ASSESSMENT OF AN OPTICAL FLOW FIELD-BASED POLYP DETECTOR FOR CT COLONOGRAPHY

B. Acar[1], C.F. Beaulieu[1], D.S. Paik[1], S.B. Göktürk[2], C. Tomasi[3], J. Yee[4], S. Napel[1]

[1]Department of Radiology, LUCAS MRS Center, 3D Lab., Stanford University, Stanford, CA, USA
[2]Electrical Engineering Department, Stanford University, Stanford, CA, USA
[3]Computer Science Department, Stanford University, Stanford, CA, USA
[4]San Francisco VA Medical Center, UCSF, San Francisco, CA, USA

Abstract- **Most current computer-aided detection (CAD) algorithms for the fully automatic detection of colonic polyps from 3D CT data suffer from high false positive rates. We developed and evaluated a post-processing algorithm to decrease the false positive rate of such a method. Our method attempts to model the way a radiologist recognizes a polyp while scrolling a cross-sectional plane through 3D CT data by quantifying the change in location of the edges in 2D plane. It uses a classifier for identification based on the Mahalanobis distance. The new method increased the ROC curve area from 0.89 to 0.98 (an increase from 34.5% to 85.0% in specificity for 100% sensitivity) in a population of 8 patients.**
Keywords - **Virtual colonoscopy, Computed Tomographic Colonography, Polyp Detection, CAD, Optical Flow Fields**

I. INTRODUCTION

Computed Tomographic Colonography (CTC) is a minimally invasive method based on the examination of the colon using CT volume data [1]. This has been achieved by examining the stack of 2D images, 3D virtual colonoscopic views, or both. Recently, research on CTC has shifted toward developing fully automatic CAD methods for polyp detection. The following methods have been proposed for computer-aided detection (CAD) of colonic polyps: shape-based detection, Hough Transform (HT)- based detection and sphere fitting [2-7]. Despite their high sensitivity, these methods suffer from large false positive rates which require radiologists to examine a large number of images manually. Reducing the number of false positives would reduce the radiologists' reading time and help make CTC cost effective.

Our goal is to develop and validate a highly sensitive and specific CAD method, intended to be used as a post-processor to reduce false positives from a high sensitivity, low specificity polyp detector. We attempted to model the way a radiologist recognizes a polyp by scrolling through a stack of 2D CT images. We used Optical Flow Fields (OFFs) to represent changes in consecutive images and a linear classifier to sort true polyps from false positives based on features extracted from the computed OFFs.

II. METHODOLOGY

A. Pre-processing

The 3D CT data was pre-processed by custom software which automatically computes a path along the central colon axis stretching from the rectum to the cecum [8], identifies the colon wall, and uses a HT based poylp detector (HT) that detects spherical surface patches on the colon wall to localize polyp-like structures [3]. For computational efficiency, we extracted subvolumes of 21×21×21 voxels centered on the HT detected points (HT_hits) to use as inputs in the next stage.

B. Post-processing

Post-processing is comprised of OFF computation to represent the changes in the location of edges in the CT images (tissue/air boundaries) as one scrolls through the volume CT data, and the morphological characterization of the computed OFFs. Let xy-plane be the image plane perpendicular to the scrolling direction, z. The basic OFF equation for a constant z value, is [9]:

$$\nabla I \cdot \mathbf{v} + \frac{\partial I}{\partial z} = 0 \qquad (1)$$

where $\mathbf{v}(x,y)$ is along the local ∇I, $I(x,y)$ representing the image. As the scrolling is done from the center of the subvolume (HT_hit) outwards in both the $-z$ and $+z$ directions, the edges of spherical, polyp-like structures, move inwards on the image plane. This consistency in their motion is required as $\mathbf{v}_Z(x,y)$'s for all z are summed and smoothed to get a single OFF, $\overline{\mathbf{v}}_D(x,y)$, associated with the current subvolume and scrolling direction, as follows:

$$\overline{\mathbf{v}}_D(x,y) = \text{Smooth}\left(\sum_Z \mathbf{v}_Z(x,y)\right) \qquad (2)$$

The smoothing filter used is a 3×3 rectangular window acting on the image plane pixelwise. This is repeated for $D \in \{coronal, sagittal, axial\}$, resulting in 3 OFFs that encode information in 3 orthogonal scrolling directions.

A parent (Pn) and eight child nodes (Cn's) are determined on each OFF for its characterization. The Pn is defined to be the minimum divergence node in a 5×5 pixel neighborhood of the HT_hit. Cn's are defined to be the points 5 pixels away on the streamlines incoming to immediate neighbors of the Pn. Fig. 1 shows three axial CT images and the associated $\overline{\mathbf{v}}_{AXIAL}(x,y)$. The local Jacobian matrix of $\overline{\mathbf{v}}_D(x,y)$ is used to compute α and β parameters defined as follows [10]:

$$\mathbf{J} = \begin{bmatrix} \dfrac{\partial v_x}{\partial x} & \dfrac{\partial v_x}{\partial y} \\ \dfrac{\partial v_y}{\partial x} & \dfrac{\partial v_y}{\partial y} \end{bmatrix}, \qquad (3)$$

$$P = -trace(\mathbf{J}) \quad , \quad Q = |\mathbf{J}| \quad , \qquad (4)$$

$$\alpha = P \quad , \quad \beta = sign(P^2 - 4Q)\sqrt{|P^2 - 4Q|} . \qquad (5)$$

Axial Image #1

Axial Image #2
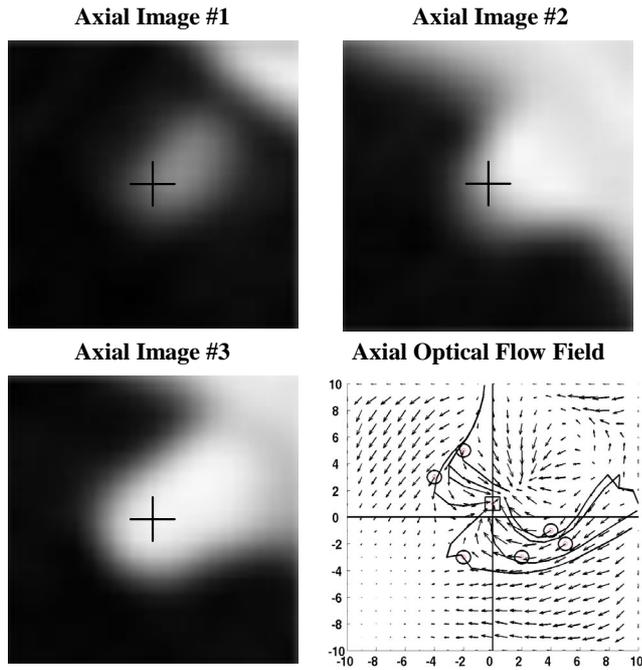
Axial Image #3

Axial Optical Flow Field

Figure 1. Three sequential axial images around a HT_hit (shown by +) and the associated axial optical flow field. The Pn is marked with a small square, the Cn's are marked with small circles on the streamlines.

The parameters α and β depend on the local properties of the OFF. Their ratio uniquely describes the local topology. They essentially carry the information in the eigenvalues of the characteristic equation of $\mathbf{J}$.

The behavior of the streamlines is quantified by using two parameters that describe the spread of the locations of Cn's around the Pn. The first parameter, d, is defined as

$$d = \frac{1}{8}\sum_i \left(\sqrt{\sum_j \theta_{ij}^2} \right), \theta_{ij} = \angle(\mathrm{Cn}_i, \mathrm{Cn}_j) \in [0, \pi] \quad (6).$$

The angle, θ_{ij}, is calculated with respect to the Pn's location.

The second parameter, Sc, is the angle of the largest sector around the Pn that does not contain a Cn.

Each parameter is calculated for the 3 orthogonal OFFs, making up a 12 dimensional feature vector.

B. Classification

A Mahalanobis distance based classifier is used for learning and classification [11]. The Mahalanobis distance of a feature vector $\mathbf{f}$ to the mean vector, $\mathbf{m}_\Gamma$, of a training set, Γ, is defined as,

$$r_{\mathbf{f}, \mathbf{m}_\Gamma} = \sqrt{(\mathbf{f} - \mathbf{m}_\Gamma)^{\mathrm{T}} \mathbf{C}_\Gamma^{-1}(\mathbf{f} - \mathbf{m}_\Gamma)} \quad (7)$$

where $\mathbf{C}_\Gamma$ is the covariance matrix of the training set. This distance, r, is a standardized measure which *(i)* automatically accounts for scaling, *(ii)* takes care of correlations between features, *(iii)* can provide linear and curved decision surfaces. Referring to the subset of polyps in the training set Γ as $\Gamma 1$ and the subset of non-polyps as
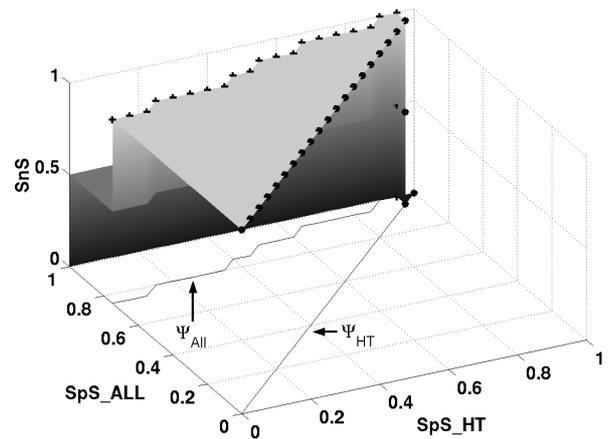

Figure 2. ROC surface for a single experiment. The points on the ROC surface corresponding to ψ_{ALL} (ψ_{HT}) are marked with + (*).

$\Gamma 0$, the feature vector $\mathbf{f} \in \Omega$, where Ω is the test set, is classified as

$$r_{\mathbf{f}, \mathbf{m}_{\Gamma 1}} - r_{\mathbf{f}, \mathbf{m}_{\Gamma 0}} + b \leq 0 \;\Rightarrow \mathbf{f} \in \Omega 1$$
$$\text{otherwise} \qquad\qquad \Rightarrow \mathbf{f} \in \Omega 0 \qquad (8)$$

where b is a bias term and $\Omega 1$ and $\Omega 0$ are polyp and non-polyp subsets of Ω respectively.

C. Preliminary Evaluation

Data were acquired from 8 patients (7 male, age 41-85, mean age 63) in supine position after colon cleansing and air-insufflation. Imaging parameters were 3mm (2.5mm) collimation, pitch 1.5-2.0 (3.0), 1.5mm (1.0-1.5mm) reconstruction interval, 200 MAs (56 MAs), and 120kVp, on a single- (multi-) detector CT system (GE Medical Systems, Milwaukee, WI). Between 236 and 403 512×512 images were reconstructed for each case with an average voxel size of 0.74mm × 0.74mm × 1.31mm. These patients contained 19 polyps (5.0 to 23.0 cm. diam.) confirmed by fiberoptic colonoscopy.

The HT detector produces a score related to the size and sphericity of each HT_hit. We excluded all HT_hits with a score below an arbitrary threshold $h = 10000$, resulting in 219 HT_hits (19 true polyps, 200 false positives). The subvolumes (21×21×21 voxels) around each HT_hit were used as the input data to the OFF-based classifier.

The preprocessed data set was divided into 10 subsets randomly such that the true positive and the false positive HT_hits were distributed uniformly. We conducted a 10-fold cross validation study in which one of these subsets was used as the test set and the rest as the training set for each experiment.

The sensitivity and specificity are defined as the percentages of the correctly detected polyps and non-polyps, respectively, within the test set. A multidimensional ROC surface is constructed for each experiment as follows: The HT_hits corresponding to a threshold, h, are fed into the OFF-based classifier (OFFC) and an ROC curve is computed

F#	Param. Used	Mean	Median	MinMax	Sorted	LeftMost
		A	B	C	D	E
HT				0.89		
f1	[α β d]	0.98	0.97	0.96	0.96	0.97
f2	[α β]	0.98	0.97	0.96	0.97	0.97
f3	[α d]	0.97	0.97	0.97	0.96	0.97
f4	[β d]	0.96	0.95	0.95	0.95	0.94
f5	[α β Sc]	0.97	0.97	0.97	0.96	0.97
f6	[α Sc]	0.98	0.97	0.97	0.97	0.97
f7	[β Sc]	0.95	0.96	0.95	0.96	0.94

Table 1. ROC areas for different feature sets: i) 'Mean' and 'Median' mean that the average or median value of each parameter among 3 directions is used, ii) 'MinMax' means that the minimum(α), minimum($|\beta|$), maximum(d) and minimum(Sc) are used, iii) 'Sorted' means that all parameters for all 3 directions are used after sorting, i.e. The feature vectors are 9D or 6D depending on the number of parameters used, iv) 'LeftMost' means that the scrolling direction with a leftmost (α,β) point (minimum $\cos(\alpha / \|[\alpha\ \beta]\|)$) on $\alpha\beta$-plane is used.

by varying b. This is repeated for all h, $h \in [10000, \max\_HT\_score]$. Varying h relates SnS to SpS_HT, where SnS and SpS_HT represent the sensitivity and specificity, respectively, of the HT detector at each h. Now consider a given SpS_HT obtained for some threshold h. Varying b results in a new ROC curve relating SnS to SpS_ALL. It represents the sensitivity and specificity of the combination HT followed by OFFC. SnS on this new ROC curve represents the sensitivity of that combination and SpS_ALL represents its specificity. Plotting these these curves for every SpS_HT constructs an ROC surface in 3D that depicts SnS as a function of SpS_HT and SpS_ALL. Note that the surface is not defined for SpS_ALL less than SpS_HT because the OFFC cannot increase the number of false positives provided by the preceding HT stage. The SnS versus the SPS_HT as one reads while walking along the (SpS_ALL=SpS_HT) line (ψ_{HT}) gives the ROC curve of HT alone (ROC_HT). If the set of points ψ_{ALL} are defined as,

$$\Psi_{ALL} = \left\{ SpS_1, SpS_2 \mid \right.$$
$$SpS_2 = \max_{SpS\_ALL} \left(SnS(SpS_1, SpS_1) = SnS(SpS_1, SpS_2) \right),$$
$$\left. \forall SpS_1 \right\} \tag{9}$$

then these points correspond to the maximum specificity one can get by post-processing the HT_hits(h), without sacrificing SnS. In other words, SnS versus the SpS_ALL as one reads while walking along ψ_{ALL} gives the ROC curve of the overall system (ROC_ALL). Further more, the area between ψ_{ALL} and ψ_{HT} depicts the improvement in specificity due to OFFC. Fig. 2 shows a ROC surface and the associated ψ_{HT} and ψ_{ALL}.

ROC_HTs and ROC_ALLs, thus computed for each of 10 experiments, are averaged in both sensitivity and specificity to get mROC_HT and mROC_ALL. Note that every point on an ROC_ALL corresponds to a (h, b) pair. However, while the corresponding points on different ROC_ALLs have the same h, this is not necessarily true for b, because for each point of an ROC_ALL, b is selected such that the specificity
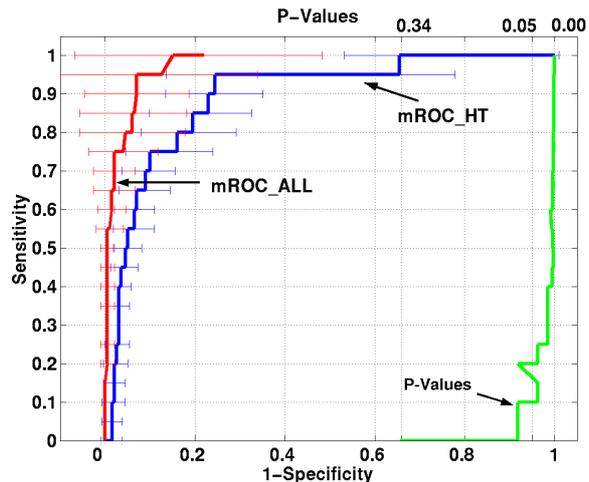

Figure 3. The final ROC curves of the HT alone (mROC_HT) and the full system (mROC_ALL), using **f**1A, averaged over 10 cross-validation experiments. The error bars indicate the variation among 10 experiments. The p-values indicate the significance of difference between the two curves (p-value<0.05 is taken to be statistically significant)

is maximum without sacrificing the sensitivity. So, despite the fact that the term "ROC curve" is not totally valid for mROC_ALL, it has similar properties in terms of allowing the comparison of sensitivity and specificity. While the true detection performance is represented by an average ROC surface, this does not allow us to assess the improvement in specificity, as further discussed in Section IV. Our assessment is based on the comparison of these curves for different feature vectors listed in Table 1 and the curve for HT alone. We compared the ROC areas and the results of the t-tests performed for each point on the mean ROC curves separately, with a null hypothesis that the mean of the specificity differences between two mean ROC curves (mROC_HT and mROC_ALL) for the corresponding sensitivity is zero over 10 cross-validation experiments. Thus the p-value for a certain sensitivity value assesses the significance of the difference between specificity values of two methods at that sensitivity.

III. RESULTS

Table 1 summarizes the ROC areas obtained for different feature vectors and the HT alone. The ROC area for the HT alone was 0.89. There is not much difference between the ROC areas for different choices of feature vectors (Mean ROC area = 0.96±0.01). However, there is an average difference of 0.07 between the areas of these curves and that of the HT alone. The point-wise t-tests between all feature sets and the HT alone resulted in mean p-values (averaged over the whole spectrum of sensitivity values) less than 0.009 (maximum mean p-value=0.009 was between HT alone and **f**1C). This means that on the average the null hypothesis (i.e. the mean difference between the specificities of the ROC curves is zero) can be rejected for each sensitivity value. Actually, a closer look at these p-values for individual comparisons show that all p-values are less than 0.05 for sensitivities greater than or equal to 0.25. Fig. 3 shows the
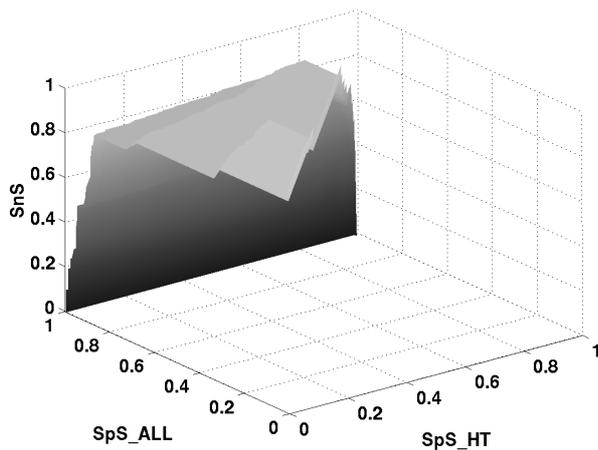
Figure 4. The average ROC surface, computed over 10 cross-validation experiments. The volume under the ROC surface is 0.82

result of one such comparison performed between the HT alone and **f**1A.

IV. DISCUSSION

These results suggest that the ROC area increases due to the OFF-based post-processing algorithm and that the difference between the ROC curves before and after post-processing is statistically significant. There is a consistent decrease in the ROC area when the parameter α is excluded (except for **f**7C), suggesting that α carries the most information. The parameters d and Sc do not seem to carry much information.

Based on these results, **f**1A seems to be the best choice because it has a large ROC area, and it uses three parameters designed to represent different qualities of the OFFs. Fig. 3 shows the mROC_HT and mROC_ALL using **f**1A, together with the p-values computed for every point on ROC curves, over the set of 10 experiments. For **f**1A, the OFF classifier increased the ROC area from 0.89 to 0.98. At a sensitivity level of 100% (95%) the specificity increased from 34.5% (75.5%) to 85% (93%), which corresponds to a decrease in the average number of false positives per test set from 13.1 (4.9) to 3.0 (1.4).

As mentioned at the end of Section II, mROC_ALL is not a conventional ROC curve. It represents the average improvement in specificity without decreasing sensitivity. It should be interpreted as the average of the maximum specificity achievable by the full system for a given sensitivity. The performance of the overall system is actually given by the ROC surfaces as in Fig. 2. The average of these surfaces for **f**1A, computed in all three dimensions point-wise, is given in Fig. 4. The ROC volume, as a percentage of the maximum possible volume (which is 0.5 as the ROC surface is defined in half of the whole volume) is 0.82.

V. CONCLUSION

Our method for computer-aided detection of colonic polyps was inspired by the processes radiologists go through

while investigating CT colonography data. Namely, they scroll through the 3D CT data in a certain direction(s), usually only perpendicular to the axial plane but also perpendicular to the coronal and sagittal planes when necessary, following the edge information in 2D images. In our algorithm, this information is represented as an OFF vector field and the resultant OFF is characterized by topology dependent parameters. We showed that our post-processing algorithm improves the specificity of a high sensitivity, low specificity CAD without sacrificing sensitivity. The design of OFFC does not assume anything about the nature of the pre-processing, and thus we expect it to be applicable to any high sensitivity, low specificity CAD output. Our results show that there is relevant information in the motion of edges in terms of polyp detection in CTC. Further research is required to validate the presented method and/or develop new ways of representing this information.

REFERENCES

[1] D.J. Vining, "Virtual colonoscopy," *Gastrointest. Endosc. Clin. N. Am.*, vol. 7(2), pp. 285-291, 1997.

[2] R.M. Summers, C.F. Beaulieu, L.M. Pusanik, J.D. Malley, R.B. Jeffrey, D.I. Glazer, S. Napel, "Automated polyp detector for CT colonography: Feasibility study," *Radiology ,* vol. 216(1), pp. 284-290, July 2000.

[3] D.S. Paik, C.F. Beaulieu, R.B. Jeffrey, J. Yee, A.M. Steinauer-Gebauer, S. Napel, "Computer aided detection of polyps in CT colonography: Method and free-response ROC evaluation of performance," *Radiology,* vol. 217(SS), pp. 370-370, November 2000.

[4] H. Yoshida, Y. Masutani, P.M. MacEneaney, K. Doi, Y. Kim, A.H. Dachman, "Detection of colonic polyps in CT colonography based on geometric features," *Radiology,* vol. 217(SS), pp. 582-582, November 2000.

[5] Y. Masutani, A.H. Dachman, P.M. MacEneaney, Y. Kim, H. Yoshida, "Computerized detection of polyps in CT colonography based on volumetric curvature information," *Radiology,* vol. 217(SS), pp. 486-486, November 2000.

[6] R.M. Summers, C.D. Johnson, L.M. Pusanik, J.D. Malley, A.M. Youssef, J.E. Reed, "Automated polyp detection at CT colonography: feasibility assessment in a human population," *Radiology,* vol. 219, pp. 51-59, April 2001.

[7] S.B. Göktürk, C. Tomasi, "A graph method for the conservative detection of polyps in the colon," *2nd International Symposium on Virtual Colonoscopy*, Boston, October 2000.

[8] D.S. Paik, C.F. Beaulieu, R.B. Jeffrey, G.D. Rubin, S. Napel, "Automated flight path planning for virtual endoscopy," *Medical Physics,* vol. 25(5), pp. 629-637, 1998.

[9] S.S. Beauchemin, J.L. Barron, "The computation of optical flow," *ACM Computing Surveys,* vol. 27(3), pp. 433-467, 1996.

[10] Y. Lavin, R. Batra, L. Hesselink, "Feature comparison of vector fields using earth mover's distance," in *Proc. IEEE/ACM Visualization'9,* North Carolina, USA, October 1998.

[11] P.C. Mahalanobis, "On the generalized distance in statistics," *Proc. Natl. Institute of Science of India,* vol. 12, pp. 49-55, 1936.