Polyp follow-Up in an Intelligent Wireless Capsule Endoscopy

Orlando Chuquimia, Andrea Pinna, Xavier Dray, Bertrand Granado

Polyp follow-Up in an Intelligent Wireless Capsule Endoscopy

Orlando Chuquimia[1], Andrea Pinna[1], Xavier Dray[2], Bertrand Granado[1]

***Abstract*—In this paper, an image processing to detect polyps in an intelligent Wireless Capsule Endoscopy (WCE) is presented. This processing will be integrated into the WCE. It is a new screening method to detect colorectal cancer (CRC). A motion estimation algorithm is used to follow a detected polyp and improve the pre-processing of our detection chain. With our methodology, the polyp detection rate is improved by up to 40% from 53% to 93.7%. The improved detection rate was validated with a large database of 20 video-colonoscopies (18,910 images).**

I. Introduction

CRC is the second highest cause of death by cancer worldwide with 880,792 deaths in 2018 and a mortality rate of 47.6%. 95% of CRC cases begin with the presence of a growth on the inner lining of the colon or the rectum, called a polyp. Multiple types of polyps exist; among them, adenoma polyps, which can degenerate into CRC. CRC is treatable in 90% of the cases if it is detected early enough [1].

Today, imaging is the modality used to analyze the colon and find polyps. A colonoscopy is the procedure used for screening, diagnosis, and therapy in the gastrointestinal tract. However, it can be painful, traumatic and poorly tolerated by patients. The colonoscopy is invasive and usually requires anesthesia, a specialist and a controlled environment. Furthermore, the colonoscopy does not allow the visibility of all the regions near the colon. Other methods exist, such as the Colorectal Tomography (CT) and the WCE. CT is non-invasive. However, this method cannot detect polyps smaller than 1 cm and exposes the patient to radiation. WCE is less invasive; it is a simple pill that the patient swallows and that transmits images of the gastrointestinal tract via a Radio Frequency communication through the body. The video feedback is then uploaded to a workstation where a specialist can review and examine these images to detect any gastroenterological pathology. The available WCE PillCam Colon 2 [2] has a length of 32.3 mm, a diameter of 11.6 mm, a battery life of 10 hours, a resolution of 256x256 pixels and an image sampling rate of approximately 2 to 4 frames per second. Ten hours is not sufficient to inspect the total intestinal tract. In addition, WCE has a low image resolution compared to a standard endoscope used in a colonoscopy, that acquires an image with a resolution of 1920x1080 pixels. A WCE generates between 144,000 and 1,260,000 images (most of these images do not contain polyps or any gastroenterological pathology). Visual analysis of this large number of images with low resolution makes the examination difficult and time-consuming for the gastroenterologist.

[1] LIP6, CNRS UMR 7606, Sorbonne Université, Paris, France.
[2]APHP, Hôpital Saint-Antoine, Sorbonne Université, Paris, France.

How can we create a new tool without the side effects and limitations of both colonoscopy and WCE, to offer a powerful screening tool and to reduce the CRC mortality rate?

We propose a new paradigm of WCE: an intelligent WCE (iWCE). Its novelty is to integrate processing capabilities directly in the capsule to automatically recognize a polyp. Our iWCE integrates a high-resolution imager as an endoscope (1920x1080 pixels) and an image processing chain to detect suspicious lesions. It solely transmits images with suspected gastroenterological pathologies. This method will reduce energy consumption and the number of images to be analyzed. The processing is defined to be integrated taking into account multiple constraints; a square surface area constraint of 8 by 8 mm^2 (due to 11.6 mm capsule diameter), a real-time constraint of 25 frames per second, and a power consumption constraint of a battery life greater than 12 hours. Our iWCE is a new screening method and a minimally invasive diagnostic tool that will overcome the current barriers that limit the use of WCE. Specialists will thus take clinical decisions in a time-effective manner without analyzing more than a hundred thousand images.

As this iWCE is a new approach, there are few existing works. However, scientific works containing off-line analysis of the images from endoscope or WCE detecting polyp lesions can be found. A review of some of the computer-assisted polyp detection methods can be found in [3].

These methods can be divided into three groups:

1) **Hand-crafted approaches** that exploit low-level image processing algorithms to detect region candidates with a polyp, inlcuding intensity valley [3] or Hessian filters [4].
2) **Machine learning approaches** based on classifiers that can be a Convolutional Neural Network (CNN) to detect polyp lesions as in [5].
3) **Hybrid approaches** which combine both methodologies. Works in [6] adopt this strategy with co-occurrence matrix and local binary patterns as Hand-craft processing. Classifiers can be, for example, Multi Layer Perceptrons or Adaboost.

All these methods run on an external computer and contribute to helping the physician in his diagnosis, but they are not useful for our purpose. They are not suitable to be integrated into a WCE and they do not take into account certain constraints, such as real-time execution, form factor of the pill and energy consumption. In particular, CNN methods use several million of synapses and neurons, which cannot be integrated in a 8 by 8 mm^2 chip in a pill, with a power consumption limitation of 158.1mW [2].

In our previous works, a hybrid approach based on an image processing chain and fuzzy trees used as classifiers [7] was

proposed. This approach was used to detect polyps and it was designed to be integrated in a WCE. Information about the integration can be found in the article [8]. The first step in the image processing chain is to extract regions of interest (ROI) from an image. These ROIs correspond to parts of the image where the *texture features* could indicate the presence of polyps. In the second step, the ROIs are classified using fuzzy trees as the machine learning algorithm.

To study the performance of our approach, we used a public database named ASU-Mayo Clinic Colonoscopy Database, which was introduced during the EndoVisSub2015-GIANA challenge at MICCAI 2015 [3]. It is composed of 18,996 images (4,278 images containing a polyp) and their respective ground-truths. The ground-truth is a binary-image that indicates the position of the polyp in the image. The images come from 20 video-colonoscopies in which 10 display a unique polyp at multiple scales and from different viewing angles.

We have manually extracted the ROIs based on the ground-truth of this database. Next, we trained the fuzzy tree to obtain two classes, a $class1$ for the presence of a polyp and a $class0$ for the absence of a polyp. We use 70% of the images to learn and 30% of the images to test.

To measure the performance, we have computed two indicators at ROI-level:

- The $sensitivity = \frac{TP}{TP+FN}$
- The $specificity = \frac{TN}{TN+FP}$

with the parameters of the equations defined as:

- **TP:** a ROI of $class1$ that contains a polyp.
- **FP:** a ROI of $class1$ that does not contain a polyp.
- **TN:** a ROI of $class0$ that does not contain a polyp.
- **FN:** a ROI of $class0$ that contains a polyp.

We have obtained a sensitivity of 93% and a specificity of 91%.

Although the results are good, a bias is present: the classification results are for manual ROI extractions. The extraction of the ROI by an image processing chain is a crucial part. By giving specific features from the image to the machine learning algorithm, it has a clear impact on the algorithm performance [2].

In this article, we analyze our image processing chain that delivers features from the image to the fuzzy trees. We make a proposition to enhance its performance and finally, we evaluate the performance to come to a conclusion.

II. Evaluation of the image processing chain

In this section, we provide a description of our first hybrid approach described in [7], which includes three steps.

1) **extraction step:** before extracting a ROI, we convert each image from an RGB model to a brightness model. This preserves the texture information and provides a better degree of integrability than an image color processing containing three channels. Then, we realize an edge detection into the image by applying a 3x3 median noise filter and a Canny filter. Finally, we use a Hough Transform to detect circular or elliptical shapes. Each circle or ellipse becomes a ROI (see Fig. 1). We chose the Hough Transform because a polyp circular or elliptical shape. It is also possible to integrate it in real-time in a WCE [9].
2) **description step:** we realize a texture analysis of the ROIs and extract 26 texture and luminosity descriptors using co-occurrence matrix algorithm [10]. Texture and luminosity are important descriptors to recognize and identify polyps [11].
3) **classification step:** in this step, we use the fuzzy trees to classify each image.

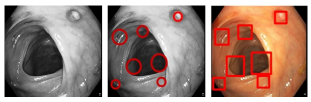

Fig. 1. Extraction step

Here, the image pre-processing part is composed by the two steps; the **extraction** step and the **description** step. To evaluate our pre-processing part, we measure the detection rate of the global processing which includes the three steps mentioned above; **extraction**, **description**, and **classification**. The measurements were made at two levels: at ROI-level and at image-level. At ROI-level, a ROI containing a polyp is labeled as $class1$ and a ROI without a polyp is classified as $class0$ [7]. At image-level, an image containing at least one ROI of $class1$ is classified as $class1$. Otherwise, it is classified as $class0$.

TABLE I
Performances of the global processing

Videos	ROI level		Image level	
20	sensitivity	Spec.	sensitivity	Spec.
Results	29.9%	95.2%	53.2%	73%

The results are shown in Table I. We notice that at the ROI-level, the sensitivity is only 30% and at the image-level, the sensitivity is 53% and the specificity is 73%. This shows the impact of the image pre-processing part. At ROI-level, the performance of the classification step falls from 93% [7] to 30%. At image-level, the sensitivity and specificity are low.

The reason for this must be examined. In the manual extraction, the ROIs contain the entire polyp according to the ground truth. Furthermore, all the polyps are contained in a ROI. In the case of automatic extraction via image processing, the ROIs do not seem to contain an entire polyp and not all polyps are in a ROI.

We have conducted a deep analysis and determined that by using the proposed image processing, 70% of polyps were extracted in at least one ROI. Also, we have shown that polyps were not always extracted completely. This impacts the sensitivity performance of our classifier that was trained to recognize ROIs that contain an entire polyp.

We have measured the percentage of polyps contained in the ROIs and obtained the following results:

- 11.5% contain between 90-100% of a polyp
- 31.6% contain between 50-90% of a polyp
- 56.9% contain between 1-50% of a polyp

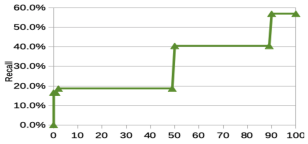

Fig. 2. Sensitivity according the percentage of polyp contained in ROI.

We have observed that a large amount of the ROIs does not contain an entire polyp and only 43% of the ROIs contain at least 50% of a polyp. We then measured the sensitivity indicator for each of these three cases. The results are visible in Figure 2.

As expected, a higher sensitivity is associated to ROIs containing higher percentages of the polyp. We need to improve the image processing to extract ROIs containing a high percentage of the polyp.

Additionally, we have analyzed the performances of the global processing at the image-level. At this level, in a first approach, we use a simple rule: an image containing at least one ROI classified as $class1$ is classified as $class1$; all other images are classified as $class0$. The specificity in this case is equal to the probability that all n ROIs are classified as $class0$; it is a binomial probability equal to:

$$Specificity_{Img\,level} = P^{n}_{(class_{ROI}=0)} = Specificity^{n}_{ROI\,level}$$

The number of ROIs of $class0$ in an image varies from 5 to 13, which could decrease the sensitivity from 78% to 53%. We measure a decrease of 73%. We notice that the rule used to define the image class is not efficient.

To solve these problems, we propose a new approach for polyp detection based on gastroenterologists' expertise.

III. Proposed Method

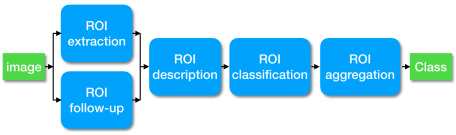

Fig. 3. Modification of our former hybrid approach with now five steps.

Our method is inspired by the gastroenterologists' psycho-visual methodology when they are performing a colonoscopy. They select the ROIs inside the gastrointestinal tract based on shape features. Once they select a ROI, they move the endoscope and follow the ROI to include a more accurate analysis and to detect the presence of an existing polyp. Based on this methodology, we modified our hybrid approach to have five steps (see in Figure 3). These steps are the following:

1) **extraction step**: the same processing as previously mentioned is used.
2) **follow-up step**: if an ROI was classified as $class1$ in the previous image, we use a motion estimation method to determine where the ROI should be in the current image and create a related ROI.
3) **description step**: the same processing as previously mentioned is used.
4) **classification step**: the same processing as previously mentioned is used.
5) **aggregation step**: Images' class is defined by an aggregation of all the ROIs of $class1$ in the last images.

In brief:

- We have added a **follow-up step**. The goal of this step is to improve the extraction of the ROIs that contain a high percentage of a polyp. In the case where a ROI is classified as $class1$ in image I_{n-1}, we use its location in the image I_{n-1} to estimate its location in the image I_n. We then place a ROI on this location. This new ROI is sent to the **description** step. In the **follow-up** step, the temporal $depth$ is denoted m to analyze the images from I_{n-1} to I_{n-m}. In this article, we have used $m = 1, 2$ and 3.
- We execute the **extraction** and the **follow-up** steps in parallel.
- We have added an **aggregation** step to determine the image's class by exploiting spatio-temporal information. An image will be considered as $class1$ if there is at least one ROI of $class1$ that was $class1$ on the last k images. In our case, we use $k = 3$.

With the **follow-up** step, we expect to increase the number of ROIs that contain more than 50% of a polyp by estimating their location in the next image.

With the **aggregation** step, we expect to increase the specificity at the image level.

In the next subsection, we describe the ROI **follow-up** step in detail.

A. *Follow-up step*

To follow a ROI validated as $class1$, we apply a motion estimation using a **block matching** algorithm described in Figure 4. Each ROI validated as $class1$ in the last image I_{n-1} is considered as a block $B_{p,q}$ of size $P * Q$. For all the pixels in the block, a **motion vector** is computed.

The motion estimation is performed by computing a similarity measurement between $I_{n-1}(B_{p,q})$ and $I_n(B_{p-i,q-j})$ using the intensity standard variation $Var(i,j)$ (equation 1). Here, we refer to the vector $I_{n-1(B_{p,q})}$ of the image I_{n-1} in the block $B_{p,q}$ as: $I_{n-1(B_{p,q})} = [I_{n-1}(p,q), ..., I_{n-1}(p+P-1,q+Q-1)]^T$.

$$Var(i,j) = \sqrt{\frac{\sum_{\forall p,q \in [BxP]} [I_{n-1}(B_{p,q}) - I_n(B_{p-i,q-j})]^2}{B*P}} \quad (1)$$

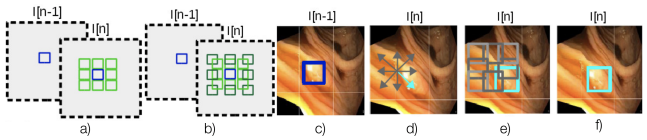

Fig. 4. For a depth=1, a) Block matching with 1 neighborhood. b) Block matching with 2 neighborhoods. c) ROI validated in the image I_{n-1}. d) 8 candidate motion vectors. e) 8 candidate blocks displaced by the candidate motion vectors. f) Candidate block having the lowest standard variation.

As an example, in the image I_n of Figure 4.a, the block B is displaced from the initial position (p,q) to $(p-i,q-j)$. This

motion corresponds to a motion vector $\overrightarrow{V} = (i, j)$. We identify this vector among 8 candidate motion vectors. To enhance the motion vector identification, we compute the candidate vector motion with different depths; in Figure 4 we see an example with a depth of 1 (a) and a depth of 2 (b). We can have a maximum depth of 10.

The motion vector $\overrightarrow{V} = (i, j)$ will be the candidate motion vector where $Var(i, j)$ is minimum. The followed ROI in the image I_n will be the validated ROI in the image I_{n-1} (see Figure 4) displaced by motion vector $\overrightarrow{V} = (i, j)$.

With this technique, we can follow the ROIs where a polyp was detected in the image I_{n-1}. Furthermore, we can increase the *temporal depth* of the motion estimation to follow the ROIs where a polyp was detected in the images $I_{n-1}, I_{n-2}, ..., I_{n-m}$.

IV. Experimental results

We have evaluated our modified hybrid approach with a *temporal depth* of $m = 0$, 1, 2 and 3. Additionally, we have evaluated the detection of the polyps at the ROI-level and the image-level (see Table II). In Table II, we notice that adding the **aggregation** step increases the specificity from 73% to 83% if a temporal depth of $m = 0$ is used. Furthermore, for a temporal depth of $m = 3$, we notice that the sensitivity increases from 30% to 57% at ROI-level and from 53% to 93.7% at image-level. The performance has increased by 40% compared to our previous approach while using processing that is compatible with integration inside an iWCE. If we analyze the specificity, we observe an important decrease at the ROI-level but a limited one of 10%, at the image-level. This can also be considered as a good result.

In addition, we have analyzed the ROIs and showed that the **follow-up** step improves the extraction of the ROIs with a larger part of the polyps:

- 15.2% of ROIs contain between 90-100% of a polyp,
- 40.4% of ROIs contain between 50-90% of a polyp,
- 44.4% of ROIs contain between 1-50% of a polyp,

TABLE II
Performances of the modified hybrid approach

	ROI-level		Image-level	
Temporal depth	sensitivity	Spec.	sensitivity	Spec.
m=0	33,48%	93.1%	66.7%	82.8%
m=1	53.8%	86.0%	93.2%	65.0%
m=2	56.8%	84.0%	93.3%	63.6%
m=3	57.1%	83.1%	93.7%	62.0%

These results show that our goals have been obtained using our new method. The amount of ROI identified that contains at least 50% of a polyp increased from 43% to 56% by using a simple Block Matching technique. Therefore, taking into account the spatiotemporal information by a ROI follow-up approach increases the performance of the detection.

By using the spatiotemporal aggregation, we increase the specificity and obtain a high score of **93.7**% for the sensitivity at the image-level.

V. Conclusions

In this paper, we have presented a hybrid approach to detect polyps that could be integrated in an iWCE [8]. This chain comes from the gastroenterologists' psychovisual methodology. The idea is to follow an extracted ROI using a simple motion estimation algorithm. With our proposition, we show an increase of the sensitivity from 53% to 93.7% at image level, validated on a large database of 18,910 images.

Follow-up is based on the analysis of the video stream and not only a per image analysis. Our results show that considering the spatiotemporal location of a polyp on a segment of video can improve the performance of the detection. The next work will be focused on a deep analysis of the spatiotemporal information and the membership degree of a ROI containing a polyp on a segment of the video. In addition, we will consider fuzzy trees to recognize ROIs with non-centered polyp to improve the detection rate. [1]

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray, "Cancer incidence and mortality worldwide: sources, methods and major patterns in globocan 2012," *International journal of cancer*, vol. 136, no. 5, pp. E359–E386, 2015.

[2] A. Karargyris and N. Bourbakis, "Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos," *IEEE transactions on Biomedical Engineering*, vol. 58, no. 10, pp. 2777–2786, 2011.

[3] J. Bernal, N. Tajkbaksh, F. J. Sánchez, B. J. Matuszewski, H. Chen, L. Yu, Q. Angermann, O. Romain, B. Rustad, I. Balasingham, K. Pogorelov, S. Choi, Q. Debard, L. Maier-Hein, S. Speidel, D. Stoyanov, P. Brandao, H. Córdova, C. Sánchez-Montes, S. R. Gurudu, G. Fernández-Esparrach, X. Dray, J. Liang, and A. Histace, "Comparative Validation of Polyp Detection Methods in Video Colonoscopy: Results From the MICCAI 2015 Endoscopic Vision Challenge," *IEEE Transactions on Medical Imaging*, vol. 36, pp. 1231–1249, June 2017.

[4] Y. Iwahori, T. Shinohara, A. Hattori, R. J. Woodham, S. Fukui, M. K. Bhuyan, and K. Kasugai, "Automatic polyp detection in endoscope images using a hessian filter.," in *MVA*, pp. 21–24, 2013.

[5] S. Y. Park and D. Sargent, "Colonoscopic polyp detection using convolutional neural networks," in *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785, p. 978528, International Society for Optics and Photonics, 2016.

[6] S. Ameling, S. Wirth, D. Paulus, and F. Vilarino, "Texture-based polyp detection in colonoscopy," in *Bildverarbeitung für die Medizin*, pp. 346–350, Springer, 2009.

[7] C. Orlando, P. Andrea, D. Xavier, and B. Granado, "Polyps recognition using fuzzy trees," in *Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on*, pp. 9–12, IEEE, 2017.

[8] C. Orlando, P. Andrea, M. Christophel, D. Xavier, and B. Granado, "FPGA-Based Real Time Embedded Hough Transform Architecture for Circles Detection," in *2018 Conference on Design and Architectures for Signal and Image Processing (DASIP)*, pp. 31–36, Oct. 2018.

[9] O. Chuquimia, A. Pinna, X. Dray, and B. Granado, "Fpga-based real time embedded hough transform architecture for circles detection," in *DASIP 2018-Conference on Design and Architectures for Signal and Image Processing*, 2018.

[10] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[11] O. Romain, A. Histace, J. Silva, J. Ayoub, B. Granado, A. Pinna, X. Dray, and P. Marteau, "Towards a multimodal wireless video capsule for detection of colonic polyps as prevention of colorectal cancer," in *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*, pp. 1–6, IEEE, 2013.

[1] Thanks to Hana Raffoul for her corrections supported in this paper.