

Automated synthesis, insertion and detection of polyps for CT colonography

Nicolas Sezille^{*a}, Robert J. T. Sadleir^a, Paul F. Whelan^a
^aVision Systems Laboratory, Dublin City University

ABSTRACT

CT Colonography (CTC) is a new non-invasive colon imaging technique which has the potential to replace conventional colonoscopy for colorectal cancer screening. A novel system which facilitates automated detection of colorectal polyps at CTC is introduced. As exhaustive testing of such a system using real patient data is not feasible, more complete testing is achieved through synthesis of artificial polyps and insertion into real datasets. The polyp insertion is semi-automatic: candidate points are manually selected using a custom GUI, suitable points are determined automatically from an analysis of the local neighbourhood surrounding each of the candidate points. Local density and orientation information are used to generate polyps based on an elliptical model. Anomalies are identified from the modified dataset by analysing the axial images. Detected anomalies are classified as potential polyps or natural features using 3D morphological techniques. The final results are flagged for review. The system was evaluated using 15 scenarios. The sensitivity of the system was found to be 65% with 34% false positive detections. Automated diagnosis at CTC is possible and thorough testing is facilitated by augmenting real patient data with computer generated polyps. Ultimately, automated diagnosis will enhance standard CTC and increase performance.

Keywords: colonoscopy, CTC, polyp synthesis, polyp detection

1. INTRODUCTION

Colorectal cancer represents 9% of all recorded cancer cases in Ireland, second only to the non-melanoma skin cancer (36%). It is the second largest cause of cancer related death (10%) second to lung cancer (19.5%) (NCRI)⁹. Colorectal cancer can be prevented by detecting and successfully resecting precursor polyps early in their course. Conventional colonoscopy is the most sensitive screening technique that is currently available. This involves a direct examination of the interior of the colon using an instrument known as a colonoscope. Conventional colonoscopy requires complete bowel preparation and colon insufflation. In addition the patient must be placed under sedation for the duration of the procedure. Conventional colonoscopy is extremely invasive and has not gained acceptance among those at risk of developing colorectal cancer. Finally, complications can occur at conventional colonoscopy the most significant being bowel perforation which can, in extreme cases, result in death.

Computed tomographic colonography (CTC), also known as virtual colonoscopy (VC)¹³, is an emerging technique for non-invasive colon imaging. Using this technique, a suitably prepared patient undergoes an abdominal computed tomography (CT) scan. In order to reduce the effects of residual material in the colon the patient is generally scanned using both prone and supine positioning³. 3D virtual reality models of the colon are then generated from the CT data. These models can then be processed to facilitate computer aided diagnosis.

This paper describes a method for the synthesis and insertion of artificial polyps into real patient CTC datasets (Sec. 4). Techniques such as colon wall tracking (Sec. 4.1.1) and colon wall angle calculation (Sec. 4.1.2) are used to derive the polyp insertion parameters. The synthesis and insertion stage of the process is described in Sec. 4.2. An automated polyp detection technique is then described in Sec. 5. This technique is based on a two stage approach: the first stage utilises a modified version of convex hull algorithm to detect colonic anomalies in successive CT slices (Sec. 5.1). The second stage involves classification of the detected anomalies as either polyps or naturally occurring features (Sec. 5.2). The results are presented in Sec. 6 and the paper is concluded in Sec. 7.

* nicolas.sezille@eeng.dcu.ie; phone: +353 1 700 7636; fax: +353 1 700 5508; Vision Systems Laboratory, School of Electronic Engineering, Dublin City University, Dublin 9, Ireland.

2. PREVIOUS WORK

Automated detection is an emerging technique in CTC. Several studies have shown the viability of CT scanning for the detection of colonic polyps. Beaulieu et al.² demonstrate the possibility of detecting polyps larger than 4 mm with sensitivity of 100%. Until recently most of the detection systems achieved high sensitivity with low specificity or low sensitivity with high specificity. More recently new approaches have emerged that are able to achieve both high sensitivity and specificity. Yoshida et al.¹⁶ use a segmentation technique to extract the colon wall and then analyse its internal structure to facilitate the detection of colonic polyps. Göktürk et al.⁵ extract morphological characteristics of potential polyps along randomly orientated orthogonal planes to generate 3D pattern histograms. These histograms are used by a training-based classifier to discriminate between polyps and naturally occurring features. There is a need for artificial synthesis and insertion of polyps in real patient data to test the polyp detection software, Karadi et al.^{6, 7} introduce a method that gives good results however this technique is not automated. The method introduced in this paper generates flat polyps and is fully automated. The user only needs to enter an insertion point and the size parameters, the orientation of the polyp is automatically detected.

3. DATA INTERPRETATION

The datasets used in this study were acquired using a Siemens Somatom 4 spirals CT scanner. Standard image acquisition parameters were used. The CTC datasets represent complete abdominal CT scans which contain a vast amount of information describing the complete anatomy of the abdomen (i.e. bones, lungs, stomach, liver, kidney and small bowel), the CTC dataset also includes room air. A segmentation process is then required prior to any analysis of the colon. Real patient CTC datasets are stored using the DICOM 3.0 standard. The data is stored slice by slice in separate files with all patient and acquisition information stored as header information. In order to reduce the processing time and remove the need to deal directly with DICOM data, the original dataset is transformed in a simple volume and stored in a custom data file. This dataset, which uses the co-ordinate system presented in Fig. 1, forms the basis of the discussion in the remainder of this paper. CTC datasets are 3D volumes which represent the density at each point in the scanned region. The density is measured using Hounsfield units (HU), the typical range of values are -1500 HU for air and 400 HU for calcium (bones).

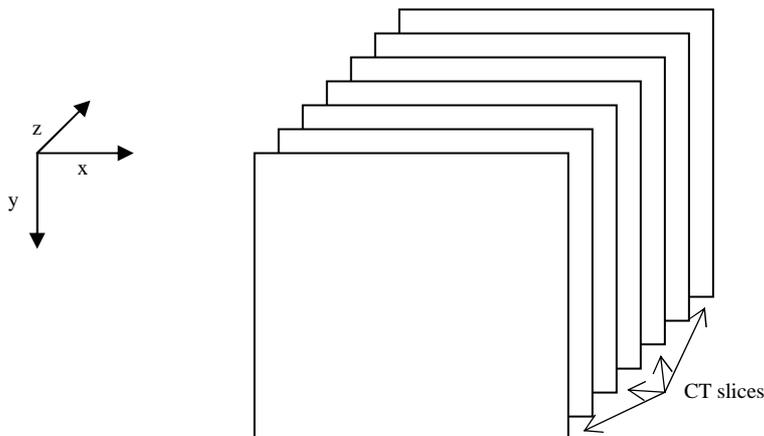


Fig. 1: Co-ordinates convention in the dataset

4. AUTOMATED POLYP INSERTION

A custom software application has been designed to introduce artificial polyps into a CTC dataset. The application GUI is used to navigate through the dataset and facilitates the user guided insertion of artificial polyps. The user must select a candidate point in the volume and enter the desired size parameters. An artificial polyp is then synthesised and inserted into the CTC dataset. The polyps are generated using a fully configurable ellipsoid model (4.1), where x , y , z

are the co-ordinates along the main axes of the ellipsoid and a, b, c its half-dimensions along the x, y and z axes respectively. The polyp insertion point is determined automatically from an examination of the section of colon wall closest to the candidate point.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1. \quad (1)$$

4.1 Insertion parameters calculation

The first step in the detection of the local parameters (i.e. colon wall local angle and density) is to find a suitable insertion point on the colon wall. The point on the colon wall closest to the user selected point is used as the actual insertion point. The segmentation of the colon lumen is achieved using a simple threshold. This is possible as air voxels have a density significantly lower than that of soft tissue. Colon wall voxels are defined as colon voxels with at least one 8-connected air voxel. In order to simplify the polyp insertion process it is required that one of the main directions of the polyp model is normal to the CT slices (i.e. in the z direction). As a result only one angle must be detected. This is the angle of the main direction of the ellipse in the central CT plane of the polyp. To compute an approximation of this angle, the colon wall must be tracked on each side of the central voxel.

4.1.1 Colon wall tracking

The colon wall tracking (CWT) algorithm exploits a property of the colon wall, the local set of colon lumen pixels is defined as the longest run of air pixels in the 8-adjacent neighbourhood of the current colon wall pixel. The wall pixels surrounding the current wall pixel would be located at the extremities of this longest run (see Fig. 2). This property is implemented in the neighbouring wall pixels (NWP) algorithm.

The NWP algorithm searches the longest run of air pixels (i.e. those pixels below the segmentation threshold) around the current colon wall pixel. The relative co-ordinates of the first and last points of the run are used as indices into a look-up table to retrieve the relative co-ordinates of the previous and the next colon wall pixels. These relative co-ordinates, combined with the co-ordinates of the current colon wall pixel give the absolute co-ordinates of the two wall pixels adjacent to the current wall pixel if they exist (i.e. if it is not a dead end).

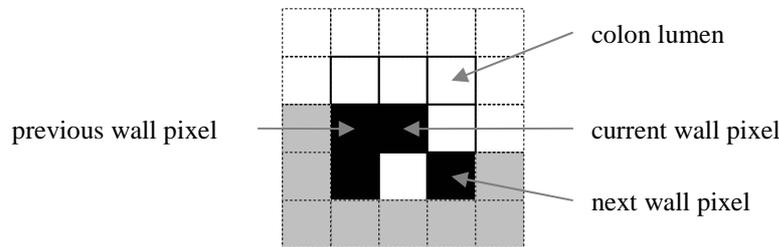


Fig. 2: Illustration of the “next wall pixels” algorithm

This algorithm is used to track the colon wall around the selected central pixel. The CWT algorithm takes the central pixel as an input and returns the co-ordinates of the successive colon wall pixels. It begins by applying the NWP algorithm on the central pixel, and selects one side to begin with (i.e. one of the returned neighbouring pixels). The colon wall is first tracked on the selected side, the “next pixel” then becomes the current pixel and the central pixel becomes the previous pixel. The algorithm is illustrated in Fig. 4. The algorithm continues until all consecutive wall pixels are identified. It terminates when the required length or a dead-end is reached. The algorithm then repeats the operations on the other side of the central pixel to retrieve the complete wall centred on that pixel. This algorithm is used prior to the local angle estimation to retrieve the wall on a length of four pixels on each side of the central pixel. The detected wall pixels are used in the estimation.

4.1.2 Angle Estimation

The angle estimation stage is the most important part of the pre-insertion analysis; the angle is used to synthesise the polyp parallel to the colon segment where it will be inserted. An exact value for the angle is not required, however if the estimation is too inaccurate the artificial polyp will not fit correctly into the colon wall. The angle estimation uses the central pixel as well as the second and fourth pixels on each side of the wall. The notation used to reference the pixels of the colon wall is introduced in Fig. 3.

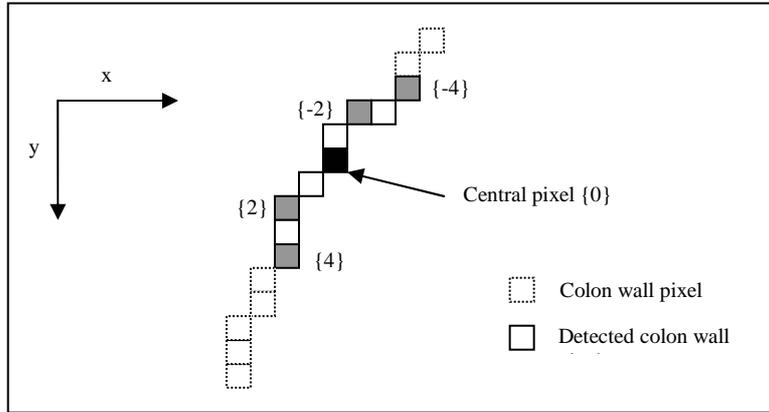


Fig. 3: Detected colon wall and key points notation

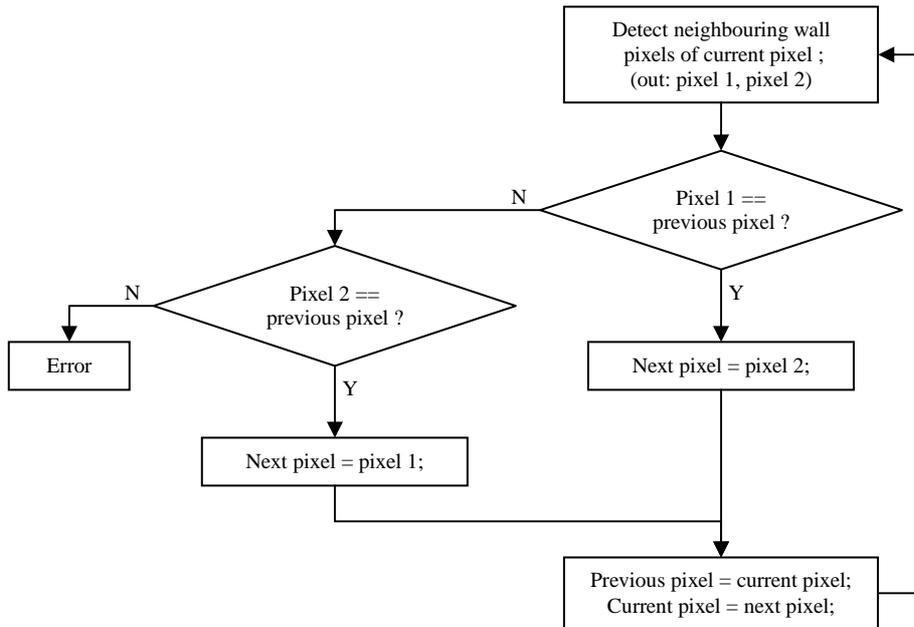


Fig. 4: Colon wall tracking (CWT) algorithm

The algorithm calculates the angles (with the x -axis) of the segments $[-4] [-2]$, $[-2] [0]$, $[0] [2]$ and $[2] [4]$. The average of these four angles gives the local angle $\alpha \pm k \frac{\pi}{2}, k \in N$. The best fit between the four possible angles resulting $(\alpha, \alpha + \pi/2, \alpha + \pi, \alpha + 3\pi/2)$ is calculated using a least mean square (lms) method. The angles are modified to stay in the range $[-\pi, \pi]$ in order to facilitate the calculation. The least mean squares error function used is:

$$lms = (\text{angle} - a_1)^2 + (\text{angle} - a_2)^2 + (\text{angle} - a_3)^2 + (\text{angle} - a_4)^2. \quad (4.2)$$

where *angle* is the currently tested angle (α , $\alpha+\pi/2$, $\alpha+\pi$ or $\alpha+3\pi/2$) and a_1, a_2, a_3, a_4 are the angles formed by the segments [{"-4"} {"-2"}], [{"-2"} {"0"}], [{"0"} {"2"}] and [{"2"} {"4"}] with the *x*-axis. The error function (4.2), calculated from the differences between the candidate angle and the four colon wall angles, gives an estimate of the suitability of the angle. The angle with the smallest *lms* error value is selected as the colon wall local angle. In addition a density value is required for polyp synthesis. This value is calculated as the mean value of the density of the colon voxels in the region surrounding the central voxel used in the synthesis and insertion stage.

4.2 Polyp Synthesis

The polyp synthesis and its insertion stages are carried out simultaneously. Size parameters are used to define a volume of interest around the insertion point. This volume is used for the polyp density calculation and in the synthesis and insertion process. The centre of the polyp (the centre of the ellipsoid model) is the central voxel $P_0(x_0, y_0, z_0)$ in the insertion slice (the central slice of the polyp). In the ellipsoid co-ordinates system (centred on P_0 , co-ordinates i, j, k along the main axes) the points belonging to the polyp will obey the inequality:

$$\frac{i^2}{a^2} + \frac{j^2}{b^2} + \frac{k^2}{c^2} \leq 1. \quad (4.3)$$

The following transformation equations (4.4) are required to migrate from the dataset co-ordinates system (O, x, y, z) into the ellipsoid co-ordinates system (O, i, j, k). They are using the angle (α) between the main direction of the central plane ellipse and the *x*-axis (i.e. the local angle of the colon wall). The equation $k=z$ is due to the restrictive condition on the ellipsoid model.

$$i = x \cdot \cos \alpha + y \cdot \sin \alpha. \quad (4.4)$$

$$j = -x \cdot \sin \alpha + y \cdot \cos \alpha.$$

$$k = z.$$

A raster scan of the voxels in the region of interest surrounding the central voxel (at co-ordinates i_0, j_0, k_0 in the ellipsoid co-ordinates system) is performed; for each voxel the transformation equations (4.4) are used on the co-ordinates. If the co-ordinates obey the inequality (4.5) the point belongs to the polyp and its density is set to the polyp's density.

$$\frac{(i - i_0)^2}{a^2} + \frac{(j - j_0)^2}{b^2} + \frac{(k - k_0)^2}{c^2} \leq 1. \quad (4.5)$$

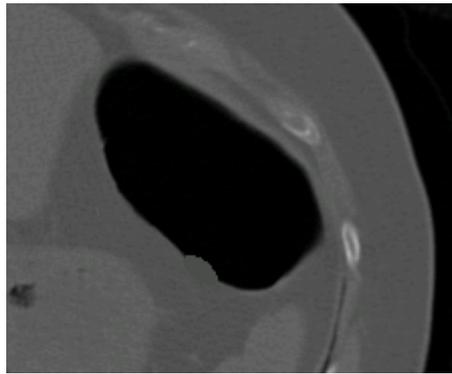


Fig. 5: Artificial polyp synthesised on colon wall

5. AUTOMATIC POLYP DETECTION

A segmented version of the colon lumen is required before the polyp detection process can begin. There are many approaches for colon segmentation including region growing^{10, 11, 12, 15}, edge detection⁸ and high level analysis techniques⁴. The technique described in this paper is based on 26-connected region growing with a boundary threshold of -800 HU. Colonic anomalies appear as bays in the segmented colon lumen. The identification process is a two stage approach: first the detection stage analyses the 2D CT slices, then the classification stage merges the 2D detected objects in 3D colonic anomalies and performs 3D shape analysis.

5.1 2D Detection

The 2D detection algorithm analyses the axial slices of the segmented colon lumen. It aims to detect the bays that may represent potential polyps. First the blobs representing the colon lumen are extracted from the dataset's current slice. The blobs are then labelled and a binary image is created for each of them. The blobs are analysed sequentially. A modified convex hull algorithm is applied to the blob binary image; the modified convex hull is then filled. The original blob image is then subtracted from the filled modified convex hull. This gives a binary image containing the anomalies detected in the current blob. The detection results of all the blobs of the current slice are merged into one binary result slice used to flag the dataset which is passed to the classification algorithm.

The modified convex hull: Due to the natural source of the blobs and to their discrete nature the borders present many imperfections and small undulations. This leads to a problem if a normal convex hull is used. It results in a *crown* effect illustrated in Fig. 7c where the original image is subtracted from the filled convex hull. To avoid this effect a modified convex hull algorithm has been developed. When the normal convex hull would have drawn a straight line it analyses the blob border between the extremities of this line and if some conditions are verified it replaces the line by the actual border.

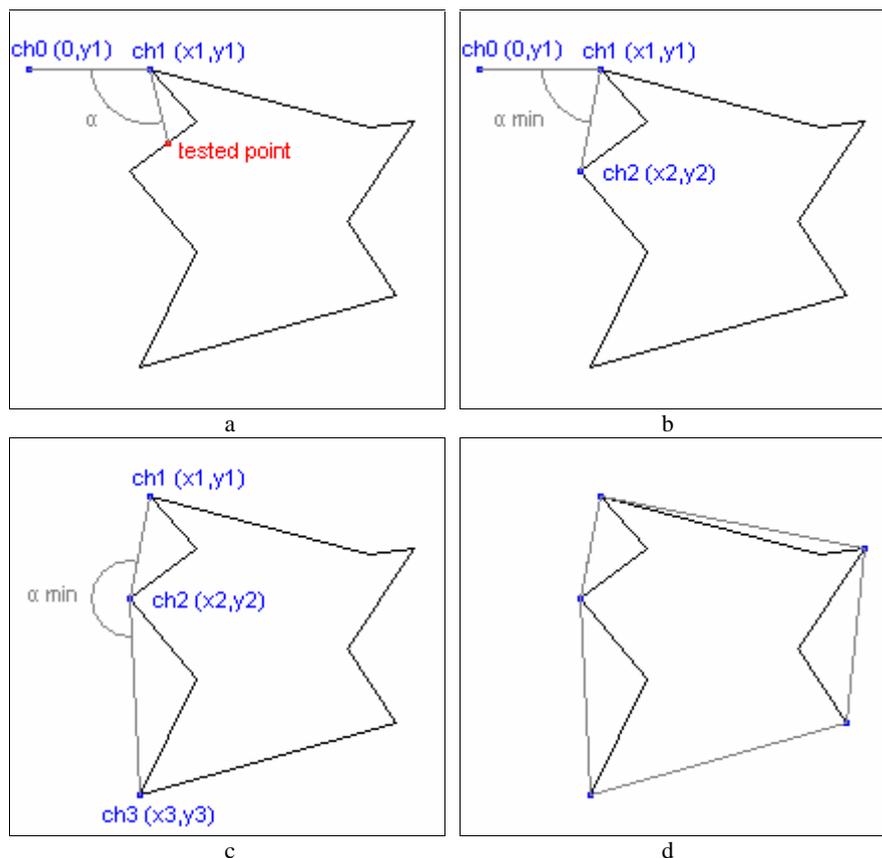


Fig. 6: Standard convex hull construction a) search of the 2^d hull point, b) 2^d hull point found, c) 3rd hull point found, d) complete convex hull

This algorithm is based on the implementation of the convex hull algorithm described by Whelan & Molloy¹⁴. The basis of this algorithm is a convex hull property: any line outside the hull, when rotated toward the hull will touch it in one of its vertex point (hull point). The algorithm is initiated at a point known to be a vertex point; the point with the smallest y value is used. Rather than processing the whole blob, only the border points are used. The algorithm analyses the points in order to find the one which forms the smallest angle with the two previous hull points. This point is a vertex point and the next hull point. For the first calculation the previous point used is the projection of the first point on the y-axis.

The modifications introduced consist of performing a basic analysis of the bay between two vertex points. If a certain condition is verified the actual border of the blob is drawn instead of the line and then it detects the next hull point. When two hull points are detected the algorithm measures the Euclidean distance between them. The border of the blob is then tracked using the approach described in Sec. 4.1.1, and measures the geodesic distance between the two points of the border. The decision between drawing the line and using the wall is then made. If the linear distance is too small (depending on the smallest object to detect) the bay cannot be a polyp. In this case drawing a line would result in the *crown* effect (see Fig. 7c), the actual border (already tracked by the algorithm) is drawn instead. If the linear distance is big enough, the algorithm calculates the ratio of the geodesic distance over the Euclidean distance. If the ratio exceeds a fixed threshold (once again depending on the detection target) then the complexity of the bay is higher than where a *crown* effect is generated. It is a real bay (a potential polyp) and therefore the line is drawn, otherwise the border is drawn. As illustrated in Fig. 7 the detection algorithm is very efficient for overcoming the *crown* effect. This approach avoids passing excessive noise to the classification stage.

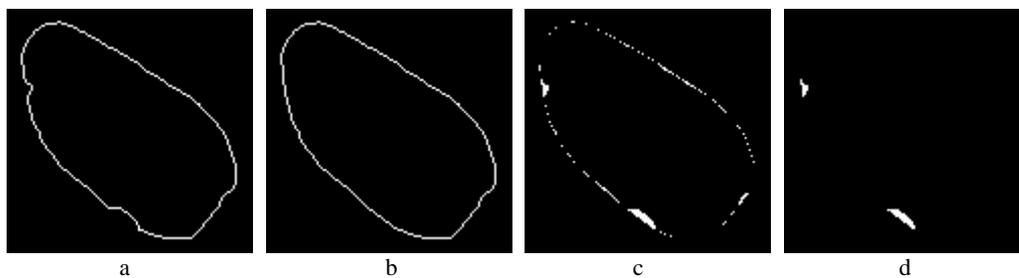


Fig. 7: Modified convex hull: a) border of the processed blob, b) modified convex hull, detection with the c) normal convex hull (*crown* effect), d) modified convex hull

5.2 Classification

The aim of the classification stage is to process the 2D objects detected in the previous stage and to differentiate between polyps and other artefacts (e.g. folds). At the end of the detection stage the detected 2D objects are flagged in the dataset. The classification step uses a 3D labelling technique on these objects to reconstruct 3D colonic anomalies. The classification of these anomalies is based on their morphology. A number of tests are performed on each anomaly, the result of these tests are used to classify the anomaly as either a polyp or a naturally occurring feature.

The first test is a threshold on the total number of points in the object, if this value is too small the anomaly is not processed further as it is below the size threshold. An overlap test is then performed between the successive slices of the object to see if two polyps could have been unintentionally merged in the labelling process (in the direction normal to the slices). If it is the case the anomaly is split and the two parts are added to the anomalies queue. As the detection program aims to detect polyps (between 5 and 20 mm) the next test is a threshold on the number of slices in the anomaly, if the thickness is below the threshold value the test is passed. The shape along the direction normal to the slices is then analysed, the anomaly has to follow a shape profile along the normal direction: in the first half of the set of slices the maximum distance in a slice must be smaller than in the next slice but above a fixed fraction of it, the inverse conditions are applied on the second half of the set of slices. Various size criteria are used next, based on the size of the slices, their number and the ratio between the maximum distance in a slice and the numbers of voxels in it. In order to differentiate between a fold and a polyp, the displacement of the centroid of the successive slices of the anomalies is analysed. If the displacement is small the anomaly is a potential polyp. The final test analyses the region of interest surrounding the anomaly and ascertains the extent of protrusion from the colon wall, if there is no difference between the level of the anomaly and the level of the surrounding wall the anomaly is classified as a naturally occurring feature. The remaining anomalies are flagged as potential polyps.

6. RESULTS

The automated detection program required an average of 13 minutes on a standard 1.5 GHz PC workstation with 512 Mb of RAM. The average size of the dataset was $512 \times 512 \times 266$. The performance of the algorithm was tested using fifteen scenarios which were generated from three different datasets (five different scenarios per dataset). Each scenario contained five polyps of random size and location which were synthesised using the polyp insertion technique described in Sec. 4. The results of the detection are presented in Tab. 1.

The detection technique has a sensitivity of 65% with an average of 1.7 false positives per dataset. This technique is unable to detect polyps on folds. These polyps represent 85% of the undetected polyps therefore an improvement to make the algorithm detect them would significantly improve sensitivity (by up to 92%). The false detections are not dependent of the number of polyps in the dataset; however they are dependent on the dataset itself, in fact the false detection rate is constant and does not depend on the number of polyps in the dataset.

scenario	polyps	true detection	non detection	False detection
1	5	4	1	0
2	5	4	1	0
3	5	5	0	0
4	5	4	1	0
5	5	5	0	0
6	5	3	2	2
7	5	4	1	2
8	5	4	1	2
9	5	2	3	2
10	5	2	3	2
11	4	2	2	3
12	5	2	3	3
13	5	3	2	3
14	5	3	2	3
15	5	1	4	3
total	74	48	26	25

Tab 1: Results of the polyp detection technique applied on 15 scenarios containing 74 polyps

7. CONCLUSION

This paper describes a technique for automatically detecting polyps. In reality there are a limited number of actual polyp configurations available from patient data. This makes software validation difficult. To overcome this problem a technique for synthesising and inserting artificial polyps into real patient data has been described. This technique automatically detects the properties of the colon wall at the location where the user wishes to insert an artificial polyp. The density of the polyp is automatically calculated by examining the local density of the colon wall. This technique can synthesise polyps of any size or shape that can be described by an ellipsoid model (4.3) (mainly round and flat polyps).

The polyp detection technique uses a two stage approach: initially a 2D analysis of the axial CTC slices is performed to detect colonic anomalies and then a 3D morphological analysis of these anomalies is used to classify them as polyps or naturally occurring features. The 2D detection stage introduces a modified convex hull algorithm. This algorithm detects the bays of interest in a shape. The 3D morphological analysis technique examines dimension, shape and shape variation profile to achieve the classification. Initial test have been performed and a sensitivity of 65% has been achieved with an average of 1.7 false positives per dataset.

ACKNOWLEDGMENT

The authors wish to thank the members of the Vision Systems Laboratory in Dublin City University. We also wish to thank members of the Gastrointestinal Unit and the Department of Radiology in the Mater Misericordiae Hospital, Dublin, particularly Dr. Helen Fenlon and Dr. Padraic MacMathuna.

REFERENCES

1. ACR-NEMA,, *Digital imaging and communication in medicine*, National Electrical Manufacturers Association, 2000.
2. C.F. Beaulieu, S. Napel, B.L. Daniel, I.Y. Chen, G.D. Rubin, I.M. Johnstone and R.B. Jeffrey, "Detection of colonic polyps in a phantom model: Implications for virtual colonoscopy data acquisition", *Journal of Computer Assisted Tomography*, **22 (4)**, pp. 656-663,1998.
3. S.C. Chen, D.S.K. Lu, J.R. Hecht and B.M. Kadell, "CT colonography: value of scanning in both the supine and prone positions", *American Journal of Roentgenology*, **172 (3)**, pp. 595-599, 1999.
4. D.Q. Chen, Z.R. Liang, M.R. Wax, L.H. Li and A.E. Kaufman, "A novel approach to extract the colon lumen from CT images for virtual colonoscopy", *IEEE Transactions on Medical Imaging*, **19 (12)**, pp. 1220-1226, 2000.
5. S.B. Göktürk, C. Tomasi, B. Acar, C.F. Beaulieu, D.S. Paik, R.B. Jeffrey, J. Yee and S. Napel, "A statistical 3D pattern processing method for computer-aided detection of polyps in CT Colonography", *IEEE Transactions on Medical Imaging*, **20 (12)**, pp. 1251-1260, 2001.
6. C. Karadi, C.F. Beaulieu, R.B. Jeffrey, D.S. Paik and S. Napel, "Synthesis and insertion of colonic polyps into patient CT data: technique and evaluation", *Radiology*, **209p (suppl)**, pp. 430 (1292), 1998, [abstract].
7. C. Karadi, C.F. Beaulieu, R.B. Jeffrey, D.S. Paik and S. Napel, "Display modes for CT colonography part I. Synthesis and insertion of polyps into patient CT data", *Radiology*, **212 (1)**, pp. 195-201, 1999.
8. Y. Masutani, H. Yoshida, P.M. MacEneaney and A.H. Dachman, "Automated segmentation of colonic walls for computerized detection of polyps in CT colonography", *Journal of Computer Assisted Tomography*, **25 (4)**, pp. 629-638, 2001.
9. NCRI, *Cancer in Ireland, 1997: incidence and mortality*, Healy & Associates, 2000.
10. Y. Samara, M. Fiebich, A.H. Dachman, K. Doi and K.R. Hoffmann, "Automated centerline tracking of the human colon", *Proceedings of the SPIE - The International Society for Optical Engineering*, **3338 (1)**, pp. 740-746, 1998.
11. Y. Samara, M. Fiebich, A.H. Dachman, J.K. Kuniyoshi, K. Doi and K.R. Hoffmann, "Automated calculation of the centerline of the human colon on CT images", *Academic Radiology*, **6 (6)**, pp. 352-359, 1999.
12. M. Sato, S. Lakare, M. Wan, A. Kaufman and M. Wax, "An automatic colon segmentation for 3D virtual colonoscopy", *IEICE Transactions on Information and Systems*, **e84-d (1)**, pp. 201-208, 2001.
13. D.J. Vining, D.W. Gelfand, R.E. Bechtold, E.S. Scharling, E.K. Grishaw and R.Y. Shifrin, "Technical Feasibility of Colon Imaging with Helical CT and Virtual Reality", *American Journal of Roentgenology*, **162**, suppl:104, 1994.
14. P.F. Whelan and D. Molloy, *Machine Vision Algorithm in Java: Techniques and Implementation*, Springer-Verlag, ISBN 1-85233-218-2, 2001.
15. C.L. Wyatt, Y. Ge, D.J. Vining, "Automatic segmentation of the colon for virtual colonoscopy", *Computerized Medical Imaging and Graphics*, **24 (1)**, pp. 1-9, 2000.
16. H. Yoshida and J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps", *IEEE Transactions on Medical Imaging*, **20 (12)**, pp. 1261-1274, 2001.