Automatic Polyp Detection Using Global Geometric Constraints and Local Intensity Variation Patterns

Nima Tajbakhsh[1], Suryakanth R. Gurudu[2], and Jianming Liang[1]

[1] Department of Biomedical Informatics, Arizona State University, Scottsdale, AZ, USA
{Nima.Tajbakhsh,Jianming.Liang}@asu.edu
[2] Division of Gastroenterology and Hepatology, Mayo Clinic, Scottsdale, AZ, USA
Gurudu.Suryakanth@mayo.edu

Abstract. This paper presents a new method for detecting polyps in colonoscopy. Its novelty lies in integrating the global geometric constraints of polyps with the local patterns of intensity variation across polyp boundaries: the former drives the detector towards the objects with curvy boundaries, while the latter minimizes the misleading effects of polyp-like structures. This paper makes three original contributions: (1) a fast and discriminative patch descriptor for precisely characterizing patterns of intensity variation across boundaries, (2) a new 2-stage classification scheme for accurately excluding non-polyp edges from an overcomplete edge map, and (3) a novel voting scheme for robustly localizing polyps from the retained edges. Evaluations on a public database and our own videos demonstrate that our method is promising and outperforms the state-of-the-art methods.

Keywords: Optical colonoscopy, polyp detection, boundary classification, edge voting.

1 Introduction

Colonoscopy is the preferred technique for colon cancer screening and prevention, during which a tiny camera is inserted and guided through the colon to detect and remove polyps—precursors to colon cancer. However, a colonoscopy is an operator dependent procedure, wherein human factors, such as fatigue and insufficient attentiveness particularly during long and back-to-back colonoscopies, can lead to the miss detection of polyps. Patients with missed polyps may be diagnosed with a late stage cancer with the survival rate of less than 10%. Computer-aided polyp detection is promising to reduce polyp miss-rate and encourage attentiveness during procedures.

Early works on polyp detection employed color and texture features. The work of Karkanis et al. [1] based on color wavelet features sets a representative example. However, effectiveness of such methods is limited by partial texture visibility of polyps and large color variations among polyps. More recent techniques have considered shape, spatio-temporal, and appearance features. Hwang et al. [2] suggested elliptical shape features for polyp detection; however, geometric features in the absence of contextual clues can be misleading. Our previous work [3] aimed to address this drawback by eliminating such misleading structures from the edge maps. Bernal et al. [4] employed valley information to localize polyps. However, as acknowledged by the authors, this method might result in false detections particularly around wrinkles and vascular structures. Finally, the spatio-temporal features suggested in [5] were suitable only for the

P. Golland et al. (Eds.): MICCAI 2014, Part II, LNCS 8674, pp. 179–187, 2014.

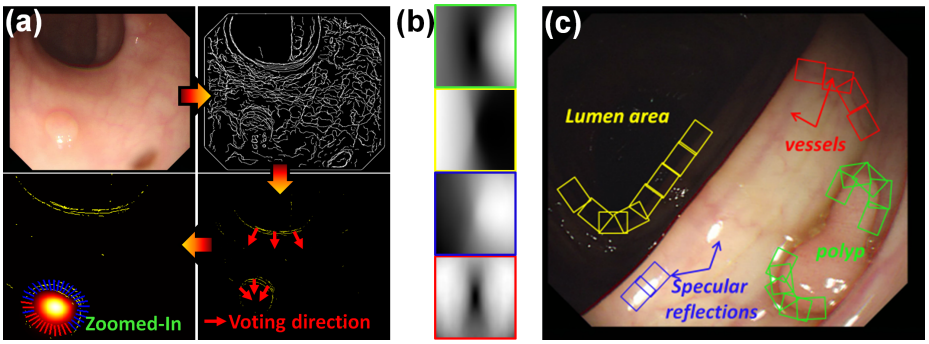

Fig. 1. (a) Illustration of our polyp detection method: a crude set of edge pixels is extracted (Sec. 2.2), the edge map is refined by effectively removing many non-polyp boundary edges through a classification scheme (Sec. 2.3 and 2.4), the retained edges vote along their voting directions determined by the classifier (Sec. 2.5), a band is placed around the candidate point with maximum vote to measure the polyp probability (Sec. 2.6). Note how the inferred voting directions help avoid a false positive at the top curvy boundary. (b) From top to bottom: average boundary appearance of polyps, lumen areas, specular reflections, and vessels. Polyp boundaries have a distinct appearance. (c) An endoluminal scene containing a polyp and other curvy structures.

off-line processing of colonoscopy videos since this method requires information from the past and future frames for polyp localization at the current frame.

The main contribution of this work is to enable effective use of shape information for polyp detection in colonoscopy videos. This is accomplished by combining shape information with intensity variation patterns (IVPs) across object boundaries, an integration that is essential because a polyp detector merely based on curvature can be easily misled by other structures with curvy boundaries in the complex endoluminal scenes. Given an overcomplete edge map, we employ IVPs (1) to remove as many non-polyp edges as possible and (2) to determine voting directions for the retained edges. The former handles misleading effects of other polyp-like structures and the latter allows for robust polyp localization. Through this work, we make the following three original contributions:

1. A new patch descriptor that quickly captures IVPs across boundaries. Our descriptor is fast, rotation invariant, and robust against linear illumination changes.
2. A 2-stage classification scheme that enhances low level image features prior to classification by learning a nonlinear similarity metric in the features space.
3. A novel voting scheme that robustly detects objects with curvy boundaries in fragmented edge maps. Our voting scheme produces a probabilistic output for each polyp candidate but does not require any predefined parametric model of shapes.

2 Proposed Method

Given an edge map, we first extract IVPs to remove as many non-polyp edges as possible, producing a refined edge map where shape and curvature can be safely used for

polyp localization. The retained edges then participate in a voting scheme to determine the locations of polyps. Fig. 1(a) illustrates how our polyp detection method works.

2.1 Intensity Variation Patterns (IVPs)

Polyps have distinct intensity variations across their boundaries. This is illustrated in Fig. 1(b) where the average appearance of hundred thousand oriented image patches around polyps, vessels, lumen areas, and specular reflections is compared. Fig. 1(c) shows how these oriented patches are extracted from one of the used images. Factors that contribute to the distinct appearance across polyp boundaries include different diffuse and specular patterns on the surface of a polyp and the depth contrast between the polyp side and background side of the boundaries.

2.2 Edge Map Construction

We apply Canny's method on the three color channels of an input image to obtain an overcomplete edge map. We then estimate edge directions for all the pixels in the map. We employ ball tensor voting as used in [3] to robustly determine edge orientations.

2.3 Feature Extraction

The goal of feature extraction is to capture IVPs in an image patch around each edge pixel. The desired feature extraction method must meet three major requirements: (1) it must be fast to handle a large volume of input patches from the edge detection stage; (2) it must provide high level of illumination invariance, since in a colonoscopy procedure, the source of light moves along with the camera, causing the same segment of a boundary to appear with varying contrast in a number of consecutive frames; (3) it must provide rotation invariance against the edge orientations because the essential information do not lie along edge directions but across the edges.

Our patch descriptor begins with extracting a patch along the orientation of an edge pixel such that the edge segment appear vertically in the middle of the patch. This presentation provides two advantages: (1) image appearance characterization independent of edge orientations, and (2) some degrees of robustness against positional variability along the horizontal axis. We then form sub-patches of size $n \times m$ all over an extracted patch with 50% overlap along horizontal and vertical directions. Each patch is then averaged vertically, resulting in a 1D intensity signal S_i, decreasing positional variability along the vertical direction. To obtain a compact and robust presentation of intensity variations, we then apply a 1D discrete cosine transform (DCT) to the extracted signal:

$$C_k = \frac{2}{n} w(k) \sum_{i=0}^{n-1} S_i \cos(\frac{2i+1}{2n} \pi k) \tag{1}$$

where $w(k) = 1/\sqrt{2}, k = 0$ and $w(k) = 1, 1 \leq k \leq n-1$. DCT has a strong energy compaction property, allowing the entire spatial information to be summarized in a few coefficients. However, such a compact presentation of the signal is not robust against

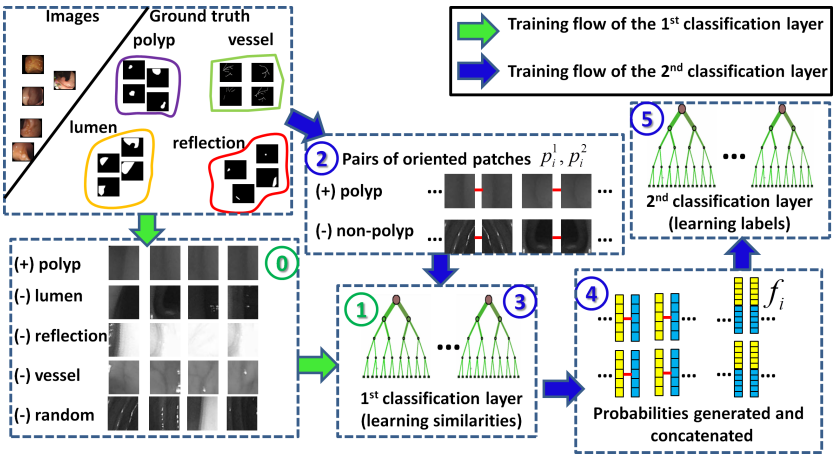

Fig. 2. Overview of the suggested classification scheme

illumination changes. A constant change in the intensity of the pixels in a patch directly appears in the DC coefficient, and illumination scaling affects both the DC and AC coefficients. To achieve invariance against constant illumination changes, we discard the DC component. To achieve invariance against linear intensity changes, we normalize the AC components using their L^2-norm. However, this scheme is not efficient given that we are interested in only a few of the AC components. We therefore compute the first few AC coefficients from each patch and use their L^2-norm for normalization, achieving a significant performance speedup. Finally, the coefficients selected from each sub-patch are concatenated to form a feature vector for the extracted patch.

2.4 Edge Classification

The classification scheme aims to (1) remove as many non-polyp edges as possible and (2) determine on which side of the retained edges a polyp exists–hereafter referred to as "voting direction" (See Fig. 1(c)). To achieve the two objectives, we design a 2-stage classification scheme. The first stage learns a non-linear metric in the low level feature space to measure the similarities between the input patches and some predefined structures. We choose such structures through a misclassification analysis. The second stage performs the main classification, removing non-polyp edges and determining voting directions for the retained edges. Fig. 2 illustrates the training process.

Step 0: We divide the space of negative patches into 4 sub-classes: vessels, lumen areas, specular reflections, and a class containing other structures in the scenes. These four categories together with the polyp class constitute the structures of interest. Given the ground truth in the database, we collect a stratified set of N_1=100,000 oriented patches around these five types of boundaries such that the structures of interest always appear on the same side of the patches. This is to ensure a consistent presentation for all structures. Step 1: We then train a 5-class classifier to learn the appearance of these 5 types of structures in the low level feature space that is generated by our patch

descriptor. The trained classifier can be viewed as a similarity learner module and its five probabilities for each input patch can be viewed as mid-level image features in contrast with the low level input features, which mainly encode IVPs across boundaries.

Step 2: We collect a stratified set of N_2=100,000 pairs of oriented patches from polyp and non-polyp boundaries in the training images. Let $\{p_i^1, p_i^2\}$ be the extracted pair of patches around i^{th} edge with the corresponding normals, $\{n_i^1, n_i^2\}$, where $\angle n_i^1 \in [0, \pi]$ and $\angle n_i^2 = \angle n_i^1 + \pi$. Note that the normal vector indicates the interpolation direction when extracting oriented patches. Based on the state of the i^{th} edge, we assign a label y_i to this pair of patches, where "0" is for a non-polyp edge, "1" is for a polyp edge with n_i^1 indicating the voting direction, and "2" is for a polyp edge with n_i^2 indicating the voting direction, $\{(p_i^1, p_i^2, y^i)|y^i \in \{0,1,2\}, i = 1,2,...,N_2\}$. Step 3 and 4: We extract low level features from each pair of patches and then apply the classifier trained in the first layer, resulting in two arrays of mid-level features per pair that are further concatenated to form a feature vector, $\{(f^i, y^i)|y^i \in \{0,1,2\}, i = 1,2,...,N_2\}$. Step 5: Once all feature vectors are collected, we train a 3-class classifier to learn both edge labels and voting directions. For both layers, we choose the random forest classifier because of its high quality probabilistic output.

Given an edge map for a test image, a pair of patches is extracted for every edge pixel and then the corresponding low level features are computed. Next, each feature vector is fed to the first classification layer. The mid level features generated from each pair of patches are concatenated and fed to the second classification layer where the label and voting direction (n_i^*) for the underlying edge are determined as follows:

$$\begin{cases} \text{"polyp" and } n_i^* \leftarrow n_i^1 & \text{if } p(y^i = 1) > 0.5 \\ \text{"polyp" and } n_i^* \leftarrow n_i^2 & \text{if } p(y^i = 2) > 0.5 \\ \text{"non-polyp"} & \text{otherwise,} \end{cases} \tag{2}$$

2.5 Voting Scheme

We refer to the edges that have passed the classification stage as the "voters" in the rest of this paper. Each voter has a polyp direction n_i^* and a classification confidence $C_{v_i} = max(p(y^i = 1), p(y^i = 2))$. Our voting scheme begins with grouping the voters into K categories according to their voting directions, V^k=$\{v_i|\frac{k\pi}{K} < \angle n_i^* < \frac{(k+1)\pi}{K}\}$, $k = 0...K$. Such edge grouping prior to vote casting minimizes vote accumulation in the regions that are surrounded by low curvature boundaries. The voters in each category then cast votes at their surrounding pixels according to their voting directions and classifications confidence. This results in K voting maps that are further multiplied to form the final voting map whose maximum vote accumulation (MVA) indicates the location of a polyp candidate. Mathematically,

$$MVA = \arg\max_{x,y} \prod_{k=1}^{K} \sum_{v \in V_k} M_v(x,y), \tag{3}$$

where $M_v(x,y)$ is the vote cast by the voter v at a receiver pixel $r = [x,y]$, which is computed as follows:

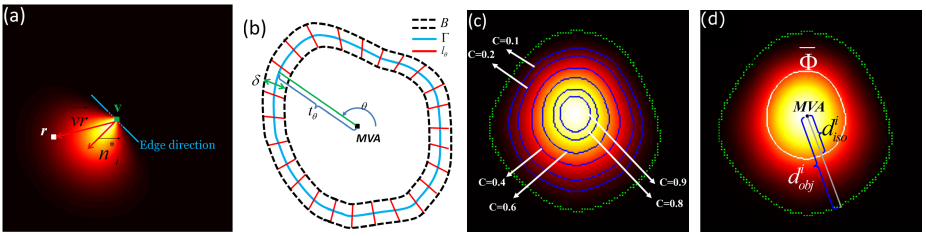

Fig. 3. (a) The voting map generated from an edge pixel lying at 135 degree. (b) The narrow band used for assigning a probabilistic score to a polyp candidate. (c) The isocontours of a voting map generated for a synthetic shape. (d) The representative isocountour $\bar{\Phi}$ of the voting map, which is used to determine the shape and width of the corresponding narrow band

$$M_v(r) = \begin{cases} C_v \exp(\frac{-\|\overrightarrow{vr}\|^2}{\sigma}) \cos(\angle \overrightarrow{n^* vr}), & \text{if } \angle \overrightarrow{n^* vr} < \pi/2 \\ 0, & \text{if } \angle \overrightarrow{n^* vr} \geq \pi/2 \end{cases} \tag{4}$$

where σ controls the size of the voting field. Fig. 3(a) shows the voting field for an edge pixel lying at 135 degree. As seen, the votes are cast only in the region pointed by the voting direction. Such selectivity arises from the condition set on $\angle \overrightarrow{n^* vr}$, which prevents the voters from casting votes in the opposite direction. The exponential and cosinusoidal decay functions enable smooth vote propagation, which we will later use to determine the likelihood of a polyp candidate.

2.6 Probability Assignment

The maximum vote accumulation at a polyp candidate depends on many factors including the size of polyps and the number of nearby voters around polyps. Therefore, we cannot use raw accumulated votes to assign a probabilistic score to a polyp candidate. Alternatively, we search for the contributing voters within a narrow band around the polyp candidate. The narrow band B, as shown in Fig. 3(b), consists of radial line segments whose extension passes through the candidate location. The line segment ℓ_θ can be parametrized as $MVA + t[\cos(\theta), \sin(\theta)]^T$ with $t \in [t_\theta - \frac{\delta}{2}, t_\theta + \frac{\delta}{2}]$ where δ is the bandwidth, and $t_\theta = \|\Gamma_\theta - MVA\|$ is the distance between the candidate location and the corresponding point on the band skeleton Γ. To form the band around a polyp candidate, one needs to determine the bandwidth δ and a set of distances t_θ. We will estimate these parameters from the corresponding voting maps in Section 2.7. Once the band is formed, the probability assigned to a polyp candidate is calculated as $\frac{2}{|S_\theta|} \sum_{\theta \in S_\theta} (I_\theta \vee I_{\theta+180})$ where S_θ denotes the set of angles along which the voters are searched for and $|S_\theta|$ is the cardinality of S_θ. We consider the discrete set $S_\theta = \{\theta | 0 \leq \theta \leq 180\}$ for probability computation. In this equation, I_θ is an indicator variable that takes 1 if the line segment ℓ_θ hits at least a voter v whose estimated polyp direction n_v^* points toward the candidate location. This equation is designed to be sensitive to both regions surrounded by continuous boundaries and those surrounded by partially segmented and discontinuous boundaries.

2.7 Narrow Band Determination

We define the isocontours of a voting map and then use them to estimate the unknown parameters of the bands. The isocontour Φ_c of the voting map V is defined as $\Phi_c = \{(x,y)|V(x,y) = cM\}$ where M denotes the maximum of the voting map and c is a constant between 0 and 1. Our experiments show that isocontours of a voting map can predict where actual object boundary is. We wish to find this mapping through a regression model. Since isocontours may get corrupted by other nearby voters in the scene, we obtain several isocontours as shown in (Fig. 3(c)) and then take their median shape as the representative isocontour $\bar{\Phi}$ of the voting map (Fig. 3(d)). Let d^i_{iso} denotes the distance between the i^{th} point on $\bar{\Phi}$ and MVA. We use d^i_{iso} to predict d^i_{obj}, the distance between the corresponding point on the object boundary and the MVA within a prediction interval. For this purpose, we use a second order polynomial regression model $d^i_{obj} = b_0 + b_1(d^i_{iso}) + b_2(d^i_{iso})^2$, where b_0, b_1, and b_2 are the regression coefficients and are estimated using a least square approach. Once the model is constructed, we take the output of the model d_{obj} at angle θ with respect to MVA as t_θ and the corresponding prediction interval as the bandwidth δ. With this information, we can form the band around the polyp candidate and then compute the probability.

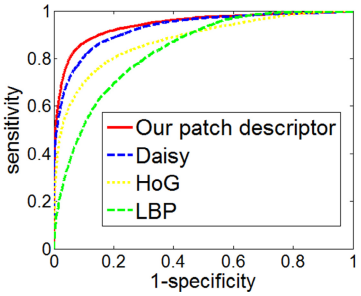

Fig. 4. Our descriptor outperforms off-the-shelf methods for patch characterization. Our closest competitor, Daisy, runs at much slower computation speed.

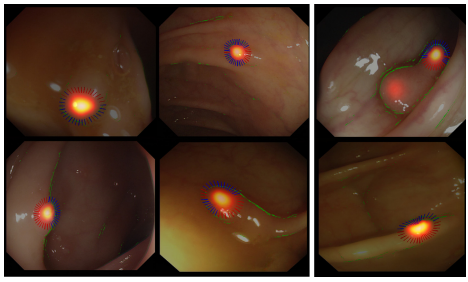

Fig. 5. Polyp localization. The edges retained after classification are shown in green. Line segments that reach polyp edges with desired voting directions are shown in blue and red otherwise.

3 Experiments

To evaluate our system, we employed *CVC-ColonDB* [4] containing 300 colonoscopy images and 8 short colonoscopy videos from our private database consisting of approximately 1700 images with polyps and 2500 images without polyps.

For feature evaluation, we collected 50,000 oriented patches around polyps and other boundaries in colonoscopy images. We selected half of the images for training, and used the rest for testing. For classification, we used a random forest classifier. Our experiments with the training set revealed that selecting 8x16 sub-patches in each image patch and extracting 3 DCT coefficients from each sub-patch yields the best performance. Fig. 4 shows the ROC curve for our patch descriptor when applied on the test set.

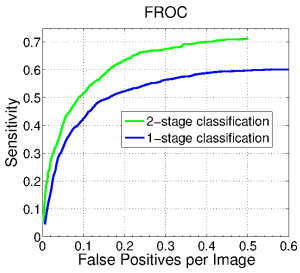

Fig. 6. The suggested 2-stage classification achieves significant performance gain for short videos from our private database

Table 1. Precision and recall of polyp detection for the proposed system, our previous work [3], and the one suggested in [4]. Our new system outperforms the rest.

Recall	Precision			
	Our method		Tajbakhsh[3]	Bernal[4]
	σ=70	σ=100		
50%	98%	100%	90%	92%
60%	97%	98%	88%	78%
70%	96%	95%	89%	65%
80%	94%	92%	86%	60%

For comparison, we used the publicly available implementations[1] of the other widely-used descriptors, such as HoG, LBP, and Daisy. As seen, our descriptor surpasses HOG and LBP with a large margin and slightly outperforms Daisy, which runs at much slower computation speed than ours.

We developed a shape generator to tune the parameters of the voting scheme and to construct the regression model for boundary localization. For this purpose, we generated 3000 objects at three different scales corresponding to small, medium, and large polyps. We then changed the number of voting categories, $K \in \{2, 3, 4, 5, 6\}$, and performed the voting scheme for all the objects. We constructed a regression model for each configuration and found out that $K = 4$ achieves the best boundary localization. We investigated the effect of the size of voting fields on the real colonoscopy images.

We employed 5-fold cross validation to evaluate our polyp detection system using *CVC-ColonDB*. A detection is "true" if it falls inside the ground truth. Table 1 compares the precision and recall rates of the proposed method with those reported in [4] and [3]. As seen our method outperforms the rest in different operating points and yields stable results over a relatively large range of σ. Examples of polyp localization are shown in Fig. 5. False candidates produced by our voting scheme mostly occur due to aggressive edge classification and inadequate clear boundary between polyps and their surrounding area. We further evaluated our system trained on the entire *CVC -ColonDB* using the eight short colonoscopy videos. The free ROC curve, as shown in Fig. 6, demonstrates that the suggested classification scheme significantly outperforms a 1-stage classification scenario where a three class classifier is used for edge classification.

4 Conclusions and Discussions

We have presented a novel polyp detection method with performance superior to the state-of-the-art, achieving an acceptable level of sensitivity and specificity for polyp detection. An important consideration is computer aided detection for colonoscopy need

[1] HOG: lear.inrialpes.fr/pubs/2005/DT05/; LBP: `www.cse.oulu.fi/CMV/Downloads/LBPMatlab`; Daisy: `cvlab.epfl.ch/software/daisy`

not to achieve perfect sensitivity to be clinically useful, because a polyp usually appears in a number of consecutive frames and if one instance of the polyp—particularly upon appearance—gets detected, the detection process is considered as a success. Our algorithm was able to detect at least one instance of polyps in both the public database and our collected videos. Noteworthy, the suggested boundary classification framework is general and can be applied to a variety of medical segmentation problems where a supervised edge classification can serve as a preprocessing stage prior to segmentation.

References

1. Karkanis, S.A., Iakovidis, D.K., Maroulis, D.E., Karras, D.A., Tzivras, M.: Computer-aided tumor detection in endoscopic video using color wavelet features. IEEE Transactions on Information Technology in Biomedicine 7(3), 141–152 (2003)
2. Hwang, S., Oh, J., Tavanapong, W., Wong, J., de Groen, P.: Polyp detection in colonoscopy video using elliptical shape feature. In: IEEE International Conference on Image Processing, ICIP 2007, vol. 2, pp. II–465–II–468 (2007)
3. Tajbakhsh, N., Gurudu, S.R., Liang, J.: A classification-enhanced vote accumulation scheme for detecting colonic polyps. In: Yoshida, H., Warfield, S., Vannier, M.W. (eds.) Abdominal Imaging 2013. LNCS, vol. 8198, pp. 53–62. Springer, Heidelberg (2013)
4. Bernal, J., Sánchez, J., Vilariño, F.: Towards automatic polyp detection with a polyp appearance model. Pattern Recognition 45, 3166–3182 (2012)
5. Park, S.Y., Sargent, D., Spofford, I., Vosburgh, K., A-Rahim, Y.: A colon video analysis framework for polyp detection. IEEE Transactions on Biomedical Engineering 59, 1408–1418 (2012)