

# Symmetric Curvature Patterns for Colonic Polyp Detection

Anna Jerebko, Sarang Lakare, Pascal Cathier\*, Senthil Periaswamy, and Luca Bogoni

Siemens Medical Solutions, CAD group, Malvern, PA 19380, USA  
{Anna.Jerebko, Sarang.Lakare, Senthil.Periaswamy,  
Luca.Bogoni}@siemens.com

**Abstract.** A novel approach for generating a set of features derived from properties of patterns of curvature is introduced as a part of a computer aided colonic polyp detection system. The resulting sensitivity was 84% with 4.8 false positives per volume on an independent test set of 72 patients (56 polyps). When used in conjunction with other features, it allowed the detection system to reach an overall sensitivity of 94% with a false positive rate of 4.3 per volume.

## 1 Introduction

As CT Colonography (CTC) is gaining broader acceptance as a screening method for colorectal neoplasia [1], [2], [3], [4] which is one of the leading causes of cancer deaths in the US, in a rather near future CAD tools could be employed as a second reader to aid in polyp detection.

In this paper, we introduce a novel approach for generating a set of features derived from properties of patterns of curvature to characterize the shape, size, texture, density and symmetry with the goal of facilitating the differentiation between polyps and other structures in the volume of interest. Our computer aided polyp detection algorithm consists of the following steps: candidate generation, feature extraction and classification. A detailed description of the feature extraction algorithm, the focus of this paper, is given in Section 2. We briefly explain the candidate generation step in Section 3, followed by introduction of the feature selection and classification methods. Error validation approach and the results on the training and test sets are reported in Section 4. The conclusion is given in Section 5.

## 2 Symmetrical Curvature Pattern Based Features for Polyp Characterization

### 2.1 Symmetric Pattern Extraction

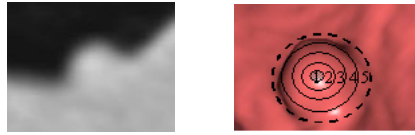
The new technique presented in this paper allows generating a set of discriminative features for a given set of candidate locations. The features aim at capturing a general class of mostly symmetrical and round structures protruding inward into the lumen

---

\* Pascal Cathier is now at SHFJ, CEA, Orsay, France.

(air within the colon), having smooth surface and density and texture characteristics of muscle tissue. These kinds of structures exhibit symmetrical change of curvature sign about a central axis perpendicular to the object's surface. Colonic folds, on the other hand, have shapes that can be characterized as half-cylinders or paraboloids and hence do not present similar symmetry about a single axis.

We first apply 3D Canny edge detection to find the boundary between tissue and colon lumen (air). Next, mean (H) and Gaussian (K) curvatures are calculated for each boundary voxel using the intensity values of the original volume [5]. Then, centering on each boundary voxel, all the connected neighbors within a maximum distance ( $r_{max}$ ), determined by the maximum targeted polyp's size (20 mm in our experiments, plus extra ~5 mm was added in order to include the part of the boundary where the polyp merges with the surrounding tissue) are considered. These neighbors are grouped radially into *concentric bands* (see Fig.1).



**Fig. 1.** Left: 2D view of a sessile polyp; Right: rendered view with overlaid concentric bands of increasing radius. The dashed circle illustrates the inflection band where polyp merges with the surrounding tissue.

Average and standard deviations of K and H are computed within each band and then, the curvature slope is computed from the top of the polyp to the place of merging. When the center is located on the top of a polypoid structure, the bands having minimal Gaussian curvature ( $K_{min}$ ) are likely to correspond to the polyp's inflection band (see Fig. 1) – the transition area between the polyp and normal healthy tissue.

Then K and H curvature slopes  $S_K$  and  $S_H$  are calculated as:

$$S_K = (K_c - K_{min}) / D_{Kmin} \quad (1)$$

$$S_H = (H_c - H_{max}) / D_{Hmax} \quad (2)$$

Where  $S_K$  – Gaussian curvature slope,

$S_H$  – mean curvature slope,

$K_c$  – average Gaussian curvature in the top central band,

$H_c$  – average mean curvature in the top central band,

$D_{Kmin}$  - distance from the top to the band having minimal Gaussian curvature,

$D_{Kmax}$  - distance from the top to the band having maximum mean curvature.

Voxels having negative Gaussian curvature slope are likely to correspond to a polyp or other protrusions on the surface. Folds or other cylindrical structures could also have similar properties, but the standard deviations of K and H curvature within the bands having minimal Gaussian curvature are significantly higher in folds than in polyps. In a specific VOI, the cluster having minimum average  $S_K$  among all clusters is then selected as most likely corresponding to the boundary of polypoid object of interest.

The following features are derived from the concentric bands centered in the voxel of this cluster, which has maximum slope  $S_K$ : minimum  $S_H$ , standard deviations of  $K$  and  $H$  in the polyp inflection band (which could be in different distances for  $K$  and  $H$ ), distance from the central voxel to the inflection band  $D_{Hmax}$  and  $D_{Kmin}$ , standard deviations of  $K$  and  $H$  in the central band, largest diameter  $d_1$ , two orthogonal diameters  $d_2, d_3$  and their ratios  $\lambda_1, \lambda_2, \lambda_3$ .

## 2.2 Cutting Surface Interpolation

Colonic polyps are attached to colon wall, folds or could be partially (or fully) surrounded by stool with no visible intensity difference, and hence detectable edges, amongst all these objects. Thus, after detecting the cluster of air/tissue boundary voxels belonging to a polypoid object, it is necessary to interpolate underlying (cutting) surface to separate polyp from the tissue it is attached to.

The segmented polyp surface is transformed to spherical coordinates using the center mass of the lesion boundary voxels as the origin for the transformation. The new coordinates  $r(\varphi, \theta)$  of every point  $(x, y, z)$  are estimated using linear interpolation. The discretization degree is computed as  $D = \text{round}(2\pi r_{max})$  so that a voxel at a maximum radial distance,  $r_{max}$  from the center of the transformation, maps to a single voxel in spherical space.

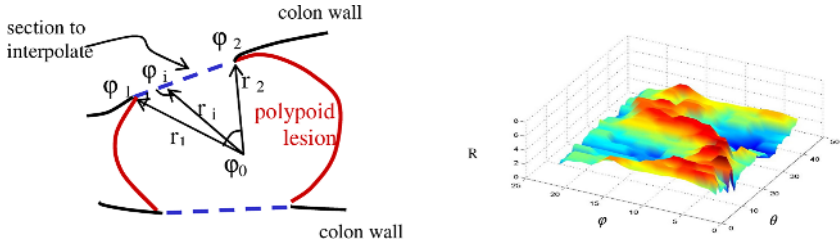
Intuitively, as a result of this step, continuous patches of surface in spherical coordinates will represent an unwrapped object boundary. The following features are derived from the object surface in spherical coordinates: the average, standard deviation, and skewness of radius. These features will help to characterize the object size and shape and its similarity to a fully symmetrical portion of a sphere.

The areas where there was no boundary voxel for the given  $(\varphi, \theta)$  correspond to the place where polyp is attached to the colon wall, or just touching other portions of the colon wall, fold or stool. Usually the largest gaps in the image  $r(\varphi, \theta)$  correspond to the base of a polyp, or attachment location. Additionally there could be other gaps where object boundary is absent or is in contact with other portion of the colon wall. The neighbors for each empty point in spherical coordinates can be much more easily located than in Cartesian space. For example, in case of an empty hemisphere, which has random holes, only the set of points from the rim are needed to interpolate the base, and other sets of points from the edge of each hole to fill the holes.

In our implementation, for each empty voxel two closest neighbors are determined from the existing surface patches first in vertical ( $\varphi$ ) and then in horizontal ( $\theta$ ) directions, (more neighbors could be used for better precision).

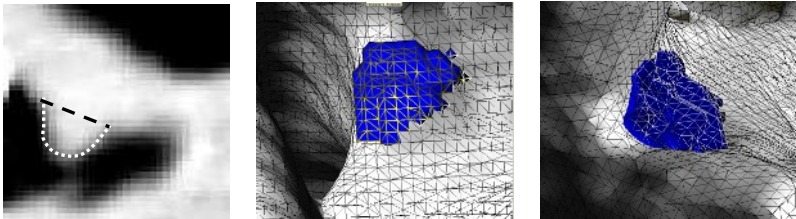
Fig. 2 shows a sketch in Cartesian coordinates for a section to interpolate where two closest neighbors, for a fixed value of  $\theta$ , have been identified. These two points (identified by the tips of the arrows from the center) and the center of coordinates form a triangle for which two sides (radii  $r_1$  and  $r_2$ ) and the angle between them  $\varphi_0$ , in  $\varphi$ -direction ( $\theta_0$ , in  $\theta$ -direction) are known. By solving this triangle relation, it is possible to interpolate all points lying along the third side (dashed line segment connected the two endpoints of radii  $r_1$  and  $r_2$ ), knowing the angle between the third side and the radial line to be interpolated. For any empty point in spherical coordinate system, the values for  $\varphi$  and  $\theta$  are known and so is the angle with the origin of the

transformation. For each value of  $\varphi$  between  $\varphi_1$  and  $\varphi_2$  each interpolating radius value spanning the linear segment can be computed. Segment interpolation is linear in Cartesian space; in spherical space it spans a concave arc (see Fig 2. Right).



**Fig. 2.** Left: Interpolation of cutting surface in Cartesian space. Right: Interpolation results in spherical space.

Final values of radii in the empty spots are computed as average of  $\varphi$  and  $\theta$  neighbors. Once the interpolation is completed in spherical space, all points of the interpolated surface are mapped back to the Cartesian coordinates.



**Fig. 3.** Left: 2D view of a rectal polyp shows no intensity difference between the polyp and colon fold it is attached to. Air/polyp boundary is shown by dotted white line and the black dashed line shows the cutting line to be interpolated. Middle: Rendered view of the top polyp surface. Right: Rendered view of interpolated cutting surface. The segmented voxels belonging to the polyp are highlighted in the rendered views. (Rendered views are not in scale with the 2D view).

### 3 A Polyp Detection System

#### 3.1 Candidate Generation

The data pre-processing phase includes colon segmentation and a transformation to render the volume isotropic. The candidate generation approach is based on the observation that, planar cross-sections of the volume at different orientations will yield roughly circular cuts only in the case of polyps. After the segmentation of the lumen (air inside the colon), our implementation uses 13 planar orientations to search for cross-sections with such characteristics around each voxels. The overall algorithm is

computationally efficient and produces a rather low number of candidates per volume, typically less than 60 candidates/volume (over 100 candidates/patient).

### 3.2 False Positive Reduction: Classification

The next step aims at filtering out most of the false positives by means of features characterizing the resemblance of the detections to our model (described in Section 2).

Greedy search method [6] was used for feature selection using the set of features described throughout the paper (see Table 1) and quadratic discriminant analysis classifier (QDA) [7] was used for false positive reduction.

**Table 1.** Set of features for classification chosen by a greedy search method

#	Feature	Feature description
1	$S_H$	Mean curvature slope
2	$StdK_c$	Standard deviation of K in the central ring
3	$StdH_c$	Standard deviation of H in the central ring
4	$StdH_b$	Standard deviation of H in the polyp inflection belt
5	$d_2$	Second maximum diameter
6	$d_3$	Third maximum diameter
7	$\lambda_1$	Ratio $d_1 / d_2$ (first to second maximum diameter)
8	MeanInt	Average intensity inside the object
9	MaxInt	Maximum intensity inside the object
10	StdInt	Standard deviation of intensity inside the object

## 4 Experiments and Results

### 4.1 Data Characteristics

In order to evaluate the discriminating power that features computed from curvature patterns provide, we used a dataset of 168 patients with, 329 prone and supine volumes (some cases did not have both prone and supine studies) from high-resolution CT scanners obtained from four different hospitals. These included both patients with polyps (positive cases) ( $n=81$ ) and patients without polyps (negative cases) ( $n=87$ ). The sensitivity and false positive rate of our system was established with respect to CTC by comparison to results from concurrent fiber-optic colonoscopy.

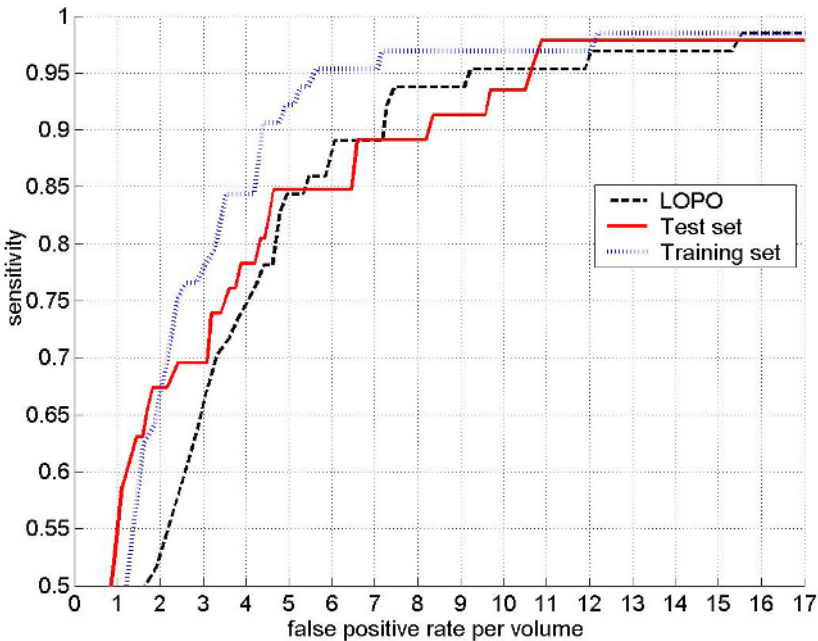
The datasets were randomly partitioned into two groups: training set ( $n=96$ , 187 volumes) and test set ( $n=72$ , 142 volumes). The training set had a total of 76 polyps: 33 small (smaller than 6 mm), 30 medium (6 to 9.99 mm), 13 large (10 to 20 mm). The test set had total of 56 polyps: 23 small, 22 medium, and 11 large. Polyps smaller or equal to 3 mm were then excluded from both training and test sets leaving 71 and 55 in the training and test sets respectively.

The test set was sequestered and only used to evaluate the performance of the final system. The training-set was used to design the classifier and to automatically select the relevant features as described above.

The system's performance was determined on the training-set using Leave-One-Patient-Out (LOPO) cross-validation procedure. In this scheme, both supine and prone views of a case are left out. The classifier is trained using the volumes from the remaining  $N-1$  cases, and tested on volumes of the "left-out" case. This process was repeated  $N$  times, and the resulting testing errors were averaged to determine the LOPO error. In the computation of the per patient sensitivity, a polyp was considered as "found" if it was detected in at least one of the volumes (supine or prone) from the same patient.

## 4.2 Results

**Results on the Training Set:** The candidate generation yielded an average of 51.7 candidates per volume while missing 6 small (2 of which were 3 mm and smaller) and 2 large polyps. LOPO cross-validation yielded a false positive rate of 4.9 per volume and overall sensitivity 84.4%. When considering polyps lost during the candidate generation phase, the overall sensitivity reduces to 77.5%.



**Fig. 4.** ROC curves for LOPO, training and test rates for false positive reduction step (without candidate generation losses taken into account)

**Results on Test Group:** The test group was used for evaluation and provided an estimate of performance on completely unseen data. The candidate generation phase yielded an average of 47.9 candidates per volume while missing 7 small (one of

which was smaller than 3 mm) and 1 medium polyps. We obtained a false positive rate of 4.8 per volume and an overall sensitivity was 84.8%. The overall sensitivity on the test set comes to 73.6% when the polyps lost during the candidate generation are taken into account.

A variation of receiver operating characteristic curve (ROC), plot of sensitivity of the QDA classifier using the set of features listed in Table 1 against its false positives rate, (without candidate generation losses taken into account) is shown in the Fig.4 below. The plot shows ROC curves for LOPO validation, training and test sets. The operating point on the ROC curve was chosen according to the desired classifier sensitivity of around 85% with false positives rate less than 5 for both test set and LOPO validation rate.

## 5 Conclusion

In this paper we presented a method for colonic polyp detection and generation of features for classification based on the symmetry of curvature patterns. The high sensitivity and low false positive rate achieved demonstrates that the proposed algorithm allows generating highly discriminative features suitable for computer aided colonic polyp detection. In fact, when used in conjunction with other features, it allowed the detection system to reach an overall sensitivity of 94% with a false positive rate of 4.3 per volume [3]. This approach could also be applied to the detection and segmentation of other abnormal anatomical structures such as aneurysms or lung nodules. The curvature pattern based polyp segmentation method used as a part of this could be used to automatically obtain accurate polyp size measurements [10]. Finally, while classification was not the focus of the paper, the false positives rate could be further reduced by using a more advanced classification such as methods described in [8], [9]; applying detection ranking scheme, and by introducing additional features.

## References

1. Pickhardt P.J, Choi J.R., Hwang I., Butler J.A., Puckett J.L., Hildebrand H.A., "Computer Tomographic Virtual Colonoscopy to Screen for Colorectal Neoplasia in Asymptomatic Adults", *NEJM*-2003, 349(23): 2191-2200.
2. Sosna, J., Morrin, M. M., Kruskal, J. B., Lavin, P. T., Rosen, M. P., Raptopoulos, V. (2003). CT Colonography of Colorectal Polyps: A Metaanalysis. *AJR* 181: 1593-1598.
3. Bogoni L., Cathier P., Dundar M., Jerebko A., Lakare S., Liang J., Periaswamy S., Baker M.E., Macari M., "Computer-aided detection (CAD) for CT Colonography: a tool to address a growing need", *The British Journal of Radiology*, 78, pp. 57-62 (2005).
4. Yoshida H, Nappi J, MacEneaney P, Rubin DT, Dachman AH. Computer-aided diagnosis scheme for detection of polyps at CT colonography. *Radiographics* - Jul 2002 (Vol. 22, Issue 4).
5. Thirion, J.P. and Gourdon, A. "Computing the Differential Characteristics of Isointensity Surfaces," *Computer Vision and Image Understanding*, vol. 61, no. 2, pp. 190-202, 1995.
6. John, G. H., Kohavi R., Pflieger K., "Irrelevant Features and the Subset Selection Problem", *International Conference on Machine Learning*, 1994.

7. Fukunaga K.: Introduction to Statistical Pattern Recognition. San Diego: Academic Press, 1990.
8. Jerebko A.K., Malley J.D., Franaszek M., Summers R.M.: "Multinetwork classification scheme for detection of colonic polyps in CT colonography data sets", *Acad Radiol* 2003; 10:154–160
9. Jerebko A.K., Malley J.D., Franaszek M., Summers R.M.: Computer-aided polyp detection in CT colonography using an ensemble of support vector machines. *CARS 2003 Vol. 1256*, 1019 - 1024 (2003).
10. Jerebko, A.K., Lakare, S., Mang, T., Mueller, C., Salovich, D., Schima, W., Wessling, J., Bogoni L.: Evaluation of an automatic polyp measurement tool compared with experienced observers: A comparative study, *ECR 2005*.