Automatic Polyp Detection of Colon Using High Resolution CT Scans

Jamshid Dehmeshki[1], Hamdan Amin[1], Wing Wong[1], Mandana Ebadian Dehkordi[1], Nahid Kamangari[1], Mary Roddie[2], John Costello[3]
[1] Medicsight Plc. 46 Berkeley Square, Mayfair, London, W1J 5AT, United Kingdom
[2] The New Victoria Hospital, 11 Briar Way, London SW15 6UD
[3] Cromwell Hospital & Kings College Hospital, 8 Pennant Mews, London, SW5 0TU

jamshid.dehmeshki@medicsight.com

Abstract

Automatic detection of polyps can be a valuable tool for diagnoses of early colorectal cancer as early detection and hence removal of polyps can save life. Polyp detection is a challenging task as polyps come in different sizes and shapes. The detection generally consists of three stages; 1) colon segmentation, 2) identification of suspected polyps and 3) polyp classification. The latter involves classifying polyps from among many suspected regions. This paper concentrates on the first two stages of the detection. For the colon segmentation, the fuzzy connectivity region growing technique is used while for the identification of suspected polyps concave region searching is applied. The method is fast, robust and validated with a number of high-resolution colon datasets.

1. Introduction

Colorectal cancer is the second leading cause of cancer in industrial western countries and the third most frequently diagnosed cancer world-wide [1]. Most colon cancers begin as polyp that grows from the mucosal surface of the colon and It takes about 10 years for a polyp to develop into cancer [2]. The risk of developing cancer is dependent on the size of polyps [3]. The slow rate of cancer growth gives sufficient time to detect polyps (often symptomless) if regular screening is performed. CT colonography (CTC) is a promising screening technique for the detection of early colorectal cancer. It produces high resolution datasets that might contain over a thousand slices per patient. This makes the polyp detection a very tiresome task for radiologists. The solution is to allow computer software to automatically detect a number of potential lesions, which would direct the radiologists to concentrate only on a limited portion of the vast information within the high-resolution datasets.

Automatic detection of colon polyps is a challenging task as polyps come in various sizes and shapes. There are a number of structures that mimic the shape and intensity of polyps such as thickened folds and stool. These lead to the detection of numerous false positives which must be reduced using different filtering techniques and classification functions. Most researchers have used shape features to identify polyps [4-6]. Majority of polyps have semi-spherical shapes and thus most algorithms are made to search for curvature variation on the 3D surface of the structures. Techniques like shape index [4], Gaussian convolution [2] and surface normal accumulation [3] are used to look for abnormalities within the colon lumen by identifying regions that posses sphere-like features. These techniques require a substantial amount of time when applied to the whole colon datasets for processing partial derivatives within the volume of interest. These techniques however can be useful if applied during the classification phase. This way, the computations can be limited to a relatively small parts of the colon lumen that are identified as potential regions.

The technique carried out in automatic detection of polyps resembles the ways a radiologist attempts to search for a polyp in a colon dataset. A radiologist first identifies the colon lumen by concentrating on the regions that are filled with air and fall under the category of colon lumen based on anatomical knowledge. He/she then looks for any object that appears as a raised structure grown from the colon wall into the lumen and evaluates the connectivity between the identified regions while traversing through the slices. The next move will involve concentrating on the region of interest to classify it as being a polyp or other polyp-like structures.

This paper provides an efficient technique for automatic detection of polyps. It consists of two stages, colon segmentation and potential polyp detection. The polyp classification, which is the last phase of detection, is currently under development and will be the subject of the future research outcomes. The aim of this research is to design a system that detects all the potential polyps prior to the classification phase.

The colon segmentation is carried out by means of the region growing technique that identifies all voxels within the colon lumen that form a connected component from a seed point. The potential areas are then identified by evaluating the contours of the segmented colon and searching for any raised structure from the colon wall. The algorithm will then follow the same step as a radiologist does by traversing through the slices and identifying the connectivity between the selected regions of abnormalities. The potential regions (3D objects) will not only include all the polyps but also

will hold a large collection of polyp-like structures which appear as raised objects from the colon wall. A large number of these false positives (FPs) will be eliminated in the last step by searching for certain known features. This last stage of the polyp detection (the classification phase) will be the subject of future publications.

The following sections will provide more information regarding the steps taken for the identification of polyps in colon datasets. The colon segmentation using fuzzy connectivity technique is presented first followed by the potential polyp detection. The results section will then provide the experimentation outcomes of applying the proposed technique on a number of colon datasets.

2. Method

The aim of the proposed method is to allow for an efficient automatic detection of polyps (without false negative) prior to the classification phase. The following sections provide the approaches taken for the colon segmentation and the potential polyps identification.

2.1 Colon Segmentation Using Fuzzy Connectivity

To extract the colon from the volume data the fuzzy connectivity technique [7] was used. This technique is based on the dynamic programming to perform region growing from a seed point. The fuzzy connectivity algorithm utilizes the fuzzy nature of images by capturing the "hanging togetherness" of the voxels that are part of one object. It takes advantage of these similarities and tries to extract the object from the image scene. Two main terminology are used in fuzzy connectivity theory, fuzzy affinity and fuzzy path.

Fuzzy path between two points is obtained based on the fact that there are numerous paths between the two points. The path with the strongest link is chosen as the best path and the strength of each path is the weakest link of adjacent points along the path. This concept is shown in Figure 1, where two possible paths between points p1 and p10 are depicted. Think of this as a set of ropes tied together with different thickness (strength). If the top set of ropes (from p1 to p10) is pulled apart, it will break at the p2-p3 link; while the bottom set of ropes will break at the link p8-p9. Therefore the bottom rope is stronger than the top one with the strength equal to the strength of p8-p9 link. The strength between two adjacent points is called fuzzy affinity.

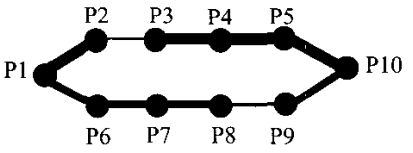


Figure 1. Two sets of ropes representing two fuzzy paths between two points (p1-p10), while two consecutive points represent the fuzzy affinity.

Fuzzy affinity is the assignment of a probability between two spels (space elements, i.e. pixel for 2D and voxel for 3D) belong to the same object. This probability is a function of Eucledian distance between spels and the image feature in the neighbourhood of the spel. Fuzzy affinity must satisfy the following conditions:

Reflectivity: for all $(x,x) \in X \times X$, $\mu(x,x) = 1$ *and*

Symmetric: for all $(x,y) \in X \times X$, $\mu(x,y) = \mu(y,x)$

Where μ ($\mu: X \times X \rightarrow [0,1]$) represents the fuzzy relation (or affinity) for a 2-ary relation; i.e. for simplicity 2-ary instead of m-ary is used. The first condition indicates that the fuzzy affinity of a spel to itself is always 1, and the second condition indicates that the fuzzy affinity from point 1 to point 2 should be the same as that of point 2 to point 1.

The general model for fuzzy relation is given by:

$$\mu_k = (c, d) = h(\mu_a(c,d), f(c), f(d), c, d)$$

where h is a scalar value with range [0,1], c and d are image locations of two spels, and $f(i)$ is the intensity of spel i. μ_a is an adjacency function based on distance between two spels which, for n-dimensional coordinate spels, is given by,

$$\mu_a(c,d) = \begin{cases} 1, & if \sqrt{\sum_{i=1}^{n}(c_i - d_i)^2} \le 1 \\ 0, & otherwise \end{cases}$$

A simplified shift invariant version was proposed in [7],

$$\mu_k(c, d) = \mu_a(c, d)[\omega_i h_i(f(c),f(d)) + \omega_g h_g(f(c),f(d))] \quad if\ c \neq d$$
and
$$\mu_k(c, c) = 1$$

where the subscripts 'i' and 'g' represents the calculations related to intensity and gradient values respectively. ω_i and ω_g are free parameter weight values whose sum is 1. The value of 0.5 for both weights was chosen to allow equal affect from both intensity and

gradient. The fuzzy affinity that was used for the current function is:

$$h_i(f(c),f(d)) = \frac{1}{\sqrt{(2\Pi)s_i}} e^{-(1/2)[(1/2)(f(c)+f(d))-m_i]^2}$$

$$h_g(f(c),f(d)) = \frac{1}{\sqrt{(2\Pi)s_g}} e^{-(1/2)[(|f(c)-f(d)|-m_g]^2}$$

where m_i, s_i, m_g and s_g are the Gaussian parameters (mean and standard deviation) for the intensity and gradient. These values are obtained by calculating the Gausaain parameters on a region that is either indicated by the user or retrieved from the seed point by applying a region growing on a small area. Our method uses the later approach to reduce the user interactivity with the computer.

2.2 Potential Polyp Detection

Once the colon is segmented, the potential areas are identified by searching for concave regions on the contour of the cross section of a segmented colon. The concave regions are identified by evaluating the angular variations on the consecutive pixels that form the contour of the segmented colon in one slice.

A difficulty arises when attempting to detect the concave regions in an area where a polyp might be surrounded by other raised objects. This occurs, for example, when a polyp has grown from the side of a fold. Such cases will lead to the formation of two or more concave regions on top of one another. This requires a more careful search engine where the folds are detected first and further concave region searching will be conducted on the same fold area.

Once the concave regions are identified, their volumetric connectivity are detected in the colon dataset to allow for the formation of 3D regions that represent the potential objects. It should be noted that there is a significant amount of overlap between two consecutive regions of the same object; furthermore the size of the object increases or decreases by relatively small amount in two slices. Threshold values for size and overlap measurement were used to create 3D objects by joining regions that have more probability of being part of a single object. This allowed for separation of folds from polyps when the two objects were attached or geometrically were very close.

Figure 2 shows the surface rendering of the three objects identified by the proposed algorithm. As can be seen the 3D formation of folds represents objects that are elongated in one direction while that of polyps or polyp-like structures, such as stool, appear as hemispherical objects. The elongation feature can be used to quickly remove many folds from the false positive list.

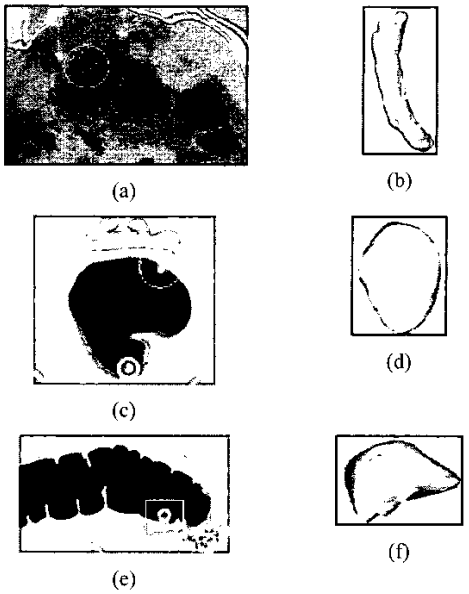


Figure 2. The 3D visualization of potential regions.; a-b) a fold, c-d) a polyp, e-f) stool.

3. Experimental Results

A set of colon data sets were used to test our algorithm. Following colon cleansing and air-insufflation, CT scans of the supine and prone positions were obtained using the GE MEDICAL SYSTEMS. The images were reconstructed on a 512×512×N matrix, where N is the number of slices ranging from 300 to 700. The resulting images had an average voxel spacing of $0.7\times0.7\times\eta$ mm where η varies between 1.25 to 4 mm. The reason for choosing such a wide range of voxel spacing was to test the robustness of the proposed approach. The total number of polyps were 30 out of 17 colon data sets. The polyps were identified on the CT scans by a qualified radiologist and verified with the colonoscopy results. The algorithm was tested on a PC with 2.8 GHz of CPU speed and 2GB of RAM.

3.1. Colon Segmentation

Colon lumens were first extracted from the datasets. In some of the scans, total colon distention was not present which caused the region growing to stop halfway due to the collapses occurred at certain locations. For these cases we needed to provide multiple seed points at different locations within the colon lumen to ensure that the entire colon was segmented. In the future, an anatomy based colon extraction will be used to eliminate the need for providing seed points which leads to a complete automation system.

Figure 3 shows the 3D surface rendering of an entire colon that was segmented from a colon dataset using the fuzzy connectivity algorithm.

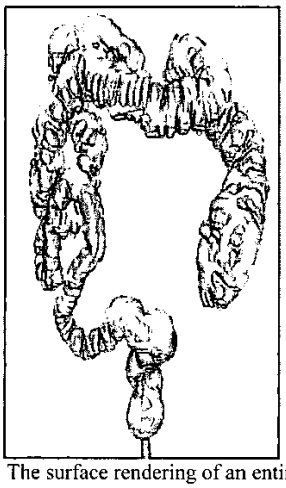
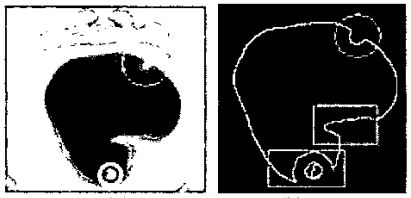


Figure 3. The surface rendering of an entire colon.

3.2. Potential Polyp Detection

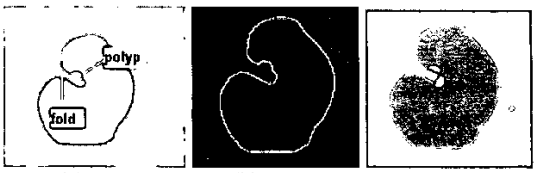
Once the colons were segmented, the algorithm was used to search for concavity of the contour of the segmented regions. Figure 4 shows an example of concave regions on the contour of a colon. In this figure, a 5mm polyp is indicated by a circle while the boxes show two false positives.



(a) (b)

Figure 4. Concave regions on the contour of a segmented colon. A 5mm polyp (of a 0.65×0.65×2mm dataset) is indicated by a circle while two boxes represent false positives.

The polyp indicated in Figure 4 is relatively easy to detect as the concavity of it's contour is clearly isolated from the surroundings. There are no other concave objects nearby to distort the detection. As mentioned earlier, a difficulty arises when trying to detect a polyp grown on the side of a fold. One such example is given in Figure 5. Here the two concave regions can be clearly seen that are located side by side. Figure 5c shows the result of the proposed method that has successfully identified the polyp as a separate potential region.



(a) (b) (c)

Figure 5. A 10mm polyp (of a 0.77×0.77×2mm dataset) on the side of a fold.

The difficulty of detecting polyps grown on fold would also exist when forming the 3D object by searching for the volumetric connectivity. This problem was alleviated by monitoring the overlap and size differences between two consecutive regions while generating the 3D objects. Two thresholds values of 60% and 30% were used for the size and overlap evaluation respectively. Figure 6 shows an example of polyp grown from the side of a fold, which can be clearly seen in Figure 6c. The middle image of Figure 6a shows that the polyp and the fold are merging as one object. Figure 6b shows the result of the proposed technique before generating 3D object where the polyp and fold are detected as being part of one structure (polyp) in the left image. This merging effect did not hamper the extraction of the polyp since the algorithm used the overlap and size differences measurement and successfully generated two separate 3D objects (one polyp and one false positive). Figure 6d shows the surface rendering of the extracted polyp by our proposed method.

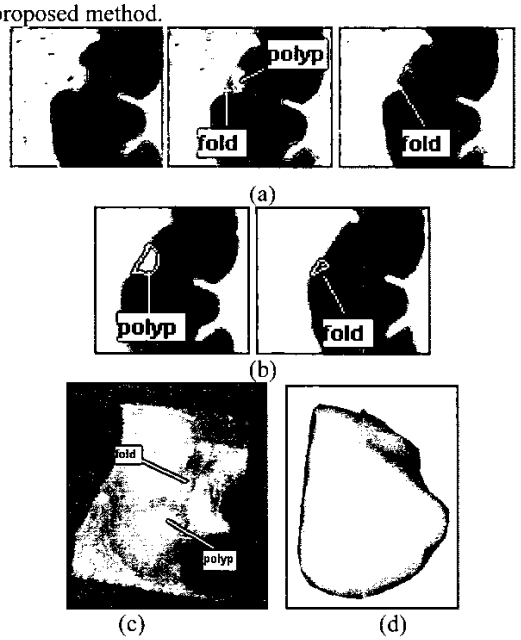


Figure 6. A 10mm Polyp attached to a fold in a 0.62X0.62X2.5mm dataset, a) a series of slices showing the attachment, b) the detected objects (the two images correspond to the two right images in (a), c) the volume rendering of the region, d) surface rendering of the extracted polyp.

All 30 polyps ranging from 3mm to 20mm were detected. The average time taken for processing each segmented colon was less than one minute running on a PC with the specification given earlier. This is a very fast technique even when dealing with high resolution data sets.

There were an average of 400 false positives per scan. No filtering was applied to eliminate the false positives. This will be the next phase of our detection system that is currently under development. Many features will be used to classify the false positives from the true positives.

4. Conclusion and Further work

An efficient automatic polyp detection system has been presented in this paper. The detection involves searching for concave regions on the contour of segmented colon. The potential objects are found by forming 3D structures through volumetric connectivity evaluation. The robustness of the technique was measured by examining various data sets with different resolution ranging from 4mm to 1.25 mm in axial direction. The experimental results showed that the proposed technique is fast, robust and reliable with the capability of not missing any polyp.
Two main research activities are currently under way, automatic colon segmentation and polyp classification. For the former anatomy information will be used to allow for the segmentation of colon based on the geometry features used by radiologists. For the classification of polyps, a number of principal features will be fed into a classifier system to distinguish polyps from among many suspected objects that would be detected using the technique described in this paper.

Acknowledgment

The Authors would like to thank the members of the Medicsight research and software engineering group for their valuable assistance and the Medicsight PLC. for supporting the project.

References

[1] "Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a UK multicentre randomised trial", The Lancet, Vol 359, April 13 2002, pp. 1291-1300, www.thelancet.com.

[2] C. Daniel Johnson and Abraham H. Dachman, "CT Colonography: The Next Colon Screening Examination?", Radiology 2000 216: 331-341.

[3] G. Scott Gazelle, Pamela M. McMahon, and Francis J. Scholz, "Screening for Colorectal Cancer", Radiology 2000 215: pp. 327-335.

[4] H. Yoshida, J. Nappi, "Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps", IEEE Transactions on Medical Imaging, No. 12, December 2001, pp. 1261-1274.

[5] R. M. Summers, C. D. Johnson, L. M. Pusanik, J. D. Malley, A. M. Youssef, and J. E. Reed, "Automated Polyp Detection at CT Colonography: Feasibility Assessment in a Human Population", Radiology 2001 219: 51-59.

[6] S. B. Goturk, C. Tomasi, B. Acar, C. F. Beaulieu, D. S. Paik, B. Jeffery, J. Yee, S. Napel, "A Statistical 3-D Pattern Processing Method for Computer-Aided Detection of Polyps in CT Colonography", IEEE Trans. Medical Imaging, Vol. 20, No. 12, Dec. 2001, pp. 1251-1260.

[7] J. K. Udupa, S. Samarasekera, "Fuzzy Connectedness and Object Definition: Theory, Algorithms, and Applications in Image Segmentation", 1996, Graphical Models and Image Processing, Vol. 58, No. 3, pp 246-261.