Polyps Flagging in Virtual Colonoscopy

Marcelo Fiori[1], Pablo Musé[1], and Guillermo Sapiro[2]

[1] Universidad de la República, Uruguay
[2] Duke University, Durham, NC 27708, USA

Abstract. Computer tomographic colonography, combined with computer-aided detection, is a promising emerging technique for colonic polyp analysis. We present a complete pipeline for polyp detection, starting with a simple colon segmentation technique that enhances polyps, followed by an adaptive-scale candidate polyp delineation and classification based on new texture and geometric features that consider both the information in the candidate polyp and its immediate surrounding area. The proposed system is tested with ground truth data, including challenging flat and small polyps. For polyps larger than $6mm$ in size we achieve 100% sensitivity with just 0.9 false positives per case, and for polyps larger than $3mm$ in size we achieve 93% sensitivity with 2.8 false positives per case.

1 Introduction

Colorectal cancer is nowadays the third leading cause of cancer-related deaths worldwide. The early detection of polyps is fundamental, allowing to reduce mortality rates up to 90%. Nowadays, optical colonoscopy (OC) is the most used detection method due in part to its high performance. However, this technique is invasive and expensive, making it hard to use in large screening campaigns.

Virtual Colonoscopy (VC) is a promising alternative technique that emerged in the 90's, which uses volumetric Computed Tomographic data of the cleansed and air-distended colon. It is less invasive than optical colonoscopy, and much more suitable for screening campaigns once its performance is demonstrated.

However, it takes more than 15 minutes for a trained radiologist to complete a VC study, and the overall performance of OC is still considered better. In this regard, Computer-Aided Detection (CAD) algorithms can play a key role, assisting the expert to both reduce the procedure time and improve its accuracy.

Flat polyps (those having $< 3mm$ of elevation above the mucosa) and "small" polyps are of special interest because these are an important source of false negatives in VC, and many authors claim that flat polyps are around 10 times more likely to contain high-grade epithelial dysplasia [1]. The goal of this work is to exploit VC to automatically flag colon regions with high probability of being polyps, with special attention to challenging small and flat polyps.

Automatic polyp detection is a very challenging problem, not only because the polyps can have different shapes and sizes, but also because they can be located in very different surroundings. Most of the previous work on CAD of colonic

J. Ruiz-Shulcloper and G. Sanniti di Baja (Eds.): CIARP 2013, Part II, LNCS 8259, pp. 181–189, 2013.

polyps consists in a segmentation step followed by a classification stage based on geometric features, some using additional texture information, but none of them takes into account the information of the tissues *surrounding* the polyp. On the other hand, for the segmentation step, not much work has been done in comparing the smoothing techniques to see which one is more adapted to polyp detection. To the best of our knowledge, no algorithm reported in the literature can detect small polyps properly, and for polyps larger than $6mm$ in size, no algorithm can achieve 100% sensitivity with less than one false positive per case.

The proposed system is illustrated in Figure 1, and consists of the following steps: colon segmentation, an adaptive-scale search of candidates in order to capture the appropriate size, computation of geometrical and textural features, and a machine learning algorithm to classify patches as polyps or normal tissue.

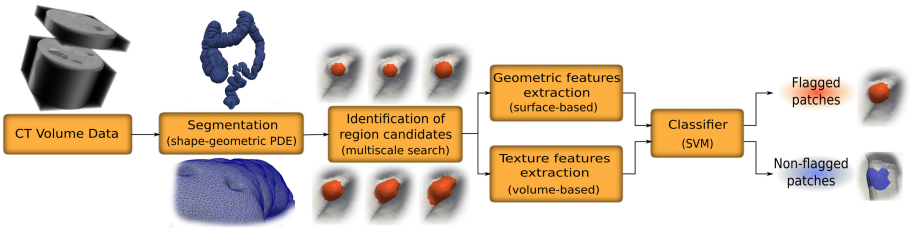

Fig. 1. Basic pipeline of the proposed polyp flagging system

2 Summary of the Colon Segmentation Method

The segmentation of the colon surface, which is critical in particular to compute geometric features, is divided into two parts: a pre-processing stage for dealing with the air-fluid composition of the colon volume, and a second stage that consists on smoothing the pre-processed image and obtaining the final colon surface by thresholding the smoothed volume. More details are available in [2].

Classifying CT Regions

All the database cases have the same preparation, which includes solid-stool tagging and opacification of luminal fluid. Figure 2 shows a CT slice and its pixel values over the highlighted vertical profile. There are 3 clearly distinguishable classes: lowest gray levels correspond to air, highest levels to fluid, and middle gray values to tissue. However, there are around 6 interface voxels between air and fluid whose gray values lie within the normal tissue range. Therefore, a naïve approach is not suitable for tissue classification. We propose to compute a volume u_0 intended to have homogeneous values in the colon interior and exterior, and a smooth transition between them. To do that, we assign to each voxel the likelihood of being air, fluid, or air-fluid interface. Air and fluid distributions are estimated using standard kernel density estimation methods; these functions are then used to assign air and fluid likelihood values to the voxels.

Note that this assignment fails on the air-fluid and air-fluid-tissue interfaces. For assigning a value to these voxels, we take advantage of the physics of the

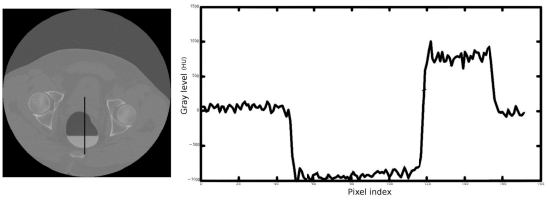

Fig. 2. CT slice and its different gray values for air, fluid and normal tissue, along the vertical profile

problem: the subject is laid horizontally so the interface between the fluid and the air is a plane parallel to the floor. The voxels situated on the interface then have a large gradient in the vertical direction. The implementation of these criteria is as follows. A cubic neighborhood around each voxel $\mathbf{x}$ is considered, and for each one of the "columns" that result of fixing the x and y coordinates, the air-likelihoods of the upper voxels and the fluid-likelihoods of the lower voxels are accumulated. The value $IC(\mathbf{x})$ that represents the confidence level of $\mathbf{x}$ being an interface voxel is then an increasing function of this accumulated measures.

We then assign to the initial segmentation u_0 the maximum of these three values, namely, the air and fluid likelihoods and the interface confidence level.

It is not rare that segmentation algorithms result in "gutter-like" shapes along the air-fluid-tissue interface. This is a critical point because of the potential of yielding several FPs in the detection step. If small oscillations occur along the "gutter" (which is expectable), artifacts with polyp-like shape are produced, thus degrading the overall performance. We paid special attention to this issue: the IC computation allows to avoid these artifacts. Figure 2 shows the comparison of our segmentation with a version of the method without the IC computation.

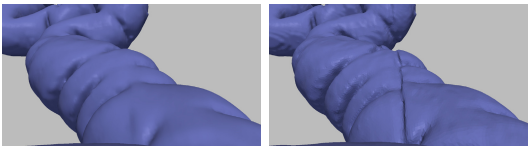

Fig. 3. Comparison of reduced artifacts in our segmentation (left) with a previously tested more standard version (right)

Smoothing and Colon Surface Computation

In order to eliminate noise and to obtain a smoother colon surface after the segmentation stage, we proceed to smooth the initial segmentation u_0. We derive a PDE-driven smoothing technique that preserves the shape of the polyps, while obtaining a smooth enough surface to reliably compute local geometric features.

We concentrate on a family of smoothing PDEs of the form

$$\frac{\partial u(\mathbf{x},t)}{\partial t} = \beta|\nabla u| \quad , \quad u(\mathbf{x},0) = u_0(\mathbf{x}) \ , \tag{1}$$

where the initial volume u_0 results form the preprocessing described in the previous section. After a few iterations of this evolution, the inner colonic wall will be extracted as a suitable iso-level surface of the resulting $3D$ image $u(\mathbf{x},T)$.

We recall that the Level Set Method [3] states that if $u(\mathbf{x},t)$ evolves according to (1), then its iso-levels (level sets) satisfy $\frac{\partial \mathcal{S}}{\partial t} = \beta \boldsymbol{\mathcal{N}}$, where $\mathcal{S}$ is any iso-level

surface and $\mathcal{N}$ its unit normal. This geometric view enables to design β to fulfill a set of requirements we will impose to the surface evolution. In particular, we are interested in motions driven by the principal curvatures κ_{max} and κ_{min}.

With the mean curvature motion ($\beta = \mathcal{H}$), and the affine motion ($(K^+)^{1/4}$), the polyps are flattened too fast [2]. As an alternative, a motion that seems to be well suited for our problem is the motion by minimal curvature. Indeed, polyps have a curve of inflection points all around it, separating its upper and lower sections. Along this curve, the minimal curvature is $\kappa_{min} = 0$, and therefore this part of the polyp does not move (or moves very slowly), so intuitively under this motion the polyps should persist longer. This PDE already yields very good results in terms of both surface smoothing and polyp enhancement.

We further derive two modifications that lead us to the proposed PDE. The first one is inspired by the exponent $1/4$ of the affine motions in dimension 3: $\frac{\partial \mathcal{S}}{\partial t} = \kappa_{min}^{1/4} \mathcal{N}$. Figure 4 shows the result after a few iterations, and Figure 5 evidences the difference between the motions by κ_{min} and $\kappa_{min}^{1/4}$ (gray and orange respectively) with a comparative image. On the polyp protrusion, the orange surface is above the gray one, while the opposite is observed in the surrounding area, showing that the evolution by $\kappa_{min}^{1/4}$ leads to better polyp enhancement.

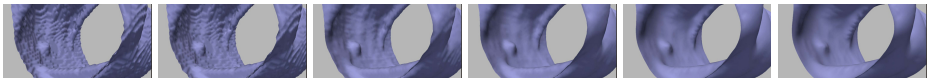

Fig. 4. Evolution by $\kappa_{min}^{1/4}$: original surface and result after 2, 8, 15, 30 and 50 iterations

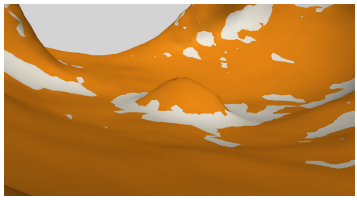

Fig. 5. Comparison between evolutions. Motion by k_{min} in light gray vs. motion by $k_{min}^{1/4}$ in dark gray. Both surfaces are overlaid, so sections that are not visible are hidden below the other surface.

The second modification is based on the idea of preserving the polyps qualities that we later use to identify them. A measure of the local shape of a surface is the so-called *shape index* SI, and the complementary *curvedness* C [4]:

$$SI := -\frac{2}{\pi} \arctan\left(\frac{\kappa_{max} + \kappa_{min}}{\kappa_{max} - \kappa_{min}}\right) \quad , \quad C := \frac{2}{\pi} \ln \sqrt{\frac{\kappa_{max}^2 + \kappa_{min}^2}{2}} \; .$$

While the value of SI is scale-invariant and measures the local shape of the surface, the value of C indicates how pronounced it is. We now include this information in order to make potential polyps evolve differently than the rest of the colon surface. We define a function of the shape index that acts as a multiplying factor to the term $\kappa_{min}^{1/4}$, making the surface evolve slower at the interest points. These function should assign low values to shape index near -1, and values close to unity to other points. A smooth function $g(SI)$ verifying these constraints is $g(SI) = \frac{1}{\pi} \arctan\left((SI - 0.75) \cdot 10\right) + \frac{1}{2}$.

The final evolution keeps all the advantages of the motion by $\kappa_{min}^{1/4}$ and in addition, polyps are flattened slower:

$$\frac{\partial \mathcal{S}}{\partial t} = g(SI)\,\kappa_{min}^{1/4}\boldsymbol{\mathcal{N}}\ . \tag{2}$$

The number of iterations can be set by experimentally choosing the value that maximizes the overall performance of the system, measured in terms of the free-response ROC curve (FROC). Alternatively, we can consider a sphere of the size of the CT resolution and compute analytically the number of iterations needed to make it vanish (see [2]). These two approaches led to the same result, namely 15 iterations, and therefore this is the chosen value for the experiments.

At this point we have a smoothed volume $u(\mathbf{x}, T)$ indicating the volume inside of the colon. We then extract the surface of the colon as the iso-value surface of level $\alpha \in [0, 1]$. The choice of the value α can be made by maximizing some criteria, in order to obtain the most contrasted surface in a given sense.

3 Polyp Delineation, Feature Extraction and Classification

All the polyp detection methods reported try to classify polyps from properties defined only within the candidate region. However, it is important to analyze the spatial context in which the candidate patch is located, not only because different sections of the colon present different characteristics, but also because polyps can be situated over different structures such as folds or plain colonic wall. In this regard, most of the features here described take into account the information of the area surrounding the candidate patch. This makes the features more robust to the local phenomena. The normal tissue of different cases may vary, so absolute thresholds lack meaning; while texture patterns differ from study to study, what does not vary is the fact that polyps have different properties than normal tissue.

Candidate Detection and Geometrical Features
Consider the shape index as a function $SI : \mathcal{S} \to [-1, 1]$, and recall that the polyps' SI are close to -1. Therefore, a region of the surface corresponding to a polyp has at least one local minimum of SI. Detection of candidate patches follows an adaptive-scale search: for each local minimum $x_0 \in \mathcal{S}$ of the function SI, several level sets of SI $(\mathcal{P}_1 \dots \mathcal{P}_n)$ around x_0 are tested, and the level set $\mathcal{P}_i$ that maximizes the distances between the histograms described below, is the considered candidate patch, denoted by $\mathcal{P}$ (Fig. 6). A total of $n = 7$ level sets are tested, corresponding to SI values from -0.8 to -0.5 with a 0.05 step. The following description is given for the final chosen patch $\mathcal{P}$, but the computations are made for all the level sets $\mathcal{P}_i$ in order to select the most appropriate one.

Given a candidate patch $\mathcal{P}$, a ring $\mathcal{R}$ around $\mathcal{P}$ is computed, in order to consider geometrical measurements with respect to the area surrounding the patch. The ring is calculated by dilating the patch $\mathcal{P}$ a certain geodesic distance, such that the areas of $\mathcal{P}$ and $\mathcal{R}$ are equal, see Figure 7.

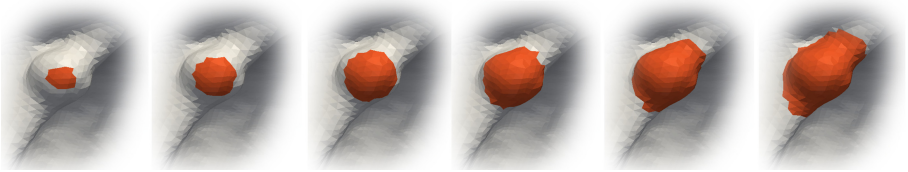

Fig. 6. $\mathcal{P}_1 \dots \mathcal{P}_n$: different sizes are tested in order to select the most appropriate patch

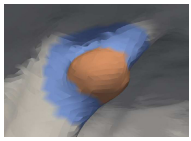

Fig. 7. Ring, in blue, surrounding a candidate polyp (actually a true polyp), in orange

Histograms of the shape index values are then computed for the patch $\mathcal{P}$ and the ring $\mathcal{R}$, and two different distances between them are computed: the L_1 distance and the symmetric Kullback-Leibler divergence. If the patch corresponds to a polyp-like shape then the values of the $\mathcal{P}$ histogram will be concentrated around -1. The histogram of $\mathcal{R}$ will be concentrated near 1 in case of a polyp on a normal colon wall (concave), or around -0.5 if the polyp is on a fold. These two features give a measure of the geometric local variation of the candidate patch $\mathcal{P}$. Although these two distances are the most discriminative features, we also consider the following ones since they help discriminating typical false positives:

- The mean value of the shape index over the patch $\mathcal{P}$.
- The area of the patch.
- The growth rate at the adaptive-size stage, meaning the ratio between the area of the chosen patch $\mathcal{P} = \mathcal{P}_i$ and the area of the immediately smaller patch $\mathcal{P}_{i-1}$; this feature measures how fast the shape of the patch is changing.
- And finally the *shape factor* $SF = \frac{4\pi \cdot Area}{Perimeter^2}$, which measures how efficiently the perimeter is used in order to gain area. It favors circle-like patches (like the polyp in Fig. 7), avoiding elongated patches (like false positives in folds).

We then end-up with a total of 6 geometric features.

Texture Features

There is evidence that the gray-level of the CT image and its texture can be very helpful for detecting polyps. This is particularly useful for flat or small polyps, where geometric information is limited [5]. Some work has been done on the inclusion of texture features (inside the candidate polyps only), in order to reduce false positives [6]. We propose both the use of new texture features and the inclusion of the information on the candidate's surrounding area.

First, for each polyp candidate $\mathcal{P}$, a volume V_1 is computed, containing the patch $\mathcal{P}$ and a portion of the inner tissue bounded by the patch. Volume V_1 is obtained by dilating $\mathcal{P}$ (in 3D) towards the inner colon tissue. A second volume V_2 surrounding V_1 is computed dilating V_1. The tissue in V_2 is intended to be

normal, to be compared with the polyp candidate tissue. The dilation is chosen as before: several distances are tested, keeping the most discriminative one.

The chosen texture features are a subset of the classical Haralick features, namely, entropy, energy, contrast, sumMean, and homogeneity. Seven co-occurrence matrices are computed with the voxels of V_1, and the five features are averaged over the seven directions. The analogous computation is made for V_2, and the differences between the two volumes, for each texture feature, are considered. Additionally, the mean gray levels in both volumes is computed, and their difference is considered as a feature. In this way, six texture features are considered.

Classification

Once the the candidates detection has been performed, the number of true polyps was much lower than the number of non-polyps patches, a relation on the order of 500:1, which is a significant problem for the learning stage of the classifier, since most classifiers are designed to maximize the accuracy, which is not adequate for imbalanced problems [7]. For instance, if we classify all candidates as "non-polyps," we would get an accuracy of 99.8% but without detecting any polyps. Three techniques were considered to overcome this problem: MetaCost, Cost Sensitive Learning (CSL), and Synthetic Minority Over-sampling TEchnique (SMOTE). The best results were obtained with CSL+SVM.

4 Results

A total of 150 patients of the Walter Reed Army Medical Center database [8] were used to test our CAD algorithm. The database contains 134 polyps detected by OC, including 12 flat polyps. Among these 134 polyps, 86 are larger than $6mm$, and 48 range from $3mm$ and $6mm$. The evaluation was carried out by splitting the dataset into halves, training and testing. Under this setting, classification with CSL+SVM yields the FROCs in Fig. 8, which shows the performance for different polyps sizes. These values are comparable with state-of-the-art results [6,9], however our study includes very small polyps. A more precise comparison is not necessarily meaningful, since in general each work considers its own database.

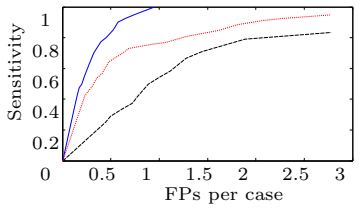

Fig. 8. FROC of our method for different polyps sizes: larger than $6mm$ (solid), smaller than $6mm$ (dashed), and all polyps (dotted)

The FROCs in Figure 9 compare the performance of our system when using different smoothing schemes (Section 2). The chosen one yields the best results.

The FROCs in Fig. 10 compare the influence of absolute and differential texture features. The classification was performed using all the geometric features, and either absolute (computed just for V_1) or differential texture features.

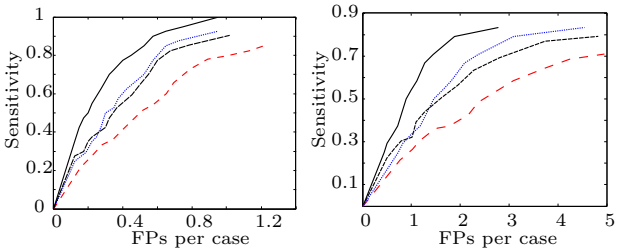

Fig. 9. FROCs comparing different smoothing methods, classifying large (left) and small polyps (right). The curve for the proposed evolution is shown in solid line, the results for the evolution by $\mathcal{H}$ and κ_{min} are shown in dotted and dashed lines respectively, and the lower curve is the result when no smoothing is performed.

The results show that differential texture features are more discriminative than the absolute ones. Finally the FROCs in Fig. 11 compare the results of different classification approaches. CSL, SMOTE, and MetaCost were used as a pre-processing stage for SVM, and C4.5 trees stabilized with AdaBoost. Parameters in all classifiers were optimized via cross validation.

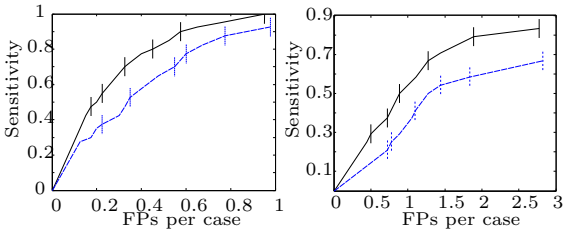

Fig. 10. FROCs (95% confidence intervals), comparing the performance with differential (solid) and absolute (dashed) texture features, for polyps larger (left) and smaller (right) than 6mm

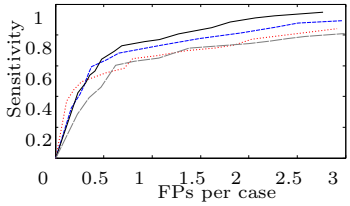

Fig. 11. FROCs comparing the performances of different classification approaches. SVM+CSL (solid), SVM+SMOTE (dashed), C4.5+AdaBoost (dotted) and plain SVM (long-dashed).

5 Conclusion

We introduced a complete pipeline for a Computer Aided Detection algorithm that flags candidate polyp regions. The segmentation stage is very simple and fast, and its main novelty is the smoothing PDE which enhances the polyps, enabling better detection rates. In addition to the incorporation of the Haralick texture features, the main yet simple novelties of the proposed features and classification stages are twofold. First, the surrounding area of candidate polyps are explicitly taken into account. Indeed, the proposed (so-called differential)

features are computed by comparing properties in the central and surrounding regions of the polyps. We show that differential features are more discriminative than the absolute ones, as they emphasize local deviations of geometry and texture over the colon. The other novelty is an adaptive-scale strategy that test regions of different sizes and automatically selects the region that best delineates each candidate polyp. The obtained quantitative results are very promising.

References

1. Fidler, J., Johnson, C.: Flat polyps of the colon: accuracy of detection by CT colonography and histologic significance. Abdom Imaging 34(2), 157–171 (2009)
2. Fiori, M., Musé, P., Sapiro, G.: A complete system for candidate polyps detection in virtual colonoscopy (2012), http://arxiv.org/abs/1209.6525
3. Osher, S., Sethian, J.: Fronts propagating with curvature- dependent speed: Algorithms based on Hamilton-Jacobi formulations. J. Comput. Phys. 79, 12–49 (1988)
4. Koenderink, J.J.: Solid Shape. MIT Press, Cambridge (1990)
5. Fiori, M., Musé, P., Aguirre, S., Sapiro, G.: Automatic colon polyp flagging via geometric and texture features. In: IEEE EMBS, pp. 3170–3173 (2010)
6. Wang, Z., Liang, Z., Li, L., Li, X., Li, B., Anderson, J., Harrington, D.: Reduction of false positives by internal features for polyp detection in CT-based virtual colonoscopy. Medical Physics 32(12), 3602–3616 (2005)
7. Martino, M.D., Hernández, G., Fiori, M., Fernández, A.: A new framework for optimal classifier design. Pattern Recognition 46(8), 2249–2255 (2013)
8. Pickhardt, P., Choi, J., Hwang, I., Butler, J., Puckett, M., Hildebrandt, H., Wong, R., Nugent, P., Mysliwiec, P., Schindler, W.: Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. New England Journal of Medicine 349(23), 2191–2200 (2003)
9. Suzuki, K., Yoshida, H., Näppi, J., Armato, S.G., Dachman, A.H.: Mixture of expert 3D massive-training ANNs for reduction of multiple types of false positives in CAD for detection of polyps in CT colonography. Medical Physics 35(2), 694 (2008)