Published in final edited form as:

Acad Radiol. 2011 August ; 18(8): 1024–1034. doi:10.1016/j.acra.2011.03.012.

Improved Curvature Estimation for Computer-aided Detection of Colonic Polyps in CT Colonography

Hongbin Zhu, PhD, Yi Fan, PhD, Hongbing Lu, PhD, and Zhengrong Liang, PhD

Department of Radiology, HSC L4 Room 120, State University of New York, Stony Brook, NY 11794 (H.Z., Y.F., Z.L.); and the Department of Biomedical Engineering, Fourth Military Medical University, Xi'an, China (H.L.)

Abstract

Rationale and Objectives—Current schemes for computer-aided detection (CAD) of colon polyps usually use kernel methods to perform curvature-based shape analysis. However, kernel methods may deliver spurious curvature estimations if the kernel contains two surfaces, because of the vanished gradient magnitudes. The aim of this study was to use the Knutsson mapping method to deal with the difficulty of providing better curvature estimations and to assess the impact of improved curvature estimation on the performance of CAD schemes.

Materials and Methods—The new method was compared to two widely used kernel methods in terms of the performance of two stages of CAD: initial detection and true-positive and false-positive classification. The evaluation was conducted on a database of 130 computed tomographic scans from 67 patients. In these patient scans, there were 104 clinically significant polyps and masses >5 mm.

Results—In the initial detection stage, the detection sensitivity of the three methods was comparable. In the classification stage, at a 90% sensitivity level on the basis of the input of this step, the new technique yielded 3.15 false-positive results per scan, demonstrating reductions in false-positive findings of 30.2% ($P < .01$) and 27.9% ($P < .01$) compared to the two kernel methods.

Conclusions—The new method can benefit CAD schemes with reduced false-positive rates, without sacrificing detection sensitivity.

Keywords

Colonic polyps; CT colonography; computer-aided detection; shape analysis; curvature; Knutsson mapping; kernel method

According to recent statistics from the American Cancer Society (1), colorectal cancer is the most common cause of both cancer deaths and new cancer cases in both men and women in the United States. Most colon cancers start from benign polyps, and the transformation from benignity to malignance often takes 5 to 10 years. Therefore, early detection and removal of colonic polyps prior to their malignant transformation can effectively decrease the incidence of colon cancer (2–4). Colon polyps are not commonly associated with any symptoms. Therefore, adequate time interval screening is recommended for people aged > 50 years by the American Cancer Society (1). As a new minimally invasive screening technique, computed tomographic colonography (CTC) or computed tomography-based virtual

colonoscopy has shown several advantages over the traditional optical colonoscopy (5). To improve the performance of CTC in detecting polyps, computer-aided detection (CAD) has shown the potential to assist physicians in finding and analyzing polyps in the colon (6–11).

In most currently available CAD systems in CTC, principal curvatures, estimated through kernel methods (12,13), and principal curvature-related measures, such as mean curvatures, Gaussian curvatures, sphericity ratio, shape index, and curvedness, are widely used to characterize the shapes of colon polyps (7,8,14–21). However, when a kernel contains two surfaces (as happens for thin slab objects such as colonic folds and spherical objects such as polyps), spurious estimations of curvatures are frequently observed, indicating false high curvature (22). This is due to the discontinuity at thin structures, where the curvatures are discontinuous because of the diminished gradient magnitude. Fortunately, this discontinuity problem can be solved using Knutsson mapping (23), which maps a discontinuous orientation field into a continuous one. In this study, we adapted the Knutsson mapping technique to accurately estimate principal curvatures and explored the possible benefits of such improved curvature estimation for CAD performance in CTC. Evaluation of the new technique using both phantom experiments and patient studies demonstrated a noticeable reduction in the false-positive (FP) rate for CAD of colonic polyps compared to two current kernel methods.

MATERIALS AND METHODS

Methods

In this section, we outline three curvature estimation techniques: Knutsson mapping and two widely used kernel methods. The details of each technique can be found in the related references. The focus of this section is on the immunity property of Knutsson mapping to the discontinuity problem for CAD of colon polyps.

Kernel methods for estimation of principal curvatures—In differential geometry theory (24), for any point on a smooth surface, any nonsingular curve through that point on the surface will have its own tangent vector $\mathbf{T}$ lying in the tangent plane of the surface orthogonal to the normal vector. The curvature associated with $\mathbf{T}$ at that point is then defined as

$$k_T = \mathbf{k} \times \mathbf{N}, \quad (1)$$

where k is the curvature vector of the concerned curve at that point, and $\mathbf{N}$ is the unit normal at that point on the surface. Nonsingular curves that have the same tangent vector $\mathbf{T}$ will have the same curvature k_T . Among all possible tangent vectors, the maximum and minimum values of k_T at that point are called the principal curvatures, k_1 and k_2 , and the directions of the corresponding tangent vectors are called principal directions.

To compute the principal curvatures for implicit surfaces (eg, isosurfaces) embedded in a three-dimensional volume image, various kernel methods were developed (12,13). In the kernel method developed by Monga et al (12), denoted as KM_1 in the following text, equation 1 was rewritten as

$$k_T = -(\mathbf{T}^T \mathbf{\Omega}_H \mathbf{T}) / \|\mathbf{g}\|, \quad (2)$$

where $\mathbf{\Omega}_H$ is the Hessian matrix of the volume image, and $\mathbf{g}$ is the gradient vector (the normal vector). The resulted principal curvatures and directions were expressed as functions

of the first and second derivatives of the image. In the kernel method developed by Thirion and Gourdon (13), denoted as KM_2 in the following text, equation (1) was rewritten as

$$k_{1,2} = H \pm \sqrt{\Delta} \text{ with } \Delta = H^2 - K, \quad (3)$$

where $H = (k_1 + k_2)$ and $K = k_1 k_2$ are the mean and Gaussian curvatures. They can be computed using the following formulas:

$$H = \frac{1}{2\|g\|^3} \left[I_x^3 (I_{yy} + I_{zz}) - 2I_y I_z I_{yz} + I_y^2 (I_{xx} + I_{zz}) - 2I_x I_z I_{xz} + I_z^2 (I_{xx} + I_{yy}) - 2I_x I_y I_{xy} \right], \quad (4)$$

$$K = \frac{1}{\|g\|^4} \left[I_x^3 (I_{yy} I_{zz} + I_{yz}^2) - 2I_y I_z (I_{xz} I_{xy} - I_{xx} I_{yz}) + I_y^2 (I_{xx} I_{zz} + I_{xz}^2) - 2I_x I_z (I_{yz} I_{xy} - I_{yy} I_{xz}) + I_z^2 (I_{xx} I_{yy} + I_{xy}^2) - 2I_x I_y (I_{xz} I_{yz} - I_{zz} I_{xy}) \right].$$

In equations 4 and 5, the notions associated with I with various subscripts indicate the corresponding partial derivatives of the image intensity I . Therefore, both methods, KM_1 and KM_2 (12,13), represent the principal curvatures as functions of the first and second spatial derivatives of the volume image.

Knutsson mapping—In the aforementioned kernel methods, the curvature is essentially evaluated as $I_{TT}/\|\nabla I\|$, with T representing the associated principal direction, where I_{TT} indicates the second derivative of the image intensity along the principal direction. Therefore, severe overestimation occurs when the gradient magnitude $\|\nabla I\|$ approaches very small values or close to zero. Rieger et al (23) presented a method based on Knutsson mapping, KM_M , to solve the discontinuity problem. The basic idea is described below using the same terms and notations as above.

The curvature k in direction T can also be defined as the magnitude of the change of the surface normal N :

$$k_T = \|\nabla_T N\|. \quad (6)$$

The normal N and the two principal directions T_1 and T_2 are aligned with the three eigenvectors e_1 , e_2 , and e_3 , with decreasing eigenvalues of the gradient structure tensor $\overline{G} = \overline{\mathbf{v}\mathbf{v}^t}$ with $\mathbf{v} = \nabla I$, where the overhead bar ($\overline{\bullet}$) stands for smoothing in a local (Gaussian) neighborhood. Therefore, two scales are involved here: the scale σ_g for evaluating the gradient vector $\nabla I = I \times \nabla G(\sigma_g)$ with Gaussian derivatives $\nabla G(\sigma_g)$ of the Gauss kernel $G(\sigma_g)$ and the scale σ_T for the Gaussian weighted tensor smoothing $\overline{G} = \overline{\mathbf{v}\mathbf{v}^t} = (\mathbf{v}\mathbf{v}^t) \times G(\sigma_T)$. ∇_T is the operator of directional derivative along T . Equation 6 essentially defines the principal curvatures by differentiating the normal with respect to the principal directions: $|k_{1,2}| = \|\nabla_{T_{1,2}} N\|$. Therefore, the gradient magnitude does not serve as denominator, and the discontinuity problem of the gradient magnitude is then avoided in theory. However, the eigenvectors of $\overline{G}$ contain only orientation, without direction information (an orientation has two directions reverse to each other), which leaves room for ambiguity in representing directional vectors (eg, $\mathbf{e}_1 = \pm N$). The ambiguous representation leads to troublesome descriptions (ie, the so-called discontinuity issue in the orientation space) (25). Direct computation of the partial derivatives within the discontinuous orientation field will lead to spurious curvatures. To establish a continuous orientation representation, Knutsson (26)

proposed that the following three properties or criteria had to be satisfied: (1) uniqueness: vectors $\pm \mathbf{e}$ of the original space should be mapped onto the same vector in the representation space; (2) polar separability: the norm of the mapped vector should be independent of the direction of the original vector; and (3) uniform stretch: the mapping should locally preserve the angle metric of the original space.

As discussed by Rieger and van Vliet (25), gradient structure tensor stands as a mapping itself satisfying the first two but violating the third criterion. The criterion of uniform stretch requires the mapping to scale linearly with the angle between two hyperplanes (interested readers are referred to Rieger and van Vliet [25] and Knutsson [26] for more details). That is to say, for a mapping M , $\|\delta M(\mathbf{v})\| = \varepsilon \|\delta \mathbf{v}\|$ holds for $\|\mathbf{v}\| = \text{const}$, where δ is the variational operator working on the original and mapped vectors, $\mathbf{v}$ and $M(\mathbf{v})$, and ε is a constant. The Knutsson mapping $M(\mathbf{v}) = \mathbf{v}\mathbf{v}'/\|\mathbf{v}\|$ with $\varepsilon = \sqrt{2}$ satisfies all the above conditions (23) and so serves the purpose of mapping the discontinuous orientation field into a continuous representation. The norm of the mapped derivatives is linearly related to the norm of the derivatives of the original orientation $\mathbf{v}$. Therefore, the principal curvatures can then be calculated as

$$|k_{1,2}| = \frac{1}{\sqrt{2}} \|\nabla_{\mathbf{e}_{2,3}} M(\mathbf{e}_1)\| = \frac{1}{\sqrt{2}} \left\| \sum_{n=1}^3 \frac{\partial M_{ij}}{\partial x^n} \mathbf{e}_{2,3}^n \right\|, \quad (7)$$

where M_{ij} is the element of the second-order tensor $M(\mathbf{e}_1)$, and x^n a coordinate. The norm is the Fröbenius norm $\|M\|^2 = \sum_{i,j} m_{ij}^2$. To be clearer, equation 7 can be laid out as

$$\begin{aligned} |k_{1,2}| &= \frac{1}{\sqrt{2}} \left\| \sum_{n=1}^3 \frac{\partial M_{ij}}{\partial x^n} \mathbf{e}_{2,3}^n \right\| \\ &= \frac{1}{\sqrt{2}} \left\| \frac{\partial M_{ij}}{\partial x} e_{2,3}^x + \frac{\partial M_{ij}}{\partial y} e_{2,3}^y + \frac{\partial M_{ij}}{\partial z} e_{2,3}^z \right\|, \end{aligned} \quad (8)$$

where $e_{2,3}^x, e_{2,3}^y$, and $e_{2,3}^z$ represent the coordinates of the vectors $\mathbf{e}_2$ and $\mathbf{e}_3$. Another scale, σ_k , is involved for evaluating first-order derivative $\nabla M_{ij} = (\frac{\partial M_{ij}}{\partial x}, \frac{\partial M_{ij}}{\partial y}, \frac{\partial M_{ij}}{\partial z}) = M_{ij} \times \nabla G(\sigma_k)$ through the Gaussian derivative $\nabla G(\sigma_k)$. The sign of the curvatures is the negative of the sign of $T^t \Omega_H T$, where Ω_H represents the Hessian matrix.

To evaluate the new curvature estimation method (equation 7) in comparison to the current methods (equations 2 and 3), a CAD system is needed that uses the curvature-based measures for the detection task. The experimental design for the evaluation is given below.

Experimental Design

CAD pipeline—The three curvature estimation methods (KM_1 , KM_2 , and KM_M) were put into a CAD pipeline to evaluate their performance. The pipeline is summarized by three stages, as shown in Figure 1.

In stages 1 and 3, previous works (19–21,27,28) were used to fulfill the specific tasks, while different curvature estimation methods were explored in stage 2 for geometric analysis. In the present study, the performance of the three curvature estimation methods was measured by their impacts on CAD of colonic polyps, which was investigated and compared on the basis of the results of the two steps of initial polyp candidate (IPC) extraction and true-positive (TP) and FP classification in the final stage. The results of these two steps were

indicated by the by-polyp detection sensitivity of the TP results and the number of FP results per patient scan.

In the second step, the TP and FP classification step is based on the support vector machine (SVM) (19,21), where the sample pool is split into two groups: the training and testing sets. Because our focus was on investigating the performance of the different curvature estimation methods, we chose various order statistics of the shape index ($SI = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_1 + k_2}{k_1 - k_2}$)

and curvedness ($CV = \sqrt{\frac{k_1^2 + k_2^2}{2}}$) (14), which are functions of curvatures, to serve as the characterizing features. These statistics are the mean, variance, skewness, and kurtosis.

Method parameter configuration—To make a meaningful comparison, we fixed all the parameters involved in stages 1 and 3 of the CAD pipeline, which had been fully explored and optimized in the related previous works. The parameters in stage 2 are summarized in Table 1.

In the methods KM_1 and KM_2 , parameters of (a_1, a_2) and (σ_1, σ_2) refer to the filter scales for estimating the first-order and second-order spatial derivatives. The parameters have been extensively investigated in the literature on CAD of colon polyps (14,15,18,22). The reported optimal values of (0.7, 0.1) and (1, 2) for (a_1, a_2) and (σ_1, σ_2) , respectively, were used in this study. σ_g in KM_M is a kernel for estimating ∇I , and its optimal value was set at 1 by Rieger et al (23). The determination of σ_T and σ_k was essentially a trade-off between noise suppression and sacrifice of estimation accuracy. These two parameters remain unoptimized in the literature on CAD of colon polyps. In our experiments, we determined their values through phantom studies (as illustrated in the following two paragraphs) and then applied them to patient data sets.

We chose a three-dimensional shell sphere to be the phantom (Fig 2c). To make it meaningful, its size, image intensities of foreground and background, and noise level were adjusted according to the patient images in the figure. For example, by a typical geometric analysis on the patient images, the colon lumen serves as the background, and the tissues adjacent to the colon wall are the foreground objects. We manually drew regions of interest (ROIs) in the areas inside the colon lumen (Fig 2a) and around the adjacent tissues (Fig 2b) on the patient images from various data sets. The mean and standard deviation of the two kinds of ROIs are shown in Table 2 in Hounsfield units. From the ROI measures, we set the background and foreground objects in the phantom to be 40 and 940 (approximating those in Table 2) and then smoothed the objects with a Gaussian kernel of $\sigma = 2$ to consider the partial volume effect. The x-ray CT image noise has been proven to be Gaussian noise (29). By adding the Gaussian noise to the background and foreground objects with standard deviations of 25 and 40 (approximating the standard deviations of colon lumen and adjacent tissue listed in Table 2), respectively, we generated a phantom image with similar characteristics as a CT image.

Because the polyp sizes are generally >6 voxel units (as listed in Table 3), the radius of the sphere phantom was sampled to be 5, 15, 35, and 45 voxel units and the shell thickness to be 3, 7, 9, and 9 voxel units, respectively. The three curvature estimation methods were then applied to the phantom image. The mean values of the resulting $|k_1|$ and $|k_2|$ at the points evenly sampled on the center spherical layers with radii of 5, 15, 35, and 45 were investigated for the four phantoms. The σ_T and σ_k of KM_M were investigated for 0.5, 1, 2, 4, 6, and 8.

CTC Database—The performance of the above three methods was evaluated on a CTC database from 67 patients with 130 CT scans from both supine and prone positions. In these scans, colon cleansing was performed with standard precolonoscopy or barium enema bowel preparation. A single dose of 2% barium (250 mL) and diatrizoate (60 mL) was given to tag residual stool and fluid. Multislice CT scanners (LightSpeed Ultra; GE Medical Systems, Milwaukee, WI) were used in helical mode to collect data with collimations of 1.25 to 5.0 mm, pitch of 1 to 2, reconstruction intervals of 1.25 to 5.0 mm, and modulated tube current–time products ranging from 50 to 200 mAs and tube voltages from 80 to 120 kVp. In-plane and cross-plane image resolution ranged from (0.6, 1) to (0.9, 1.2). One polyp in different scans was counted as different polyps, and there were 208 clinically significant polyps (>5 mm) and masses (>30 mm) confirmed by both optical colonoscopy and virtual colonoscopy. Polyps ≤ 5 mm were not considered in this study. The size distribution of the polyps and masses is shown in Table 3, and the shape distribution is listed in Table 4. For simplicity, the term “polyp” in the following text refers to both polyps and masses.

Performance evaluation for patient studies—To evaluate the performance on the basis of the patient studies in the above CTC database, we incorporated the three curvature estimation methods into the CAD pipeline (Fig 1). The performance on CAD of colonic polyps was investigated by exploring the results of the two steps of IPC extraction and TP and FP classification in stage 3 (Fig 1).

In the IPC extraction step, the detection sensitivity and the FP rate served as the measures of performance estimation. In addition, typical false-negative (FN) results were investigated to compare the performance of the three curvature estimation methods.

In the TP and FP classification step, there were three IPC sets, based on which of the sample pools of feature vectors were extracted to classify the final TP and FP findings. Generally, the IPC extraction and feature calculation go hand in hand in a typical CAD system. However, keeping in mind that we were evaluating the performance of the three curvature estimation methods by implicitly exploring the classification abilities of the curvature-related features, we needed to ensure that the features in comparison were extracted from the same objects (ie, the same IPCs) before being input into the SVM classifier. Considering that the IPC volume and even different IPCs may be generated in a patient’s colon with different curvature estimation methods (20), each of the three groups of IPCs from the three methods, respectively, was chosen to serve as the sample pool. This led to nine separate studies with various combinations of IPCs and features. With each IPC group, the principal curvatures from the three methods were used to calculate the features, and the generated feature vectors formed three sample pools for TP and FP classification. We used free-response receiver-operating characteristic (fROC) curves, where the vertical and horizontal axes indicate the detection sensitivity and FP rate, to measure the performance of each TP and FP classification. It should be noted that sensitivity in this step was estimated in terms of the input of it, i.e., the extracted IPCs in the previous step. The fROC curves were parameterized by the decision thresholds in the SVM testing procedure (20). For a given IPC group, it was randomly split 200 times to obtain 200 pairs of equally sized training and testing sets by IPCs, and the number of patient scans in the testing process was assumed to be 65 when evaluating the FP rates. The final fROC curves were averaged from the testing procedure of the 200 runs of the SVM classification. The average was calculated according to the detection sensitivity by sampling at 0.05, 0.1, 0.15, ..., 0.95.

RESULTS

Phantom Study

By inspecting the plots in Figure 3, the error of a small object (theoretically with a larger curvature) is higher than that of a large object, which agrees with the results of Rieger et al (23). For objects with different sizes, the optimal parameters of σ_T and σ_K may vary. The optimal values are about $\sigma_T = 4$ and $\sigma_K = 1$.

So far, for the three curvature estimation methods, we have obtained their optimal parameters, which are (0.7, 0.1) for (a_1, a_2) , (1, 2) for (σ_1, σ_2) , and (1, 4, 1) for $(\sigma_g, \sigma_T, \sigma_K)$. We applied these values to the shell sphere phantom with radius of 15 for comparison. Curvatures of points on the spherical layers with different radii are sampled by using the bicubic interpolation technique. The means and variances of the estimations on three layers ($r = 14, 15$, and 16) are listed accordingly in Table 5. Obviously, in different layers, KM_M consistently provides the best estimations, and KM_1 and KM_2 overestimate the magnitudes of the first and second principal curvatures. Larger overestimations are given by both KM_1 and KM_2 at the center layer ($r = 15$) than at the other layers ($r = 14$ and 16), which is due to the higher degree of symmetry, indicating smaller magnitude of the gradient, at the center layer.

Patient Studies

The phantom study above indicates the accuracy on the curvature estimation by the three methods. In this section, we incorporate the three curvature estimation methods into the CAD pipeline (Fig 1), and their performance on CAD of colonic polyps is investigated by exploring the performance of the two steps of IPC extraction and TP and FP classification in stage 3 (Fig 1).

Results of IPC extraction—In the first step of IPC extraction, we kept the configuration of Zhu et al (20) for all the three curvature estimation methods. The detection sensitivity and FP rate were assessed according to the whole database. The results are shown in Table 6.

The KM_M method yielded fewer FP findings in the experiment, which might be due to its ability to provide better principal curvature estimation. The FN findings are summarized as follows: for KM_1 , seven flat polyps (range, 6–12 mm) and five sessile and one pedunculated polyp (<8 mm); for KM_2 , seven flat and five sessile polyps (the same as for KM_1 , without the pedunculated polyp); and for KM_M , seven flat and five sessile polyps (the same as for KM_1 , without the pedunculated polyp).

As shown in Figure 4, the area indicated by the arrows is on a thin haustral fold, where the segmented colon wall includes voxels with small gradients. Either KM_1 or KM_2 estimated the first and second principal curvatures at about (0.03, 0.02) for some voxels in the area. The curvature values imply that the two curvature radii are about 33 and 50 voxel units, in disagreement with visual perception, because the fold is quite flat. The curvatures yielded shape index and curvedness of about (0.0628, 0.0255) for the voxels. The voxels were then labeled as seeds and grew to form an IPC according to previously reported methods (18–20), generating FP results. The KM_M method estimated the curvatures at about (0.000123, 0.000111), a much more accurate indication for the flat areas, which yielded shape index and curvedness of about (0.016, 0.000117) for those voxels. Therefore, those voxels were not labeled, and no IPC was generated accordingly by the previous methods (18–20).

From the IPCs, our next task was to differentiate the TP and FP findings. The results will show the difference of the three methods.

Results of TP and FP classification—Figure 5 shows the final results on the basis of the evaluation strategy mentioned above in the section “Performance Evaluation for Patient Studies.” At a detection sensitivity level of 90%, the error bars in Figure 5 indicate the 95% confidence intervals of the FP rates. Table 7 tabulates the FP rates and their 95% confidence intervals in the pairs of parentheses. By using z tests on the sampled FP rates for the operating points in Figure 5, the relative P values are listed in Table 8.

By inspecting each row of Table 7, in which the same IPC group was used for the three different curvature estimators, the KM_M yielded the smallest FP rates. The KM_M method generated 3.15 FP findings per scan from the KM_M method’s IPC group (95% confidence interval, 2.95–3.35), which is the best performance among the three separate experiments, each with a corresponding IPC group.

Table 8 indicates that the performance of KM_M is significantly different from that of KM_1 and KM_2 ($P < .01$) regardless of IPC group. However, the difference in performance between KM_1 and KM_2 is not significant with all IPC groups ($P > .05$), which is also implied by the overlaps between the related error bars in Figure 5.

DISCUSSION

Phantom Study

For two traditional kernel methods, KM_1 and KM_2 , the optimal parameters have been extensively explored in previous works. In this study, we focused on investigating the optimal configuration for a new method, KM_M . Bearing in mind that the purpose of the investigation was for CAD in CTC, we constructed phantom images with similar image contrast and noise level to those of clinical patient images. The phantom sizes are specified analogously to the sizes of typical polyps in the CTC database (Table 3). From the results shown in Figure 3, the optimal values are about $\sigma_T = 4$ and $\sigma_k = 1$ for objects with different sizes. Theoretically, the parameters should be adaptive to the object sizes. However, the object sizes are unknown prior to the curvature estimation process. Furthermore, the variation of the parameters to the object sizes was shown small by the phantom studies. Therefore, the obtained values of about $\sigma_T = 4$ and $\sigma_k = 1$ are practically valuable for CAD in CTC.

With the optimized parameters, the three methods were applied to a spherical shell phantom image, and the resulting curvatures were compared to the theoretical values. From the results listed in Table 5, two conventional kernel methods, KM_1 and KM_2 , overestimated the magnitudes of the first and second principal curvatures because of the gradient discontinuity problem. The overestimation becomes more severe when approaching the middle symmetric layer of the object, where the gradient magnitude gradually approaches zero. The presented new method, KM_M , could solve the problem and has generated very good estimations of both principal curvatures at various degrees of gradient discontinuities.

Patient Study

The three curvature estimation methods were embedded in an existing CAD pipeline, and their performance on CAD of polyps in patient computed tomographic colonographic images were investigated in the two steps of IPC extraction and TP and FP classification.

In the step of IPC extraction, shape index and curvedness were thresholded to generate IPCs by assuming that colon polyps are morphologically similar to spherical objects (19,20). The detection sensitivities in this step were comparable for the three methods, which is about 94% as listed in Table 6, while the presented new method provided slightly better results. The FN findings were almost the same, most of which were flat polyps and small sessile

polyps (<8 mm). For most current CAD pipelines, flat polyps are missed because they violate the assumption of these CAD pipelines that colon polyps are roughly spherical objects (14,21). A curvature estimator, such as KM_1 , KM_2 , and even KM_M , cannot help avoid the violation. Similar to KM_1 and KM_2 , KM_M also faces some challenge in accurately estimating the curvatures for small objects, as evidenced in the phantom study. Therefore, small sessile polyps may be missed with the three curvature estimators. However, the FP rate of KM_M is smaller than that of KM_1 and KM_2 . This is due to a higher accuracy in curvature estimation delivered by KM_M .

In the step of TP and FP classification, we focused on investigating the classification ability of the curvature-related features to explore whether there are benefits with the curvatures estimated by KM_M . Because curvature estimators are also involved in IPC extraction, three comparisons were conducted on the basis of the three IPC groups from the three curvature estimators, respectively. In each comparison, we ran the SVM program 200 times on the basis of 200 random selections of the training and testing sets. Figure 5 shows the averaged fROC curves from the 200 runs. Through visual inspection, in terms of both sensitivity levels and IPC groups, the fROC curves from KM_M consistently outperformed the other two kernel methods. It is noted that a more rigorous comparison between different fROC curves may be conducted by using the jackknife fROC method (30), which is a topic of our future work.

The results in Figure 5 and Tables 7 and 8 show that in the step of TP and FP classification, KM_M 's performance is superior to that of KM_1 and KM_2 . The reason is that curvatures with higher accuracy provide enhanced classification. Figure 6 shows two examples of FN detections for the CAD pipeline with KM_1 or KM_2 but TP detections with KM_M in a certain condition. The two pedunculated polyps have round tips, and the curvatures were accurately estimated by all the three methods. Therefore, in the IPC extraction step, two IPCs are successfully generated, representing the polyps by the three methods. However, some voxels of the IPCs locate near the center of the fold (where the polyp in top row of Figure 6 sits on) or around the axis of the polyp itself. The symmetric property gives rise to the severely overestimated curvatures with KM_1 and KM_2 , which may flaw the related features, especially the mean and variance of curvedness. The issue was solved by the CAD based on KM_M , because of the improved curvature estimations. In our experiment, the two polyps in Figure 6 might be missed (ie, FN findings) by KM_1 and KM_2 but successfully detected as TP results by KM_M at 80% and 90% sensitivity levels. This implies that to classify the IPC as a true detection at a specified sensitivity level, the classifier has to bear more non-polyp objects as true detections, leading to an increased FP rate. This is evidenced by the higher FP rates listed in Table 7.

CONCLUSION

With the two widely used kernel methods, KM_1 and KM_2 , spurious calculations in curvature estimation were frequently observed (22) because of the gradient discontinuity problem, indicating false high curvature. In this study, we applied a new method, KM_M , to improve the curvature computation and investigated the potential benefits of KM_M for CAD of colon polyps on CTC.

From the results of the phantom study, the new method with optimized parameters greatly improved the curvature estimation for the voxels on the thin shell objects. The improvement was not limited to the voxels at the center of the objects.

The three curvature estimators, KM_1 , KM_2 and KM_M , were incorporated into a CAD pipeline, and experiments were conducted on a database including 130 patient computed

tomographic colonographic scans with 208 confirmed significant polyps. The number of polyps is not small for evaluation purposes compared to those in previous studies (averaging about 32) (31). Despite this, we noticed that the estimation accuracy of KM_M depends partially on the size of the objects. Therefore, the parameters of KM_M could be refined by a larger polyp database.

The results of the IPC extraction step show that the KM_M does not help much for IPC extraction upon detecting flat polyps, although it can reduce FP findings to some degree. The FP rates after IPC extraction were 24.3, 23.8, and 22.5 for KM_1 , KM_2 , and KM_M respectively. However, in the final step of TP and FP classification, the improved curvatures from KM_M enhanced the classification ability. From our experiments, the FP rates for KM_1 , KM_2 , and KM_M were 4.51, 4.37, and 3.15 (ie, the numbers on the main diagonal in Table 7), at 90% detection sensitivity on the basis of the input of this step. The percentages of FP reduction by KM_M were 30.2% and 27.9% over KM_1 and KM_2 , respectively. Therefore, we conclude that KM_M can improve curvature estimation and enhance the shape analysis, leading to improved CAD in CTC.

We note that radiologists really want to know the overall performance. For example, at the operating points in Figure 5, after including the missed polyps in the IPC extraction step, the overall detection sensitivities were 84.4%, 84.8%, and 84.8% with FP rates of be 4.51, 4.37, and 3.15, for KM_1 , KM_2 , and KM_M respectively. It should be noted that in the current study, only functions of curvatures were used as features in the TP and FP classification step. It would be interesting to investigate the potential benefits when combined with other features, such as the projection features presented by Zhu et al (21), to yield better overall performance.

Acknowledgments

This work was partially supported by grants CA082402 and CA120917 from the National Cancer Institute (Bethesda, MD). Dr Lu is supported by the National Nature Science Foundation of China under grant 60772020.

We would like to acknowledge the use of the Viatronix V3D-Colon Module (Viatronix, Inc, Stony Brook, NY).

References

1. American Cancer Society. Cancer facts & figures 2008. Atlanta, GA: American Cancer Society; 2008.
2. Eddy D. Screening for colorectal cancer. *Ann Intern Med.* 1990; 113:373–384. [PubMed: 2200321]
3. Gluecker T, Johnson C, Harmsen W, et al. Colorectal cancer screening with CT colonography, colonoscopy, and double-contrast barium enema examination: prospective assessment of patient perceptions and preferences. *Radiology.* 2003; 227:378–384. [PubMed: 12732696]
4. Levi B, Brooks D, Smith R, et al. Emerging technologies in screening for colorectal cancer: CTC, immunochemical fecal occult blood tests, stool screening using molecular markers. *CA Cancer J Clin.* 2002; 53:44–55.
5. Pickhardt P, Choi J, Hwang I, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med.* 2003; 349:2191–2200. [PubMed: 14657426]
6. Kiss G, Cleynenbreugel J, Thomeer M, et al. Computer-aided diagnosis in virtual colonography via combination of surface normal and sphere fitting methods. *Eur Radiol.* 2002; 12:77–81. [PubMed: 11868078]
7. Summers R, Beaulieu C, Pusanik L, et al. Automated polyp detector for CT colonography: feasibility study. *Radiology.* 2000; 216:284–290. [PubMed: 10887263]

8. Summers R, Yao J, Pickhardt P, et al. Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population. *Gastroenterology*. 2005; 129:1832–1844. [PubMed: 16344052]
9. Taylor S, Halligan S, Burling D, et al. Computer-assisted reader software versus expert reviewers for polyp detection on CT colonography. *AJR Am J Roentgenol*. 2006; 186:696–702. [PubMed: 16498097]
10. Bielen D, Kiss G. Computer-aided detection for CT colonography: update 2007. *Abdom Imaging*. 2007; 32:571–581. [PubMed: 17690932]
11. Yoshida, H.; Nappi, J. CAD in CT colonography: past, present, and future. Presented at: 11th International Conference of MICCAI, Workshop on Computational and Visualization Challenges in the New Era of Virtual Colonoscopy; New York, NY. September 6, 2008;
12. Monga O, Ayache N, Sander PT. From voxel to intrinsic surface features. *Image Vis Comput*. 1992; 10:403–417.
13. Thirion JP, Gourdon A. Computing the differential characteristics of isointensity surfaces. *Comput Vis Image Understand*. 1995; 61:190–202.
14. Yoshida H, Nappi J. Three-dimensional computer-aided diagnosis scheme for detection of colonic polyps. *IEEE Trans Med Imaging*. 2001; 20:1261–1274. [PubMed: 11811826]
15. Wang Z, Liang Z, Li L, et al. Reduction of false positives by internal features for polyp detection in CT-based virtual colonoscopy. *Med Phys*. 2005; 32:3602–3616. [PubMed: 16475759]
16. Konukoglu E, Acar B, Paik D, et al. Polyp enhancing level set evolution of colon wall: method and pilot study. *IEEE Trans Med Imaging*. 2007; 26:1649–1656. [PubMed: 18092735]
17. van Wijk C, van Ravesteijn VF, Vos FM, et al. Detection and segmentation of colonic polyps on implicit isosurfaces by second principal curvature flow. *IEEE Trans Med Imaging*. 2010; 29:688–698. [PubMed: 20199908]
18. Wang S, Zhu H, Lu H, et al. Volume-based feature analysis of mucosa for automatic initial polyp detection in virtual colonoscopy. *Int J Comput Assist Radiol Surg*. 2008; 3:131–142. [PubMed: 19774204]
19. Zhu H, Duan C, Pickhardt P, et al. Computer-aided detection of colonic polyps with level set-based adaptive convolution in volumetric mucosa to advance CT colonography toward a screening modality. *Cancer Manage Res*. 2009; 1:1–13.
20. Zhu H, Fan Y, Lu H, et al. Improving initial polyp candidate extraction for CT colonography. *Phys Med Biol*. 2010; 55:2087–2102. [PubMed: 20299733]
21. Zhu H, Liang Z, Pickhardt P, et al. Increasing computer-aided detection specificity by projection features for CT colonography. *Med Phys*. 2010; 37:1468–1481. [PubMed: 20443468]
22. Campbell S, Summers R. Analysis of kernel method for surface curvature estimation. *Int Congr Ser*. 2004; 1268:999–1003.
23. Rieger R, Timmermans FJ, van Vliet LJ, et al. On curvature estimation of ISO surfaces in 3D gray-value images and the computation of shape descriptors. *IEEE Trans Patt Anal Mach Intell*. 2004; 26:1088–1094.
24. Lipschutz, M. *Differential geometry*. New York: McGraw-Hill; 1969.
25. Rieger, B.; van Vliet, LJ. Representing orientation in N-dimensional spaces. In: Petkov, N.; Westenberg, MA., editors. *Proceedings of the 10th International Conference on Computer Analysis of Images and Patterns*. 2003.
26. Knutsson H. Representing local structures using tensors. *Proc Sixth Scand Conf Image Anal*. 1989:244–251.
27. Liang Z, Wang S. An EM approach to MAP solution of segmenting tissue mixtures: a numerical analysis. *IEEE Trans Med Imaging*. 2009; 28:297–310. [PubMed: 19188116]
28. Wang S, Li L, Cohen H, et al. An EM approach to MAP solution of segmenting tissue mixture percentages with application to CT-based virtual colonoscopy. *Med Phys*. 2008; 35:5787–5798. [PubMed: 19175136]
29. Lei T, Sewchand W. Statistical approach to x-ray CT imaging and its applications in image analysis—part I: statistical analysis of x-ray CT imaging. *IEEE Trans Med Imaging*. 1992; 11:53–61. [PubMed: 18218356]

30. Chakraborty D, Berbaum K. Observer studies involving detection and localization: modeling, analysis, and validation. *Med Phys.* 2004; 31:2313–2330. [PubMed: 15377098]
31. Sundaram P, Zomorodian A, Beauleu C, et al. Colon polyp detection using smoothed shape operator: preliminary results. *Med Image Anal.* 2008; 12:99–119. [PubMed: 17910934]

Figure 1.

The evaluation computer-aided diagnosis pipeline. CTC, computed tomographic colonographic; FP, false-positive; KM, kernel method; TP, true-positive.

Figure 2.
(a) Region of interest (ROI) in the colon lumen. (b) ROI in tissue area adjacent to colon wall. (c) Cross-section of the three-dimensional phantom sphere shell.

Figure 3.

Plots of error (vertical axis) (ie, $|k_{1,2}| - k_{\text{theory}}$) according to objects with various sizes ($r = 5, 15, 35$, and 45 from top to bottom rows) and various values of σ_T and σ_k (two horizontal axes) with the Knutsson mapping kernel method. The left and right columns indicate the errors of k_1 and k_2 .

Figure 4. A false-positive (*arrow*) finding detected in the first step of initial polyp candidate extraction by the computer-aided diagnosis pipeline with either kernel method 1 or kernel method 2, but removed by that with the Knutsson mapping kernel method. (*Left*) Three-dimensional endoscopic display. (*Right*) Two-dimensional display in axial slice.

Figure 5.

Plots of the true-positive (TP) and false-positive (FP) classification results on the basis of the three initial polyp candidate (IPC) sets: **(a)** from kernel method 1 (KM_1), **(b)** from kernel method 2 (KM_2), **(c)** from the Knutsson mapping kernel method (KM_M). The sensitivity levels are evaluated on the basis of the input of this stage. In each panel, there are three curves representing the performance of the features on the basis of the three curvature estimation methods (KM_1 , KM_2 , and KM_M), as indicated by the legends. The error bars at the operating points indicate the 95% confidence intervals of the averaged FP rates at 90% detection sensitivity.

Figure 6.

Examples of false-negative (FN) findings (*arrows*) of kernel method 1 (KM_1) or kernel method 2 (KM_2) and true-positive (TP) findings of the Knutsson mapping kernel method (KM_M). (Top row) A 7-mm pedunculated polyp sitting on a fold in hepatic flexure. (Bottom row) A 12-mm-long pedunculated tubular adenoma, with the diameter measured at about 7 mm. (Left and middle columns) Three-dimensional endoscopic displays in different view angles. (Right column) Two-dimensional display in axial and sagittal slices. At the voxels indicated by the dashed crosses in the right column, the principal curvatures with KM_1 and KM_2 were overestimated at about 3.35 and 1.49 for the top and bottom polyps, while KM_M gave more reasonable estimations of about 0.012 and 0.027. For the computer-aided diagnosis with either KM_1 or KM_2 , the top polyp was an FN finding at both 90% and 80% detection sensitivity, and the bottom polyp was an FN finding at 80% sensitivity but a TP finding at 90% sensitivity. However, both polyps were always TP findings for the computer-aided diagnosis with KM_M at both sensitivity levels.

TABLE 1

Parameters Involved in Stage 2 of the Computer-Aided Diagnosis (CAD) Pipeline

Method	Parameters
Kernel method 1	a_1, a_2
Kernel method 2	σ_1, σ_2
Knutsson mapping kernel method	$\sigma_g, \sigma_T, \sigma_k$

TABLE 2

Means and Standard Deviations of Regions of Interest in Colon Lumen and Tissue Adjacent to Colon Wall

	Mean	Standard Deviation
Lumen (Hounsfield units)	41.54	23.64
Adjacent tissue (Hounsfield units)	937.81	38.77

TABLE 3

Size Distribution of the Polyps

Size (mm)	n (%)
6–9	118 (56.7)
10–30	82 (39.4)
>30	8 (3.8)

TABLE 4

Shape Distribution of the Polyps

Shape	n (%)
Sessile	103 (49.5)
Pedunculated	92 (44.2)
Flat	13 (6.3)

TABLE 5

Statistics of $|k_1|$ and $|k_2|$ Resulting From the Three Methods

KM ₁			KM ₂			KM _M		
	Mean	Variance	Mean	Variance		Mean	Variance	
$r = 14$ ($ k_1 = k_2 \approx 0.0714$)								
$ k_1 $	2.24	1.01	1.88	0.86		0.0684	0.000210	
$ k_2 $	3.15	1.15	2.87	0.97		0.0672	0.00722	
$r = 15$ ($ k_1 = k_2 \approx 0.0667$)								
$ k_1 $	23.04	302	21.53	402		0.0641	0.000204	
$ k_2 $	31.21	411	27.31	433		0.0640	0.00249	
$r = 16$ ($ k_1 = k_2 \approx 0.0625$)								
$ k_1 $	1.94	0.58	1.77	0.77		0.0620	0.000332	
$ k_2 $	2.58	0.89	2.33	0.95		0.0614	0.00745	

KMM, Knutsson mapping kernel method; KM₁, kernel method 1; KM₂, kernel method 2.
Theoretical values are listed in the first column.

TABLE 6

Performance of Initial Polyp Candidate Extraction According to the Three Methods

Variable	KM ₁	KM ₂	KM _M
Sensitivity (%)	93.8	94.2	94.2
False-positive rate	24.3	23.8	22.5
Number of false-negative findings	13	12	12

KM_M, Knutsson mapping kernel method; KM₁, kernel method 1; KM₂, kernel method 2.

TABLE 7

False-positive Rates at 90% Detection Sensitivity on the Basis of the Input of the True-positive and False-positive Classification Stage

IPCs	Features		
	KM ₁	KM ₂	KM _M
KM ₁	4.51 (4.21–4.81)	5.07 (4.72–5.42)	3.28 (3.07–3.49)
KM ₂	5.01 (4.64–5.38)	4.37 (4.02–4.72)	3.38 (3.14–3.62)
KM _M	4.61 (4.34–4.88)	4.57 (4.31–4.83)	3.15 (2.95–3.35)

IPC, initial polyp candidate; KM_M, Knutsson mapping kernel method; KM₁, kernel method 1; KM₂, kernel method 2.

In each row, the IPC set is from the same curvature estimation method, but the characterizing features are from various methods.

TABLE 8*P* Values of the Operating Points in Figure 5

IPCs	Features		
	KM₁ vs KM₂	KM₁ vs KM_M	KM₂ vs KM_M
KM ₁	0.14	0.00317	0.000211
KM ₂	0.21	0.000531	0.00353
KM _M	5.87	0.000342	0.000355

IPC, initial polyp candidate; KM_M, Knutsson mapping kernel method; KM₁, kernel method 1; KM₂, kernel method 2.