Automatic polyp detection from a regional appearance model and a robust dense Hough coding

Lina Ruiz, Luis Guayacán, and Fabio Martínez

⋆ Biomedical Imaging, Vision and Learning Laboratory (BIVL^{2}ab)
Motion Analysis and Computer Vision (MACV)
Universidad Industrial de Santander (UIS)
Bucaramanga, Colombia

***Abstract*—Polyps are the main biomarker to diagnose colorectal cancer. These polyps are protuberant masses that namely appear around intestinal tract. Currently, the automatic polyp detection is a challenging task because the high shape and appearance variability. Additionally, the videos captured during colonoscopies are namely distorted because the clinical procedure, presenting strong camera motions and illumination changes. This work introduces an automatic strategy to detect polyps by using a pixel level characterization of orientation and curvatures, which are coded using a dense Hough transform.**

Because the complexity of polyp modelling, an additional appearance model was herein implemented, allowing to filter and to find potential polyp regions. The proposed strategy was evaluated over real sequences of ASU-Mayo Clinic Colonoscopy Video dataset, reporting in average 81% of precision on polyp detection task. This strategy is also able to online track polyps in long sequences, requiring only a first frame delineation.

***Index Terms*—Colonoscopy, Polyp Detection, Dense Hough Transform, appearance polyp model**

I. Introduction

Colorectal cancer is the third most aggressive type of cancer disease around world, with more than one million and a half new cases recently detected and more than eight hundred thousand associated deaths [1].A dramatic increase of 80% of new cases is expected by 2030 in the Americas. Colonoscopy is the first clinical procedure to detect polyp masses, which are abnormalities on colon epithelium associated with this cancer disease.

Nevertheless, the characterization of these polyps is a challenge because the wide variability in shape and appearance w.r.t. neighborhood regions, being in many cases very difficult to detect even for expert gastroenterologists [2]. In addition, polyp detection task could be influenced by other factors such as inter-patient variability, state of disease (according to stage polyp can pass to [1-5] millimeters to 1 centimeters or more [3]). Likewise, colonoscopy resolution, abrupt camera motion during exploration and strong frame variation of one frame to other difficult the recognition task.

In fact, studies have reported that during colonoscopy they may overlook 26% of polyps because their plane shape and very similar appearance regarding the intestinal tract [4]. Hence, computational tools are demanding to support clinical polyp detection and characterization, to reduce the number of false negatives in the cancer diagnosis. This is fundamental, since the early detection of the disease increases the probability of success in treating the disease by up to 93% [2].

In current state-of-the-art has been proposed several strategies to detect, segment and classify polyps in colonoscopies sequences [5], [6], [7]. Regarding polyp detection, Bernal *et. al.* proposed a strategy based on a deep analysis of salient valleys and ridges along digestive tract to find polyps [8]. However, the natural tract folds, as well as, remaining artifacts produce a lot of false positive detections. Also, on [9] was implemented an AdaBoost classifier that uses Hard negative descriptors, allowing to detect polyps in real time. In this work was also increased the polyp detection by including positive and negative samples but with an associated computational limitation. Additionally, such descriptor is limited to a proper representation on polyp shapes. Likewise, Tajbakhsh *et al.* [10] proposed a hybrid system based on shape and context polyps characterization to limit false artifact detections. Additionally, an appearance model was implemented to detect surrounding polyp regions. This approach is however sensible to small changes in polyps between consecutive frames along colonoscopy sequence. In [11] was implemented a mixture of Gaussian model that together with a classical Hough transform allows to detect polyps in different sequences. This approach however is limited to the number of Gaussians that represent the appearance, which could be fail in typical illumination changes caused by camera motions. Also, the use of classical Hough descriptor is limited to represent the polyp as a cycloid, which in much of the cases is no a proper shape model for these masses. This shape assumption, also requires a proper perspective of camera to capture the polyp, requiring controlled sequences.

The main contribution of this work is an automatic polyp detection strategy based on a dense Hough descriptor that allows to characterize non-parametric shapes. This descriptor coded pixel gradient and curvatures in recursive tables that are fixed on a region of interest at the first frame. Then, for consecutive frames of colonoscopy sequence a voting map is generated to predict the most probable location of polyps.

978-1-7281-1491-0/19/$31.00 ©2019 IEEE

Because the complex structure of tract, an appearance model is associated to discard false positive regions. The proposed strategy is computationally efficient and allows to recover polyps along sequence with a negligible training performed only at first frame. Next sections fully describe the proposed method (section 2) and report some results obtained over public datasets (section 3). Finally in section 4 are presented some conclusions and perspectives.

II. Proposed approach

This work introduces an automatic polyp detection strategy using a per-pixel multi-scale Hough transform and a robust color model. The Hough transform achieves a non-parametric shape characterization at each frame, while the color model complements the characterization and help with polyp searching. On one hand, a RoI template that bound the polyp is herein obtained in first frame as reference for the rest of the sequence. This RoI is then characterized by the set of orientation gradients and pixel curvatures into a multi-scale approach, and effectively stored as cumulative tables. For rest of the colonoscopy sequence, the polyp is found by following a voting process w.r.t the learned template. On the other hand, a complementary color polyp model was implemented to decide the most probable region among several maximums. In next subsection are explained in detail the step of the proposed approach.

A. Learning polyp template

Video colonoscopies are part of clinical routine in patients with potential prognosis of colorectal cancer. The polyp is a mass that strongly changes in shape and appearance regarding the patient stage, the colonoscopy light and the camera position that navigate along track during the clinical exam. Such variations difficult the learning of polyp non-parametric structure to automatically track and segment in colonoscopy sequences.

In this work, a particular polyp structure is learned at first frame $I_1(\mathbf{x})$, with pixels as $\mathbf{x} = (x, y)^T$. Then, a polyp region $P_r(\mathbf{x})$ is manually selected around the mass. This $P_r(\mathbf{x})$ region is fully characterized using local structural primitives, namely orientation gradients (ϕ) and curvatures (κ), coded on special tables. For doing so, a multi-scale framework is herein implemented by computing several incremental Gaussian filters over the region of interest $P_r^\sigma(\mathbf{x})$. Then, gradients over each scale image is computed to obtain the respective gradient magnitude $\|\nabla I\| = \sqrt{I_x^2 + I_y^2}$ and angle $\phi = \arctan \frac{I_y}{I_x}$. Then, curvature κ is estimated at each pixel using second order derivatives, expressed as: $\kappa = \frac{(I_{xx} I_y^2 - 2 I_{xy} I_x I_y + I_{yy} I_x^2)}{||\nabla I||^3}$. An invariant non-parametric shape of the polyp can be represented by locally computing angles and curvatures (ϕ, κ) at each scale. This dense representation is performed for each of the pixels into P_r, and coding each of these values (ϕ, κ) into recursive (R)-tables, as $RT_\sigma(\phi, \kappa)$. In this case, the angles were defined into a range of $[-90, 90]$, while κ was defined into a range of $[-1, 1]$. The coding of these primitives along different scales define the recursive nature of R-tables.

The bi-dimensional R-tables $RT_\sigma(\phi, \kappa)$ are coded with respect to each scale σ and the geometric primitives (ϕ, κ). A summary of the algorithm to index these values is described in algorithm 1. For all pixels $\mathbf{x}$ that belongs to P_r with norm gradient largest than zero, is computed the respective orientations and curvatures. For first gradient dimension, a distance w.r.t to centre $(\delta_x = \frac{w_T}{2} - px, \delta_y = \frac{h_T}{2} - py)$ is stored and weighted with respect to the norm $\|\nabla I\|$. The curvature index also store the distance w.r.t the centre but weighted with Frobenius norm of Hessian matrix, defined as: $\|H_I\|_F = \sqrt{I_{xx}^2 + 2I_{xy}^2 + I_{yy}^2}$.

This $RT_\sigma(\phi, \kappa)$ is robust to illumination change and invariant to rigid transformations. Hence a compact representation of polyp is achieved by the coding of this geometric primitives, computed at pixel level, for different scales. To reduce the computational complexity on $RT_\sigma(\phi, \kappa)$ representation, vectors with negligible weights (according to $\Psi(\|\nabla I\|, \|H_I\|_F)$) are removed from table. In this case, the Ψ term denotes a kernel that is activate according to the respective norms for α and κ.

Algorithm 1 Algorithm R-table

```
for σ ∈ {σ_1, ..., σ_N} do
  for x ∈ P_r do
    ∇I ← (I_x^σ(x), I_y^σ(x))
    H_I ← ( I_xx^σ(x)  I_xy^σ(x) )
          ( I_xy^σ(x)  I_yy^σ(x) )
    if ||∇I|| > 0 then
      compute (φ, κ)
      RT_σ(φ, κ) ←
      RT_σ(φ, κ) ∪ (δ_x, δ_y, Ψ(||∇I||_φ, (||H_I||_F)_κ))
    end if
  end for
end for
```

B. Detection from dense Hough transform

Once a polyp RoI is learned in first frame, for consecutive frames each pixel is characterized by computing the gradient orientation and curvature (ϕ, κ) and performing a voting process. The resulting score is stored into a Γ matrix, summing up the weight value ω defined as $\omega = \Psi(\|\nabla I\|_\phi, (\|H_I\|_F)_\kappa)$, for (ϕ, κ) respectively. These weights are indexed on the respective scale of R-table: $RT_\sigma(\phi, \kappa)$ for (px, py) values of the frame. If several pixels point up to the same location, then such position will have the maximum score, indicating the most probable center of the learned RoI. As expected, the Γ map has image size $(w \times h)$, and the cumulative process is carried out over the different scales σ_i.

C. Color colonoscopy representation

During colonoscopies, the expert explore exhaustively along digestive track leading to abrupt changes of camera, changes of illumination. Also high variance on track appearance representation and artifacts in scene produce challenge conditions

Algorithm 2 Algorithm to recover Γ matrix

for $\sigma \in \{\sigma_1, ..., \sigma_N\}$ **do**
  for $\mathbf{x} \in I(\mathbf{x})$ **do**
    $\nabla I \leftarrow (I_x^\sigma(\mathbf{x}), I_y^\sigma(\mathbf{x}))$
    $H_I \leftarrow \begin{pmatrix} I_{xx}^\sigma(\mathbf{x}) & I_{xy}^\sigma(\mathbf{x}) \\ I_{xy}^\sigma(\mathbf{x}) & I_{yy}^\sigma(\mathbf{x}) \end{pmatrix}$
    if $\|\nabla I\| > 0$ **then**
      compute (ϕ, κ)
    end if
    for all $(\delta_x, \delta_y, \omega) \in \mathrm{RT}_\sigma(\phi, \kappa)$ **do**
      $\Gamma(\delta_x + px, \delta_y + py) \leftarrow \Gamma(\delta_x + px, \delta_y + py) + \omega$
    end for
  end for
end for

to perform an automatic evaluation. All these variations imply confusion, even for expert observations, among polyps and other regions artificially created by light patterns.

For all these reasons is mandatory to perform a video normalization and correction that alleviate the characterization of polyps during the clinical routine. A opponent color representation was herein implemented to mitigate such illumination problems [12]. In this representation, color are normalized as:

$$O_1 : \frac{R-G}{\sqrt{2}} \quad O_2 : \frac{R+G-2B}{\sqrt{6}} \quad O_3 : \frac{R+G+B}{\sqrt{3}}$$

Because subtraction on O_1 and O_2 channels, they result relative shift-invariant with respect to light. Such color representation reduce problem of change of illumination and allows a better polyp color representation, from the rest of the scene. In this work was implemented whole strategy using only O_1 channel, even to compute R-tables.

Additionally, this color characterization allows to decide among several maximums in voting map. To avoid an increasing of false positives at each frame, the learning RoI P_r is also characterized with a histogram h_{P_r} in channel O_1. Then, on spatial location where Γ map is maximum, is computed an histogram h_i. Finally a cosine metric allows to measure the most probable region that correspond to polyp, regarding color information, coded into the histograms as:

$$RoI = \max_{RoI(h_i)} cos(h_{P_r}, h_i) = \frac{h_i \cdot h_{P_r}}{\|h_i\|_2 \|h_{P_r}\|_2}$$

In this way is compactly integrated the geometrical polyp information, together with appearance information. It should be noted, that this color information only operates into the maximums regions indexed on Γ table.

III. Evaluation and Results

Evaluation of the proposed approach was carried out over ASU-Mayo Clinic Colonoscopy Video Database, that counts with 20 short colonoscopy sequences, being ten control sequences, while the rest ten report only one polyp along the sequence [10]. The R-tables were computed with different set of Gaussians, being the best configuration: $\sigma_i = \{3, 5, 7\}$.

In Figure 1 (a)-(c) is observed the complexity of colonoscopy sequences, regarding the polyp shape, appearance and the wide variability from one frame to other. The green circle is the obtained detection by the proposed approach, regarding the voting map illustrated in second row for each frame, respectively. The yellow regions represent the maximums findings. As expected, the proposed approach achieve a very good identification in much of the frames, but with some limitations when there exist previous abrupt camera motions and specular reflections. It is worth nothing that even in false final detection frames (see on Figure 1 (f)), the true polyp is mark as a salient yellow region on the voting map. The limitation is reported for the decision of maximum region.

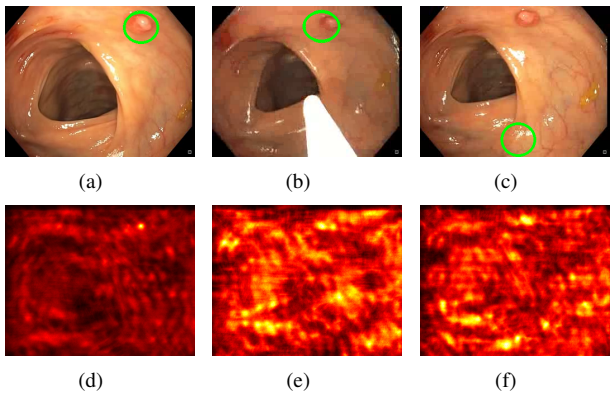

Fig. 1. Detection of the polyp in the image and its respective Γ accumulation matrix. (a) Good detection; (b) regular detection; (c) bad detection; (d)-(e) Γ matrix for each case.

A. Short sequence detection

For polyp detection task, the proposed strategy was evaluated in a total of 15 sub-sequence videos (more than 1180 frames) where is always visible the polyp. For each video, in first frame is mark the region that contains the polyp, from which a recursive learning table is learned.

In table 1 is reported the quantitative results achieved by the proposed approach in term of four different indexes: True positives (TP), False positives (FP), Precision and Recall. The TP correspond to detected regions with complete or partial identification of the polyp, and the FP correspond to detected regions with not evidences of polyp masses. The Precision represents how many polyps were properly selected from whole dataset, while the recall counts the TP into the subset predicted by the proposed approach. The proposed approach achieved a total of 74.91% of correct polyp detection in such sequences, requiring only the learning on first frame. This performance result promising since the marked polyp is detected even the change abruptly of position in some times. Few false positive detection are also reported in frames where the polyp is partially occluded or there exist strong changes on background because residuals and camera reflections. Also the reported recall of 40% is obtained since in several frames the polyp disappear completely.

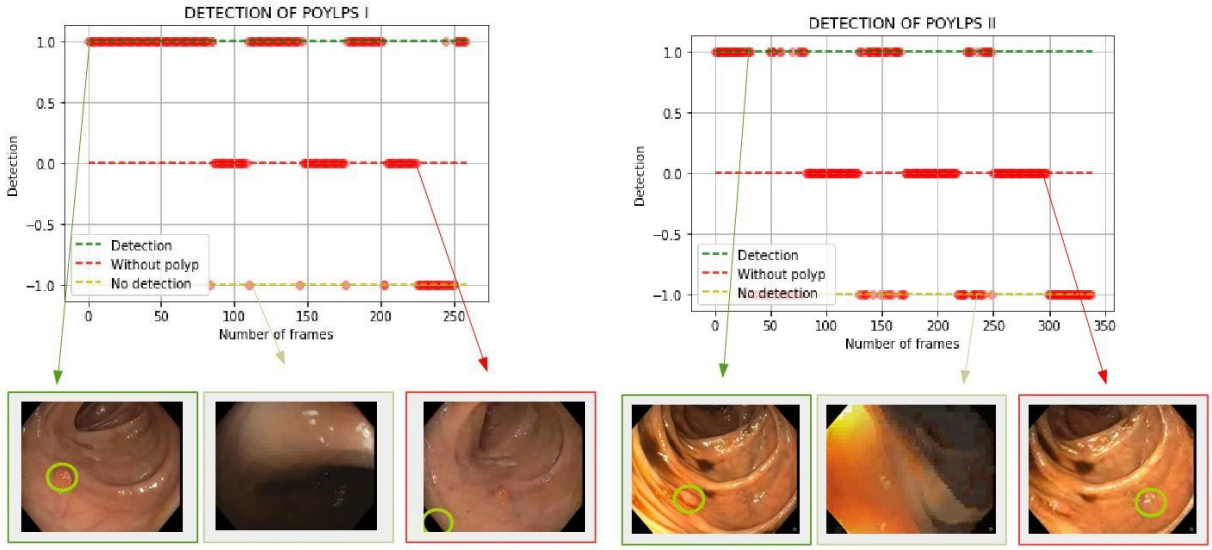

Fig. 2. Polyp detection on video sequences. As observed the polyp change along the sequence, and in some cases, result almost impossible the identification, even for experts.

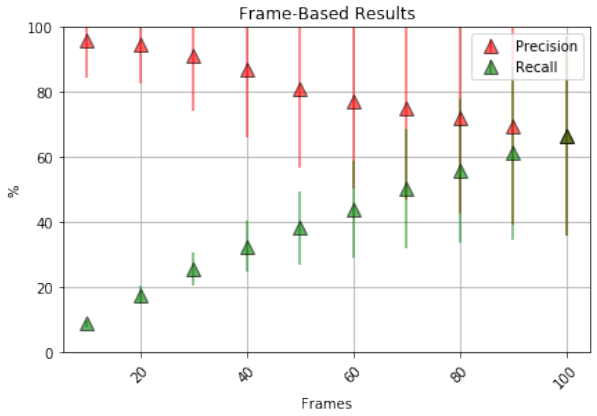

Fig. 3. Frame-Based Results. In plot is observed the Recall and Precision computed recursively along the frame sequence.

Measure	%
Precision	81.009
Recall	40.145
TP	74.916
FP	25.084

Table 1. Quantitative results of the proposed approach. In general the proposed strategy achieves a proper detection of polyps during the real sequences.

B. Long sequence detection

Because the capabilities of the proposed approach to robustly represent and track the polyp along the sequence, an additional experiment was carried out to detect such masses during long colonoscopy sequences. In figure 2 is illustrated the online performance of the proposed approach, being (1-label) the true detection of polyps, the ($-$1-label) the false positive detection and frames where polyp was totally occluded was label with 0.

As observed, in much of the frames the proposed approach achieves a true polyp detection, being much more effective in left sequence because video resolution. Also, it can be observed for both sequences the strong changes in background among consecutive frames, as well as the close similarities between polyps and some intestinal folds. Also, as shown in Figures after frame occlusions, the proposed approach is able to recover again true polyp detections because the independent analysis carried out in recursive tables. Finally on Figure 3 are defined and quantified two inline metrics for polyp recognition at frame level. This metrics are defined as: ($P(t) = TP/\#\text{polyps(t)}$) and Recall ($R(t) = TP/\#\text{frames(t)}$), in which $x-axis$ represents the evaluated frames, and $y-axis$ the obtained score. In this temporal evaluation, $P(t)$ counts the rate of true polyps detected with respect to the increasing samples of polyps in such interval of video. Also the $R(t)$ counts the rate of true polyps but with respect to the interval of the video. As expected, in first frames the precision score achieve almost perfect scores and remain stable for the rest of the sequence. Also the temporal recall has a natural increasing at each frame iterated.

IV. Conclusions

This work presented a dense Hough polyp detection strategy that allows to recover abnormal masses in short and long sequences, even when occlusion is present on intermediate frames. Also a regional histogram, into an opponent color representation, was carried out to reduce false positive detections. Results show a robust performance in many different sequences, requiring only the first frame to learn polyp gradient, curvature, and color properties. Some limitation are reported because the complex structure of track, that difficult the camera navigation and the multiple folds with close similarity

with polyps. Future works include an additional primitives to codify robustly the polyp detection as well as the extension of the proposed strategy on the segmentation task.

ACKNOWLEDGMENTS

The authors acknowledgments to the *Vicerrectoría de Investigación y Extensión* of the Universidad Industrial de Santander for supporting this research registered by the next project: *Reconocimiento continuo de expresiones cortas del lenguaje de seas registrado en secuencias de video*, with SIVIE code 2430.

REFERENCES

[1] S. I. M. C. P. D. P. M. Z. A. B. F. Ferlay J, Colombet M, "Estimating the global cancer incidence and mortality in 2018: Globocan sources and methods doi: 10.1002/ijc.31937. [epub ahead of print] pubmed pmid: 30350310," *Int J Cancer.*, 2018 Oct 23.

[2] E. Perez, *Gastroenterologa.* McGraw Hill Mexico, 2012.

[3] J. L. Vleugels, Y. Hazewinkel, P. Fockens, and E. Dekker, "Natural history of diminutive and small colorectal polyps: asystematic literature review," *Gastrointestinal Endoscopy*, vol. 85, no. 6, pp. 1169 – 1176.e1, 2017.

[4] Q. Angermann, A. Histace, and O. Romain, "Active learning for real time detection of polyps in videocolonoscopy," *Procedia Computer Science*, vol. 90, pp. 182–187, 2016.

[5] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks," in *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pp. 79–83, IEEE, 2015.

[6] M. Ganz, X. Yang, and G. Slabaugh, "Automatic segmentation of polyps in colonoscopic narrow-band imaging data," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2144–2151, 2012.

[7] T. Stehle, R. Auer, S. Gross, A. Behrens, J. Wulff, T. Aach, R. Winograd, C. Trautwein, and J. Tischendorf, "Classification of colon polyps in NBI endoscopy using vascularization features," in *Medical Imaging 2009: Computer-Aided Diagnosis* (N. Karssemeijer and M. L. Giger, eds.), vol. 7260, (Orlando, USA), SPIE, February 7–12 2009.

[8] J. Bernal, J. Sánchez, and F. Vilarino, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognition*, vol. 45, no. 9, pp. 3166–3182, 2012.

[9] Q. Angermann, J. Bernal, C. Sánchez-Montes, M. Hammami, G. Fernández-Esparrach, X. Dray, O. Romain, F. J. Sánchez, and A. Histace, "Towards real-time polyp detection in colonoscopy videos: Adapting still frame-based methodologies for video sequences analysis," in *Computer Assisted and Robotic Endoscopy and Clinical Image-Based Procedures*, pp. 29–41, Springer, 2017.

[10] N. Tajbakhsh, S. R. Gurudu, and J. Liang, "Automated polyp detection in colonoscopy videos using shape and context information," *IEEE transactions on medical imaging*, vol. 35, no. 2, pp. 630–644, 2016.

[11] G. Tarik, A. Khalid, K. Jamal, and D. A. Benajah, "Polyps's region of interest detection in colonoscopy images by using clustering segmentation and region growing," in *Information Science and Technology (CiSt), 2016 4th IEEE International Colloquium on*, pp. 455–459, IEEE, 2016.

[12] K. Van De Sande, T. Gevers, and C. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1582–1596, 2010.