Eurographics Workshop on Visual Computing for Biomedicine (2008)
C. P. Botha, G. Kindlmann, W. J. Niessen, and B. Preim (Editors)

Surface Curvature Line Clustering for Polyp Detection in CT Colonography

Lingxiao Zhao[1], Vincent F. van Ravesteijn[2], Charl P. Botha[1], Roel Truyen[4], Frans M. Vos[2,3] and Frits H. Post[1]

[1]Data Visualization Group and [2]Quantitative Imaging Group, Delft University of Technology, Delft
[3]Dept. of Radiology, Academic Medical Center, Amsterdam
[4]Philips Healthcare, Best
The Netherlands

Abstract

Automatic polyp detection is a helpful addition to laborious visual inspection in CT colonography. Traditional detection methods are based on calculating image features at discrete positions on the colon wall. However large-scale surface shapes are not captured. This paper presents a novel approach to aggregate surface shape information for automatic polyp detection. The iso-surface of the colon wall can be partitioned into geometrically homogeneous regions based on clustering of curvature lines, using a spectral clustering algorithm and a symmetric line similarity measure. Each partition corresponds with the surface area that is covered by a single cluster. For each of the clusters, a number of features are calculated, based on the volumetric shape index and the surface curvedness, to select the surface partition corresponding to the cap of a polyp. We have applied our clustering approach to nine annotated patient datasets. Results show that the surface partition-based features are highly correlated with true polyp detections and can thus be used to reduce the number of false-positive detections.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation Line and curve generation

1. Introduction

CT colonography is a minimally invasive technique that enables effective screening of the large intestine [vGVS*02, Bar05, KPT*07]. The main task is to identify adenomatous colonic polyps, which are well-known precursors of colon cancer. Typically, colonic polyps are visible as protrusions on the interior colonic surface. In virtual CT colonography, the iso-surface of the colon wall is first extracted as a triangle mesh, or as an implicit iso-surface using volume ray casting techniques. Unfortunately, a complete visual inspection of the colon wall is rather time consuming and important areas are inevitably missed. To maintain high effectiveness and efficiency, computer-aided diagnosis (CAD) has been proposed as a helpful addition to the CT colonography pipeline.

The majority of existing CAD schemes [YN01, SYP*05] for automatic polyp detection are based on calculating image features, e.g. surface principal curvatures, at discrete points on the colon wall. These image features are then filtered using hysteresis thresholding to identify suspected regions that belong to colonic polyps. Polyp candidates are detected by fuzzy clustering of these connected components. Such approaches mainly focus on point-wise surface shape characteristics. Large-scale shapes of the colonic surface are not captured. Point-sampling features are in general more sensitive to CT data noise and surface irregularities. Aggregation of shape information within a certain area is necessary to enhance the robustness of polyp surface characterization.

Instead of calculating point-sampling features on the colon wall, we first partition the colonic surface into regions that exhibit consistent behaviour with respect to surface geometry. We perform this partitioning by clustering surface-constrained curvature lines. Curvature lines are curves everywhere tangent to one of the two principal curvature direction vector fields on the surface. Geometric features are then computed within each surface partition, over the clustered curvature lines, and used to classify these surface partitions into various predefined types. These features are also used to identify surface partitions that contain polyps.

We integrated our colonic surface partitioning and aggregated surface shape feature calculation with an existing CAD pipeline. Polyp candidates are pre-detected using an existing automatic polyp detection scheme [vWvRV*06], which yields all true polyp detections for the data set we used, as confirmed by an experienced radiologist, and also a large number of false detections. We make use of techniques from [ZBB*06] to generate surface-constrained curvature lines in the area around each detected candidate position. Then each candidate area is partitioned into geometrically homogeneous regions by clustering curvature lines. New features are derived as aggregation of surface shape information within each surface partition. These features are integrated at curvature line points of each cluster that covers a surface partition. Our experimental study showed that these features can be used to discriminate between clusters that contain polyps and clusters that do not.

Our contribution is two-fold:

- We present surface-constrained curvature line clustering as a new method for surface partitioning in polyp detection. This idea can also be applied in other problems where larger scale surface shapes have to be captured and analyzed.
- We show that features calculated over surface partitions could be used to discriminate effectively between polyps and non-polyps.

The remainder of the paper is structured as follows. Related research is briefly reviewed in Section 2. Section 3 explores the relation between curvature line clustering and colonic surface partitioning. In Section 4, our curvature line clustering method is described. We discuss how to compute geometric features within each surface partition in Section 5. An experimental study to analyze the discriminative power of these features is given in Section 6. We draw conclusions in Section 7, as well as outline our future work.

2. Related Work

Many existing polyp detection schemes make use of surface shape features derived from scalar principal curvatures, which are important concepts in differential geometry. It was shown that the per patient sensitivity of computer-aided polyp detection in an asymptomatic screening population is comparable to that of optical colonoscopy for polyps of 8 mm or larger and is generalizable to new CT colonography data in [SYP*05]. Yoshida et al. [YN01] developed a method that started by computing volumetric shape index and curvedness to characterize polyps, folds and colonic walls at each voxel in the extracted colon. Afterwards, polyp candidates were obtained using fuzzy clustering to connected components filtered based on these geometric features. Huang et al. [HSH05] introduced a two-stage curvature estimation approach based on cubic spline fitting on triangulated surface meshes for automatic polyp detection.

There are also methods not or indirectly based on scalar surface curvatures. Hong et al. [HQK06] presented an automatic polyp detection pipeline that integrated texture and shape analysis with volume rendering and conformal colon flattening. Van Wijk et al. [vWvRV*06] proposed a shape and size invariant approach in which a "protrusion"measure was used to find polyp candidates. In [ZBB*06], the potential of surface principal curvature direction fields for automatic polyp detection was explored. Techniques were developed for computing surface-constrained curvature lines on the colon wall. Features strongly correlated with true positive polyp detections are derived from the geometry of curvature lines and used to reduce the number of false positive detections in a complete polyp detection protocol.

Techniques have been developed for surface partitioning. Based on Morse-Smale complex, Natarajan et al. [NWB*06] proposed a method using topological analysis of a scalar function defined on the surface. Partition-based techniques are also widely-used in flow visualization. A survey article on this topic was recently published by Salzbrunn et al. [SJWS08]. Chen et al. [CML*07] provided a technique for vector field modification.

Many line clustering methods can be found in the literature. In flow visualization, Chen et al. [CCK07] presented a streamline similarity distance metric that considered not only Euclidean distance but also shape and directions. A well-known application is the clustering of fibers to obtain bundles in diffusion tensor imaging (DTI) [MVvW05]. An important issue of clustering fibers (or streamlines) is to define a good similarity (or distance) metric between fibers. Brun et al. [BKP*04] mapped fibers to a Euclidean feature space and used a Gaussian kernel for pairwise fiber comparison. They proposed a clustering method in which a normalized cut criterion was used to partition a weighted undirected graph derived from their fiber similarity measure. Corouge et al. [CGG04] computed closest point pairs to define distance between fibers. Based on this distance measure, they propagated cluster labels from fiber to neighboring fibers. Klein et al. [KBL*07] calculated an affinity matrix based on the use of a reconstructed 3D grid to represent fiber similarity. The number of clusters were automatically determined by performing a linear eigenvalue regression of this affinity matrix. Brain white matter fibers were clustered using a spectral clustering method based on multiple eigenvectors.

3. Colonic Surface Partitioning

Point-sampling features are not capable characterizing large-scale surface shapes, which are necessary to identify polyps. It is important for automatic polyp detection schemes to aggregate surface shape information within a certain region to avoid localization. This requires a partitioning of the colonic surface before feature calculation.

By applying curvature lines [ZBB*06], colonic surface

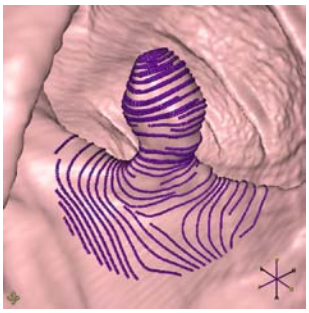


Figure 1: *Maximum curvature lines on a polyp surface.*

shape is well outlined (Figure 1). We calculate principal curvature values and directions on implicit iso-surfaces using the method developed by van Vliet et al. [vVYV98]. This method is based on the eigen analysis of Hessian matrix, of which entries are second order partial derivatives. Gaussian convolution is used to compute these 3D image derivatives. We choose $\sigma = 2.0$ for the Gaussian kernel with regard to iso-surface smoothing and computational efficiency. Curvature lines are traced directly in 3D volume and constrained on the colonic iso-surface using an iso-projection approach. The seeding and spacing distance is determined by principal curvature values. In rare cases, there are umbilics on the colonic iso-surface. Curvature line integration stops at these isotropic points.

We use principal curvature direction vectors to cluster coherent surface shapes for surface partitioning. For arbitrarily close positions on the surface, they have coherent shapes when their principal curvature directions are sufficiently parallel. Otherwise, even their shape types are analogical, their shapes are not coherent and belong to different structures. By definition, curvature lines are curves that everywhere follow one principal curvature direction field. It is intuitive that parallel curvature lines link coherent shapes over a surface area and demonstrate meaningful structures. Such a structure presents homogeneous geometry since its embedded shapes are coherent. This leads to the idea that the colonic surface can be partitioned by grouping parallel curvature lines.

In our surface partitioning approach, each partition corresponds with the surface area covered by a single group of parallel curvature lines. Thus the surface partitioning problem is converted into a curvature line clustering problem.

4. Clustering Curvature Lines

This section decribes our curvature line clustering method for colonic surface partitioning, using a new curvature line similarity measure and a multiple-eigenvector spectral clustering algorithm.

4.1. Symmetric Curvature Line Similarity Metric

A number of clustering algorithms exist, for example agglomerative hierarchical clustering and spectral clustering. By using different similarity metrics between clusters, variations are devised for the same clustering method. In our work, a similarity measure based on parallelism between curvature lines is desired. The Euclidean distance should of course also contribute to the similarity.

We define a symmetric line-to-line similarity measure in our application, based on the work by Chen et al. [CCK07]. As a symmetric measure, the similarity between two curvature lines is unchanged no matter it is calculated from one curvature line to the other or the other way around. In our definitions, the similarity between two curvature lines is always calculated from the shorter line to the longer line.

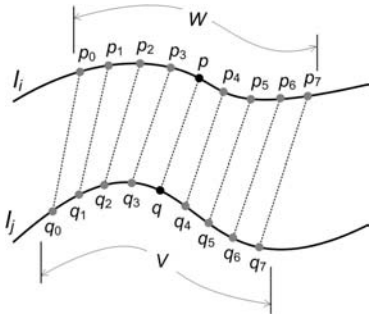


Figure 2: *Similarity measure between curvature lines l_i and l_j: point pairs are formed by corresponding window w and v. Solid black dots indicate window centers and grey dots indicate point pairs. This similarity is direction-independent.*

For a point p on a curvature line l_i, a window W of size $m+1$ centered at p is computed. On both sides of point p, $\frac{m}{2}$ neighboring points of p on l_i are included in W. W is a portion of l_i. On the other curvature line l_j, we find the point q, closest to p in an Euclidean sense. Another window V of size $m+1$ is centered at q on l_j. Thus we yield $p_0, p_1, \cdots, p_{m-1}$ in window W about p and $q_0, q_1, \cdots, q_{m-1}$ in window V about q. A correspondence between these two windows W and V is defined such that p_i and q_i $(i = 0, 1, \cdots, m-1)$ form one point pair. There are now $m+1$ point pairs including the center point pair $\{p, q\}$. The similarity distance d_{sim} from point p to curvature line l_j is computed as:

$$d_{sim} = \|p-q\| + \alpha \frac{\sum_{i=0}^{m-1} (\|p_i - q_i\| - \|p-q\|)}{m} \quad (1)$$

This computation is described in Figure 2. d_{sim} is computed from every point on curvature line l_i to curvature line l_j. Then the average value of d_{sim} is taken as the overall similarity distance d_{ij} from l_i to l_j. Since we only consider the parallelism and Euclidean distance between curvature lines, note that this similarity distance is direction-independent, as described in Figure 2. This is different from the similarity distance defined by Chen et al. [CCK07].

Larger similarity distance d_{ij} indicates less similarity between curvature lines l_i and l_j. This distance is transformed to a similarity measure using:

$$S_{ij} = \exp(-d_{ij}) \quad (2)$$

This similarity measure can be affected by both the Euclidean distance between curvature lines as well as their parallelism. A larger Euclidean distance increases the term $\|p-q\|$ in Eq.1. This leads to a smaller similarity measure. A larger deviation of point pair distances within the window from the center point pair distance indicates a less parallelism between two curvature lines. This increases the second term in Eq.1 and results in a smaller similarity. The coefficient α is a scale factor that can be used to strengthen the effect of parallelism. The window size m also influences this similarity measure. A zero window size will result in a similarity metric purely based on the Euclidean distance. In our experience, $\alpha = 2.0$ and $WindowSize = 20$ yielded the best result as suggested by Chen et al. [CCK07].

4.2. Spectral Clustering Algorithm

We applied our new similarity metric described in Section 4.1 in a spectral clustering algorithm [KBL*07] to cluster curvature lines. There are also alternative hierarchical clustering methods [CHH*03]. A disadvantage of hierarchical clustering is that the number of clusters is application dependent and sometimes difficult to determine.

The algorithm proposed by Klein et al. [KBL*07] uses a linear eigenvalue regression technique to compute a reasonable number of clusters. It takes an affinity matrix as input. Based on our new curvature line similarity metric, the affinity matrix M is computed as:

$$\begin{pmatrix} S_{0,0} & S_{0,1} & \cdots & \cdots & S_{0,N-1} \\ S_{1,0} & \ddots & & & \vdots \\ \vdots & & S_{i,j} & & \vdots \\ \vdots & & & \ddots & \vdots \\ S_{N-1,0} & \cdots & \cdots & \cdots & S_{N-1,N-1} \end{pmatrix} \tag{3}$$

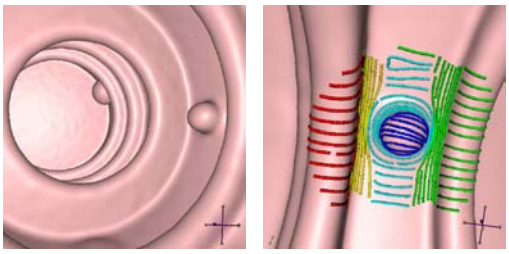
where N is the total number of curvature lines generated on the colon wall at the candidate position and in its vicinity. The entry $S_{i,j}$ $(i,j \in \{0,1,\cdots,N-1\})$ of the affinity matrix is the similarity measure defined by Eq. 2 between curvature lines l_i and l_j. Since by definition $S_{i,j} = S_{j,i}$, this affinity matrix is symmetric. It has N eigenvalues and eigenvectors.

The affinity matrix M is normalized by:

$$S_{i,j} = \frac{S_{i,j}}{S_{max}} \tag{4}$$

where S_{max} is the maximum value of entries in matrix M, except diagonal items. These diagonal items are set to 1. Eigenvalues of the normalized matrix $\hat{M}$ are indexed and plotted against their indices in descending order. At each index, these eigenvalues are splitted. The resulting two halves are fitted with two linear functions. The index at which the smallest fitting error occurs corresponds with the number of clusters. This number K enables the identification of inner clusters within coarse structures.

The K largest eigenvectors of $\hat{M}$ are selected and assembled as columns in a $N \times K$ matrix X. Rows of X are considered as K-dimensional vectors. Each of these vectors corresponds with a curvature line with the same index. A hierarchical clustering method [MVvW05] is performed on these N K-dimensional vectors to create K clusters. The indices of rows of X in each resulting cluster correspond to the indices of curvature lines that belong to the same cluster. In our experiments, using the complete linkage in the hierarchical clustering provided the best result.



(a) (b)

Figure 3: *Clustering curvature lines: (a) A synthetic colonic surface; (b) Curvature line clusters on a fabricated polyp, different colors represent different clusters.*

There are two advantages of this spectral clustering algorithm. First, a reasonable number of clusters is autmatically determined, considering the inter-cluster connectivity. Second, using multiple eigenvectors for clustering leads to more accurate and robust results [AKY99]. Figure 3 shows the result of our curvature line clustering algorithm on a synthetic 3D volume data. The simulated colonic surface is rendered as an implicit iso-surface using volume ray casting. Curvature lines that follow maximum curvature direction are traced. Parallel curvature lines are clustered into several groups, each of which corresponds with a surface partition. This clustering algorithm can also be used in other applications, e.g. clustering white matter fibers to find bundles in DTI of the human brain.

5. Geometric Features for Surface Shape Analysis

In this section, we describe how geometric features are calculated to characterize the large-scale shape of each surface partition. These features are used to identify the surface partition that captures a colonic polyp.

To describe the overall shape of a surface area, a traditional way is to calculate the average or mean values of geometric features in most existing automatic polyp detecion approaches. Unfortunately, important surface shape information could be neglected in such a way. Therefore we need to aggregate surface shape information as much as possible within the surface partition for shape analysis.

Volumetric shape index and curvedness of iso-surfaces are two well-known features used for automatic polyp detection. They were firstly introduced by Yoshida et al. [YN01] to

CAD schemes in CT colonography. These two features are derived from surface principal curvatures:

$$SI = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_{min} + k_{max}}{k_{min} - k_{max}} \quad (5)$$

and

$$CV = \sqrt{\frac{k_{min}^2 + k_{max}^2}{2}} \quad (6)$$

where k_{min} is the minimum curvature value and k_{max} is the maximum curvature value. However, *SI* and *CV* were mostly used as point-sampling measures for shape analysis and large-scale shapes were difficult to be described in existing approaches. We define a new feature based on the aggregation of *SI* and *CV* at curvature line points corresponding to a surface partition to characterize its overall shape.

5.1. Feature Definition

Curvature line points of a cluster are treated as sampling points of the corresponding surface partition. To identify the overall shape of a surface partition, surface shape information is aggregated through these sampling points. *SI* and *CV* are calculated and aggregated within each surface partition.

The most important feature is the shape index *SI*. A unique value of *SI* describes a distinct surface shape [YN01]. The value of *SI* varies in the closed interval $[0.0, 1.0]$. Five predefined shape types are represented by their characteristic shape index values (Table 1). These values divide the shape

Table 1: *Five predefined shape types and their corresponding shape index values.*

Shape type	Shape index value
Cup	0.0
Rut	0.25
Saddle	0.5
Ridge	0.75
Cap	1.0

index value range into four subintervals. Values that fall into one of these subintervals represent transitional shapes from one predefined shape type to another. Such a shape transition is continuous over the whole domain of *SI*.

The iso-surface curvedness *CV* is a complementary feature to the shape index *SI*. The shape index only measures the shape type of a surface. The curvedness can represent how significant the surface shape is. The value of *CV* varies in the interval $[0.0, +\infty)$, where a small value implies a more flat surface and a large value implies a sharp edge or peak. Using the mean value of *SI* and *CV* as an aggregation for surface analysis has a significant drawback. For example, if a surface partition has most of its sampling points where *SI* values fall into the intervals $[0.0, 0.25]$ and $[0.75, 1.0]$, the mean value of its *SI* is close to 0.5. Thus the dominant shape

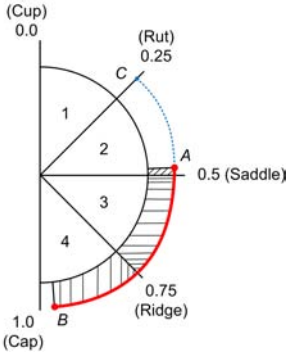


Figure 4: *Using SI subintervals to characterize the overall shape of a surface partition: a surface partition has its SI value range a (red line) overlapped with 2nd, 3rd and 4th SI subintervals. In this case, the surface partition has overall cap-like shapes.*

of this surface partition is represented as the saddle, which is obviously incorrect.

We define new features, which account not only for values of *SI* and *CV* but also for the shape types at sampling points over the surface partition. This definition is demonstrated in Figure 4. Values of *SI* over a surface partition are mapped to the four subintervals of the *SI* value domain. Assuming a geometrically homogeneous surface partition has analogical shapes over it in general, we intend to use shape subintervals, in which shapes described by the majority of sampling points are included, to characterize the overall shape of a surface partition. Values of *SI* and *CV* are calculated through curvature line points of the single cluster that corresponds with a surface partition. The value ranges *a* and *b* of these discrete *SI* and *CV* values can be determined. Then we compute four overlaps of range *a* with the four subintervals of *SI* (Figure 4):

$$F_i = \frac{n_i}{\overline{N}} \times 100\%, (i = 0, 1, 2, 3) \quad (7)$$

where n_i is the number of curvature line points that have their *SI* values included in the *i*th *SI* subinterval. $\overline{N}$ is the total number of curvature line points that belong to the cluster corresponding to the surface partition. If there is no overlap of *a* with a *SI* subinterval, then a negative value is calculated as:

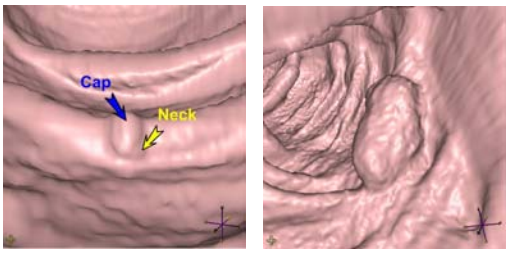
$$F_i = -\frac{\|C - A\|}{\|A - B\|} \times 100\% \quad (8)$$

where *A* and *B* are the two values at the border of range *a* and *C* is the closest value of the non-overlapped *SI* subinterval to *a*. In the case shown in Figure 4, *C* and *A* are the closest domain border values. The overlap of range *b* with a proper value domain is also computed for the iso-surface curvedess of the surface partition using similar functions. This proper *CV* value domain is determined as suggested by Yoshida et al. [YN01]. Reasonably curved surface partitions are significant for shape analysis in our automatic polyp detection

scheme, whereas approximately flat regions are much less meaningful. For finding surface significant overall shapes, the *CV* value domain is $[0.08, +\infty)$. We compute a percentage F_{CV} of the number of points with their *CV* values of at least 0.08 against the total number of points to characterize the general curvedness of a surface partition. Now we have a 5-dimensional vector feature $F = \{F_0, F_1, F_2, F_3, F_{CV}\}$ for each surface partition. It can be used to identify the surface partition that captures a polyp surface.

5.2. Shape of Colonic Polyps

The vector feature defined in Section 5.1 has four *SI* components and one *CV* component. If a *SI* component of this vector has a larger value than other *SI* components, the overall shape of the corresponding surface partition is described by that shape subinterval. As shown in Figure 4, the *SI* range *a* of an example surface partition has larger overlaps with the 3rd and 4th *SI* subintervals, while just a very small overlap with the 2nd *SI* subinterval and no overlap at all with the 1st *SI* subinterval. Its feature vector *F* has a negative 1st component and a small and positive 2nd component. Its 3rd and 4th components have larger positive values. This indicates that the overall shapes are ridge and cap-like.



(a) (b)

Figure 5: *(a) A polyp surface consists of a cap and a neck. (b) Irregular shapes of a large polyp.*

In most current CAD schemes, colonic polyps are usually modeled as approximately spherical or ellipsoidal protrusions. A polyp surface consists of the head and the neck. As shown in Firgure 5, the polyp head is a cap-like surface while the polyp neck is an anticlastic (or saddle) surface. Since irregularities are included on most significant polyps, sometimes not all points on a polyp head present a cap shape. Ridge-like or even saddle-like shapes could also be presented on the polyp head. The polyp neck presents saddle-like shapes in general. Therefore, overall shapes on a polyp surface have significant overlaps with the 3rd and 4th *SI* subintervals. As discussed by Yoshida et al. [YN01], the vector feature *F* for a surface partition that captures a polyp surface also has a significant overlap of its *CV* range *b* with the interval of $[0.08, +\infty)$.

Summarizing these issues, we hypothesize that overall shapes of a polyp are mainly included in the 3rd and 4th *SI* subintervals, i.e. these overall shapes correspond to *SI* values in $[0.5, 0.75]$ and $[0.75, 1.0]$. Importantly, the *SI* component F_3 has a larger value to represent the polyp head. Otherwise, the corresponding surface partition does not capture the polyp cap. We calculate $\{F_0, F_1, F_2, F_3, F_{CV}\}$ for each surface partition of a pre-detected polyp candidate. If the sum of its 3rd and 4th *SI* components (F_2, F_3) is larger than the value of F_0 or F_1, and F_3 is significant by itself, such a surface partition is considered to correspond with a polyp surface. On the other hand, its *CV* component should be able to indicate that it is a sufficiently curved surface partition. For each polyp candidate, we pick one of its surface partitions that most likely captures a polyp surface. The vector feature *F* of this partition is used to represent corresponding polyp candidate. We use this feature to discriminate between true polyps and false detections.

6. Results

This section documents the results of an experimental study in which our surface partition-based features were tested on nine real patient data sets.

6.1. Materials and CAD approach

Our patients underwent CT colonography before optical colonoscopy, which is considered as the golden standard. Each patient has two CT scans in prone and supine directions. The average resolution of these CT data is $512 \times 512 \times 267$ and the average voxel size is $0.77\text{mm} \times 0.77\text{mm} \times 1.60\text{mm}$.

Our automatic polyp detection system consists of three fully automatic steps. First an existing approach [vWvRV*06] was used to pre-detect polyp candidates. These candidates included all true polyps and a large number of false detections. In second step, we took pre-detected candidate positions as the input and applied our approach to calculate surface partition based features for the vicinity of each candidate. The thrid step is supervised pattern recognition, by which these features were processed and the number of false detections was reduced.

6.2. Colonic Surface Partitioning

Curvature lines are directly traced in 3D volume data and rendered as polylines [ZBB*06]. In our study, curvature lines that follow the maximum principal curvature direction were used for colonic surface partitioning. The average time cost for curvature line computation per candidate area (centered at the pre-detected position with a radius of 15mm) is 16.871 seconds. For curvature line clustering, the average time cost per candidate area (on average 221 lines and $7-9$ clusters) is 1.082 seconds. We show our colonic surface partitioning results at some true polyp detection areas in our CAD pipeline (Figure 6). Visual inspection of our results shows that the

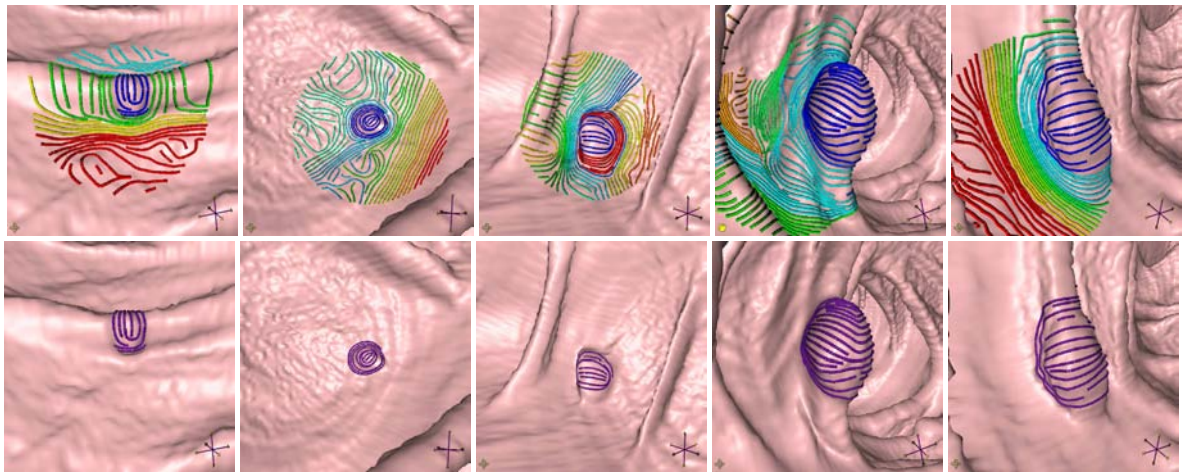


Figure 6: *Images in the upper row: Colonic surface partitioning based on clustering curvature lines. Each surface partition corresponds with the surface area covered by a single cluster. Different colors indicate different curvature line clusters. Images in the lower row: The curvature line cluster that corresponds with a colonic polyp.*

curvature line clustering method provides a reasonable pre-segmentation of the polyp candidate area. Most importantly, as shown in images on the lower row of Figure 6 a surface partition that corresponds well with a true polyp surface can be obtained. This indicates that our overall shape features calculated in such a surface partition can be used to discriminate between true and false detections.

6.3. Discriminating Candidate Detections

The polyp candidate pre-detection step based on the method of van Wijk et al. [vWvRV*06] returned 4036 candidates in total 18 scans. Protrusions are calculated on the surface throughout the entire colon. Candidate detections are found by hysteresis thresholding of protrusions as suggested in [vWvRV*06]. Medical diagnosis confirmed that these polyp candidates included all polyps of the 9 patients. 30 polyps (larger than or equal to 6mm) were annotated using an expert opinion of an experienced radiologist. A polyp was counted as a true positive if it was classified as a polyp in at least one of the two scans. There were 4006 false-positive detections, which needed to be significantly reduced.

In order to estimate the discriminating power of the five-dimensional vector feature, we first performed basic statistic analysis on all pre-detected polyp candidates. In univariate logistic regression analyses, each of the five components were separately found to be significantly related to the polypness of clusters ($p = 0.000, 0.000, 0.001, 0.000$ and 0.011). However, in a multivariate logistic regression analysis, only F_3 was found to be independently associated with the polypness of clusters (p=0.000). In other words, each of the other features does not significantly improve the prediction of polypness over that attainable with F_3 by itself. This analysis does show that F_3 could be used to discriminate between clusters over polyps and non-polyps.

The features F_3 and F_{CV} were used in our pattern recognition step to classify polyps out of false-positive candidates. Only candidates with a mean internal intensity of around that of tissue were retained. This excluded all candidates that were detected as a result of the partial volume effect and due to artifacts of CT. A nine-fold cross-validation was used to compute the system's performance. A logistic classifier [Web02] was trained on the train data set consisting of 16 scans from 8 patients, and was used to classify the two scans of the remaining patient to obtain the classification results. The results for all patients were summed up to get an estimate of the performance of the CAD system.

The system achieved 95% sensitivity for polyps while presenting on average less than 11 false positives per scan and 80% sensitivity with on average 7 false positives per scan. This means that more than 95% of the false-positive detections, i.e. 3808 false positives, were discarded while retaining high sensitivity that 28 true positives were correctly detected. To conclude, this analysis shows that the features derived from the surface partitions correlate quite well with the polyp annotations.

7. Conclusions and Future Work

In this paper, we presented surface-constrained curvature line clustering as a new method for surface partitioning in polyp detection. These surface regions are able to capture polyp characteristics. They could also be used to capture and analyze other larger scale surface shapes. We proposed a new direction-independent, line-to-line and symmetric similarity metric between curvature lines. This similarity metric was applied in a spectral clustering algorithm, in which a reasonable number of clusters is automatically determined. Results showed that polyp surfaces can be captured by corresponding surface partitions. A 5-dimensional feature based on the

volumetric shape index and iso-surface curvedness was proposed to describe overall surface partition shapes. Statistical analysis showed that one of the five components could be used to discriminate between clusters over polyps and non-polyps. We demonstrated that our new proposed features can be used in a polyp detection system. Visual inspection of the clustered curvature lines showed a strong correlation with expected polyp areas.

We plan to investigate the possibility of using our colonic surface partitioning approach for polyp segmentation. Optimizing our scheme will be an important avenue for future work. Our approach will be further tested with a larger number of clinical data sets and other clustering algorithms will be explored, as well as more partition-based discriminative features for polyp detection.

References

[AKY99] ALPERT C. J., KAHNG A. B., YAO S.-Z.: Spectral partitioning with multiple eigenvectors. *Discrete Applied Mathematics (Special volume on VLSI) 90*, 1-3 (Jan. 1999), 3–26.

[Bar05] BARTZ D.: Virtual endoscopy in research and clinical practice. *Computer Graphics Forum 24*, 1 (Mar. 2005), 111–126.

[BKP*04] BRUN A., KNUTSSON H., PARK H.-J., SHENTON M. E., WESTIN C.-F.: Clustering fiber traces using normalized cuts. In *MICCAI 2004* (2004), pp. 368–375. Lecture Notes in Computer Science.

[CCK07] CHEN Y., COHEN J. D., KROLIK J. H.: Similarity-guided streamline placement with error evaluation. *IEEE TVCG 13*, 6 (Nov-Dec 2007), 1448–1455.

[CGG04] COROUGE I., GOUTTARD S., GERIG G.: Towards a shape model of white matter fiber bundles using diffusion tensor MRI. In *Proceedings of IEEE ISBI* (Apr. 2004), pp. 344–347.

[CHH*03] CO C. S., HECKEL B., HAGEN H., HAMANN B., JOY K. I.: Hierarchical clustering for unstructured volumetric scalar fields. In *Proceedings of IEEE VIS'03* (Oct. 2003), pp. 325–332.

[CML*07] CHEN G., MISCHAIKOW K., LARAMEE R. S., PILARCZYK P., ZHANG E.: Vector field editing and periodic orbit extraction using morse decomposition. *IEEE TVCG 13*, 4 (Jul-Aug 2007), 769–785.

[HQK06] HONG W., QIU F., KAUFMAN A.: A pipeline for computer aided polyp detection. *IEEE TVCG 12*, 5 (Sep-Oct 2006), 861–868.

[HSH05] HUANG A., SUMMERS R. M., HARA A. K.: Surface curvature estimation for automatic colonic polyp detection. In *Proceedings of SPIE MI* (2005), vol. 5746, pp. 393–402.

[KBL*07] KLEIN J., BITTIHN P., LEDOCHOWITSCH P., HAHN H. K., KONRAD O., REXILIUS J., PEITGEN H.-O.: Grid-based spectral fiber clustering. In *Proceedings of SPIE MI* (Feb. 2007), pp. 65140V–1–65140V–10.

[KPT*07] KIM D. H., PICKHARDT P. J., TAYLOR A. J., LEUNG W. K., WINTER T. C., HINSHAW J. L., GOPAL D. V., REICHELDERFER M., HSU R. H., PFAU P. R.: CT colonography versus colonoscopy for the detection of advanced neoplasia. *NEJM 357*, 14 (Oct. 2007), 1403–1412.

[MVvW05] MOBERTS B., VILANOVA A., VAN WIJK J. J.: Evaluation of fiber clustering methods for diffusion tensor imaging. In *Proceedings of IEEE VIS'05* (Oct. 2005), pp. 65–72.

[NWB*06] NATARAJAN V., WANG Y., BREMER P.-T., PASCUCCI V., HAMANN B.: Segmenting molecular surfaces. *CAGD 23*, 6 (Aug. 2006), 495–509. Special issue: Applications of geometric modeling in the life sciences.

[SJWS08] SALZBRUNN T., JÄNICKE H., WISCHGOLL T., SCHEUERMANN G.: The state of the art in flow visualization: Partition-based techniques. In *Proceedings of SimVis'08* (Feb. 2008), pp. 75–92.

[SYP*05] SUMMERS R. M., YAO J., PICKHARDT P. J., FRANASZEK M., BITTER I., BRICKMAN D., KRISHNA V., CHOI J. R.: Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population. *Gastroenterology 129*, 6 (Dec. 2005), 1832–1844.

[vGVS*02] VAN GELDER R. E., VENEMA H. W., SERLIE I. W. O., NIO C. Y., DETERMANN R. M., TIPKER C. A., VOS F. M., GLAS A. S., BARTELSMAN J. F. W., BOSSUYT P. M. M., LAMÉRIS J. S., STOKER J.: CT colonography at different radiation dose levels: Feasibility of dose reduction. *Radiology 224*, 1 (May 2002), 25–33.

[vVYV98] VAN VLIET L. J., YOUNG I. T., VERBEEK P. W.: Recursive gaussian derivative filters. In *Proceedings of the 14th International Conference of Pattern Recognition (ICPR'98)* (Aug. 1998), pp. 509–514.

[vWvRV*06] VAN WIJK C., VAN RAVESTEIJN V. F., VOS F. M., TRUYEN R., DE VRIES A. H., STOKER J., VAN VLIET L. J.: Detection of protrusions in curved folded surfaces applied to automated polyp detection in CT colonoscopy. In *MICCAI 2006* (2006), pp. 471–478.

[Web02] WEBB A. R.: *Statistical Pattern Recognition*, second ed. John Wiley and Sons Ltd., 2002.

[YN01] YOSHIDA H., NÄPPI J.: Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps. *IEEE Trans. Med. Imag. 20*, 12 (Dec. 2001), 1261–1274.

[ZBB*06] ZHAO L., BOTHA C. P., BESCOS J. O., TRUYEN R., VOS F. M., POST F. H.: Lines of curvature for polyp detection in virtual colonoscopy. *IEEE TVCG 12*, 5 (Sep-Oct 2006), 885–892.