Global Analysis of Haustral Fold Ridges for the Reduction of False Positives in CTC-CAD

Tahir Nawaz
tahir.nawaz@eecs.qmul.ac.uk
Queen Mary University of London
London, UK

Greg Slabaugh
Greg.Slabaugh@medicsight.com
Medicsight PLC
London, UK

Abstract

Haustral folds are a major source of false positives (FPs) in the Computer Aided Detection (CAD) of colorectal lesions. In this paper, a novel bottom-up data-driven approach is presented to *globally analyze* haustral folds. Unlike the earlier haustral fold detection methods that rely solely on local differential geometric computations, which can be unreliable, particularly for thicker folds, our method builds a global model of the fold ridge to characterize the fold on a much larger scale. We propose a novel geometric feature, the *density of ridge candidates* (DRC), to remove CAD false positives. The method is tested on a dataset of 333 CAD marks detected using a state-of-the-art CAD system [7] on a data from 34 patients, and a significant FP reduction of **14.13%** is achieved while maintaining **100%** sensitivity.

1 Introduction

There has been a growing research in the past decade in the area of the detection of colorectal lesions, the second leading cause of cancer deaths in developed countries. The standard method for screening of the colorectal lesions is Optical Colonoscopy (OC). However, it has several disadvantages such as invasiveness, difficulties imaging the entire colon, use of sedation, and risk of perforation which has led to the evolution of CT Colonography (CTC). In CTC, CT images of the entire colon, cleansed and insufflated, are obtained and then examined by a radiologist using a visualization software to detect colorectal lesions. Computer Aided Detection (CAD) systems employ image processing and pattern recognition algorithms to help the radiologist analyze the CT data and automatically identify lesions which may have otherwise been overlooked. However, haustral folds are a common source of false positives [4] in CAD systems. Therefore, their accurate detection can significantly reduce false positives (FPs), thereby enhancing the CAD performance.

Some work has been previously presented for the detection of haustral folds. Huang et al. [2] detected folds by estimating surface curvatures using patch methods. Lamy and Summers [3] relied on local estimation of principal curvatures and morphological operations to detect folds. Another method was proposed in [9] which also detected folds by estimating curvatures on the colonic wall and used that information to extract the teniae coli. A difference filter was proposed in [6] which computed the difference of CT values along the centerline of colon and used it to detect folds. Recently Chowdhury et al. [1] presented an

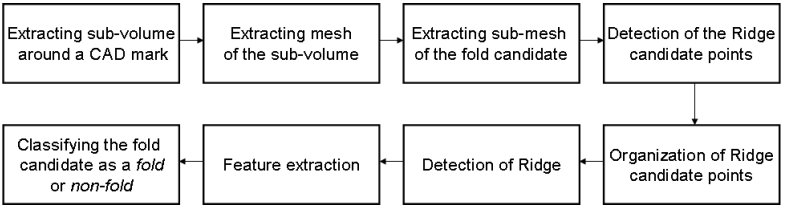


Figure 1: Block-diagram of the haustral fold detection method

approach which combined two methods (the first method used level sets, and the second used heat diffusion and fuzzy c-means algorithm) to detect folds.

Previous solutions to haustral fold detection rely primarily on local computations around a voxel in the image or mesh vertex, typically on the order of $1-2mm$ of resolution. However, haustral folds are elongated objects that are significantly larger, often having lengths in the range of $20-50mm$. In this paper, we demonstrate the advantage of our global approach that analyzes folds at this scale. We geometrically model the fold ridge, from which we derive a novel feature, the *density of ridge candidates* (DRC), which is shown to be highly effective in reducing FPs in CAD system. Unlike [10] which employed the ridge-line detection for the segmentation of folds, the proposed method is aimed more towards the analysis of folds, and additionally, it can filter out FPs significantly with no impact on the sensitivity.

2 Proposed method

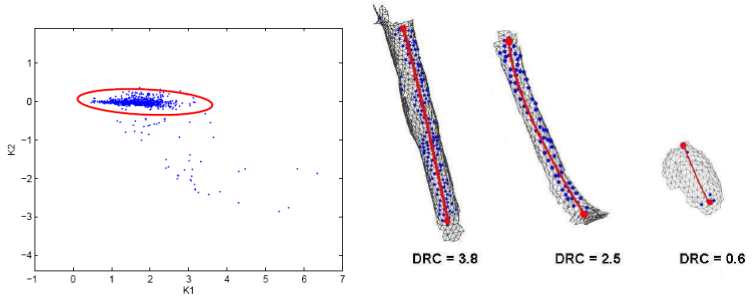
2.1 Extraction of fold candidates

Fig. 1 shows the block diagram of the proposed fold detection method. Given a CAD mark, a detected polyp location (a point) in the CT volume, the method first extracts a $50mm^3$ sub-volume around it for further processing. With a threshold of -800, we apply the marching cubes algorithm to the extracted image sub-volume, resulting in a triangular mesh. We work with the triangular mesh for two reasons. Firstly, since the folds are outward (when viewed endoluminally) protrusions of the colon surface, the change in their shape can easily be studied on a 2D manifold rather than the 3D volumetric CT data. Secondly, [8] showed that it is more accurate to compute surface curvatures on mesh than on the volumetric data.

Inspired by [5], for each vertex $\mathbf{v}$ of the mesh, the average distance of the points in the 2-ring of $\mathbf{v}$ from the plane running through $\mathbf{v}$ and orthogonal to its surface normal is computed. The distance is large where the mesh is curved, and low where the mesh is locally flat. Hysteresis thresholding is applied to this distance, producing fold candidates, and the one closest to the CAD mark is retained. In hysteresis thresholding, a fixed value of $lower\ threshold = 0.07$ is used and upper threshold is computed adaptively. The largest average distance value among that of all the vertices, which are less than five voxels away from the CAD mark, is chosen as the upper threshold.

2.2 Detection of ridge candidate points and modeling the fold ridge

To detect ridge candidate points on a fold candidate, a data-driven probabilistic model of the fold ridge is created by analyzing the behavior of principal curvatures (K_1 and K_2) on real fold ridges. This is done by manually annotating 732 points on 17 fold ridges and plotting



(a) (b)

Figure 2: (a) Distribution of K_1 and K_2 for annotated ridge candidate points (b) Examples of detected ridges. Ridge candidate points rendered in blue and ridges rendered in red color. Left and middle sub-meshes are haustral folds having greater DRCs, and the right sub-mesh is a polyp having a smaller DRC.

their K_1 and K_2 values (see Fig. 2(a)). There are some outliers in Fig. 2(a) which have been filtered out and the resulting distribution is fitted into a 2D Gaussian and is given by

$$P(K_1,K_2) = exp\left\{\frac{-1}{2(1-\rho^2)}\left[\left(\frac{K_1-m_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{K_1-m_1}{\sigma_1}\right)\left(\frac{K_2-m_2}{\sigma_2}\right) + \left(\frac{K_2-m_2}{\sigma_2}\right)^2\right]\right\}, \quad (1)$$

where $P(K_1,K_2)$ represents the probability of a mesh vertex to be a ridge candidate point given its K_1 and K_2, $(m_1,m_2) = (1.6787,-0.0175)$ is the mean point of the 2D Gaussian distribution, σ_1 and σ_2 are the standard deviations of K_1 and K_2 respectively, ρ is the correlation coefficient of K_1 and K_2, and the computed $Covariance = \begin{bmatrix} 0.2163 & -0.0023 \\ -0.0023 & 0.0063 \end{bmatrix}$. The vertices with $P(K_1,K_2) > \tau$ are considered as ridge candidate points where $\tau = 0.3$ and is chosen empirically. A decrease in the value of the τ leads also to the selection of those vertices as ridge candidate points that lie outside the vicinity of the ridge and an increase in the value of τ tends to discard vertices lying in the close vicinity of the ridge, both of which are undesirable. Ridge candidates points are unorganized 3D points. They are first organized and then fitted into a quadratic space curve to obtain the ridge.

We adopted a three-step strategy to organize these points. Firstly, we find the *start point* and the *end point* of a ridge based upon the Seeded Distance Transform. To find the start point, a ridge candidate point is randomly selected as a seed and the seeded distance transform is applied on the mesh. Seeded distance transform assigns a distance value to all other points depending upon their geodesic distance from the seed point. The farthest point from the seed is selected as the *start point*, as it will lie at one end of the ridge. Then, we repeat the same procedure again, selecting the *start point* as a seed now, to get the *end point* that lies at the other end of the ridge. Secondly, the geodesic path between the start point and the end point is computed based upon the A*-algorithm. Thirdly, all points are organized based upon their interpolated distance values along the geodesic path, which is found by projecting them on the geodesic path (closest projection is taken). The organized ridge candidate points can then be fitted to a quadratic space curve to obtain the geometric ridge model (see Fig. 2(b)).

2.3 Feature extraction and classification

From the ridge model, we extract two values that are combined to form the proposed feature, the *density of ridge candidates* (DRC). Specifically, we determine the ridge length L_R which is computed along the quadratic space curve, and the number of ridge candidates points N_c. Both L_R and N_c are very useful to analyze folds. L_R is typically higher for folds (as they are elongated structures) than that of polyps. Also, N_c is significantly higher for folds as compared to polyps. Thus, combining the information of L_R and N_c, we define the DRC as

$$DRC = \begin{cases} 0, & L_R < \tau_L; \\ \frac{N_c}{L_R}, & \text{otherwise.} \end{cases} \tag{2}$$

DRC is quite a discriminative feature as its value is considerably higher for folds as compared to polyps (see Fig 2(b)). It is important to note that DRC is set to zero for fold candidates with $L_R < \tau_L$, where $\tau_L = 27$ in this paper. This is done to minimize the chances of misclassifying polyps, as the fold candidates with smaller ridge lengths ($L_R < \tau_L$) are more likely to be polyps, hence, are allowed to pass as true positives (TPs).

3 Experimental results

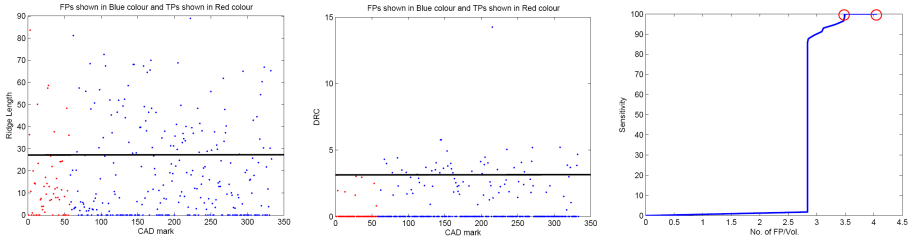
The proposed method is tested on 333 CAD marks containing both TPs (polyps) and FPs (including folds). CAD marks are detected by the CAD system described in [7], on a data from 34 patients for which we have 68 volumes (each having one prone and one supine volume). The dataset is obtained from seven different institutions around the world.

Using L_R alone as a feature can provide a significant FP reduction, but at the expense of some sensitivity, which infrequently results from large or long polyps. Fig. 3(a) demonstrates that some TPs have large ridge lengths and are more likely to be misclassified as folds. That is why, we combine with L_R the N_c in the DRC formulation as it is more robust in preserving the sensitivity while reducing a significant amount of FPs (Fig. 3(b)). Both in Fig. 3(a) and Fig. 3(b), red points are TPs and blue points are FPs. The FROC curve computed with the DRC feature is plotted in Fig. 3(c). The results are quite encouraging and a FP reduction of 14.13% is achieved with 100% sensitivity (see Fig. 3(c)).

The performance of the proposed method depends on how well the meshes of fold candidates are extracted, which is accurate for the majority of CAD marks in our experiments. However, in the case, where the extracted mesh of a fold candidate contains a polyp attached to a fold, DRC is found to misclassify it as a fold. This is because, in such a case, DRC value is high due to the detection of a large number of ridge candidate points N_c on the *fold-part* of the extracted mesh of the fold candidate.

4 Conclusions

In this paper, we have presented a method for haustral fold detection that can significantly reduce false positives (FPs). Unlike most of the existing methods that depend solely on *local* differential geometric computations, we have studied haustral folds on a much larger scale. We fit a global shape model to the local detections to estimate a ridge model of haustral folds capable of robust characterization. The overall performance of 14.13% FP reduction with no impact on sensitivity has been demonstrated on a sufficiently large dataset of 333 CAD marks resulting from 68 CT volumes. The focus of our future work is to further study the



(a) (b) (c)

Figure 3: (a) Ridge Lengths (L_R) of CAD marks (b) DRC values of CAD marks (c) FROC curve with the DRC feature ($\tau = 0.3$). The two points used to compute the 14.13% FP reduction are encircled in red color

generalization of the method so as to address the problem of misclassification of the cases where polyps are attached to folds.

References

[1] A. S. Chowdhury, S. Tan, J. Yao, and R. M. Summers. Colonic fold detection from computed tomographic colonography images using diffusion-FCM and level sets. *Pattern Recognition Letters*, 31(9):876–883, 2010.

[2] A. Huang, R. M. Summers, and A. K. Hara. Surface curvature estimation for automatic colonic polyp detection. In *Proc. of SPIE*, volume 5746, pages 393–402, 2005.

[3] J. Lamy and R. M. Summers. Teniae coli detection from colon surface: Extraction of anatomical markers for virtual colonoscopy. In *Proc. of ISVC*, pages 199–207, 2007.

[4] E. M. Lawrence, P. J. Pickhardt, D. H. Kim, and J. B. Robbins. Colorectal polyps: Stand-alone performance of computer-aided detection in a large asymptomatic screening population. *Radiology*, 256(3):791–798, 2010.

[5] K. Li, L. Guo, J. Nie, C. Faraco, Q. Zhao, S. Miller, and T. Liu. Gyral folding pattern analysis via surface profiling. In *Intl. Conf. on MICCAI*, 2009.

[6] M. Oda, T. Kitasaka, K. Mori, Y. Suenaga, T. Takayama, H. Takabatake, M. Mori, H. Natori, and S. Nawano. Haustral fold detection method for CT colonography based on difference filter along colon centerline. In *Proc. of SPIE*, volume 7260, 2009.

[7] G. Slabaugh, X. Yang, X. Ye, R. Boyes, and G. Beddoe. A robust and fast system for CTC computer-aided detection of colorectal lesions. *Algorithms*, 3(1):21–43, 2010.

[8] P. Sundaram. *Geometry Processing for Colon Polyp Detection*. PhD thesis, Stanford University, 2008.

[9] Y. Umemoto, M. Oda, T. Kitasaka, K. Mori, Y. Hayashi, Y. Suenaga, T. Takayama, and H. Natori. Extraction of teniae coli from CT volumes for assisting virtual colonoscopy. In *Proc. of SPIE*, volume 6916, 69160D, pages 1–10, 2008.

[10] H. Zhu, L. Li, Y. Fan, and Z. Liang. Haustral fold segmentation of CT colonography using ridge line detection. In *Proc. of MICCAI Workshop*, pages 33–39, 2010.