Automatic Classification of Non-Informative Frames in Colonoscopy Videos

Cristian Ballesteros, María Trujillo, Claudia Mazo

Multimedia and Computer Vision Group
Universidad del Valle, Cali-Colombia

Keywords: Frame Classification, Edge Detection, Brightness Segmentation

Abstract

Colonoscopy is the most recommended test for prevention of colorectal cancer. Nowadays, digital videos are recorded during colonoscopy procedures and used for training machine learning algorithms. Machine learning algorithms are used for automatically recognizing lesions based on supervised learning. Moreover, annotation of lesions is a difficult and time-consuming process that is manually made by gastroenterologists. Those annotations may contain frames that have not useful information, called Non-Informative frames. The presence of Non-Informative frames in a group of frames labelled as lesion affects the accuracy of machine learning algorithms.
In this paper, a method based on edge detection is proposed to automatically classify a frame – from a colonoscopy video – into either Informative and Non-Informative. Non-Informative Frames usually do not contain many edges. However, brightness regions produce false edges. Therefore, the proposed method includes a technique for brightness segmentation to identify false edges. The proposed method is evaluated using videos annotated by gastroenterologists. Elimination of *Non-Informative* frames may reduce significantly the number of frames to be annotated by gastroenterologists and may improve the accuracy of machine learning algorithms. Experimental evaluation showed that the accuracy and the precision of the proposed method are over 95%.

1 Introduction

Every year, more than one million people are diagnosed with colorectal cancer in the world [5]. In Colombia, colorectal cancer is the fourth most common cancer [3]. The exact causes of colorectal cancer are unknown, however some factors are established, such as: age, family history, fatness and alcoholism [11]. An early detection of cancer is crucial for a successful treatment. Colonoscopy is the most recommended test for prevention and early detection of colorectal cancer. During a colonoscopy, a gastroenterologist uses a colonoscope to examine interior walls of the colon and check anomalies. The colonoscope is inserted into the rectum and advanced through the large intestine. It is a long flexible tube with a lens at one end and a video camera at the other. The end with the lens is inserted into a patient. Light passes down the tube to illuminate the area, and the video camera magnifies the area and projects it onto a television screen, in this way the gastroenterologist observes what is there. Videos may be captured during colonoscopies.
Automatic detection of polyps and cancer in colonoscopy videos is based on machine learning algorithms with supervised learning [6, 7, 2]. Machine learning algorithms build a model based on a training set. Annotations manually made by gastroenterologists on colonoscopy video are used to form training sets. Annotations of colonoscopy video may contain frames with blur, low contrast, noise or brightness. Those frames — called *Non-Informative* frames — do not contribute to the diagnosis. Since machine learning is an inverse and ill-posed problem, there is a set of assumptions that a learning algorithm makes about the true function that it is trying to learn a model off. If annotations – of cancer or polyps – contain *Non-Informative* frames, it may affect the learning of a model and, consequently, produce a classifier with low accuracy.
Research has been conducted on classifying Non-Informative frame in colonoscopy videos. Two techniques to classify colonoscopy frames into Informative and Non-Informative frames are proposed in [8]. The first technique is based on edge detection and the second one is based on the Discrete Fourier Transform. In edge-based frame classification, edges are detected using the Canny edge detector. *Isolated Pixel* (IP) is a term defined to identify edge pixels that are not connected to any other pixels. *Isolated Pixel Ratio* (IPR) is a term defined to measure the percentage of IP in a frame. Frames with large IPR value are labelled as *Non-Informative*. Otherwise, frames are labelled as *Informative*. The main advantage of the edge-based frame classification is the easy implementation. The main disadvantage of the edge-based frame classification is the sensitivity to the Canny threshold values.
A modification of the edge-based method proposed by [8] in order to distinguish between *Informative* and *Non-Informative* frames and reduce the number of frames to be transmitted, in a tele-surgery application, is presented in [10]. The proposed modification uses the Sobel edge detector and a high-pass filter for detecting edges. The IP and the IPR are used to classify frames into three categories: *Informative*Informative, *ambiguous* and *Non-Informative*. The *ambiguous* frames are filtered using a pixel count to classify into *Informative* or *Non-Informative*. The main advantage of this method is the simplic-

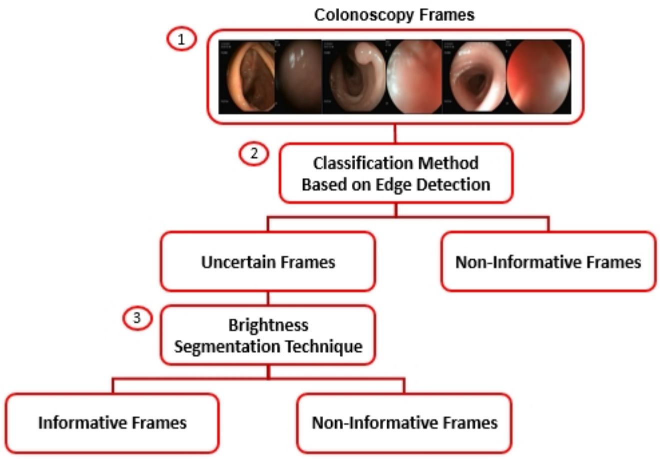

Figure 1: The process diagram of the proposed method.

ity. The main disadvantage of this method is that the required high-pass filter kernel values may vary.

In this paper, a method based on edge detection to automatically classify *Informative* and *Non-Informative* frames, in colonoscopy videos, is proposed. The proposed method uses a brightness segmentation technique to recognize false edges. The classification method is proposed in order to eliminate *Non-Informative* frames from training data sets used in machine learning algorithms. The remainder of the paper is organized as follows: In section 2, the proposed method is described. In section 3, experimental evaluation is presented. Finally, section 4 contains the conclusions.

2 Automatic Classification of Non-Informative Frames

The purpose of the classification is to identify *Non-Informative* frames. A brief description of Informative and Non-Informative frames is presented.

Informative Frames: information content is well defined and spread over whole image.

Non-Informative Frames: information content is captured out-of-focus with bubbles inside, light reflection artifact due to wall contact, light reflections on water used to clean the colon wall and motion blur.

Figure 1 illustrates the process diagram of the proposed method. First step, frames are extracted – using the FFmpeg Multimedia Framework [4]. Second step, edges are calculated and used to classify into two groups: *Uncertain* and *Non-Informative* frames. The *Uncertain* frames group contains *Informative* and *false-Informative* frames. Third step, the brightness segmentation (BS) is used to refine the classification of the *Uncertain* frames group by distinguishing *Informative* frames from *false-Informative* frames.

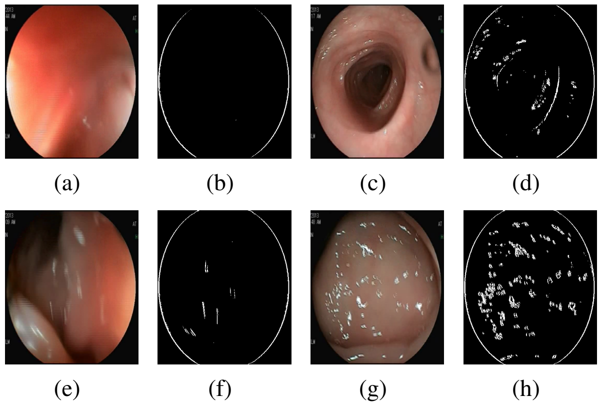

Figure 2: (a) Frame with light reflection artifact due to wall contact, (b) Edges from (a), (c) *Informative* frame, (d) Edges from (c), (e) Frame with motion blur, (f) Edges from (e), (g) Frame with light reflections on water, (h) Edges from (g).

2.1 Classification based on Edge Detection

Edges are an indicator of the amount of content in a frame. Commonly, an *Informative* frame has more edges than a *Non-Informative* frame. Figure 2 shows a frame with light reflection artifact due to wall contact, an *Informative* frame, a frame with motion blur, and a frame with light reflections on water, along with edges extracted from them using the Sobel edge detector. Sobel uses a pair of 3×3 convolution masks for calculating the gradient magnitude, in the X-direction and in the Y-direction. The gradient magnitude in both directions is used to detect edges [1]. The amount of edges extracted from a frame with light reflection artifact, due to wall contact, and a frame with motion blur is less than the amount of edges extracted from an *Informative* frame.

The classification method is based on the assumption that an *Informative* frame has larger amount of edges than a *Non-Informative* one. The metric called *Edge Pixel Percentage* (EPP) is introduced to calculate the percentage of edge-pixels in a frame.

$$EPP = \frac{Number\ of\ edge\ pixels}{Total\ number\ of\ pixels} \times 100. \tag{1}$$

A frame with an EPP larger than an experimental threshold (**Th**) is labelled as *Uncertain*. Otherwise, the frame is labelled as *Non-Informative*. The classification process based on edge detection is presented in Figure 3.

2.2 Brightness Segmentation

Brightness regions due to light reflections on water used to clean the colon wall have a higher luminance value as illustrated in Figure 4.

Since light reflection produces false edges and frames are wrongly classified as *Informative*, a segmentation of brightness regions is proposed in order to detect false edges. The segmentation of brightness regions is illustrated in Figure 5.

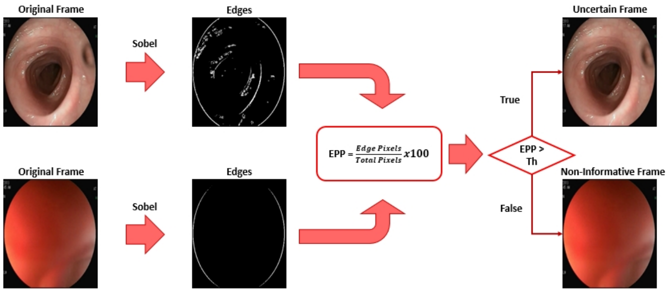

Figure 3: Illustration of Classification based on Edge Detection.

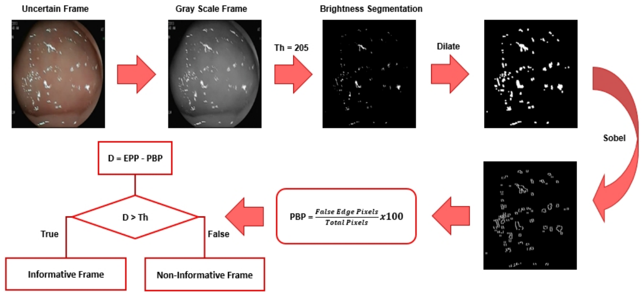

Figure 5: Illustration of Segmenting Brightness Regions.

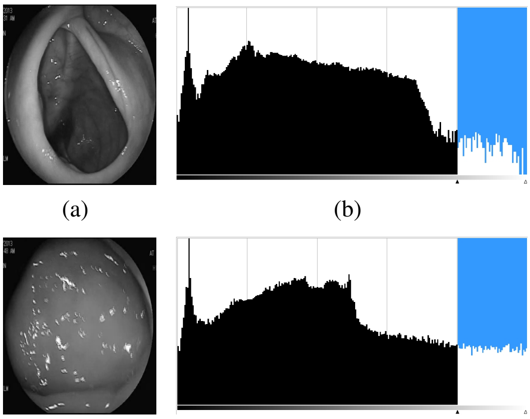

Figure 4: *Uncertain* frames: (a) *Informative* frame, (b) Histogram from (a), (c) *False-Informative* frame with brightness regions due to light reflection, (d) Histogram from (c) – where frequency values from 205 to 255 are in white color and highlighted with blue color.

A brief description of the brightness segmentation is presented as follows: A frame is converted into gray scale and a threshold of 205 is applied to obtain a binary image. Pixels segmented as brightness regions have value of one and are morphologically modified using a dilation, with a rectangular 5×5 structuring element, in order to enhance edges. Finally, edges – called false edges – are obtained using the Sobel edge detector. The percentage of brightness-edge pixels (PBP) is calculated as:

$$PBP = \frac{Number\ of\ false\ edge\ pixels}{Total\ number\ of\ pixels} \times 100. \quad (2)$$

The difference between EPP and PBP is calculated and used for eliminating the effect of false edges in the frame classification. The obtained result is compared to the threshold (**Th**). If the difference is larger than **Th**, the frame is labelled as *Informative*. Otherwise the frame is labelled as *Non-Informative*.

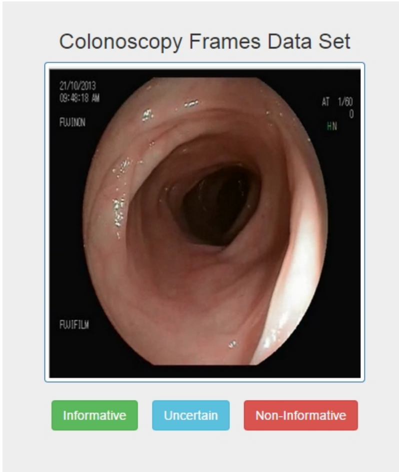

Figure 6: Interface for Annotating Colonoscopy Data Set.

3 Experimental Evaluation

The aim of these experiments is to evaluate the performance of the proposed method and compare to the Canny edge method proposed in [8]. In the following data set, experimental settings and evaluation criteria are explained. All tests are conducted on a Laptop with Windows 8 Pro X64, Intel (R) Core (TM) i5 @ 2.60 GHz and 4,00 GB RAM.

3.1 Data Sets

The performance of the proposed method is evaluated using 3 colonoscopy videos from the Hospital Univesitario del Valle. Each video has length of 6, 12 and 14 minutes, respectively, frame resolution of 636×480 and was recorded at 10 fps, using MP4 format and H264 compression.

The evaluation data set consists on frames extracted from videos, a total of 2000 frames – containing 1000 Informative and 1000 Non-Informative – were used. The extracted frames were manually annotated by a gastroenterologist using the web interface in Figure 6.

3.2 Performance Metrics

A set of metrics commonly used to evaluate the performance of a binary classification is employed [9].

The confusion matrix is calculated using the following terms:

- *True Informative* (TI): Informative frame correctly classified,
- *False Informative* (FI): Non-Informative frame incorrectly classified as Informative frame,
- *True Non-Informative* (TNI): Non-Informative frame correctly classified, and
- *False Non-Informative* (FNI): Informative frame incorrectly classified as Non-Informative. frame.

The selected metrics for evaluating performance are as follows:

- *Sensitivity* (Sen): A measure of correct classification of frames as *Informative*. This metric is defined as:

$$Sen = \frac{\sum TI}{\sum TI + \sum FNI} \tag{3}$$

- *Specificity* (Spec): A measure of correct classification of frames as *Non-Informative*. This metric is defined as:

$$Spec = \frac{\sum TNI}{\sum TNI + \sum FI} \tag{4}$$

- *Accuracy* (Acc): The percentage of frames correctly classified as *Informative* and *Non-Informative*. This metric is defined as:

$$Acc = \frac{\sum TI + \sum TNI}{\sum TI + \sum FI + \sum TNI + \sum FNI} \tag{5}$$

- *Precision* (Prec): The ratio of the number of frames correctly classified as *Informative* to the total number of frames correctly classified as *Informative* and incorrectly classified as *Informative*. This metric is defined as:

$$Prec = \frac{\sum TI}{\sum TI + \sum FI} \tag{6}$$

- *F-Measure* (FM): The harmonic mean of precision and sensitivity. The ideal value for this metric is 1. F-Measure is defined as:

$$FM = 2 \cdot \frac{Prec \cdot Sen}{Prec + Sen} \tag{7}$$

3.3 Results

The evaluation is to compare different classification conditions, with and without the Brightness Segmentation (BS) using a threshold equal to 2.5. Moreover, the proposed approach is compared to the Canny edge method proposed in [8]. The Canny edge method was implemented using T-low= 20 and T-high= 50 and the decision criterion was adjusted accordingly to the assumption that an *Informative* frame has larger amount of edges than a *Non-Informative* one. A frame with IPR larger or equal to than 0.08 is classified as *Informative*.

Tables 1 and 2 show obtained results using the proposed method and the Canny edge method. As expected, the use of BS reduced the number of frames wrongly classified as *Informative* and increase the specificity, the accuracy and the precision values. The computer performance of the proposed method is presented in Table 3.

Classification	TI	FI	TNI	FNI
without the BS	936	113	887	64
with the BS	925	11	989	75
Canny	793	95	905	207

Table 1: Number of frames classified by the proposed method and the Canny edge method in [8].

Classification	Sen	Spec	Acc	Prec	FM
without the BS	0.94	0.89	0.91	0.89	0.91
with the BS	0.92	0.99	0.96	0.99	0.96
Canny	0.79	0.90	0.85	0.89	0.84

Table 2: Performance metrics calculated using the proposed method and the Canny edge method in [8].

4 Conclusions

In this paper, a method based on edge detection – using the Sobel edge detector – was proposed to classify colonoscopy frames into two categories: *Informative* and *Non-Informative*. In addition, the proposed method includes a segmentation of brightness regions due to light reflection that produce false edges.

The edge based classification method is able to correctly detect frames without relevant information to colon-lesson diagnosis that exist in data-sets used for training machine learning algorithms. In the same way, the proposed method may be used to significantly reduce duration of videos – frames classified as *Non-Informative* are deleted from the colonoscopy video – before being analysed by gastroenterologists.

The metrics used to evaluated the performance of the proposed method show that accuracy and precision are over than 95% when false edges originated by light reflection are removed. And the specificity of the proposed method was 99%.

Classification	Total Time [sec] (2000 Frames)
without the BS	113.93
with the BS	135.67
Canny	107.24

Table 3: Computer performance of the proposed method and the Canny edge method in [8].

Acknowledgements

The authors would like to express their sincere appreciation to Dr. Jesus A. Hoyos (Hospital Universitario del Valle) for recording and annotating colonoscopy videos. Thank to Ms. Vivian Sandoval and Ms. Diana Gutierrez for helping to accomplished the evaluation of the propose method.

References

[1] T. Acharya and K. Ray. *Image Processing Principles and Applications*. John Wiley and Sons, Inc., New Jersey, first edition, 2005.

[2] J. Bernal, D. Gil, C. Snchez, and F. J. Sanchez. Discarding non informative regions for efficient colonoscopy image analysis. *Lecture Notes in Computer Science*, 8899:1–10, 2014.

[3] Liga Contra el Cancer. Cancer colorrectal, 2015. [Web; accessed on 06-11-2015 to http://www.ligacontraelcancer.com.co/cancer-colorrectal/magnitud-del-problema/].

[4] FFMPEG. Ffmpeg, multimedia framework, 2015. [Web; accessed on 06-11-2015 to https://www.ffmpeg.org/].

[5] WCRF International. Colorectal cancer statistics, 2012. [Web; accessed on 06-11-2015 to http://www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/colorectal-cancer-statistics].

[6] S. Manivannan, R. Wang, E. Trucco, and A. Hood. Automatic normal-abnormal video frame classification for colonoscopy. *2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, pages 644 – 647, 2013.

[7] S. Manivannan, R. Wang, M. P. Trujillo, J. A. Hoyos, and E. Trucco. Video-specific svms for colonoscopy image classification. *Lecture Notes in Computer Science*, 8899:11 – 21, 2014.

[8] J. Oh, S. Hwang, J. Lee, W. Tavanapong, J. Wong, and Piet C. de Groen. Informative frame classification for endoscopy video. *Medical Image Analysis*, 11(1):110–127, 2007.

[9] D. M. W. Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 02(1):37 – 63, 2011.

[10] N. Rungseekajee, M. Lohvithee, and I. Nilkhamhang. Informative frame classification method for real-time analysis of colonoscopy video. *ECTI-CON 2009. 6th International Conference*, 02(1):1076 – 1079, 2009.

[11] Mayo Clinic Staff. Diseases and conditions colon cancer, 2013. [Web; accessed on 06-11-2015 to http://www.mayoclinic.org/diseases-conditions/colon-cancer/basics/causes/con-20031877].