

# A statistical approach for robust polyp detection in CT colonography

Tarik A. Chowdhury, Ovidiu Ghita and Paul F. Whelan

**Abstract**—In this paper we describe the development of a computationally efficient *computer-aided detection (CAD)* algorithm based on the statistical features derived from the local colonic surface that are used for the detection of colonic polyps in computed tomography (CT) colonography. The candidate surface voxels were detected and clustered using the surface normal intersection, convexity test, region growing and Hough Transform. The main objective of this paper is the selection of the statistical features that optimally capture the convexity of the candidate surface and consequently provide a high discrimination between local surfaces defined by polyps and folds. The developed polyp detection scheme is computationally efficient (typically takes 3.9 minute per dataset) and shows 100% sensitivity for phantom polyps greater than 5mm and 87.5% sensitivity for real polyps greater than 5mm with an average of 4.05 false positives per dataset.

## I. INTRODUCTION

Colon cancer is the second leading cause of cancer deaths in the developed nations [1], [2], [3]. Early detection and removal of colorectal polyps via screening is the most effective way to reduce colorectal cancer (CRC) mortality [4], [5], [6]. Colonography (CTC) [7], [8], [9] is a rapidly evolving noninvasive technology for the detection of colorectal polyps. In the last decade research has been focused on developing automated *computer aided detection CAD* polyp detection techniques and several approaches have been proposed. In this regard, Vining et al. [10] proposed a method to detect the colonic polyps based on surface extraction and curvature analysis and they indicated that a 73% sensitivity with 9 to 90 false positives (FP) per dataset was achieved. Summers et al. [11] developed a method that uses the curvature of colon surface computed by partial derivative and local shape criteria. One problem with this approach is the fact that the sensitivity and specificity of the system depend on the filter chosen to evaluate the shape and as a result their system performed only modestly. Yoshida et al. [12] proposed to evaluate the shape index and curvedness values from a small volume of interest and polyp detection was achieved by employing a fuzzy clustering scheme. They reported 89% sensitivity with 2.0 FP per dataset, but FP per polyp increased with a factor of 1.5 when sensitivity was increased to 100%.

This work was supported under an Investigator Programme Grant (02/IN1/1056) by Science Foundation Ireland (SFI)

Tarik A. Chowdhury is with the Vision Systems Group, School of Electronic Engineering, Dublin City University, Dublin, Ireland. [tarik@eeng.dcu.ie](mailto:tarik@eeng.dcu.ie)

Dr. Ovidiu Ghita is with the Vision Systems Group, School of Electronic Engineering, Dublin City University, Dublin, Ireland. [ghitao@eeng.dcu.ie](mailto:ghitao@eeng.dcu.ie)

Prof. Paul F. Whelan is with the Vision Systems Group, School of Electronic Engineering, Dublin City University, Dublin, Ireland. [whelanp@eeng.dcu.ie](mailto:whelanp@eeng.dcu.ie)

Kiss et al. [13] combined the surface normal distribution and sphere fitting to produce 90% polyp sensitivity for polyps higher than 6mm with 2.82 FPs/dataset. Recently, Kiss et al. [14] also used surface normal and slope density function to produce 85% polyp sensitivity for polyps higher than 6mm with 2.48 FPs/dataset (it is worth mentioning that these results were obtained when the polyp detection algorithm has been applied to datasets with a 0.8mm reconstruction interval). More recently, Paik et al. [15] developed a new technique based on surface normal overlap where the sensitivity was 100% with 7.0 FPs/dataset. Acar et al. [16] method detects the spherical patches using the *Hough Transform* (HT) method [12] and the spherical patches were validated using the optical flow to decide if they are polyps or not. The sensitivity rate of their method was 100%, specificity was 85% and the FPs/dataset was 3. Gokturk et al. [17] also used the HT method for detecting spherical colonic patches and the features collected from a large number of cross sectional images of suspicious structures were supplied to a *support vector machine (SVM)* classifier to perform polyp detection. Acar et al. [18] proposed an edge displaced field-based classification scheme that achieved a 30% reduction of false positive per dataset.

In this paper we propose a computationally efficient method to detect colonic polyps based on statistical features, surface normal concentration, 3D histogram and Hough Transform. The main contribution of this paper is the inclusion of statistical features that maximize the discrimination between the folds and polyps.

## II. MATERIALS AND METHOD

Prior to their scheduled examination all patients were instructed to take a low-residue diet for 48 hours followed by clear fluids for 24 hours. Prior to the day of examination, patients were prescribed one sachet of Pixcolax at 8.00, a second sachet of Pixcolax at 12.00, a sachet of clean prep in a litre of cold water at 18.00 and a Senokot tablet at 23.00. Before the CT scan, a rectal tube is inserted and the colon is gently insufflated with room air to the maximum level tolerated by the patient. All scans were performed on a commercially available Siemens Somatom 4 multi slice Spiral CT scanner. The scanning parameters were 120kVp, 100mAs, 2.5mm collimation, 3mm slice thickness, 1.5mm reconstruction interval, 0.5s gantry rotation. The scanning time range from 20 to 30s, and the image acquisitions were performed in a single breath-hold. The procedure was first performed with the patient in the supine position and then repeated with the patient in the prone position.

### III. CAD ALGORITHM

The developed polyp detection algorithm consists mainly of four steps that include 3D data formatting, extraction of candidate surfaces, calculation of statistical features from detected candidate surfaces and polyp/fold classification. An overview of our algorithm is illustrated in Figure 1.

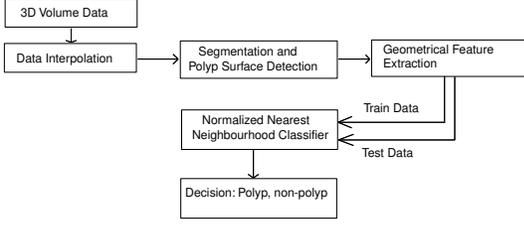


Fig. 1. Overview of the proposed CAD-CTC system.

#### A. Interpolation and Segmentation

Initially the non-isotropic patient data was converted to isotropic data by using cubic interpolation. As the CTC images offer a good contrast between the gaseous and lean tissues the colon can be successfully segmented by using a standard seeded region growing algorithm [20]. It is useful to note that in some datasets remaining residual material and water can create collapses in the colon and the region growing algorithm may require multiple seed points to segment the entire colon. The region growing algorithm is iterative and the threshold value for segmentation was set to -800HU [10]. The adjacent voxels having HU values higher than -800HU define the *colonic wall* (CW).

#### B. Surface Analysis

The normal vector for each colonic wall voxel was calculated using the Zuker and Hummel operator [21]. Each voxel of the CW generates 8 *Hough points* (HP) in the normal direction from 2.0mm to 10mm by varying the  $t$  value in Eq. 1. In Eq. (1)  $p_1$  is the colon wall voxel under investigation and  $n$  is the normal vector at the voxel position.

$$p = p_1 + t \times n \quad (1)$$

The intersections between the normal vectors are obtained by evaluating the 3D histogram for each point in the HP set. As the normal vectors are determined using 3D local operators their orientation is sensitive to abrupt changes in the 3D structure of the CW and to reduce the level of noise in the 3D histogram a weighted smoothing procedure is applied. After smoothing, all HP's having histogram values higher than 5.0 intersections are considered as *initial candidate center points* (ICCP) for polyp. At this stage each point in ICCP creates a cluster of surface points. This cluster of surface points was created by inclusion of the Hough points and their corresponding surface points from the HP within a certain distance (10mm to 16mm). A minimum distance of 10mm was experimentally selected in initial clustering to include the highest possible number of surface points in

the clustered surface. The distance threshold is adaptive and varies from 10mm to 16mm depending on the histogram value for each center point in ICCP. It is useful to note that the candidate surface cluster may include surrounding non-convex surface points or disconnected surfaces. Thus, in order to remove the non-convex surface points and associated HP points from the initial clusters, a convexity test was performed. Convex voxel detection can be performed using different techniques including curvature analysis [12] and normal intersection [13]. In this paper, the non-convex surface points from each candidate cluster were eliminated using the simple convexity test described in [13]. After the removal of the non-convex surface voxels, each cluster was further processed to evaluate discontinuities in the surface under examination. If discontinuities exist in the surface area, the cluster is divided into multiple clusters and they are also considered as potential candidate surfaces. The center of each cluster is calculated using a Gaussian distribution (see Eq. 2) and the Hough point with the highest Gaussian distribution was set as the center of the clustered surface.

$$GM_i = \sum_{j=1}^N e^{(-x^2/2.0 \times \sigma)} \quad (2)$$

In Eq. 2  $x$  is the distance between the Hough points,  $\sigma$  is the standard deviation and is set to 1.0,  $N$  is the number of Hough points in the cluster and  $j$  takes values between  $1 \dots N$ .

### IV. FEATURE EXTRACTION

Our aim is to extract the statistical features associated with each cluster surface that offer optimal discrimination between polyps and folds. The features that we compute for each cluster surface are: standard deviation (SD) of surface variation, SD of the three axis of the ellipsoid, SD of ellipsoid fit error, SD of sphere fit error, Gaussian distribution, major axis of the ellipsoid and the Gaussian sphere radius.

Let  $S_{init}$  be the *surface number* (SN) in the candidate cluster (CS) and  $d_{max}$  the maximum distance from the cluster center to the surface normal (see Figure 2). For each iterated radius  $R_i$  (see Eqns. 3,4) the new cluster surface  $S_i$  was calculated from the candidate surface (CS). The standard deviation  $SN_{SD}$  of the cluster surface variation was calculated using the Eqns. 5,6,7. Experimental results indicate that for folds the SD of the surface variation is significantly higher than the SD of the surface variation for polyps. This can be observed in Figure 3 (note that polyp and fold classes are ordered by size in the diagram) where the SD of the surface variation for a large variety of polyps and folds is plotted.

$$Step = (D_{max} - 1.0)/N \quad (3)$$

$$R_i = D_{max} - Step_i \quad for \quad i = 1, \dots, N, \quad (4)$$

$$SN_{mean} = \sum_{j=1}^N SN_i \quad (5)$$

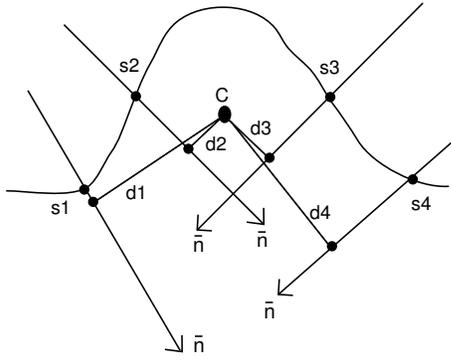


Fig. 2. Cluster Center to surface normal distance ( $d_N$ ).  $s_1...s_4$  are the surface voxels,  $n$  is the normal at each surface voxels and  $d_1..d_4$  are distances from cluster center  $C$  to normal  $n$ .

$$SN_{inorm} = \frac{SN_i}{SN_{mean}} \quad \text{for } i = 1, \dots, N, \quad (6)$$

$$SN_{SD} = \sqrt{\frac{1}{N} \sum_{j=1}^N (SN_{inorm} - SN_{mean})^2} \quad (7)$$

where  $Step$  is the step size for radius change and  $N$  is the number of steps required to reduce the radius  $R_i$  from  $D_{max}$  to 1.0mm. In Eq. 5  $SN_i$  is the surface number ( $SN$ ) calculated for each radius  $R_i$  and  $SN_{mean}$  is the mean of  $SN_i$  for the cluster  $CS$ . In Eqns. 6,7  $SN_{inorm}$  is the normalized surface number and  $SN_{SD}$  is standard deviation of the cluster  $CS$ .

As additional features we calculated the least square ellipsoid fitting [22] for each surface  $S_i$  of the cluster  $CS$  and calculate the SD for each axis of the ellipsoid. In similar way, we calculated the SD of the ellipsoid fitting error and SD of the sphere fitting error [22]. In line with aforementioned features we have also included the maximum value of the Gaussian distribution for each cluster that was calculated as indicated in the Surface analysis section and the major axis and sphere radius of each cluster surface. All these features are used to classify the candidate surfaces into polyps or folds using a feature normalised nearest neighbourhood classification scheme [23]. The classifier was trained with 64 polyps and 354 folds that were selected as true positives by a radiologist.

## V. RESULTS

Thirty six patients' supine and prone data with 57 polyps, five patients' data with 33 synthetic polyps [24] and a phantom data with 47 polyps of various sizes [25] were tested using the proposed method. Overall sensitivity for real polyp detection was 70.175% with a false positive rate of 4.05 per dataset (Table I). Sensitivity for polyps greater than 5mm was 87.5%. Sensitivity for polyps less or equal to 5mm was 61.76% and for masses the detection was 71.429%. When the algorithm was applied to synthetic polyp detection, the overall sensitivity was 84.08% and the false positive level per dataset was 3 (Table II). Sensitivity for polyps higher than 5mm was 100% and less than 5mm was 33.33%. For

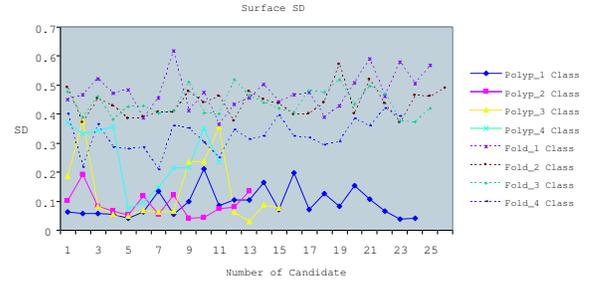


Fig. 3. Standard deviation of the surface variation for different classes of polyps and folds (classes are sorted by size).

phantom data, overall sensitivity was 89.58% (Table III). Sensitivity for polyps  $\leq 5mm$ , 5 to 10mm,  $> 10mm$  and flat polyps were 100%, 100%, 100% and 44.44% respectively (Table III). For comparative testing purposes we have made the phantom data available from the following web page: <http://www.eeng.dcu.ie/~whelanp/cadctc>

To determine whether a polyp was correctly detected by the proposed algorithm, we cross correlated the detected polyps' location with the CTC reports performed by the radiologists. Also we compared the result with the Colonoscopy reports for both supine and prone views. It is important to mention that approximately 20% of the polyps are visible in only one view and as a consequence there was only one chance to detect these polyps.

The average size of typical interpolated CT dataset was 300MB for each view. The average time required for processing each volume of data was approximately 3.9 min on a Pentium-IV 1.6 GHz processor machine with 1GB memory.

TABLE I  
PERFORMANCE ANALYSIS FOR REAL POLYP DATA

Type	Number	True Positive	Sensitivity
$\leq 5mm$	34	21	61.765%
$> 5mm$	16	14	87.5%
Mass	7	5	71.429%
False Positive			4.05
Total	57	40	70.175

TABLE II  
PERFORMANCE ANALYSIS FOR SYNTHETIC POLYP DATA

Type	Number	True Positive	Sensitivity
$\leq 5mm$	6	2	33.33%
$> 5 - \leq 10mm$	17	17	100%
$> 10mm$	9	9	100%
Flat	1	0	00.00%
False Positive			3
Total	33	27	84.85

### A. Discussion and Conclusion

The proposed statistical feature-based CAD system for colonic polyp detection provides high sensitivity yet maintaining a low false positive incidence per dataset. Our polyp

TABLE III  
PERFORMANCE ANALYSIS FOR PHANTOM DATA

Type	Number	True Positive	Sensitivity
$\leq 5mm$	5	4	80%
$> 5 - \leq 10mm$	19	19	100%
$> 10mm$	14	14	100%
Flat	9	4	44.44%
False Positive			0
Total	47	42	89.36

detection scheme was not able to correctly classify the polyps that are adjacent to fold or on fold. The statistical features derived from small and medium polyps when positioned adjacent to folds or on folds show similar characteristics as generic folds, and the classifier detected them as folds. When the CAD system was applied to real datasets, 30% (5 out of 15) of the undetected small polyps were placed adjacently to folds and the classifier failed to identify them correctly. However, a better surface detection technique can be envisioned and this will increase the polyp detection rate when they are situated adjacently to folds. Approximately 10% of the false positives were generated by the residual material attached to the colonic surface. These false positives can be eliminated by texture analysis [19] as they have a different density than folds and polyps.

By using surface normal intersection and statistical features calculated from polyps/folds morphology and least square fitting, we aimed to extract the optimal features that capture the surface convexity. In fact, SD of surface variation, SD of three axis of ellipsoid fitting, SD of sphere fitting, SD of ellipsoid fitting error and Gaussian distribution provide a detailed morphological description of the candidate surface.

The experimental data indicate that our polyp detection technique shows better results when compared to other existing techniques. Another advantage of our method is its low computational overhead and more importantly it shows high sensitivity for medium (6 – 9mm) and large (> 9mm) polyps while the false positive rate is maintained at low levels. The experimental results indicate that our CAD polyp detection technique is a suitable tool to be utilized in clinical studies.

## VI. ACKNOWLEDGMENTS

We would like to acknowledge the contributions of our clinical partners in this project: Dr. Helen Fenlon (Department of Radiology) and Dr. Padraic MacMathuna (Gastrointestinal Unit) of the Mater Misericordiae Hospital, Dublin. This work was supported under an Investigator Programme Grant (02/IN1/1056) by Science Foundation Ireland (SFI).

## REFERENCES

- [1] S. Parker, T. Tong, S. Bolden, P. Wingo, "Cancer statistics 1997," *Journal for Clinicians*, vol. 47, pp. 5-27, 1997.
- [2] NCRI, "Cancer in Ireland, 1997: Incidence and mortality", *Healy & Associates*, 2000.
- [3] Cancer Research UK, "Bowel cancer factsheet", April 2003.
- [4] F.R. David, S.S. Robert, "Screening for colorectal cancer", *N Eng J Med*, 346, no. 1, January 3, 2002.

- [5] American Cancer Society, "Cancer facts and figures", 1999
- [6] National Cancer Institute, "Working guidelines for early cancer detection: Rationale and supporting evidence to decrease mortality", *Bethesda: National Cancer Institute*, 1987.
- [7] D.J. Vining, D.W. Gelfand, R.E. Bechtold, "Technical feasibility of colon imaging with helical CT and virtual reality", *AJR*, vol. 162 pp. 104, 1994.
- [8] C.D. Johnson, A.K. Hara, J.E. Reed, "Virtual endoscopy: What's in a name?", *AJR*, vol. 171, pp. 1201-1202, 1998.
- [9] A.K. Hara, C.D. Johnson, J.E. Reed, D.A. Ahlquist, H.Nelson, R. L. Ehman, "Detection of colorectal polyps by CT colonography: Feasibility of a novel technique", *Gastroenterology*, vol. 100, pp. 284-290, 1996.
- [10] D. J. Vining, G. W. Hunt, D. K. Ahn, D. R. Stelts, P. F. Helmer, "Computer-assisted detection of colon polyps and masses", *Radiology*, vol. 219, pp. 51-59, 2001.
- [11] R.M. Summers, C.D. Johnson, L.M. Pusanik, J.D. Malley, A.M. Youssef, J.E. Reed, "Automated polyp detection at CT colonography: Feasibility assessment in a human population", *Radiology*, vol. 219, pp. 51-59, 2001.
- [12] H. Yoshida, Y. Masutani, P. MacEneaney, D. T. Rubin, A. H. Dachman, "Computerized detection of colonic polyps at CT colonography on the basis of volumetric features: Pilot study", *Radiology*, vol. 222, pp. 327-336, 2002.
- [13] G. Kiss, J. Cleynenbreugel, M. Thomeer, P. Suetens, G. Marchal, "Computer-aided diagnosis in virtual colonography via combination of surface normal and sphere fitting methods", *European Radiology*, vol. 12(1), pp. 77-81, 2002.
- [14] G. Kiss, J. Cleynenbreugel, P. Suetens, G. Marchal, "Computer aided diagnosis for CT colonography via slope density functions", *MICCAI 2003*, pp. 746-753, 2003.
- [15] D. S. Paik, C. F. Beaulieu, G. D. Rubin, B. Acar, R. B. Jeffrey, J. Yee, J. Dey, S. Napel, "Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical CT", *IEEE Trans Med Imaging*, vol. 23(6) pp. 661-75, 2004.
- [16] B. Acar, S. Napel, D. Paik, S. B. Gokturk, C. Tomasi, C. F. Beaulieu, "Using optical flow fields for polyp detection in virtual colonoscopy", *Proc of Medical Image Computing and Computer-Assisted Intervention*, Utrecht, Holland, 2001.
- [17] S.B. Gokturk, C. Tomasi, B. Acar, C. F. Beaulieu, D.S. Paik, R.B. Jeffrey, J. Yee, Napel S., "A statistical 3D pattern processing method for computer aided detection of polyps in CT colonography", *IEEE Trans. Med. Imaging*, vol. 20(12), pp. 1251-1260, 2001.
- [18] B. Acar, C.F. Beaulieu, S.B. Gokturk, C. Tomasi, D.S. Paik, R.B. Jeffrey Jr, J. Yee, S. Napel, "Edge displacement field-based classification for improved detection of polyps in CT colonography", *IEEE Trans Med Imaging*, vol. 21(12), pp. 1461-7, 2001.
- [19] Z. Wanga, L. Liab, J. Anderson, D. Harrington, Z. Lianga, "Colonic polyp characterization and detection based on both morphological and texture features", *International Congress Series*, vol. 1268, pp. 1004-1009, 2004.
- [20] R.C. Gonzalez, R.E. Woods, *Digital image processing*, Reading MA: Addison-Wesley, 1993.
- [21] S.W. Zucker, R.A. Hummel, "A Three-Dimensional edge operator", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 3(3), pp. 324-331, 1981.
- [22] *Least square approximation*, <http://www.magic-software.com/Documentation.html>.
- [23] O. Ghita, P.F. Whelan, "A bin picking system based on depth from focus", *Machine Vision and Application*, vol. 13, pp. 234-244, 2003.
- [24] N. Sezille, R.J.T. Sadleir, P.F. Whelan, "Automated synthesis, insertion and detection of polyps for CT colonography", *Opto-Ireland - SPIE's Irish Meeting on Optoelectronics, Photonics and Imaging*, Galway, September 5th-6th, 2002.
- [25] T.A. Chowdhury, P.F. Whelan, H. Fenlon, P. MacMathuna, "Evaluation of radiation dose on automatic polyp detection at CT colonography: Experiments with a synthetic phantom," *Association of Physical Scientists in Medicine*, Annual Scientific Meeting, Galway, Ireland, 2005.