CT COLONOGRAPHY COMPUTER-AIDED POLYP DETECTION USING TOPOGRAPHICAL HEIGHT MAP

Jianhua Yao[1], *Jiang Li*[1,2], *Ronald Summers*[1]
1. Diagnostic Radiology Department, the National Institutes of Health, Bethesda, Maryland 20892
2. ECE and VMASC, Old Dominion University, Norfolk, VA 23529

ABSTRACT

CT Colonography (CTC) is an emerging noninvasive technique for screening and diagnosing colon cancers. Computer Aided Detection (CAD) techniques can increase sensitivity and reduce false positives. We propose to employ topographical height maps in our CAD pipeline. For every detection, a height map is computed using a ray-casting algorithm. Since colonic polyps are protrusions outward from the colon wall and are round in contour, their height maps present concentric patterns. The projection direction is optimized through a multi-scale spherical search. We derive several topographic features from the map, and also compute texture features from the Haar wavelet coefficients. We send the selected features to a committee of support vector machines for classification. We have tested our method on 1186 patients with 226 polyps. Results showed that the height map features can reduce false positives by about 50%.

Index Terms— CAD, CT Colonography, Height map

1. INTRODUCTION

Colorectal cancer is the second leading cause of cancer death in the United States, estimated to claim 56,290 lives in 2005 alone [1]. CT colonography (CTC) is a minimally invasive technique and can detect both pre-cancerous polyps and colon cancers. The accuracy of CTC is reported by some investigators to be close to that of traditional optical colonoscopy (OC) [2].

Computer aided detection (CAD) systems to detect polyps using CTC have been under investigation in the past decade. Most of these systems used features computed from the original CT data set, such as surface curvature [3], optical flow, deformable models [4], and surface normal overlap [5]. One investigation suggested that radiologists can improve their interpretation accuracy by navigating through the 3D colon surface reconstructed from the CT data. Therefore, Li et al [6] proposed a method based on the 2D endoluminal projection image. They computed the projection image using graphical snapshot through an optimized viewpoint and derived wavelet coefficients for classification.

Colonic polyps are small growths that protrude outward from the colon wall, and are characteristically round in contour. In contrast, haustral folds and other normal colonic structures tend to be circumferential and ridge-shaped. Based on these observations, we can employ the height map in the polyp detection. The main idea is that the colon surface can be viewed as a terrain and polyps as bumps on the terrain. The unique topographic features of bumps can assist us in distinguishing polyps from the rest of the colon.

Height maps are commonly used in geographic information systems (GIS), where they are also called digital elevation models [7]. Remote sensing techniques such as radar or LIDAR are applied to determine the distance to an object or surface. The distance can then be converted to the elevation data. In the case of colonography, we can survey the colon surface by placing a camera above the detections. Using the ray casting technique, we can compute the distance from the camera to the surface. The inverted distance can be converted to the elevation height from the base of the polyp. The height data is then digitized and stored in a 2D grid. Due to the special characteristics of polyps, their height map will present concentric patterns and other topographic features. Figure 1 shows one CT slice of a polyp, its 3D endoluminal view and height map.

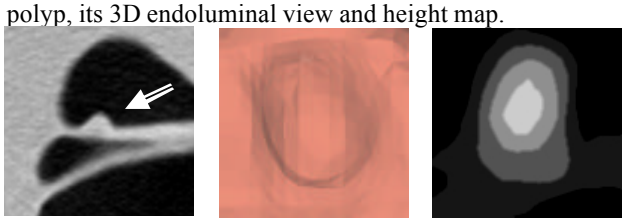

Figure 1. Colonic polyp and height map
a) 2D CT slice of a 10mm polyp, b) 3D endoluminal view, c) Height map

2. METHODS

The pipeline of our CAD system is as follows: First a CTC data set is read into the program. The colon surface is extracted using fuzzy connectness and iso-surface techniques [3]. A curvature filter is applied to every vertex on the surface. The passed vertices are then clustered according to their connectivity, and the centroid of each cluster is treated as one detection. The detections are then sent to a segmentor for computation of boundaries, 3D

shape, and attenuation features. These features are then fed to a classifier. A height map is generated for each detection passed through the classifier. Topographic features and wavelet coefficients are computed and used for further classification. This pipeline has three stages of filtering: surface curvature filter, CT feature classifier and height map feature classifier. This paper is focused on the height map stage. The details of the first two stages can be found in [8].

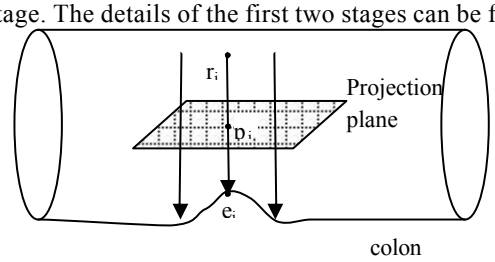

Figure 2. Height map generation

2.1 Height map generation

The height map generation is illustrated in figure 2. The method is based on a ray casting technique. We use orthogonal projection (figure 2). For every point p_i on the projection plane, a ray r_i is cast through p_i. The point that ray r_i encounters on the colon surface is e_i. The distance d_i between p_i and e_i is recorded at p_i. When locating point e_i, we resort to the 3D image grid directly and use the implicit iso-surface associated with the colon surface. The iso-value is -700 HU. The distance map can be converted to a digitized height map using the following equation,

$$h_i = (\max D - d_i) * 256 / \max D \qquad (1)$$

Where maxD is the maximum distance in the distance map, and 256 is the grayscale in the height map. The size of the height map is 25mm * 25mm (128*128 image grid). A typical height map is shown in figure 1c.

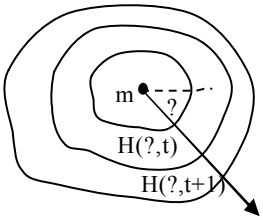

Figure 3. Directional slope

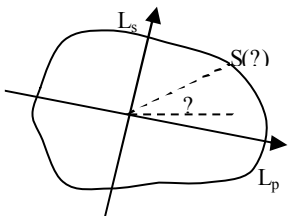

Figure 4. Slope contour

2.2 Topographic features

The directional slopes in a height map describe many attributes of the terrain surface. In the case of ideal semi-spherical shape polyps, the slope from the apex to the base should be homogeneous and big in every direction, and the height map presents concentric patterns. In the case of haustral folds, the slopes are big in the directions perpendicular to the ridge and small in the directions parallel to the ridge. In the normal flat colon surface, the slopes are uniformly small in all directions.

To compute the directional slope, first we locate the apex in the center of the height map. In the case where there is a plateau, the center of the plateau is used. We then compute the accumulative slope in a direction *r* defined by the angle *?*,

$$s(\theta) = \sum_t (H(\theta,t) - H(\theta,t+1))$$
$$H(\theta,t) = h(m + r(\theta) * t) \qquad (2)$$
$$r(\theta) = (\cos\theta, \sin\theta)$$

where $s(?)$ is the accumulative directional slope at the angle *?*, m is the apex, t is the time step, $h()$ is the height map. Figure 3 illustrates how the directional slopes are computed. Several topographic features can be derived from the directional slopes. The mean $M(s)$ and standard deviation $STD(s)$ of directional accumulative slopes in all directions are among the most useful ones.

We then transform the directional accumulative slopes into a contour in a polar coordinate system, where the angular coordinate is the slope direction (angle) and the radial coordinate is the accumulative slope (Figure 4). We name this contour the slope contour. We perform a principal component analysis [9] on the slope contour and compute the principle and secondary axis of the contour. The lengths of the principle and secondary axis (L_p, L_s) and the slope aspect ratio $AS=L_p/L_s$ are computed as topographic features.

The concentric pattern is one of the characteristic features of a polyp presenting in the height map. We design a concentric index to gauge the pattern,

$$CI = M(s) - STD(s) \qquad (3)$$

where CI is the concentric index. A height map with higher and more homogeneous directional slopes in all directions tends to have higher concentric index.

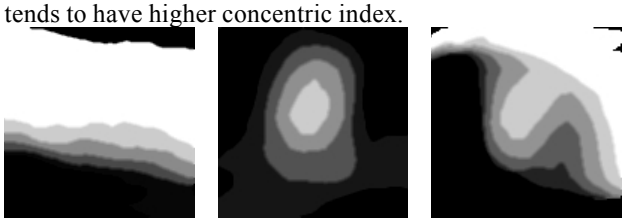

Figure 5. Height maps from different projection directions

2.3 Projection direction optimization

In the ray casting method, different projection directions could generate very different height maps. Figure 5 shows height maps from different projection directions for the same detection. Since the polyps are distinguished from false positives based on the concentric pattern of the height map, we optimize the projection direction by maximizing the concentric index. The projection direction can be obtained as the direction from points on a unit sphere looking at the center of the sphere which is the detection location. The center of the projection plane is located at 12mm from the detection center. If the projection plane is out of the lumen, the projection direction is excluded from consideration. We

adopted a spiral-point technique [10] to generate uniformly distributed points on a sphere and used the spherical coordinates (*?, f*), *0= ?=p, 0= f =2p*, to define the projection directions,

$$h_k = -1 + \frac{2(k-1)}{(N-1)}$$
$$\theta_k = \arccos(h_k) \qquad (4)$$
$$\varphi_k = \left(\varphi_{k-1} + \frac{3.6}{\sqrt{N(1-h_k^2)}} \right) \bmod 2\pi$$
$$v_k = (\sin\theta_k \cos\varphi_k, \sin\theta_k \sin\varphi_k, \cos\theta_k)$$

here N is the total number of projection directions, $1 = k = N$, v_k is a projection direction. The angle interval between adjacent directions approximately equals $\frac{218.4}{\sqrt{N}}$ (in degree). It requires 1907 evaluations for all directions at a 5 degree angle interval. Currently the system takes about 0.05 second for one height map evaluation. It takes about 100 second to exhaustively search all directions. The CAD system typically generates about 50 detections per data set at this stage. It will take 5000 seconds (about 80 minutes) for just the projection optimization process alone.

In order to improve the time efficiency of the projection direction optimization, we propose a multi-scale spherical search scheme. In this scheme, the search starts with a big angle interval and decreases the angle interval in following iterations. A full range search is conducted at the first scale. After that, a local search is conducted in the neighborhood of the direction returned from the previous scale. Currently we use a 3-scale searching scheme, with 30, 10 and 5 degree intervals at each scale. It takes at most 53+25+9=87 height map evaluations for the 5 degree angle resolution.

2.4 Wavelet features

Texture analysis using wavelet transformation has been applied in painting forgery identification [11] and polyp detection [6]. We conduct a 5-level wavelet transformation on the height map. Each level has sub-bands in vertical, horizontal and diagonal directions. We compute 6 features for each sub-band coefficients (mean, variance, skewness, kurtosis, energy and entropy). We also compute the prediction errors for all features in levels 1-3. The prediction errors are based on the difference of the magnitude of one wavelet coefficient and that predicted by its neighbors. We therefore obtain 150 wavelet features in total for a height map, including 96 (16*6) coefficient statistics and 54 (9*6) error statistics.

2.5 Feature selection and classifier

We use a committee of support vector machines (SVM) as our classifier. We have 156 features (6 topographic and 150 wavelet features) in the feature space. We employed a forward stepwise feature selection method to choose a pool of good performing feature vectors. Then we applied another stepwise method to form a SVM committee. The configuration of the committee was determined by an ANOVA analysis. We use a five-member committee, and each member has four features. The details of the classifiers used in this paper can be found in [12].

2.6 Data specification and experiments

The CTC data were acquired as follows. Patients underwent 24-hour colonic preparation that consisted of oral administration of contrast agents. Each patient was scanned in supine and prone positions. CT scanning parameters included 1.25- to 2.5-mm section collimation and 1-mm reconstruction interval. The patient also underwent optical colonoscopy on the same day of the CTC. After the CTC data was transferred to a workstation, manual segmentation was performed by a consensus panel consisting of a research trainee and an experienced radiologist to delineate the border of every polyp in an axial view of the CTC. The manual segmentation was used as the reference standard.

We have separate training and test data sets. The classifier was generated using the training set and a 10-fold cross validation was conducted. The classifier was also independently validated using a separate test set. FROC curves are used to evaluate the performance. The evaluations were conducted in two polyp size categories: >=10mm, and between 6mm and 9mm. In order to test the effectiveness of the height map, the performances were compared with and without the height map classification phase.

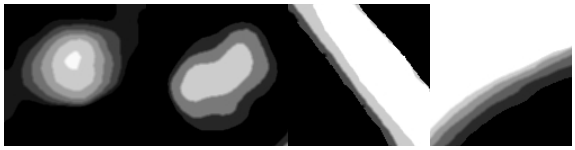

Figure 6. Height map samples
a, b) polyps; c) Haustral fold; d) normal colon surface

3. RESULTS

The patient population consisted of 1186 adults between 40 and 79 years of age at 3 medical institutions. Table 1 lists the distribution of CTC studies and polyps in the training and test sets.

	Total patients	Polyps >=6mm	>=10mm	6-9 mm
Training	394	71	25	46
Test	792	155	43	112

Figure 6 shows the height map of some typical polyps and false positives. Figure 7a-b shows the FROC curves of the training and test sets. The operating point is indicated on the plot (big markers). The sensitivity and specificity were calculated for only the detections that passed the initial

filter. For polyps>=10mm, the FP rate was reduced from 4.3 to 2.4 at 100% sensitivity in the training set and from 5.6 to 2.4 at 95% sensitivity in the test set. For polyps between 6-9mm, the FP rate was reduced from 3.9 to 2.2 at 74% sensitivity for polyps in the training set, and from 5.8 to 3.1 at 75% sensitivity in the test set.

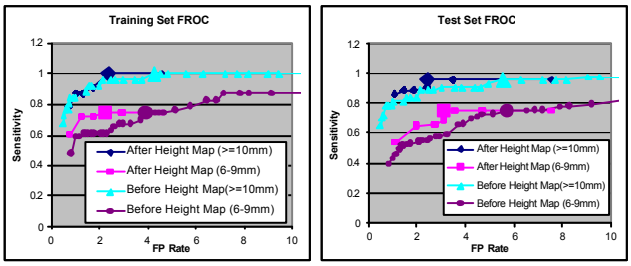

Figure 7. FROC curves

4. DISCUSSION AND CONCLUSION

We proposed a novel technique to distinguish colonic polyps from normal colon surfaces based on the topographical height map. The novelty of this technique is that in addition to using the original 3D CT image, it uses derived height maps which manifest some natural properties of polyps. The technique can reduce false positives by more than 50%, and work well for both big polyps (>=10mm) and medium sized polyps (6-9mm).

One method that is closest to our height map method is the 2D endoluminal projection image technique [6]. 2D endoluminal projection is more realistic compared to height map and mimics the way radiologists interpret the image during a 3D fly-through inside the colon. However, the endoluminal projection has several limitations. Firstly, lighting plays an important role in the image formation and is difficult to control. Secondly, excessive and complex background could confuse the recognition of polyps in the center of the image. Thirdly, some features from the wavelet coefficients are dependent on the in-plane rotation and translation of the projection plane. The height map handles these issues. It does not rely on lighting. An elevation threshold can be applied to eliminate most of the background. In addition to the wavelet features, the height map method uses topographic features based on the directional slopes from the apex to the base of the polyp, which are affine invariant (scale, translation and rotation).

The height map is generated using the ray casting technique. Both orthogonal and perspective projections were explored. We chose an orthogonal projection since it is invariant to the projection distance and can preserve polyp size information, and it also avoids distortion between tip and base of polyp. Only the projection direction needs to be set in the orthogonal projection.

The concentric pattern of the polyp is obvious when the projection direction is parallel to the polyp elevation direction. Using the correct projection direction is essential. There are metrics such as image entropy to maximize the information in the map. However, in our case, more information does not equal stronger differentiating power. The concentric index based on the directional slope is proven to be effective.

In future work, more topographic features such as profile curvatures, ridge lines and incident radiation could be computed and applied in the classification. A more effective concentric index will also be investigated.

5. ACKNOWLEDGEMENTS

The authors thank Perry J. Pickhardt, William R. Schindler and Richard Choi for providing computed tomographic colonography and supporting data. This research was supported by the Intramural Research Program of the National Institutes of Health, Clinical Center.

6. REFERENCES

1. Jemal, A., *et. al.*, *Cancer statistics.* CA Cancer J Clin, 2005. **55**: p. 10-30.
2. Pickhardt, *et. al.*, *Computed Tomographic Virtual Colonoscopy to Screen for Colorectal Neoplasia in Asymptomatic Adults.* the New England Journal of Medicine, 2003. **349**(23): p. 2191-2200.
3. Summers, R.M., *et. al.*, *Colonic Polyps: Complementary Role of Computer-Aided Detection in CT Colonography.* Radiology, 2002. **225**: p. 391-399.
4. Yao, J., *et al.*, *Colonic Polyp Segmentation in CT Colonography Based on Fuzzy Clustering and Deformable Models.* IEEE Trans. Med. Imag., 2004. **23**(11): p. 1344-1352.
5. Paik, D.S., *et. al.*, *Surface normal overlap: A computer-aided detection algorithm, with application to colonic polyps and lung nodules in helical CT.* IEEE Trans. Med. Imag., 2004. **23**: p. 661-675.
6. Li, J., *et. al.*, *Wavelet Method for CT Colonography Computer-Aided Polyp Detection.* in *IEEE ISBI*. 2006. Arlington, VA.
7. http://en.wikipedia.org/wiki/Height_map.
8. Summers, R.M., *et. al.*, *Computed Tomographic Virtual Colonoscopy Computer-Aided Polyp Detection in a Screening Population.* Gastroenterology, 2005. **129**: p. 1832-1844.
9. http://en.wikipedia.org/wiki/Principal_component_analysis.
10. Saff, E.B. and A.B.J. Kuijlaars, *Distributing Many Points on a Sphere.* The Mathematical Intelligencer, 1997. **19**(1): p. 5-11.
11. Lyu, S., *et. al.*, *A digital technique for art authentication.* Proceeding of National Academy of Science, 2004. **101**: p. 17006-17010.
12. Yao, J., *et. al.*, *Optimizing the Support Vector Machines (SVM) Committee Configuration in Colonic Polyp CAD System*. in *SPIE Medical Imaging*. 2005. San Diego, CA.