

# An open source implementation of colon CAD in 3D Slicer

Haiyong Xu<sup>\*a,c</sup>, H. Donald Gage<sup>b,c</sup>, Pete Santago<sup>a,c</sup>

<sup>a</sup>Dept. of Biomedical Engineering, Wake Forest Univ. Health Sciences, Medical Center Blvd.  
Winston-Salem, NC, USA 27157;

<sup>b</sup>Dept. of Radiology, Wake Forest Univ. Health Sciences, Medical Center Blvd. Winston-Salem,  
NC, USA 27157;

<sup>c</sup>Virginia Tech – Wake Forest University School of Biomedical Engineering and Sciences

## ABSTRACT

Most colon CAD (computer aided detection) software products, especially commercial products, are designed for use by radiologists in a clinical environment. Therefore, those features that effectively assist radiologists in finding polyps are emphasized in those tools. However, colon CAD researchers, many of whom are engineers or computer scientists, are working with CT studies in which polyps have already been identified using CT Colonography (CTC) and/or optical colonoscopy (OC). Their goal is to utilize that data to design a computer system that will identify all true polyps with no false positive detections. Therefore, they are more concerned with how to reduce false positives and to understand the behavior of the system than how to find polyps. Thus, colon CAD researchers have different requirements for tools not found in current CAD software. We have implemented a module in 3D Slicer to assist these researchers. As with clinical colon CAD implementations, the ability to promptly locate a polyp candidate in a 2D slice image and on a 3D colon surface is essential for researchers. Our software provides this capability, and uniquely, for each polyp candidate, the prediction value from a classifier is shown next to the 3D view of the polyp candidate, as well as its CTC/OC finding. This capability makes it easier to study each false positive detection and identify its causes. We describe features in our colon CAD system that meets researchers' specific requirements. Our system uses an open source implementation of a 3D Slicer module, and the software is available to the public for use and for extension (<http://www2.wfubmc.edu/ctc/download/>).

**Keywords:** CT Colonography, CAD, 3D Slicer, visualization

## 1. INTRODUCTION

Colorectal cancer (CRC) is the third most common and the second most deadly form of cancer in men and women in the United States<sup>1</sup>. CRC largely can be prevented by detection and removal of adenomatous polyps, and survival is significantly better when CRC is diagnosed while still localized<sup>2</sup>. Optical colonoscopy (OC) is a widely performed medical procedure in the United States as a screening method for CRC<sup>2</sup>. It is an invasive procedure and requires sedation. CT Colonography (CTC), also called virtual colonoscopy (VC), is another screening method, which is minimally invasive. Based on CT images of the cleansed and air-distended colon, 2D and 3D projections are interpreted by radiologists to determine if polyps are present, their location, and their size. Several clinical improvements in patient preparation, technical advances in CT, and new developments in evaluation software have allowed CTC to develop into a powerful diagnostic tool<sup>3</sup>. The use of computer aided detection (CAD) in CTC is an active research field<sup>4,5</sup>, and several CAD systems are commercially available. In these schemes, computer software should automatically locate all clinically relevant polyps (high sensitivity), while minimizing the number of false positives (high specificity).

Most commercial colon CAD products are designed to assist radiologists in finding polyps in a clinical environment. However, colon CAD researchers, many of whom are engineers or computer scientists working with CT studies with polyps identified in CTC and/or OC, are more concerned about discovering causes of false positive detections and designing methods to reduce them. Thus, the researchers have different requirements for the tools not found in current commercial CAD products. We have implemented a 3D Slicer module to assist researchers in colon CAD. In this module, we integrate results from colon segmentation, polyp candidate generation, classification, and CTC/OC findings, into one workbench, and visualize colon wall, polyp candidate surface, and voxels in one user interface. This integration and visualization give researchers a powerful tool to study false positives and to discover the causes, which are not easily perceptible from scattered information.

3D Slicer (Slicer)<sup>6</sup> is a free, open source software package for visualization and image computing. It provides functionality for segmentation, registration, and 3D visualization of multi-modal medical image data. Slicer began as an image-guided surgery system developed at the MIT AI Lab in collaboration with the Surgical Planning Laboratory at the Brigham and Women's Hospital in 1999<sup>7</sup> and has evolved to support a wide variety of clinical applications. Slicer includes routines to read and write various file formats, manipulate 2D and 3D coordinate systems, and present a consistent user interface paradigm and visualization metaphor<sup>8</sup>. Slicer's component architecture is based on the concept of modules that can provide new user interface components, new software services, or a combination thereof. Almost all functionalities of Slicer are implemented as modules (command-line module or interactive module). In this paper, we describe our implementation of an interactive module in Slicer to assist CTC researchers in discovering causes of false positive detections and designing methods to reduce them.

To the best knowledge of the authors, the most recent work of CTC in Slicer was published by Nain, et al.<sup>9</sup> The proposed 3D virtual endoscopy system allows the user to interactively explore the internal surface of a 3D anatomical model and to create and update a fly-through trajectory through the model to simulate endoscopy. The goal in that work was to combine the strength of 2D imaging techniques with 3D visualization in order to simulate the surgical environment and provides the user with navigational and path creation options. Like most commercial colon CAD products, that system assists radiologists in finding polyps in CTC rather than helps CTC researchers in discovering causes of false positive detections. In our colon CAD, we developed a free, open source colon CAD system to meet the researchers' requirements.

## 2. METHODOLOGY

### 2.1 Overview of colon CAD

A colon CAD system is a multi-step procedure, typically consisting of (1) colon wall segmentation, (2) intermediate polyp candidate generation, (3) classification for detection of final candidates, and (4) final polyp candidate presentation<sup>10</sup>. The purpose of segmentation is to limit the search area for polyps to reduce primary processing time, and to reduce causes of false positive detections coming from the small bowel and extra colonic structures. Generating intermediate polyp candidates is applied to find regions which are likely polyps; each intermediate polyp candidate is represented by features gathered from the candidate region. The goal of polyp candidate generation is to identify as many true polyps as possible while minimizing false positives. However, since high sensitivity is important at this stage, many false positives are generated. Thus, the intermediate polyp candidates are fed into a classifier that is trained to reduce false positives. Positive polyp candidates filtered by a classifier are presented to users as final polyp candidates. In this paper, we focus on the last step (final polyp candidate presentation) and describe a Slicer module which shows a list of polyp candidates and display 3D view for each polyp candidate.

### 2.2 Segmentation, intermediate polyp candidate generation, and classification

The details of segmentation, generating intermediate polyp candidates, and classification are published in <sup>11,12,13,14</sup>. In this section, we briefly describe the procedure in each step.

A fully automatic segmentation method that has been proposed in <sup>11</sup> was used in our colon CAD. The algorithm uses the geometry of the colon as features to detect and segment the lumen. First, segmentation of the volume background is done using a region grower. The resulting connected region is used as a mask to eliminate the background during further processing. Next, a set of thresholds is applied to generate a binary image of gas and tissue. A distance transform is then used to locate a point inside the gas regions with a maximal distance from other tissues. This seed point is used by the region growing algorithm to segment the gas-filled portion of the colon. An estimate of the amount of elongation of the gas-filled object is obtained during the segmentation process and is used with the location of the seed point to decide if the object is bowel or stomach. If the object is relatively elongated or occurs below the top one-quarter of the volume, it is considered part of the bowel and is added to the segmentation. The algorithm then proceeds by masking the previous grown region from the distance transform image and searching for another seed point in the remaining gas-filled segments. Besides gas, the colon is also filled with retained residue after bowel preparation, which is homogeneous contrast enhanced fluid (CEF). To include these CEF-filled lumen sections, the mean and Gaussian curvature at each point on the surface are computed and relatively large areas of low curvature indicate a fluid boundary. Selective dilation across the boundary is used to extend the gas segmentation in the residual CEF.

In generating intermediate polyp candidates, a polyp candidate consists of a group of connected voxels, from which geometric features are calculated and classified to identify a polyp. These candidates are initially detected using a subset of features and segmented roughly using various heuristic means to form candidate polyp regions, over which statistics can be computed per polyp. At each colonic wall voxel, principal, Gaussian, mean curvatures, shape index (SI) and curvedness (CV) values are computed respectively. Voxels that have SI and CV values within predefined range are extracted as *seed voxels* by thresholding (SI: [0, 0.11], CV [0.075, 0.2]). From those seed voxels, a region growing procedure then extracts the major part of a polyp. It is reasonable to relax the ranges (SI: [0, 0.22], CV [0.05, 0.25]) around polyps' peripheral region. Then a fuzzy c-means clustering<sup>15</sup> is applied to remove some polyp candidates due to image noise and effectively group voxels belonging to the same polyp in a large cluster. The clustering is based on voxels' feature values, which include SI, CV, magnitude of the gradient, CT intensity value, and spatial coordinates. Further features such as gradient concentration, directional gradient concentrate, sphericity, compactness, wall/region density and polyp radius are computed after the clustering step. These features are further characterized by statistical operations (mean, max, min, variance, etc.) to form polyp candidates' features. An intermediate polyp candidate is considered as a true positive detection if the distance between it and the center of a polyp (detected in CTC/OC) is equal to or less than the polyp size measured in CTC/OC.

In classification for detection of final candidates, several algorithms have been investigated in our CAD system, including support vector machines (SVMs), AdaBoost, artificial neural network, C4.5 decision tree, etc. We use the classification results from SVMs in this paper for its balanced performance in terms of sensitivity, specificity, and area under ROC curve (AUC)<sup>13,14</sup>. For the classification task in a colon CAD system, the training data is extremely imbalanced because there are many more false positive detections than true positive ones in the intermediate polyp candidates. Simply training a classifier with this imbalanced data will produce an uninformative result since the classifier can reach very high accuracy by classifying all polyp candidates as false positive detections. In order to overcome this imbalanced data problem, we employed the SMOTE oversampling technique<sup>16</sup> during classifier training. Another issue in classification task is to determine the value for parameters for each classification algorithm, e.g., what kernel function should be used in SVMs. In our colon CAD, we tuned the parameters experimentally. A classifier with different values for each parameter is trained and tested in a computer cluster environment. Trained classifiers are evaluated using 10-fold cross-validation method and are ranked according to different combinations of criteria: sensitivity, specificity, and AUC. The output of each classifier is mapped to a prediction value ranging from 0 to 1. In the implementation, we use the open source machine learning software WEKA<sup>17</sup>.

### 2.3 Presentation of the polyp candidates – a Slicer module to integrate and visualize polyp candidates

The input to the Slicer module includes three parts: 1) the segmentation file of the colon wall, which is *label map* indicating inside and outside of colon wall. 2) The intermediate polyp candidates' feature files. As mentioned above, each intermediate polyp candidate is generated from clustering *seed voxels*. The features for intermediate polyp candidates and features for seed voxels are stored separately, and the relationship between each intermediate and its seed voxels are established through an assigned polyp id. 3) The classification file that includes the predication value for each intermediate polyp candidate and the CTC/OC finding ("1" indicates an intermediate polyp candidate is true positive; "0" indicates false positive). Normally these three inputs are stored in different files making it difficult to analyze the false positive intermediate polyp candidates. In the next sections, we describe a Slicer module that integrates this information into one workbench and allows one to the visualize colon wall, polyp candidate surface, and voxels in one user interface. In this way, allowing discovery of the basis for the many false positives.

Figure 1 illustrates how to use this Slicer module. This includes the following steps: load CT study, load intermediate polyp candidates (including its seed voxels), load segmentation result, locate a polyp candidate, observe the polyp candidate in 2D views (sagittal, coronal, and axial), and visualize the polyp candidate in a 3D view. The first step, load CT study, is done through Slicer's *volumes* module. In loading intermediate polyp candidates, the filenames for polyp candidates' features, seed voxels' features, and classifier's prediction value are specified in a panel as shown in figure 2. Each file is a text file; the appendix describes its format. After this step, a list of intermediate polyp candidates is shown as in the lower part of the figure 2. The 1st column is the assigned polyp id. The next three columns are DICOM coordinates for each intermediate polyp candidate. The 4th column, *pred 0*, is the classifier's predication value for this intermediate polyp candidate. A value close to 1 means this intermediate polyp candidate is most likely to be a negative detection. In the 5th column, a value close to 1 means this intermediate polyp candidate is most likely to be positive. The last column, *target*, is the CTC/OC finding; A value of "1" indicates that there is a true polyp found near the location of that intermediate polyp candidate in CTC/OC, while "0" indicates no nearby polyp. In the third step, load segmentation

result, a label map file (in VTK format as described in appendix) is loaded and a contour is applied to generate the colon surface. Since Slicer uses a RAS (Right, Anterior, Superior) coordinate system, which is different from a DICOM coordinate system (Left, Posterior, Superior), the colon surface must be transformed in order to properly locate each polyp candidate on the colon wall.

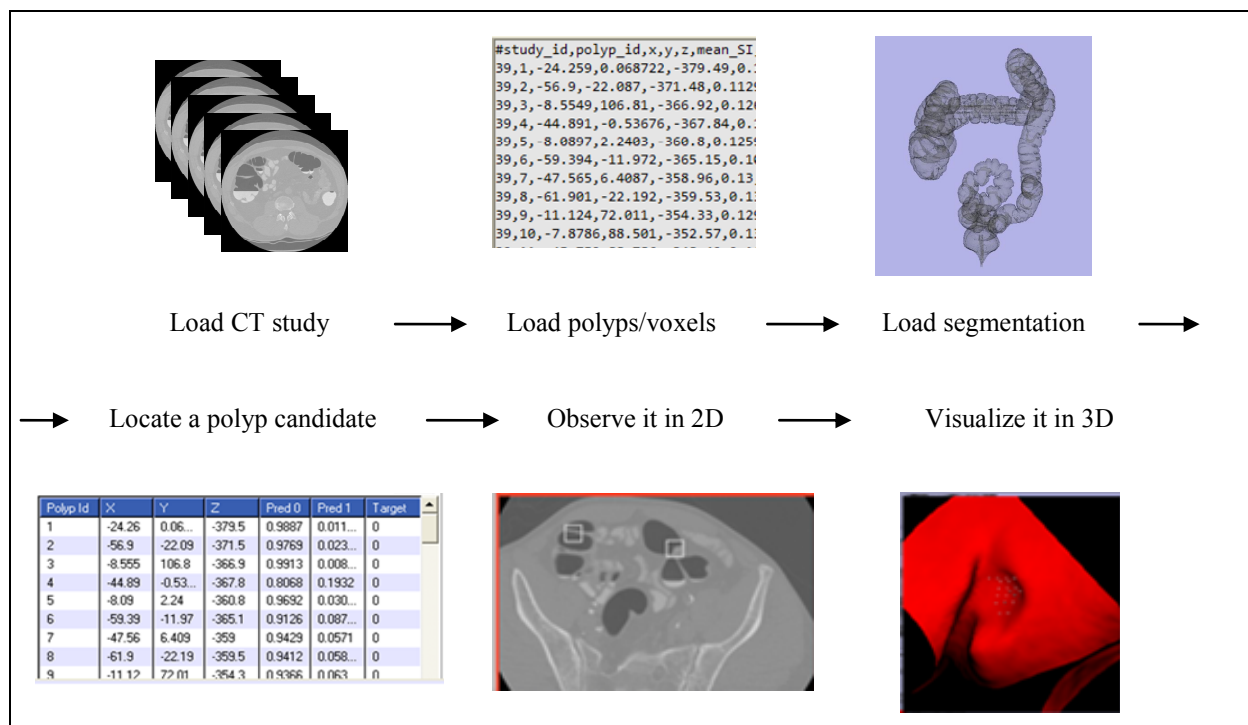


Figure 1. Overall procedure of using the Slicer module to integrate and visualize polyp candidates

Figure 3 shows the user interface of the Slicer module which is designed to present polyp candidates. To locate an intermediate polyp candidate in step four, users click on a polyp candidate in the list (figure 2). The three 2D views of the CT image (sagittal, coronal, and axial) will automatically display the correct slice according to the polyp candidate's location, and a square is placed at the location of the polyp candidate (figure 3, right bottom). At the same time, a dot is shown on the colon surface in the 3D view (figure 3, right top). Users can either observe the details of the polyp candidate in 2D views or visualize it in 3D view by clicking on the dot on the colon surface. To display the 3D view of the polyp candidate, a small segment of CT volume is extracted and a contour operation is applied to generate the polyp candidate surface. All seed voxels for that polyp candidate are shown as small dots inside the polyp candidate region (figure 5, 6). As we can see in figure 3, all distributed information from different steps in a colon CAD is integrated into this user interface. Easy access to both 2D and 3D views of each polyp candidates gives CTC researchers a good opportunity to discover the root causes of false positive detections. Moreover, the distribution of all intermediate polyp candidates for a CT study could be seen in one diagram as shown in figure 4.

### 3. DATA AND RESULTS

We conducted an experiment on WRAMC CTC data acquired by Pickhardt<sup>18</sup> and provided by the NIH. We selected five patients with 12 polyps ( $\geq 6\text{mm}$ ) found in CTC/OC. Our CAD detected 66 polyps, 12 being true positives and 54 being false positives. Among the false positives, 33 are on haustral folds (figure 6a and 6b), 5 are small bump clusters (figure 6c), 4 are small polyps (figure 6d), and 1 is rectal tube (figure 6e).

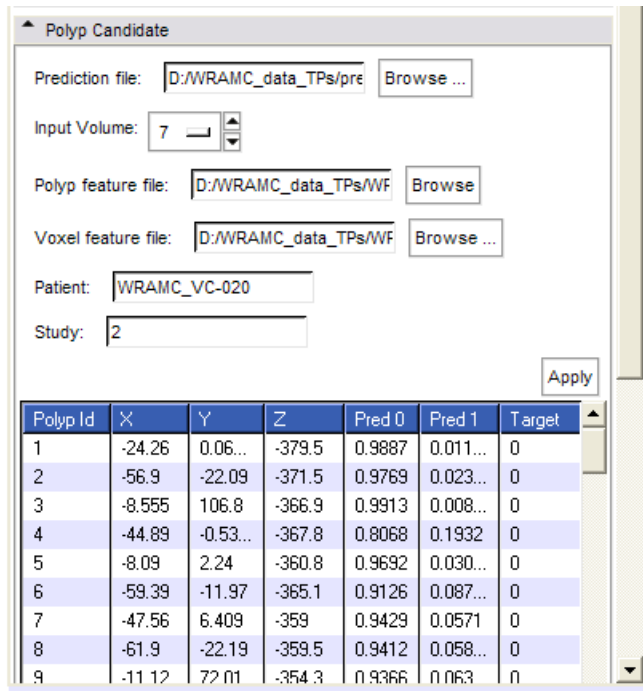


Figure 2. A panel to load and display intermediate polyp candidates

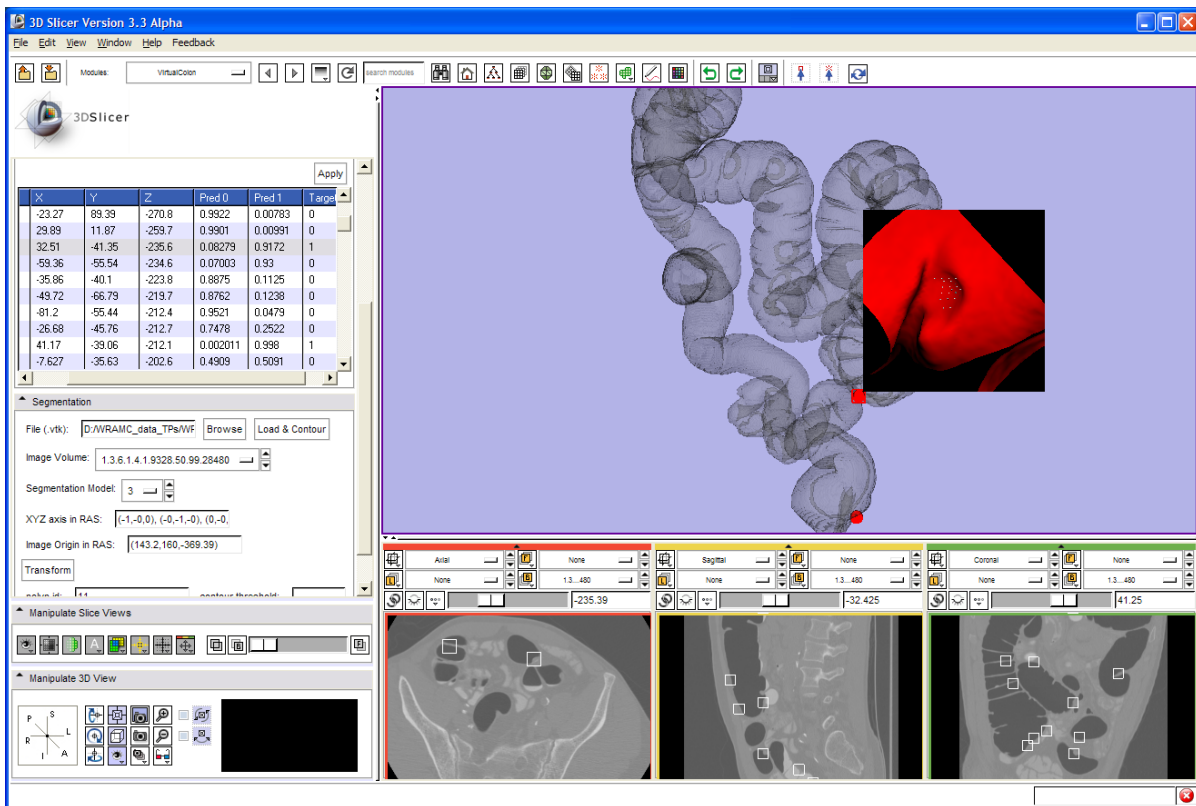


Figure 3. The user interface of the Slicer module to present polyp candidates

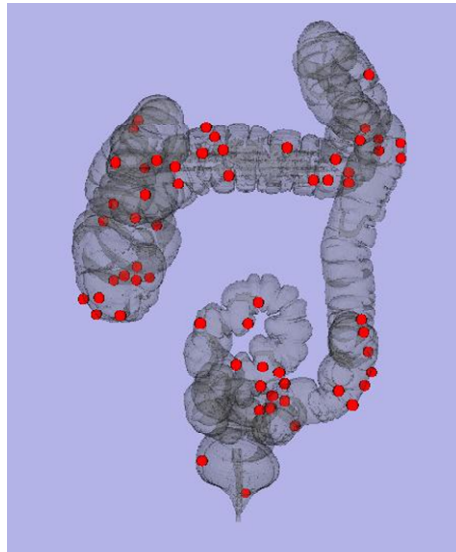


Figure 4. All intermediate polyp candidates for a CT study which are indicated by dots on a colon wall

### 3.1 True positive detections

Figure 5 shows examples of polyps that are successfully detected in our CAD. The prediction values of these polyp candidates are close to 1, which strongly suggests that these are true positive detections. The seed voxels for these polyp candidates are closely clustered inside of the polyp. The pattern of spatial distribution of voxels in true positives (figure 5d) is a strong contrast to the one in false positives (figure 6b): the voxels for true positives form a dense cloud of white dots inside a polyp whereas voxels belonging to false positives (on haustral folds) are adjacent to colon wall. This difference in spatial distribution of seed voxels implies that this distribution could be a useful feature in false positives reduction.

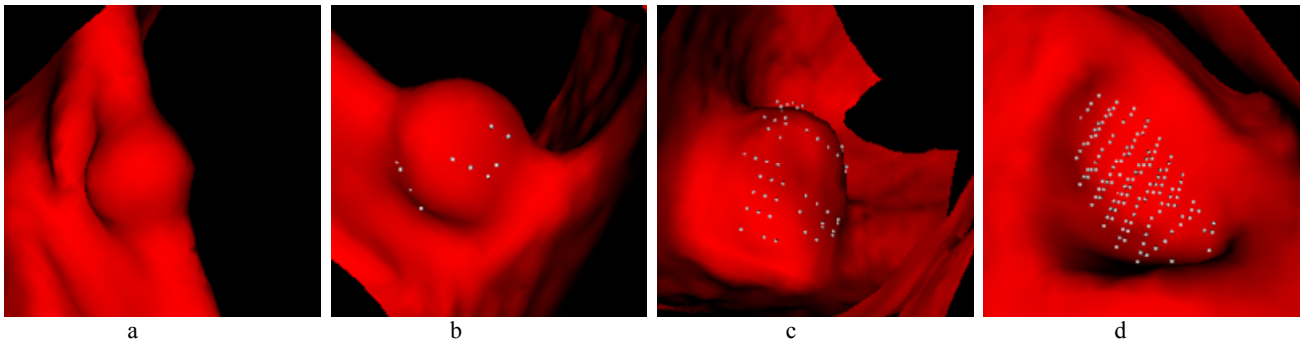


Figure 5. Examples of true positive detections. The red surface is colon wall around polyp candidate region and white dots are seed voxels for that polyp candidate.

### 3.2 False positive detections

Figure 6 shows four typical causes of false positives in our CAD. The majority of false positives are on haustral folds, which comprises up to 61% of all false positives in our experiment. In the majority of these false positives, voxels are close to colon wall as in figure 6b. Another type of false positives (figure 6c) consists of several clusters of small bumps. Although each individual bump could be recognized by the classifier as a non-polyp because of its small size, three closely located small bumps may cause a false positive. Figure 6d is a small size polyp, and figure 6e is a rectal tube.

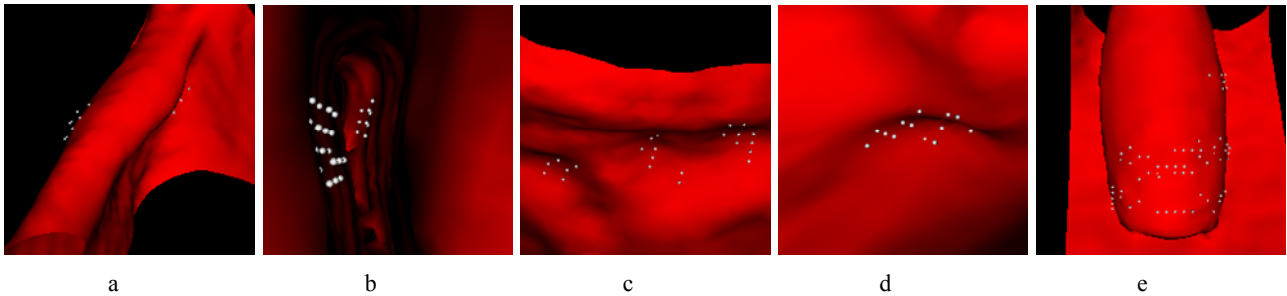


Figure 6. Four causes of false positives: (a, b) haustral fold, (c) clusters of small bump (d) small polyp (e) rectal tube

## 4. DISCUSSION AND CONCLUSIONS

We implemented a Slicer module based on version 3.3 of Slicer on the Windows XP platform. Porting it to Linux and Mac platforms shouldn't be difficult since platform independence is seriously considered in our implementation.

In our colon CAD, we present all intermediate polyp candidates rather than final candidates for two reasons. First, even though some intermediate polyp candidates are correctly classified as negative detections, the prediction value for them are close to 0.5 which implies that the classifier is not very confident to put it as a negative one. Only presenting final polyp candidates, as most commercial colon CAD products do, prevents researchers from studying some intermediate polyp candidates that are similar to polyps. Therefore, we list all intermediate polyp candidates to researchers and let them select one to study. Second, presenting all intermediate polyp candidates makes it easier to evaluate the performance of each step in a colon CAD, i.e., segmentation, generating intermediate polyp candidates, and classification. And it encourages researchers to design a better subsystem in a colon CAD.

In our colon CAD system, several parameters need to be configured experimentally. Please see previous publication<sup>11,12,13,14</sup> for parameter configuration in each individual subsystem. In this Slicer module, two parameters are set in according to the visual effect of polyp candidates. The first parameter is the size of CT volume for each polyp candidate. We crop a CT volume of about 30mm x 30mm x 30mm to get a better coverage of polyp candidates. The second parameter is the threshold value to contour polyp surface. We use -800 as the threshold because it provides the best view of the polyp surface.

In conclusion, we implemented an open source colon CAD system to meet the requirements of researchers in CTC - finding causes of false positive detections and design methods to reduce them. We use Slicer as a platform and integrate results from segmentation, intermediate polyp candidate generation, and classification into one user interface. The visualization of colon surface, polyp candidates, and seed voxels of each polyp candidate gives researchers more power in studying false positive detections. Using this colon CAD we identified four typical causes of false positives in our experimental data. In the future, we will analyze more false positives and design methods to improve the detection performance of our colon CAD.

## 5. APPENDIX

### 5.1 Format of segmentation file:

In the VTK file, the following code scheme is used to represent the type of each voxel.

0: background: not in colon lumen or adjacent to colon wall.

1: lumen: voxel on the interior of the colon

100: colon wall: any voxel adjacent to lumen, may or may not be a polyp candidate.

100+ID: polyp id: polyp candidate voxel which belongs to polyp candidate number ID.

### 5.2 Format of intermediate polyp candidate feature file:

There are 94 columns with 89 of which are polyp features. Each row represents an intermediate polyp candidate:

#Study\_id,polyp\_id,centerx,centery,centerz,mean\_SI,max\_SI,min\_SI,var\_SIskew\_SI,kurt\_SI,contrast\_SI,...

1,1,4.0347,-40.2952,-350.4048,0.1546,0.2485,0.0232,0.0042,-0.568,-0.6806,0.0932,...

### 5.3 Format of seed voxels feature file:

There are 16 columns with nice of which are voxel features. Each row represents a seed voxel:

```
#polypid,centerx,centery,centerz,index_x,index_y,index_z,SI,CV,DGC,GC,Gradient_mag,Gaussian_curvature,Mean_curvature,Min_curvature,Max_curvature
```

1,4.0347,-40.2952,-350.4048,111,212,336,0.1052,0.1164,0.0953,0.467,54.4643,0.0107,0.1101,0.0724,0.1479

### 5.4 Format of classifier's prediction file:

There are five columns with each row representing an intermediate polyp candidate:

```
#study_id,polyp_id,CTC/OC finding,prediction value for negative, prediction value for positive
```

WRAMC\_VC-008/2/15,12,0,0.9318,0.0682

## ACKNOWLEDGEMENTS

We would like to thank Christopher Wyatt, Ph.D. and Yuan Shen from the Department of Electrical and Computer Engineering at Virginia Tech for their work in colon segmentation and polyp candidate generation. We would also like to thank James Perumpillichira, M.D. from the Department of Radiology at Wake Forest University Health Sciences for his help in reading CT studies and finding the location and size of polyps in our experimental data. This research was supported in part by NIH grant 5R01CA114492-03.

## REFERENCES

- [1] Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T., and Thun, M. J., "Cancer statistics, 2008," *CA Cancer J Clin* (58), 71-96(2008).
- [2] Levin, B., Lieberman, D. A., McFarland, B., et al. "Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology," *CA Cancer J Clin* 58(3), 130-160 (2008).
- [3] Mang, T., Graser, A., Schima, W., and Maier, A., "Ct colonography: Techniques, indications, findings," *European Journal of Radiology*, 61(3), 388-399 (2007).
- [4] Yoshida, H. and Dachman, A. H., "CAD techniques, challenges, and controversies in computed tomographic colonography," *Abdominal Imaging* 30(1), 26-41 (2005).
- [5] Yoshida, H. and Nappi, J., "CAD in CT colonography without and with oral contrast agents: progress and challenges," *Computerized medical imaging and graphic* 31(4-5), 267-284 (2007).
- [6] <http://www.slicer.org>
- [7] Gering, D., Nabavi, A., Kikinis, R., Grimson, W., Hata, N., Everett, P., Jolesz, F., and Wells, W., "An Integrated Visualization System for Surgical Planning and Guidance using Image Fusion and Interventional Imaging", *Int Conf Med Image Comput Comput Assist Interv*, 2, 809-819 (1999).
- [8] Pieper, S., Halle, M., and Kikinis, R., "3D SLICER," *Proceedings of the 1st IEEE International Symposium on Biomedical Imaging: From Nano to Macro* 1, 632-635 (2004).
- [9] Nain, D., Haker, S., Kikinis, R., Eric, W., and Grimson, L., "An interactive virtual endoscopy tool," in *Satellite Workshop at the Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'2001)*, 55-60 (2001).
- [10] Bielen, D. and Kiss, G., "Computer-aided detection for CT colonography: update 2007," *Abdominal Imaging*, vol. 32(5), 571-581 (2007).
- [11] Wyatt, C.L., Ge, Y., Vining, and D.J., "Automatic segmentation of the colon for virtual colonoscopy," *Computerized medical imaging and graphics*, 24(1), 1-9 (2000).



- [12] Shen, Y. and Wyatt, C. L., "Open implementation of feature extraction methods for computer aided polyp detection with principal component analysis," in 11th International Conference on Medical Image Computing and Computer Assisted Intervention, 141-147 (2008).
- [13] [http://www.siiim2009.org/abstracts/SS2\\_Xu/SS2\\_Xu.html](http://www.siiim2009.org/abstracts/SS2_Xu/SS2_Xu.html)
- [14] Xu, H., Santago, P., Wyatt, C. L., Shen, Y., and Gage, H. D., Computer-aided polyp detection classifier design in CT colonography. *Journal of Digital Imaging* 2009 (In Review).
- [15] Yoshida, H. and Nappi, J., "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps," *IEEE Trans. Med. Imag.*, 20, 1261-1274 (2001).
- [16] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P., "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, 16, 321-347 (2002).
- [17] Witten, I. H. and Frank E., [Data Mining: Practical Machine Learning Tools and Techniques], 2nd ed., Morgan Kaufmann, (2005).
- [18] 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., and Schindler, W.R., "Computed Tomographic Virtual Colonoscopy to Screen for Colorectal Neoplasia in Asymptomatic Adults", *The New England Journal of Medicine*, 349(23), 2191–2200 (2003).