber Segmentation from 3D CT Data Using Marginal space Learning and Steerable Features, *ICCV*, 2007.