

Computer Aided Detection in CT Colonography via Geometric Modeling

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

Katholieke Universiteit Leuven - Center for Processing Speech and Images



Kasteelpark Arenberg 10
B-3001 Heverlee, Belgium

Telephone: +32-(0)16-32.17.13 Fax: +32-(0)16-32.17.23
<http://www.esat.kuleuven.ac.be/psi>



Computer Aided Detection in CT
Colonography
via Geometric Modeling

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, B-3000, Leuven, Belgium
Email: `Gabriel.Kiss@uz.kuleuven.ac.be`

March 3, 2003

Abstract

The paper presents a technique for Computer Aided Detection in Virtual Colonography based on model fitting. Our approach tries to model polyps using their shape. Locations which have a high probability of being colonic polyps are labeled and presented in a user-friendly way to the reading radiologist. The method was tested on a study group of 50 data sets. Using normal colonoscopy as standard of reference, true positive and false positive findings were determined. The detection rate for polyps larger than 6mm was above 85%. The main advantage of the method is its speed, while preserving a high detection rate for polyps. Initial results show that Computer Aided Detection is feasible and that our method holds potential for screening purposes.

1 Introduction

Colorectal cancer is amongst the leading causes of cancer related death in the industrialized world, with a 4-6% lifetime risk in the general population [1]. Its precursors are the colorectal neoplasms, which are benign, uncontrolled growths of tissue. Neoplasms can be classified into polypoid and non-polypoid types. There is some reference in literature [2] that adenomatous polyps are the primary cause of cancer, 95% of the cases. The nonpolypoid neoplasms [3] are mainly classified into three groups: small flat adenomas, laterally spreading and depressed. While the first two types are thought to develop into protruded polyps, depressed lesions grow endophytically and become advanced cancers.

Fortunately colorectal cancer is characterized by slow growth (5-10 years), likely evolution including thus the appearance of adenomatous polyps. The malignancy of these structures increases with increasing size. As a result early detection and treatment of colonic neoplasms can prevent colonic cancer. Studies [4] show that the survival rate after five years is 92% when early treatment is received.

Many detection methods are available, including fecal occult blood testing (FOBT), barium enema examinations, sigmoidoscopy, colonoscopy, virtual colonography and lately genetic testing. When thinking of screening some criteria should be fulfilled by a method to be considered as a valid candidate: cost-effectiveness, high sensitivity/specificity and patient compliance (to ensure patient cooperation in screening studies).

From the above mentioned methods only FOBT was used in screening. Although safe and inexpensive, it has a low sensitivity and thus effectiveness. Colonoscopy on the other hand has the highest accuracy and it is widely considered as a gold standard, but it is invasive and costly. Lately virtual colonography has received increasing attention as a possible alternative for screening. Introduced in 1994 by Vining et. al. [5], it is a method for exploring the colonic area hinging on CT data. Using advanced graphical software and/or hardware the acquired data is visualized. After the introduction of virtual colonography, CT was the preferred imaging technique due to its high spatial resolution. Today, after inherent technical innovations MR emerges as a viable alternative for data acquisition thus eliminating the presence of ionizing radiation and offering diagnostically valuable images.

Although many visualization techniques are available, including multi planar reformatted images, virtual double contrast imaging, sliding thin slabs, virtual colonoscopy, 3D unfolded cube, flattened/unfolded colon, radiologists prefer the axial images as primary support for diagnosis. The other techniques are mostly seen as problem solving tools.

With the appearance of the 16 slice multi-slice CT machines the number of acquired slices is well beyond 1000 per patient, especially if two scans per patient (prone and supine) are acquired. It is not difficult to see that perceptual errors due to human errors can affect the performance of CT colonography. Additional pitfalls include reader inexperience (due to the learning curve involved in the process), satisfaction of search [6] (failure to detect multiple polyps in the same patient), lack of original color clues to guide the detection and the different performances of the visualization techniques. To overcome these problems, automated methods for polyp detection were developed. Computer Aided Detection (CAD) is one possible approach to improve reading efficiency and accuracy. It consists in automatic detection of conspicuous masses that resemble polyps. Due to low sensitivity values of CAD its results are considered as a “second opinion” and the human reader makes the final diagnosis. Apart from reducing interpretation time, a positive side effect of CAD is the increase of reader confidence.

The remaining of the paper is organized as follows: first our method for CAD is presented followed by results on 50 colonoscopic data-sets. The advantages and pitfalls of the method compared to other known techniques are presented in the discussion section.

2 Method

In this paper we present a new approach for polyp detection in CT colonography. Although it is a new research field several approaches have been already proposed [7] [8] [9] [10]. The problem remains the same for the computer as for the radiologist, who is trying to find the needle (polyp location) in the haystack (colonic wall), since it is well known that over 90% of the colonic wall represents normal structures. What our method tries to achieve is to eliminate as much of the colonic wall as possible, as quickly as possible and without losing interesting locations (polyps). It does that by simple geometric analysis and then applies a more elaborated algorithm on the remaining structures to come up with the final candidates. Our primary goal is to detect protruded, polypoid type neoplasms larger than 5 mm.

Most of the CAD algorithms contain two steps: a generation step (polyp candidates are generated) and a testing step (final polyp candidates are selected based on the previously determined set). In fact the testing step tries to detect and eliminate the false positive findings generated by the first step. Our method respects the same pattern.

The novelty of the method is twofold: first it introduces a fast method for wall elimination and second it tries to model polyps based on their shape. The

classification step of our method is therefore a comparison of the presented candidate with a database of polypoid models learned from examples. The top-level architecture of our method is presented in figure 1, and elaborated in the following subsections.

2.1 Generation

The generation step has the following steps: segmentation and colonic wall elimination.

2.1.1 Segmentation

In this step the colonic wall is determined. Since CTC images have a large contrast between (insufflated) colonic air and colonic wall, classic region growing algorithms [12] can be used successfully. CTC preparation may influence the segmentation technique. In our case wet preparation was used and thus a substantial amount of fluid remained in the colon. Also in some patients collapsed colonic regions (due to inadequate distension) were present. Therefore multiple seed points were used to initialize the region-growing algorithm.

Previously [11] the threshold value for the region-growing algorithm was established interactively. Now we use the method of Wiemker [13], to determine it automatically. This method computes the cumulative Laplacian histogram of the image volume and then assumes that important changes occur around its local maxima. Due to the large volumes, we are dealing with, a computation of the cumulative Laplacian histogram on the whole volume is not feasible that is why it was computed on slices situated around the seed points. Also an initial guess T_{guess} for $T_{segment}$ (based on 18 sample data sets) was computed. The final segmentation threshold $T_{segment}$ is considered to be the local maxima around T_{guess} . The search interval is considered as $[T_{guess} - \delta * 2, T_{guess} + \delta * 2]$, where δ is the standard deviation of the $T_{segment}$ values on the sample data sets.

Given this $T_{segment}$, for all the seed voxels a region A of connected air voxels (CT value smaller than $T_{segment}$) is grown outwards in a breadth first manner. The colonic wall is defined as the set W of voxels adjacent to A . For each voxel of A 6 direct neighbors are taken into consideration.

The final result of the segmentation algorithm is thus the set W of disjunctive regions representing voxels on the colonic wall. It has to be mentioned that some voxels in W represent in fact colonic fluid and not actual colonic wall.

2.1.2 Colonic wall elimination

At first sight the colonic wall can be divided into two major regions: convex and concave regions. By nature polyps are spherical thus convex, while the colonic wall is mostly concave. The only exceptions to this rule are the haustral folds, which have a convex but cylindrical appearance.

To estimate curvature properties one can use methods based on differential geometry, such as principal, mean and Gaussian curvatures or can derive the local shape of the wall by looking at the intersection pattern of the surface normals situated close to a region of interest. However we employ a simple geometrical approach to estimate the local curvature.

The principle of our method is presented in figure 2. For each voxel the colonic wall is intersected with the plane P perpendicular to the local surface normal and situated at a distance d from the surface of the wall. The resulting patterns of intersection can be seen in figure 3. As highlighted in the figure the colonic wall gives completely filled planes while polyps and folds give a smaller number of voxels in the reformatted plane. Thus to eliminate the colonic wall we propose a thresholding method based on the number of voxels in the reformatted plane (in fact in a squared region of n pixels). The required number of pixels in the reformatted plane (expressed as a percentage of n^2) and the distance d were determined empirically in order to ensure maximal response on polyps of 5 mm or larger.

The process mentioned above is described analytically as follows (see figure 2 (right) for details). Given $P_0 = [x_0, y_0, z_0]^T$ as the current point on the surface of the colon, compute $P_1 = [x_1, y_1, z_1]^T$ as $P_1 = P_0 + d.G$, where $G = [g_x, g_y, g_z]$ is the local gradient in P_0 . To compute the local gradient a Zucker-Hummel operator [12] is used and the voxel size is taken into account. The obtained gradient is perpendicular to the colonic surface and oriented from air towards soft tissue. Next we define a local coordinate system centered in P_1 , in order to be able to sample voxels around that point. The coordinate system is situated in the plane α and fully determined by the points P_1 , P_2 and P_3 . The plane α is computed using the vector $\vec{P_0P_1}$ as its normal and considering that $P_1 \in \alpha$. The point P_2 is determined as the intersection of the plane α with one of the axes of the global coordinate system Oxyz. The vector $\vec{P_1P_2}$ is one of the axes of the local coordinate system. The next one is given by the vector $\vec{P_1P_3}$, where P_3 is a point along the normal in P_1 to the plane ($P_0P_1P_2$). The vectors $\vec{P_1P_2}$ and $\vec{P_1P_3}$ are normalized and the local coordinate system $O'ij$ with $O' \equiv P_1$ is obtained. The x, y, z coordinates of a point having the local coordinates u, v are obtained as follows:

$$O\vec{V} = O\vec{u} + O\vec{v} - OO'$$

2.2 Testing

After the concave colonic wall was eliminated the remaining regions are considered for further investigation. For that a shape based classification is used. The actual statistical model building is described in the next section, here it will be shown how polyps are modeled and how a set of existing models are used for discriminating between normal and polypoid structures.

2.2.1 Slope density functions

We are modeling polyps based on their slope density function (SDF) [12]. Given a basis curve $\psi - s$ the SDF is defined as the histogram or frequency distribution of ψ collected over the boundary. The boundary is obtained by region growing in the reformatted plane. The SDF can be seen as the histogram of gradient orientations and while it is smooth for circular and elliptic objects, it shows peaks for elongated structures. That is why it can be used for discriminating between polyps and haustral folds.

At this point one has to remember that no interpolation was done when extracting the reformatted planes, the closest image voxel to the computed 3D point was always chosen, the reason being speed. This however can lead to irregularities of the extracted boundary, thus influencing the gradient orientations when computed with classic filters. To compensate for this, the gradients are computed using the recursive filtering method of R. Deriche [14]. The method uses the notion of separable filters and allows the estimation of smoothed partial derivatives and has the additional advantage of being computationally inexpensive.

Using this method an additional deficiency present in the reformatted planes is eliminated. As visible in figure 4 apart from the cut through the polyp or fold the plane contains information related to the colonic wall as well. The colonic wall has a disturbing effect on the SDF's shape, mostly smoothing thus reducing the difference between folds and polyps. This however can be eliminated computing the curvature of each point along the boundary and eliminating voxels with negative curvature values.

The curvature is computed using the formula presented in [15]:

$$k = \frac{-t^T H t}{\|\vec{g}\|}$$

where t is the unit tangent vector to the boundary and H the Hessian of the image in $I(u, v)$. Their formulas are:

$$H = \begin{bmatrix} I_{uu} & I_{uv} \\ I_{uv} & I_{vv} \end{bmatrix} \text{ and } t = - \begin{bmatrix} I_v \\ I_u \end{bmatrix}$$

The advantage of using recursive filtering is that it offers the majority of the elements needed for the Hessian computation, and offers for free the tangent orientation. After these steps a correct estimation of the slope density function is obtained and can be used for modeling and classification respectively.

2.2.2 Sample modeling

In the training stage a set of SDF samples are presented to the classification algorithm which will try to find similarities between them and group them into different clusters. From each cluster of samples a representative model is built, thus we will have model SDF's for polyps, haustral folds and remaining colonic wall. In the testing step the distances between the SDF of the current point and the SDF's of the models are computed and the type of the current point is assigned to the type of the closest SDF.

The distance between SDF's is the χ^2 statistic defined as:

$$D_{ij} = \frac{1}{2} \sum_{k=1}^K \frac{(h_i(k) - h_j(k))^2}{h_i(k) + h_j(k)} \quad (1)$$

where $h_i(k), h_j(k)$ denote the k -th element in the SDF histogram of the sample and the model respectively. Although SDF's ensure invariance under scaling and translation they do not ensure invariance with regards to rotation. That is why the distance between two samples shown in figure 5 (according to the formula 1) is large. In order to cope with this situation the χ^2 statistic is minimized over a range of possible orientations. Equation 1 is rewritten as follows:

$$D_{ij} = \min \left(\sum_{k=1}^K \frac{[h_i((k+l)\%K) - h_j(k)]^2}{h_i((k+l)\%K) + h_j(k)} \right), \quad l = 1..K \quad (2)$$

the result being that D_{ij} is minimized over the discretized orientations.

2.2.3 Polyp extraction

After the previous step some positions on the colonic surface are labeled as polyps. Of course multiple responses for the same polyp are obtained and on the other hand some responses are generated by noise (false positives). To eliminate these inconveniences a connected component extraction of final polyp candidates is employed.

In this step clusters of polyp labeled positions are identified. The connectedness is not limited to the first order neighbors but to all neighbors situated at a distance $d_{neighbor} < T_{neighbor}$, where $T_{neighbor}$ is a predefined constant.

Finally the mean position of the elements present in a cluster, having a number of components higher than $T_{component}$ is returned as a polyp candidate. The corresponding axial and 3D volume rendered positions are presented to the reading radiologist as shown in figure 6.

3 Polyp models

In this section the process of obtaining an SDF model is described. The algorithm receives as input a set of training samples and gives as output a set of SDF models. The user assigns a label to the output models, instead of assigning a label to each training sample individually. The used classifier is based on an unsupervised learning algorithm, and it is preferable in this case due to the huge amounts of training data. Starting from the training data, clusters containing elements of similar training samples are generated. Each of these clusters will generate a particular model used in the testing step of the algorithm.

An additional problem to be solved is the computational complexity of the classifier. To handle this the classifier is modified to have a hierarchical behavior. The next subsections will present first the way the models are built given a cluster of training samples, and second the changes required by the classifier in order to have a hierarchical behavior.

3.1 Model building

To achieve a clustering of the training SDF samples a technique that produces a partition of the input according to a given distance is needed. The unsupervised learning algorithm proposed by Hutchinson in 1994 [16] can be considered a viable choice. It can be seen as a statistical clustering since it is based on numerical similarity within the object descriptions, in our case the SDF's. All one has to keep in mind at this moment is that the algorithm provides clusters of similar SDF's given a set of training samples. The pseudo-code of the algorithm and some additional details are presented in the appendix.

After clusters of training SDF's are obtained, representative models have to be built. In our case each of the output clusters will generate a model. Inside a cluster each samples contributes to the final model in a gaussian manner. By that centrally placed samples gain a higher weight than samples situated at the extremities of the cluster.

Before presenting the formula for the gaussian weights and the one used for the actual model computation, let's consider n the number of samples in

the current cluster. The following notations can be introduced:

$$avgDist(i) = \frac{1}{n} \sum_{j=1}^n D_{ij}; \quad minDist = \min(avgDist(i)), \quad i = 1..n$$

$$totalDist = \sum_{i=1}^n avgDist(i); \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^n (avgDist(i) - minDist)^2$$

where $avgDist(i)$ is the average distance from the current sample to the remaining samples in the current cluster, and D_{ij} defined by 2.

The gaussian weight of each sample is given by:

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

The model is computed using the equation:

$$M_c(k) = \sum_{i=1}^n \frac{G(i)}{\sum_{j=1}^n G(j)} * h_i((k + l_i) \% K), \quad k = 1..K \quad (3)$$

where l_i is a rotation coefficient used to minimize $totalDist$ over the current cluster c .

3.2 Hierarchical classifier

On our data-base of 22 significant polyps the mean number of training samples per polyp was 503.36 with a minimum of 219, and a maximum of 976. When applying the classifier to the entire training data the total number of pairs between the training samples can be evaluated as follows:

$$\#pairs = \frac{N \cdot (N - 1)}{2} = \frac{(22 * 503.36)(22 * 503.36 - 1)}{2} \approx 60 * 10^6$$

Having such a large number of components it is not feasible due to both memory and computation limitations to handle this amount of data. Our solution was to split up the training data in groups of sn samples apply the clustering technique to each of them and then consider the resulting models as input to the classification scheme. The previous sequence is repeated until no changes occur. In our experience, if a cluster size of 500 elements is used then a two-level hierarchical scheme is sufficient to obtain a consistent clustering.

If the number of pairs is recomputed for the first level in the hierarchy we have the following number of pairs, considering (for simplicity) that each group has 503.36 elements and the total number of groups is $\#gr = 22$:

$$\#pairs = \#gr. \frac{N.(N-1)}{2} = 22 \frac{503.36(503.36-1)}{2} \approx 22 * 126000 \approx 2.8 * 10^6$$

thus a 23 times reduced pairs number. Also each sorting has to be done on approximately 126.000 pairs instead of the original 60 million pairs.

One problem that arises is that initial clusters and thus generated models have a different number of elements. This has to be accounted for when building models at higher levels. Considering $Ell(i)$ as the number of pairs from which a certain model was attained, the modified model building equation can be rewritten as:

$$M_c(k) = \sum_{i=1}^n \frac{G(i).Ell(i)}{\sum_{j=1}^n G(j). \sum_{j=1}^n Ell(j)} * h_i((k+l_i)\%K), k = 1..K \quad (4)$$

Also the algorithm presented in the appendix was slightly modified, not to allow the merging of models which are situated at a large distance one from the other. If the original algorithm would have been used the obtained result would have been a single cluster.

We are aware of the fact that when using a hierarchical scheme some of the coarse details are lost, however this approach can be seen as a good compromise between accuracy and computational efficiency.

4 Results

Fifty data-sets, 25 normal and 25 with 40 polyps of various sizes (Table 1) 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.

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 breadth holds of 20 to 30 seconds. This resulted in 50 data-sets which were considered as input for our CAD system. On average the size of the acquired data sets was 246.47 MB. The image processing was done on a dual processor, Intel Pentium 2.4 GHz system, having 2GB of RAM.

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

Type	No	Submerged	Detectable	TP	Sensitivity
Flat	6	1	5	1	16.67%
< 5 mm	8	2	6	2	33.33%
6-9 mm	10	3	7	6	85.7%
> 9 mm	12	1	11	10	90.90%
Tumor	4	0	4	4	100%
Total	40	7	33	23	69.69%

As a first step seed points were defined manually and then the whole CAD process was completed automatically. The mean segmentation threshold was 582 (CT number) and had a standard deviation of 31 units. The typical shape of the cumulative Laplacian histogram for abdominal images is shown in Figure 7. The polyp models were generated using the leave one out technique.

Using conventional colonoscopy as standard of reference true positive (TP) and false positive (FP) findings were determined for each patient. The total number of polyps was 40, from these 7 were submerged under residual fluid and were not considered as false negative cases, since our method was not designed to detect them. The detection rate differentiated on polyp size is presented in Table 1. The average computation time for the whole CAD process as well as for different steps is shown in Table 2.

Table 2: Average computation times, expressed in minutes.

Segmentation	0:11
Generation + Testing	4:07
Polyp extraction	0:7
Overall	4:25

The total number of false positives was 124, which gives us a mean value of 2.49 false positive findings per data-set. The main causes for false positives are presented in Table 3.

Table 3: False positives causes.

Cause	Procent
Haustral fold	60.68 %
Colonic wall	18.80 %
Stool or fluid	11.97 %
Insufflation tube	5.13 %
Ileocecal valve	3.42 %

5 Discussion

In this paper a fast method for CAD in CT colonography was presented. The fastness of the method comes from reducing the shape based analysis from a 3D space into a 2D space while preserving 3D clues. Its main purpose is the detection of polypoid lesions larger than 5mm. On the relatively small data-base of patients it was proven that it has a high accuracy in detecting sessile and pedunculated polyps that protrude inwards into the lumen. It also can detect tumors, but has a low sensitivity for small polyps and for flat lesions.

All our experiments were done using data collected with the Volume Zoom 4 Siemens machine, it is expected that on the new generation scanners (e.g. Somatom Sensation 16) the number of polypoid voxels will increase thus resulting in a better detection rate implicitly. 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. When using a low-dose setup the mAs can be reduced from 60 mAs to 10 mAs, obtaining an effective dose value acceptable in a screening environment. Our method can be applied to this kind of data as well, without further changes. The relation between dose and CAD sensitivity is the subject of on-going work. Also initial testing was carried out on MR data-sets. Although the segmentation step has to be changed due to noise in the data, the CAD algorithm can be used successfully.

Looking at the draw-backs of the method one can observe that the segmentation process is not fully automatic and that the pixel values around or inside the polyp are not taken into account. The reason for both is patient preparation and more explicitly fluid tagging. For the segmentation we preferred a semi-interactive step to ensure that a correct segmentation is achieved, and as much of the small bowel as possible is eliminated. The reason for not using voxel intensity based features is that some of our polyps

are close to tagged colonic fluid or even semi-submerged and thus missed by such schemes.

Improvements of our scheme will include the generation of new polypoid models, learning from mistakes and the extension of the distance presented in equation 2, to include more 3D features. The easiest way to achieve this is the inclusion of the distance to wall of each point laying on the extracted plane. Also experiments with the use of a parabola instead of the cut plane were carried out. Initial results show that no difference between polyps and haustral folds is obtained. The main goal of our developments will be to further reduce the number of false positive findings.

6 Conclusion

The results of our experiments show that CAD in CT colonography is feasible, and a high sensitivity and specificity can be obtained. However to be relevant the tests have to be confirmed on a larger number of cases. Once its efficacy is proven, CAD can be integrated into clinical practice, improving on current accuracy and cost values.

7 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.

8 Appendix

The unsupervised learning algorithm performs bottom-up clustering. The algorithm starts with a few small clusters and enlarges them to create new clusters. It is non-incremental and simple to implement. The third step has the highest computational intensity, since it does a distance based sorting of all the training samples presented as input.

Unsupervised learning algorithm:

1. Start with a list of unordered input samples, SDF's:
[x, y, z, ...]
2. Generate the list of ALL unordered pairs of distinct input samples, and compute the distance between the two members of the pair:

```

    [(x, y, d1), (x, z, d2), (y, z, d3), ....]
3. SORT this list of pairs, into increasing order of
    distance between the pairs.
4. Do classification:
    WHILE the list is not empty take the first pair {x,y}
        from the remaining of the list and

        IF neither x nor y is as yet in a cluster
        THEN
            form a new cluster C. to start with, C={x, y}.

        IF one of the two points (x), is in some cluster C
            AND the other point y is not yet in any cluster
        THEN add y to C.

        IF x and y are in different clusters C1 and C2
            AND x is nearly central in C1
            AND y is nearly central in C2
            AND the distance from x to y is less than
                the average diameter of C1 and C2
        THEN amalgamate C1 and C2 into a new combined cluster

        (Otherwise, do nothing to the clusters)

```

The distance between samples is defined in equation 2. The algorithm requires two additional measures: cluster diameter and nearly central samples. The cluster diameter is defined as the largest distance between any two points in a given cluster, while a sample is considered “nearly” central if the distance from it to every other point in the cluster is less than $\frac{2}{3}$ of the cluster diameter.

References

- [1] G. Stevenson, "Radiology in the Detection and Prevention of Colorectal Cancer", *Eur J Cancer*, 31A:1121-1126, 1995
- [2] B.C. Morson, "Factors Influencing the Prognosis of Early Cancer in the Rectum", *Proc R Soc Med*, 59:607-8, 1966
- [3] S. Kudo et al., "Colonoscopic Diagnosis and Management of Nonpoly-poid Early Colorectal Cancer", *World J Surg*, 24:1081-1090, 2000

- [4] "Colorectal cancer - Oncology Channel", <http://www.oncologychannel.com/coloncancer/>
- [5] D.J. Vining et al., "Virtual colonoscopy", *Radiology*, pp:193:446 (abstract), 1994
- [6] R.N. Strickland, "Image-Processing Techniques for Tumor Detection", *Marcel Dekker Inc.*, pp:47:70, 2002
- [7] R.M.Summers et al., "An Automated Polyp Detector for CT Colonography - Feasibility Study", *Radiology*, 216:pp:284-290, 2000
- [8] S.B. Gokturk et al., "A Statistical 3-D Patter Processing Method for Computer-Aided Detection of Polyps in CT Colonography", *IEEE Transactions on Medical Imaging*, 20:pp:1251-1260, 2001
- [9] H. Yoshida, J. Nappi., "3-D Computer-Aided Diagnosis Scheme for Detection of Colonic Polyps", *IEEE Transactions on Medical Imaging*, 20:pp:1261-1274, 2001
- [10] G. Kiss et al., "Computer Aided Detection of Colonic Polyps via Geometric Features Classification", *Proceedings 7th International Workshop on Vision, Modeling, and Visualization*, pp:27-34, 2002
- [11] G. Kiss et al., "Computer-aided diagnosis in virtual colonography via combination of surface normal and sphere fitting methods", *European Radiology*, vol. 12, no. 1, pp. 77-81, 2002
- [12] D.M. Ballard, C.M. Brown, "Computer Vision", *Prentice Hall*, pp.:123-166, 1982
- [13] R. Wiemker, V. Pekar, "Fast Computation of Isosurface Contour Spectra for Volume Visualization", *Proceedings Computer Assisted Radiology and Surgery CARS 2001*, 2001
- [14] R. Deriche, "Fast Algorithms for Low-Level Vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:1pp.:78-87, 1990
- [15] O. Monga et al., "From Partial Derivative of 3D Density Images to Ridge Lines", *CVPR-Computer Vision and Pattern Recognition*, pp.:354-359, 1992
- [16] "Unsupervised learning" <http://www.cs.mdx.ac.uk/staffpages/serengul/ML/unsupervised.html>

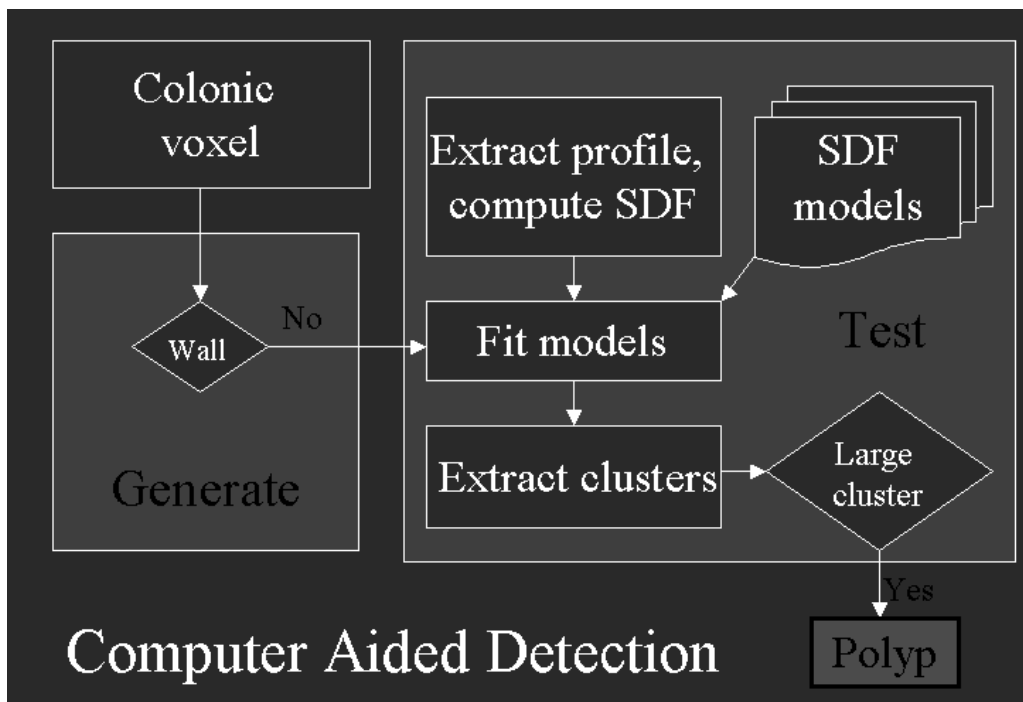


Figure 1: Diagram showing the data flow inside the CAD algorithm, includes the two main steps: generation and testing.

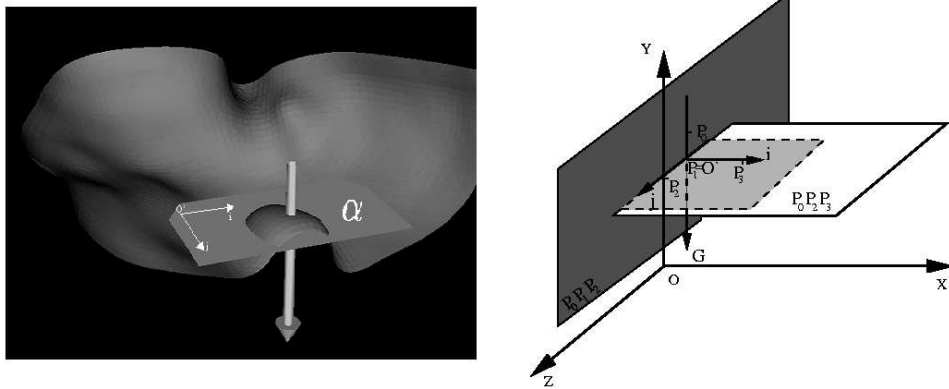


Figure 2: Part of the surface of the colon intersected with the plane alpha, on the right the coordinate system and main vectors used for extracting the voxels of the cut plane.

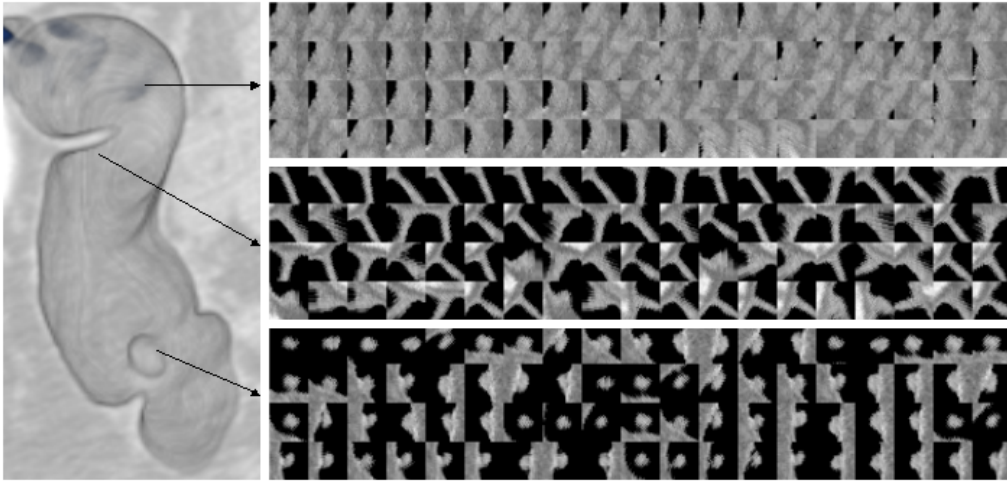


Figure 3: Patterns of intersection on the surface of the colon corresponding to: wall (top), fold (middle) and polyp (bottom)



Figure 4: Smoothed cut through a polyp and a haustral fold respectively. Second row illustrates the boundary of the two structures. Third row shows the eliminated parts of the boundary due to negative curvature values (marked dark grey).

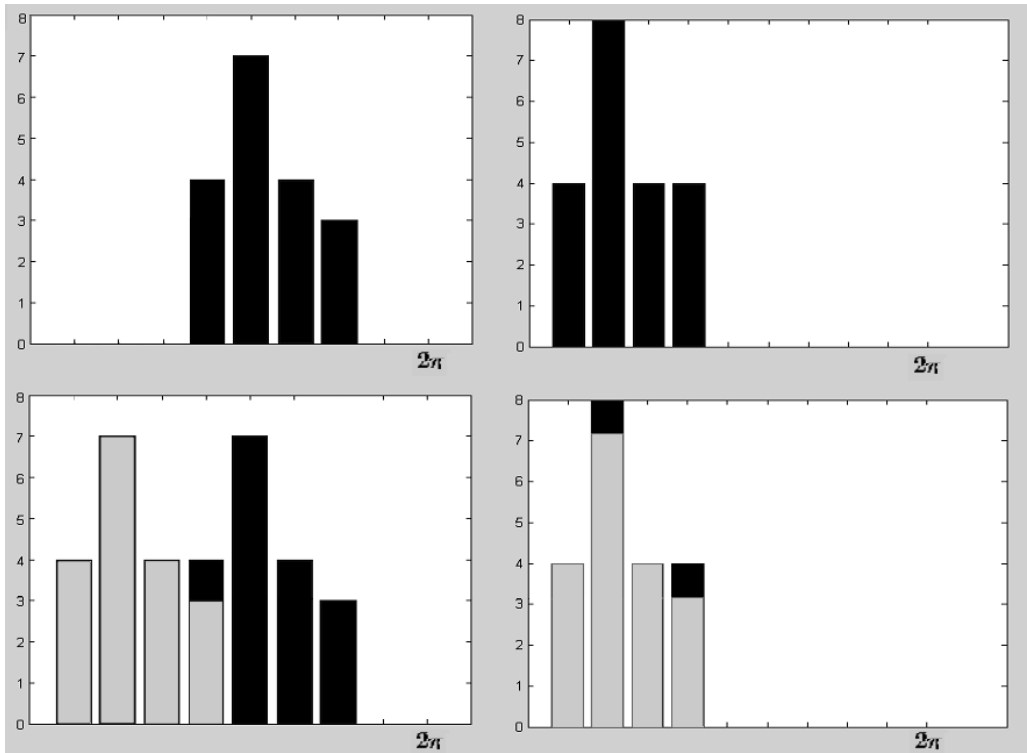


Figure 5: Minimization of the SDF over all possible rotation values. Top row presents two similar SDF functions, while the bottom row shows the rotated SDF (left) and a superposition of the rotated SDF's.

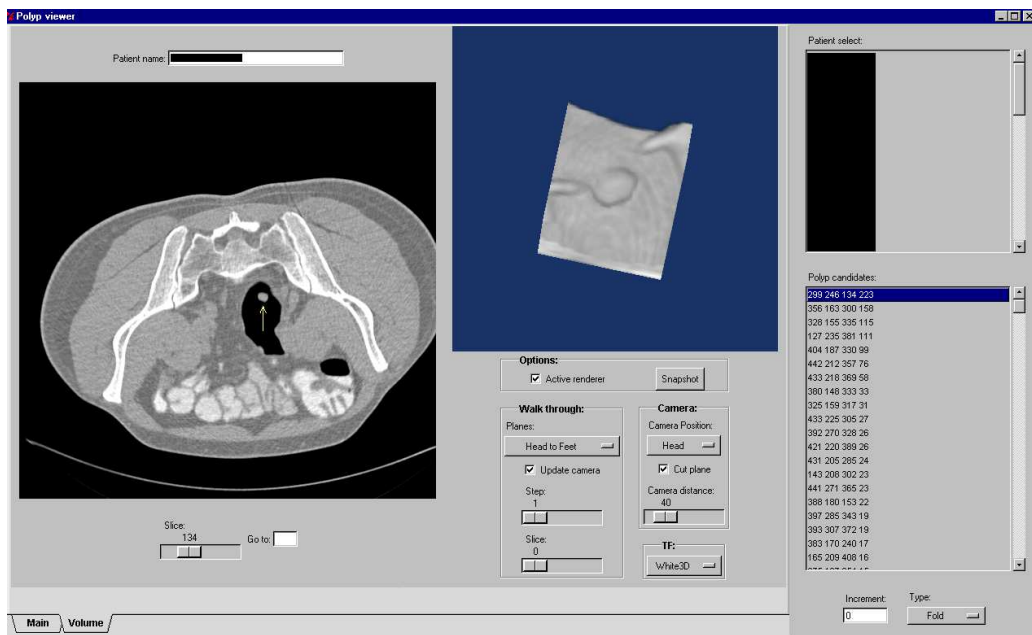


Figure 6: Presentation software, polyp indicated with an arrow on the axial slice, a 3D volume rendering around the structure of interest is present. On the right a list of available patients and suspected polypoid positions.

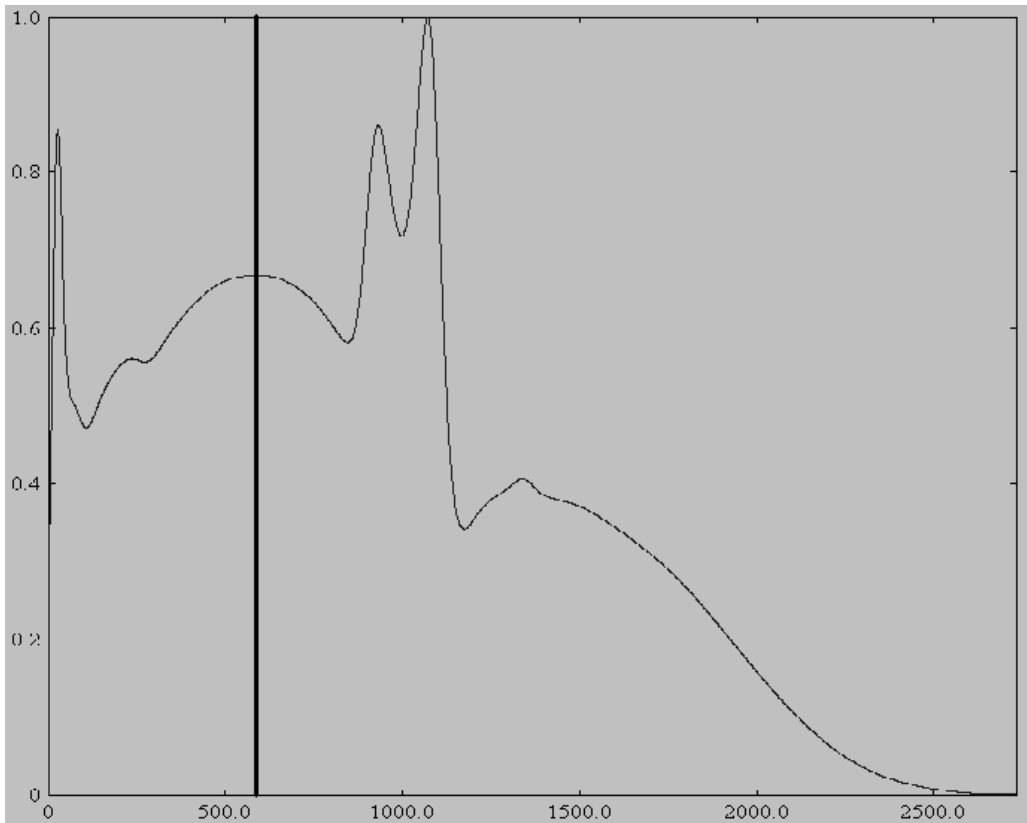


Figure 7: The cumulated Laplacian histogram for an abdominal image. The vertical line indicates the value of the extraction threshold.