

Computer Aided Detection in CT Colonography, via Spin Images

Gabriel Kiss, Johan Van Cleynenbreugel, Guy Marchal, and Paul Suetens

Faculties of Medicine & Engineering
Medical Image Computing (Radiology - ESAT/PSI)
University Hospital Gasthuisberg
Herestraat 49, B3000 Leuven, BELGIUM
`Gabriel.Kiss@uz.kuleuven.ac.be`

Abstract. A technique for Computer Aided Detection (CAD) of colonic polyps, in Computed Tomographic (CT) Colonography is presented. Following the segmentation of the colonic wall, normal wall is identified using a fast geometric scheme able to approximate local curvature. The remaining structures are modeled using spin images and then compared to a set of existing polypoid models. Locations with the highest probability of being colonic polyps are labeled as final candidates. Models are computed by an unsupervised learning technique, using a leave one out technique on a study group of 50 datasets. True positive and false positive findings were determined, employing fiber optic colonoscopy as standard of reference. The detection rate for polyps larger than 6mm was above 85%, with an average false positive detection rate of 2.75 per case. The overall computation time for the method is approximately 6 minutes. Initial results show that Computer Aided Detection is feasible and that our method holds potential for screening purposes.

1 Introduction

Alongside lung cancer in males and breast cancer in females colorectal cancer is one of the most frequent causes of cancer related deaths in the industrialized world. Studies [1] show that it has a mortality rate of 20-25 cases/10⁵/year. The typical sequence is from adenoma to carcinoma and the evolution includes in 95% of the cases the appearance of adenomatous polyps [2]. Luckily, this process is a lengthy one taking from 5 up to 10 years. Furthermore, the survival rate after 5 years is 92%, when early treatment is received and polyps are removed endoscopically [3]. The lasting prophylactic effects of polyp removal make screening for adenomatous polyps a viable alternative to reduce mortality rate due to colorectal cancer.

Numerous methods are available for colonic polyps detection, well-known methods include: fecal occult blood testing (FOBT), barium enema examinations, sigmoidoscopy and conventional colonoscopy. When looking for the most effective methods for polyp detection the choice is between conventional colonoscopy and barium enema examinations. It has been proven that conventional colonoscopy has the highest accuracy in detecting polyps and furthermore,

once found polyps can be immediately removed. Although, considered as the gold standard conventional colonoscopy has its pitfalls, when patient screening is taken into consideration. It is quite invasive, there is a small risk of wall perforation and in some cases the examination of the entire colon is not attainable.

That is why lately other methods usable for screening were investigated. Among these FOBT is the only one used for clinical screening. Although, safe and inexpensive it has a low sensitivity and thus effectiveness. Lately, CT colonography has emerged as an alternative for screening. Introduced in 1994 by Vining et. al. [4], it is a method for exploring the colonic area hinging on CT data. When over 1000 axial slices have to be interpreted it is not difficult to see that perceptual errors will have a negative effect on the performance of CT colonography. To overcome this problem computer aided detection (CAD) methods have been developed; they automate the process of finding polypoid structures.

In the literature some methods have been proposed [5] [6] [7]. All of the cited methods use shape properties of the colonic wall while trying to identify polyps. Polyps are viewed as spherical or semi-spherical structures protruding inwards into the colonic lumen. The mathematical description of polyps differs from author to author. While some choose curvature as their shape descriptor, others use information given by the local normals to differentiate between polyps, colonic wall and haustral folds (the main structures contained in the colon).

The remaining of the paper is organized as follows: first our method for CAD is explained, followed by results on 50 colonoscopic data-sets. The advantages and pitfalls of the method are presented in the discussion section. Finally, some conclusions on the future of CAD are given.

2 Method

Since it is well known that over 90% of the colonic surface represents normal wall, this has to be identified and disregarded as quickly as possible. That is why most of CAD algorithms contain two steps a generation step (colonic wall is identified) and a filtering step (final polyp candidates are selected from the remaining structures). The novelty of our method is twofold, first it introduces a fast method for wall identification and second, it models polyps using spin images. The following sections will present the method in detail. A brief overview of our segmentation is given, followed by the colonic wall identification process. Finally the filtering step and the algorithm for building polypoid models are detailed.

Segmentation. During segmentation the colonic wall is determined. A classic region growing algorithm [8], with multiple seed points (to overcome collapsed regions) is used. The threshold value is determined automatically, using the cumulative Laplacian histogram of the volume [9]. A detailed description of the segmentation process can be found in [7]. Figure 3 (left) illustrates the normal and the Laplacian histograms for abdominal images. As depicted in the figure

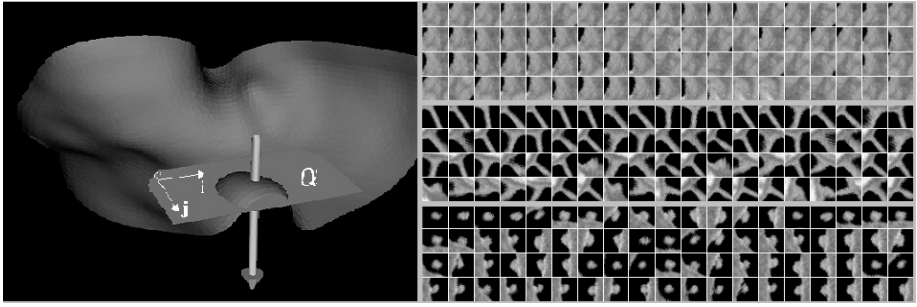


Fig. 1. Part of the colonic surface intersected with the rectangle α . On the right the resulting intersection patterns for colonic wall, haustral folds and colonic polyps.

the optimal threshold value for the transition between colonic air and tissue can be easily identified using the cumulative Laplacian histogram.

The final result of the segmentation algorithm is a set of disjunctive regions representing voxels on the colonic wall. It has to be mentioned that some voxels represent in fact colonic fluid.

Normal colonic wall identification. When detecting voxels belonging to the colonic wall we take advantage of its concavity. The principle of our method is presented in figure 1 (left). For each voxel, the colonic wall is intersected with the rectangle α , perpendicular to the local surface normal and situated at a distance d inside the surface of the wall. The resulting patterns of intersection can be seen in figure 1 (right). As highlighted in the figure the colonic wall gives completely filled planes while polyps and folds give a smaller number of voxels in the rectangle.

Thus, to identify colonic wall a thresholding method based on the number of voxels in the reformatted rectangle (in fact in a squared region of n pixels) can be used. The size of the rectangle is taken as $n = 12mm$ and the distance from the surface is $d = 1.5mm$. A voxel is classified as normal colonic wall if the number of pixels, N_{plane} situated in the rectangle α and having a CT value higher than the segmentation threshold (expressed as a percentage of n^2) is higher than $T_{plane} = 70\%$. The relation between n , d and the radius of detectable polyps (r_p) is given by equation 1, while the number of the voxels N_{plane} , in the worst-case scenario, is given by equation 2. With the selected parameter values the theoretical range of detectable polyp diameters ($r_p * 2$) is from 5 mm up to 18.16 mm.

$$r_p = \frac{(\frac{n-2}{2})^2 + d^2}{2d} \quad (1) \qquad N_{plane} = \frac{\pi d(2r_p - d)}{2} + \frac{n^2}{2} \quad (2)$$

Spin images. The voxels remaining after colonic wall identification are modeled using spin images. Spin images were introduced by A.E. Johnson et al. [10] as a way of matching 3D surfaces. They are a local description of the global shape

of the object, invariant to rigid transformations. Given a basis (P, \vec{G}_1) , P a 3D point and \vec{G}_1 the image gradient in P , a spin map $S_{(P, \vec{G}_1)}$ that projects 3D points T to 2D coordinates (α, β) can be defined as follows, considering that \bullet is the scalar product:

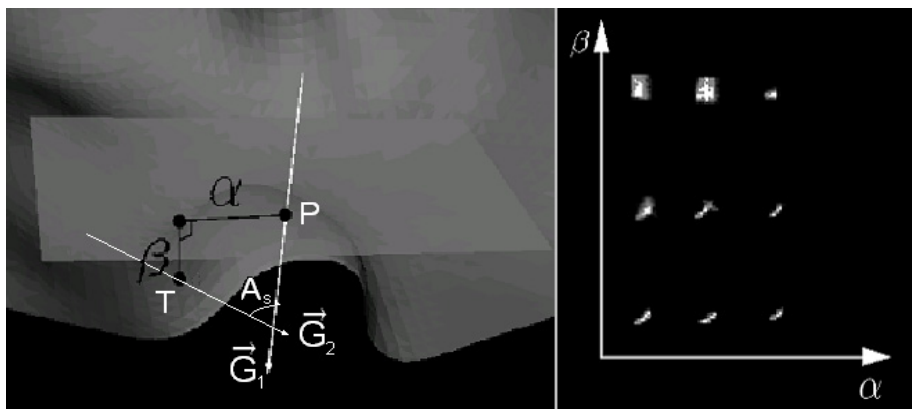


Fig. 2. On the left the process of spin image generation is depicted, on the right some example spin images are given. The first row corresponds to haustal folds, the second to real colonic polyps, the last row represents the spin images given by perfect spheres of diameters 10, 8, and 6 mm respectively

$$S_{(P, \vec{G}_1)} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$$

$$S_{(P, \vec{G}_1)}(T) \rightarrow (\alpha, \beta) = (\sqrt{\|T - P\|^2 - (\vec{G}_1 \bullet \overrightarrow{T - P})^2}, \vec{G}_1 \bullet \overrightarrow{T - P}) \quad (3)$$

Supposing that \vec{G}_2 is the image gradient in T , a support angle is introduced as:

$$A_s = \text{acos}(\vec{G}_1 \bullet \vec{G}_2)$$

Given a basis (P, \vec{G}_1) a spin image I , with N bins is generated as follows:

foreach (T, \vec{G}_2) , $T \neq P$, $\text{distance}(P, T) < N$ and $A_s < T_{\text{support}}$:
 compute (α, β) as indicated by (3)
 $I(\alpha, \beta) = I(\alpha, \beta) + 1$

The distance between two spin images $I_1(\alpha, \beta)$, $I_2(\alpha, \beta)$, each having N bins, is computed using the confidence C of the normalized correlation coefficient $R(I_1, I_2)$. Their analytical formulas are:

$$R(I_1, I_2) = \frac{N^2 \sum_{\alpha, \beta=1}^N I_1(\alpha, \beta) \cdot I_2(\alpha, \beta) - \sum_{\alpha, \beta=1}^N I_1(\alpha, \beta) \cdot \sum_{\alpha, \beta=1}^N I_2(\alpha, \beta)}{\sqrt{\left(N^2 \sum_{\alpha, \beta=1}^N I_1(\alpha, \beta)^2 - \left(\sum_{\alpha, \beta=1}^N I_1(\alpha, \beta) \right)^2 \right) \left(N^2 \sum_{\alpha, \beta=1}^N I_2(\alpha, \beta)^2 - \left(\sum_{\alpha, \beta=1}^N I_2(\alpha, \beta) \right)^2 \right)}}$$

$$C(I_1, I_2) = (atanh(R(I_1, I_2)))^2 - \lambda \frac{1}{N-3}; \lambda = constant; \sum_{\alpha, \beta=1}^N \equiv \sum_{\alpha=1}^N \sum_{\beta=1}^N$$

Polyp labeling using spin images. For each remaining voxel (after the normal colonic wall was identified) its spin image is generated. The spin image has a number of bins $N = 12$, equal to the size of the reformatted rectangle used during the generation step. To limit the negative effects of surrounding wall the support angle of the spin image is limited to values below $T_{support} = 90^\circ$.

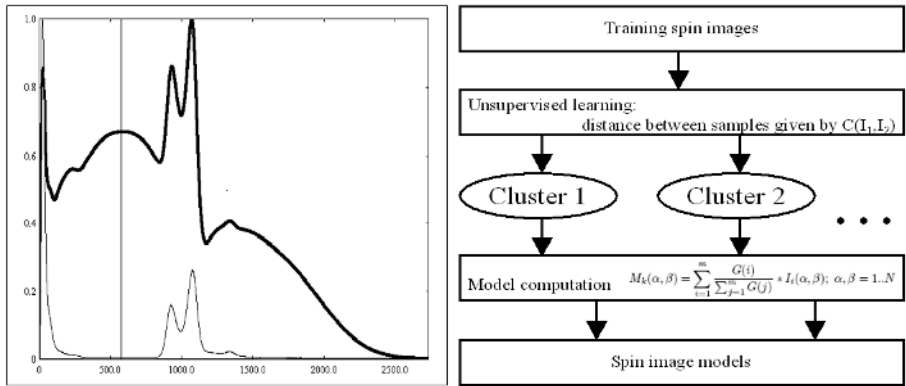


Fig. 3. The cumulative Laplacian histogram for abdominal images. The vertical line corresponds to the threshold used during region growing. On the right, the process of generating spin image models starting from a set of unlabeled samples.

Once a spin image $I(\alpha, \beta)$ is generated it is compared to a set of learned model spin images $M_k(\alpha, \beta), k = 1 \dots N_m, N_m$ the number of models. The models are representative spin images for polyps, haustral folds and remaining wall structures; their generation is described in chapter 3. The similarity measure between a spin image and a model is the confidence $C(I, M_k)$. The current voxel is labeled with the class of the closest model.

Finally, clusters of polyp labeled positions are identified using a connected component extraction algorithm. The connectedness is not limited to first order neighbors but to all neighbors situated at a distance $d_{neighbor} < T_{neighbor}$, where $T_{neighbor} = 2$ is a defined constant. The mean position of the elements belonging to large clusters (having a number of components higher than $T_{component} = 40$) is returned as a polyp candidate. The corresponding axial and 3D volume rendered positions are presented to the reading radiologist.

3 Polyp Models

In this section the process of obtaining model spin images is described, figure 3 (right). The model-building algorithm receives as input all the locations present in the training data, which were not labeled as normal colonic wall. It is not feasible to label each position (approximately 7000/case) as polyp, fold, colonic fluid or wall and thus a supervised classifier is not a valid choice. Instead an unsupervised learning algorithm is employed to cluster the input spin images and the resulting clusters are labeled. The method of Hutchinson [11] is the current classifier. The method can be seen as a statistical clustering since it is based on numerical similarity (given by the confidence C) within object descriptors (spin images $I(\alpha, \beta)$). The classifier provides as output a set of clusters, containing similar spin images.

Inside a cluster, the most representative spin image is considered to be the one for which the average correlation coefficient C , towards all the remaining elements of the cluster is the smallest. This spin image will have the highest contribution to the cluster's model. The weight of an element is computed in a gaussian manner as given by equation 4. Finally, the model for a given cluster is built using equation 5. Let's consider m the number of samples in the current cluster and N the number of bins in the spin image. The following notations can be introduced:

$$\begin{aligned} avgDist(i) &= \frac{1}{m} \sum_{j=1}^m C(I_i, I_j); \quad minDist = \min(avgDist(i)), \quad i = 1..m \\ \sigma^2 &= \frac{1}{m} \sum_{i=1}^m (avgDist(i) - minDist)^2 \end{aligned}$$

The Gaussian weights $G(i)$ for each sample and the model M_k representing the current cluster are computed as:

$$G(i) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{(avgDist(i) - minDist)^2}{2\sigma^2}}; \quad i = 1..m \quad (4)$$

$$M_k(\alpha, \beta) = \sum_{i=1}^m \frac{G(i)}{\sum_{j=1}^m G(j)} * I_i(\alpha, \beta); \quad \text{for each } \alpha, \beta = 1..N \quad (5)$$

Due to computational limitations the previous scheme was modified to have a hierarchical behavior, as described in [12].

4 Patient Data and Results

Fifty data-sets belonging to 26 patients, 25 normal and 25 with 40 polyps of various sizes (Table 1) were considered as input for our CAD scheme. Seven polyps were submerged under residual fluid and were not considered as false negative cases, since our method was not designed to detect them. All patients underwent CT colonography prior to conventional colonoscopy. Informed consent was

obtained from all patients. The patient preparation consisted in the oral administration of 3 to 5 liter of precolon, an in-house developed tagging agent. In some cases the use of polyethylene glycol electrolyte solution was preferred. Immediately before CT colonography a bowel relaxant was injected intravenously. CO₂ was insufflated using a bag system.

Table 1. Polyp distribution and detection results. (No = total number of polyps, TP = true positives)

Type	Submerged	Detectable	TP	Sensitivity
Flat	1	5	3	60.00%
< 5 mm	2	6	2	33.33%
6-9 mm	3	7	6	85.70%
> 9 mm	1	11	10	90.90%
Tumor	0	4	4	100.00%
Total	7	33	25	75.75%

CT colonography was performed on a multi-detector CT (Multi Slice Helical CT; Volume Zoom, Siemens, Erlangen, Germany) using 4x1 mm detector configuration, 7 mm table feed per 0.5 s tube rotation, 0.8 mm reconstruction increment as well as 60 effective mAs and 120 keV. Patients were scanned in both supine and prone positions, in breath holds of 20 to 30 seconds.

The CAD process was carried out as follows: after the CT data was transferred to an offline workstation (Intel Pentium 2.4 GHz system), seed points were selected manually to ensure complete colonic segmentation. For the model building step a leave one patient out technique was employed. The reason for it was that the same polyp in prone and supine positions could be considered similar. Using the remaining data-sets as training data models were generated. These models are then used when running the CAD as described in section 2. This process is repeated for each of the 26 patients.

Using conventional colonoscopy, as standard of reference, true positive (TP) and false positive (FP) findings were determined. The detection rate differentiated on polyp size is presented in Table 1. The average computation time for the whole CAD process (excepting seed point selection) as well as for different steps is shown in Table 2 (left). The total number of false positives was 137, which gives us a mean value of 2.74 false positive findings per data-set. The main causes for false positives are presented in Table 2 (right)

5 Discussion

In this paper a fast yet robust method for CAD in CT colonography is presented. Using spin images the complexity of the colonic wall is reduced to a 2D space, while preserving 3D clues. The method gives better classification results (especially on flat lesions) than our earlier experiments [12] involving slope density functions (SDF). This we believe is due to the fact that spin images contain

Table 2. Average times for completing the CAD algorithm (left). Main causes of false positives, both number and percent (right).

Step	Average time	Cause	Number	Percent
Segmentation	0:11	Haustral fold	75	54.74 %
CAD	5:46	Colonic wall	34	24.82 %
Polyp extraction	0:07	Stool or fluid	13	9.49 %
		Insufflation tube	9	6.57 %
		Ileocecal valve	6	4.38 %
Overall	6:04		137	

information coming from different depths, while SDF's were built at a fixed depth. However, the improvements come at the expense of some computation time, although 6 minutes/data-set is still a clinically acceptable time.

Intrinsic improvements due to the use of newer scanners (e.g. Somatom Sensation 16) are expected. Other advantages of the new scanner relate to the possibility of quicker scan times thus reducing motion artifacts and to low-dose scanning, with or without the use of edge preserving filters. Initial experiments with the new scanner are underway.

Looking at the drawbacks of the method one can observe that polyps are only geometrically modeled. The reason is patient preparation and more explicitly fluid tagging. Some of our polyps are close to tagged colonic fluid or even semi-submerged and thus missed by schemes taking into account intensity values.

A possible debate point is the choice of the unsupervised learning algorithm for clustering the training space. However, we feel that due to the large number of training points (on average 7000/polyp) it is not feasible to label each position a priori, that is why we prefer to label the generated models. The effects of different support angles on the outcome of our method are currently evaluated.

Compared with other CAD methods, this method is faster due to the colonic wall identification step. Opposed to other methods [5],[6] our technique employs a simple method for identifying the wall and then applies more elaborated (computationally costly) modeling. The presented results are comparable with the ones present in the literature and they can be considered very good when looking at clinically relevant polyps (polyps of 6 mm or larger).

To conclude, if our early results are proven on a larger data-base, the method can become clinically useful and can be considered as a "second reader", helping to improve the accuracy and efficiency of the reading process.

Acknowledgement. This work is part of the GOA/99/05 project: "Variability in Human Shape and Speech", financed by the Research Fund, K.U. Leuven, BELGIUM. We would like to thank Maarten Thomeer for his involvement in the project and Dirk Vandermeulen for his idea of using spin images.

References

1. Tumori Aparato Gastroenterico, Dati per la pianificazione dell'assistenza Edited by CPO Piemonte, May 1998
2. Morson, B.C.: Factors Influencing the Prognosis of Early Cancer in the Rectum. Proc R Soc Med (1966) 59:607-8
3. Colorectal cancer - Oncology Channel. [http:// www.oncologychannel.com /colon-cancer/](http://www.oncologychannel.com/colon-cancer/)
4. Vining, D.J., et al.: Virtual colonoscopy. Radiology (1994) 193:446 (abstract)
5. Summers, R.M., et al.: An Automated Polyp Detector for CT Colonography - Feasibility Study. Radiology (2000) 284-290
6. Yoshida, H., Nappi, J.: 3-D Computer-Aided Diagnosis Scheme for Detection of Colonic Polyps. IEEE Transactions on Medical Imaging (2001) 1261-1274
7. Kiss, G., et al.: Computer Aided Detection of Colonic Polyps via Geometric Features Classification. Proceedings 7th International Workshop on Vision, Modeling, and Visualization (2002) 27-34
8. Ballard, D.M., Brown, C.M.: Computer Vision. Prentice Hall (1982) 123-166
9. Wiemker R., Pekar, V.: Fast Computation of Isosurface Contour Spectra for Volume Visualization. Proceedings Computer Assisted Radiology and Surgery CARS (2001)
10. Andrew J.: Spin-Images: A Representation for 3-D Surface Matching. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, Aug. 1997.
11. Unsupervised learning [http://www.cs.mdx.ac.uk /staffpages /serengul /ML /un-supervised.html](http://www.cs.mdx.ac.uk/staffpages/serengul/ML/un-supervised.html)
12. Kiss, G., et al.: Computer Aided Detection in CT Colonography. MICCAI (2003) 746-753