# Matching Colonic Polyps from Prone and Supine CT Colonography Scans Based on Statistical Curvature Information

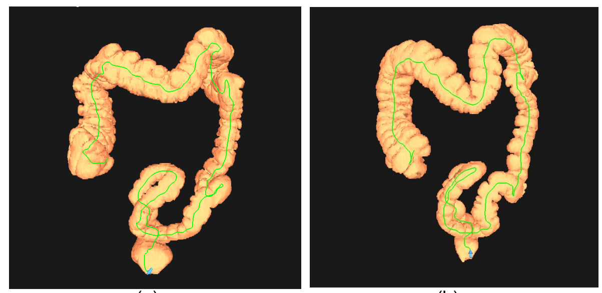
Shijun Wang, Jianhua Yao and Ronald M. Summers*
*Diagnostic Radiology Department, National Institutes of Health, Bethesda, MD 20892, USA*
*{wangshi, jyao}@cc.nih.gov,rms@nih.gov*

## Abstract

*Computed tomographic colonography (CTC) provides a feasible way for the detection of colorectal polyps and cancer screening. In the clinical practice of CTC, a true colonic polyp will be confirmed with high confidence if a radiologist can find it in both the supine and prone scans. To assist radiologists in CTC reading, we propose a new colonic polyp matching method based on statistical curvature information of polyp candidates. We first extract histograms of curvature-related features (HCF) from each polyp candidate, then use diffusion map to embed the original high dimensional data into a low-dimensional space. Experimental results show that by using our HCF method, we can improve the sensitivity from 0.58 to 0.74 at false positive rate 0.1 compared with a traditional method that uses only means of curvature-related features.*

## 1. Introduction

Computed tomographic colonography (CTC) is a feasible and minimally invasive method for the detection of colorectal polyps and cancer screening. When CTC is performed in conjunction with computer aided detection (CAD) software, the screening becomes easier, less time-consuming, and more accurate for the radiologist [1,2]. In current practice, a patient will be scanned twice - once supine and once prone (Fig. 1) which improves the sensitivity by reducing extent of uninterpretable collapsed or fluid-filled segments. In order to assist the radiologist in evaluating colon polyp candidates detected by the CAD in both scans, it would be helpful to provide not only the locations of suspicious polyps, but also the possible matched pairs of polyps in supine and prone scans.



(a) (b)

Figure 1. Two typical three-dimensional colon CTC surface reconstructions of a male patient of 50 years old. (a) Prone scan and (b) supine scan. Green lines inside the segmented colon indicate centerlines.

In recent years, feature extraction based on statistical information of basic image descriptors showed promising results in many image processing and computer vision applications. The basic image descriptors include edge descriptor, gradient descriptor, etc. Representative methods include Scale-Invariant Feature Transform (SIFT) [3], Shape Context [4] and Histograms of Oriented Gradient (HOG) descriptor [5], etc. The SIFT descriptors focus on the local appearance of the object at particular interesting points [3]. By using orientation assignment at key points and matching of orientation histograms, the SIFT features are robust to changes of illumination and viewpoint and to the presence of noise and occlusion. Shape Context [4] describes the shape of an object by using log-polar bins to capture the relative location distribution of other edge points in relation to the central edge point. The key idea of Shape Context is that, randomly choosing one key point, the distribution of relative positions of other key points to the chosen key point is a robust, compact and highly discriminative descriptor that can handle non-rigid transformation. Histogram of Oriented Gradient descriptors, or HOG descriptors, count occurrences of gradient orientation in localized portions of an image [5]. For the detection of colonic polyps, CT images have lower resolutions compared with that of optical

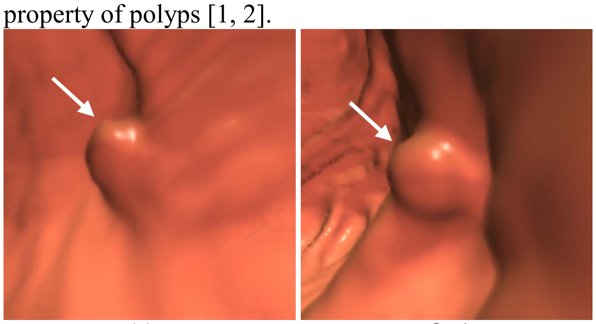
978-1-4244-2175-6/08/$25.00 ©2008 IEEE

images. Consequently, SIFT and Shape Context are not suitable descriptors because they rely on key or edge points. Usually in CT images there are no distinct local structures which can serve as markers. HOG does not need to locate key points but it is sensitive to image rotation.

Inspired by the idea of HOG, in this paper we propose a set of new shape descriptors called Histograms of Curvature Features (HCF) and use it to match polyp candidates from different views. HCFs are statistical descriptors that can capture the intrinsic properties of true polyps. We utilize curvature-related descriptors as basic image descriptors, i.e. shape index, curvedness, Gaussian and mean curvatures, etc. The advantage of these curvature related descriptors is that they are rotation, translation and scale invariant. After feature extraction, non-linear dimensionality reduction is applied to HCF features to find the intrinsic dimensions of polyp pair candidates.

## 2. Histograms of Curvature Features

Colonic polyps appear as bulbous protrusions that adhere either to the inner wall of the colon or to colonic folds that have elongated, ridgelike structures. Fig. 2 shows a typical colonic polyp from prone and supine scans. How to extract distinct features from polyp candidates is a key factor in the detection of colonic polyps. Perhaps the most widely used features to describe polyp candidates are Gaussian and mean curvatures because they can capture semispherical property of polyps [1, 2].



(a) Supin

Figure 2. 3D endoluminal surface reconstructions of an 8 mm adenomatous polyp (arrows). (a) Prone and (b) supine scans.

Intuitively, curvature measures the extent that a geometric object deviates from flat. For a two-dimensional iso-surface embedded in $R^3$, the intersection of the surface with a plane containing the normal vector and one of the tangent vectors at a point on the surface is a plane curve and has a curvature called normal curvature. The maximum and minimum values of the normal curvature at a point are called the principal curvatures, $k_1$ and $k_2$. The directions of the corresponding tangent vectors are called principal directions. The Gaussian curvature is defined as the product of the principal curvatures: $k_{Gaussian} = k_1 k_2$. A surface is locally convex when Gaussian curvature is positive; it is locally saddle when the Gaussian curvature is negative. The mean curvature is one-half of the sum of the principal curvatures: $k_{mean} = \left(k_1 + k_2\right)/2$. Besides Gaussian and mean curvatures, shape index (SI) and curvedness (CV) can also describe the shape of a polyp candidate [6]. At a given voxel $p$, the SI and CV features can be defined as follows: $SI(p) = \frac{1}{2} - \frac{1}{\pi}\arctan\frac{k_1(p)+k_2(p)}{k_1(p)-k_2(p)}$, $CV(p) = \sqrt{\left(k_1^2(p)+k_2^2(p)\right)/2}$, where $k_1(p) > k_2(p)$ are two principal curvatures.

Traditional colonic polyp detection methods focus on the mean of the surface points' curvatures of a polyp candidate [1,2]. To make full use of curvature information and capture internal texture information of polyp candidates, we utilize histograms of curvature features to represent polyp candidates. In Table 1, we list six curvature-related features used in our HCF method. For each feature, we choose a range and divide it into 98 equally-spaced bins. Voxels whose feature values are smaller than the lower limit or larger than the upper limit are counted in two additional bins. For curvedness, Gaussian curvature and mean curvature, the binning is done on the natural logarithm value of them. We used absolute value of negative curvatures for logarithm calculation and shifted logarithm values of positive curvatures with +20 to distinguish them from negative curvatures. In addition, previous research shows that the CT value is also an informative feature [1]. We also compute the histogram of CT values for each polyp candidate. We concatenate the seven histograms and get a feature vector with 700 dimensions for each polyp candidate.

In Fig. 3 we show comparisons of a true polyp pair and a false polyp pair. From it we can find that the true polyp shows similar histograms of SI and CV whereas a false pair will have different statistical features.

## 3. Dimensionality reduction by diffusion map

By using Histograms of Curvature Features and CT values, we can extract about 1400 features from a polyp candidate pair. To reduce the dimensions of polyp candidate pair, in this paper we employed diffusion map (DM) algorithm because of its high resistance to noise. DM is a non-linear dimensionality reduction method based on the Markov random walk on the data [7]. Assume that a system contains $n$

samples $X = \{x_1, x_2, ..., x_n\}$. The weight function $w(x,y)$ depicts the distance relationship between samples $x$ and $y$. The transition probability of going from node $x$ to $y$ in one step is $p_1(x,y) = w(x,y)/d(x)$, where $d(x) = \sum_{z \in X} w(x,z)$ is called the degree of node $x$. The one step Markov transition matrix $P$ has a set of right eigenvectors corresponding eigenvalues $1 = |\lambda_0| \geq |\lambda_1| \geq ... \geq |\lambda_{n-1}|$ : $P\psi_j = \lambda_j \psi_j$. The diffusion map is defined as

$$\Psi_t : x \mapsto \left(\lambda_1^t \psi_1(x), \lambda_2^t \psi_2(x), ..., \lambda_{q(t)}^t \psi_{q(t)}(x)\right),$$

where $t$ represents the steps of Markov random walk on the graph and $q(t)$ is the number of dimensions of embedding space.

Table 1. Six curvature-related features used in our HCF descriptor. Lower limits, upper limits and numbers of bins are listed for each feature and CT value. These limits are selected according to the distributions of features.

| | Lower limit | Upper limit | Number of Bins |
|---|---|---|---|
| Shape Index | 0 | +1 | 100 |
| Curvedness | -9 | +5 | 100 |
| Gaussian Curvature | -15 | +25 | 100 |
| Mean Curvature | -10 | +30 | 100 |
| Max Curvature | -0.4 | +0.6 | 100 |
| Min Curvature | -0.8 | +0.2 | 100 |
| CT Value | 0 | 1500 | 100 |

# 4. Centerline calculation

To reduce the number of false pairs, the position of polyp candidates along the centerline of the colon is used for the matching of polyp candidates between supine and prone images. We used a subvoxel precise centerline extraction method to determine the polyp's position along the centerline [8]. Initially, segmentation of the colon is performed to obtain a subvoxel precise representation of the colon. The discrete segmentation is used as an initial surface for a narrow band level set segmentation to more accurately determine the location of the colon inner wall and smooth the boundary between the air and fluid-filled regions of the colon. From the level set segmentation, a subvoxel precise distance field is computed using the fast marching method. The centerline of a colon is computed based on the distance field. A polyp's position along the centerline then can be computed by finding the minimum distance between the polyp's position and the points along the centerline.

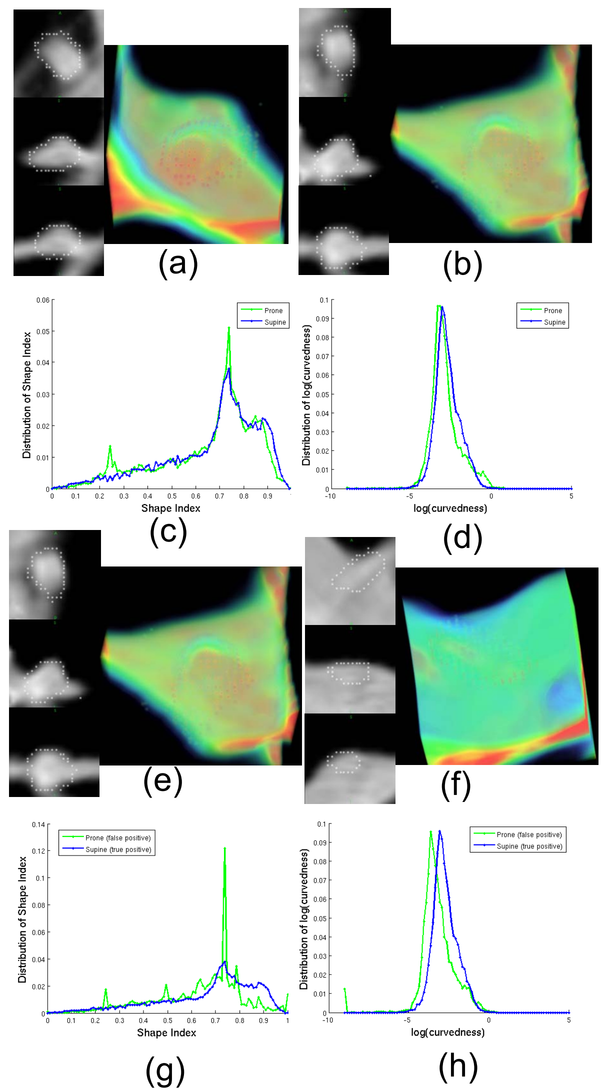



Figure 3. (a) and (b) show a 10 mm true polyp with 3D endoluminal CTC volumetric reconstructions. Their distributions of shape index and curvedness are shown in (c) and (d) respectively. (e) and (f) show a true polyp and a false positive. Their distributions of shape index and curvedness are shown in (g) and (h) respectively. For each polyp candidate, we show its axial, frontal and sagittal slices in which white dots mark the boundary of the candidate.

# 5. CTC data

Our dataset consists of CTC examinations of 37 patients collected from three medical centers. Each patient was scanned in the supine and prone positions. Each scan was done during a single breath hold using a 4-channel or 8-channel CT scanner. CT scanning parameters included 1.25- to 2.5-mm section collimation, 15 mm/s table speed, 1-mm reconstruction interval, 100 mAs, and 120 kVp.

The dataset of matched polyp candidates is constructed as follows: For each polyp candidate, a vote or prediction value is assigned. The vote value (in range [0, 1]) comes from a support vector machine (SVM) committee classifier [9]. Two polyp candidates from different scans of the same patient are matched if the difference between their normalized centerline positions is within 0.1. But only candidates coming from the same physical true polyp are regarded as a true pair. Other pairs are treated as false pairs. Here we just consider the false polyps whose SVM votes are greater than 0.4 because they look very much like true polyps and may confuse radiologists. We got 984 candidate pairs that included 19 true adenomatous polyp pairs whose sizes are in the range of 6 to 30 mm. Then we used summation and absolute difference of HCF features of two members of a polyp pair candidate as the feature vector for the pair.

## 6. Experimental results

To evaluate the effectiveness of our method for classification of colonic polyps, we compared the performance of SVM with HCF, HOG and means of curvature features (MCF) which are widely used in colonic polyp detection [1]. The experimental settings are as follows: for each test we randomly selected 9 true pairs and 9 false pairs which compose the training set. Other polyp candidates were treated as testing samples. After dimensionality reduction, we reduced the 1400 HCF features to 15 dimensions. For HOG descriptor, the optimal parameters were found to be 2x2x2 cell blocks of 8x8x8 pixel cells with 6x4 histogram channels. Fig. 4 shows the FROC curves on the testing sets. Each point on the FROC is the average result of 100 random tests. For clinical purposes, radiologists are mostly concerned about the sensitivity of the CAD system at low false positive rate. When the false positive rate is 0.1, our HCF method achieves 0.74 in average sensitivity, whereas the average sensitivity is 0.58 when only means of curvature features are used ($p < 0.01$). The areas under the ROC curves (AUC) of HCF, HOG and MCF are 0.91, 0.85 and 0.81 respectively. t-test shows that there is significant difference between the AUCs of HCF and HOG ($p < 0.01$).

## 7. Conclusion

In this paper, we propose a new method for colonic polyp matching. Our HCF method utilizes the distributions of various curvature-related features inside polyp candidates. In order to capture the intrinsic dimensions of the high-dimensional HCF feature, DM is employed after feature extraction. Experimental results on a CTC dataset of 37 patients show that the HCF method is superior to the traditional method which just uses the means of curvature-related features.

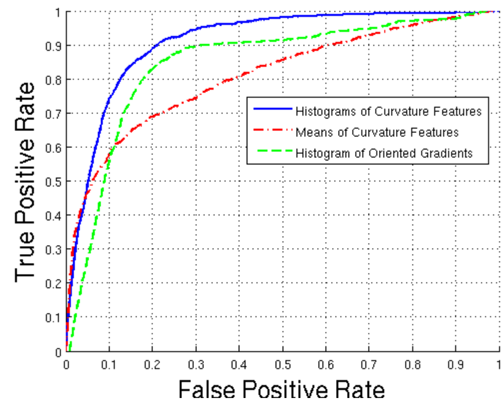



Figure 4. Comparisons of ROCs of HCF method, means of curvature features method and HOG.

**Acknowledgement**: This research was supported by the Intramural Research Program of the NIH Clinical Center.

## References

[1] R.M. Summers, et al., Automated polyp detection at CT Colonography: Feasibility assessment in a human population, Radiology 219, 51–59, 2001.

[2] R.M. Summers, et al., Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population, Gastroenterology, 129, 1832-1844 (2005).

[3] D. Lowe. Distinctive image features from scale-invariant keypoints. Internation Journal of Computer Vision (IJCV), 60(2):91–110, 2004.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape Matching and Object Recognition Using Shape Contexts. IEEE Trans. PAMI 24 (24): 509-521, 2002.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. IEEE Conference on Computer Vision and Pattern Recognition, 886-893, 2005.

[6] H. Yoshida and J. Nappi, Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps, IEEE Trans. Med. Imaging 20, 1261–1274, 2001.

[7] R.R. Coifman, S. Lafon, A.B. Lee, M. Maggioni, B. Nadler, F. Warner and S.W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data. Proceedings of the National Academy of Sciences, vol. 102(21), May 2005.

[8] R.L. Van Uitert and R.M. Summers, Automatic Correction of Level Set Based Subvoxel Precise Centerlines for Virtual Colonoscopy Using the Colon Outer Wall, IEEE Trans. Medical Imaging, 26, 1069-1078 (2007).

[9] J. Yao, R.M. Summers, and A.K. Hara, Optimizing the committee of support vector machines (SVM) in a colonic polyp CAD system, In Medical Imaging 2005 of SPIE (2005).