Informative Frame Automated Extraction from Colonoscopy Videos

Juan C. Arcila-Diaz, Victor Tuesta-Monteza, Heber I. Mejia-Cabrera
Señor de Sipán University
Peru
diarcilaju@crece.uss.edu.pe, vtuesta@crece.uss.edu.pe, hmejia@crece.uss.edu.pe

Maria P. Trujillo-Uribe
Universidad del Valle
Colombia
Maria.trujillo@corrounivalle.edu.co

Kim Jeong-Hyun
Pusan National University
Korea, Republic Of
mare@pusan.ac.kr

ABSTRACT: *Colonoscopy videos contain blurred, non-informative frame sequences due to the rapid movements of the endoscope during the exploration which need to be excluded to allow the expert physician to carry out his work in less time. In this paper two methods of artificial vision are proposed for the automated extraction of informative frames based on detectable characteristics of them. The first method allows for frame sorting into informative and non-informative based on the number of contours detected in each frame. The second method makes use of the dense optical flux to determine the percentage of individual frame motion, for group the frames whit K-Means algorithm by your motion in three groups: mean motion (informative frame), large motion and little motion (non-informative frame). Both methods were successful in filtering out blurred frames from colonoscopy video samples with the first method outperforming the second, i.e., 76.7% accuracy versus 74.7%, respectively.*

Keywords: Motion Estimation, Blurry Frames, Colonoscopy, Frame Classification

Received: 27 September 2017, Revised 3 November 2017, Accepted 9 November 2017

1. Introduction

The advancement of technology serves as a support in several professional fields, in the medicine the images diagnostic allows physicians to observe inside the human body to look for inklings about a medical condition, there are currently several devices and techniques that can represent digital images of the structures and activities within the organism [1].

During an endoscopic procedure, which is a medical imaging diagnostic technique, a small camera included in the tip of the endoscope is inserted gradually into the human body to inspect some abnormalities, this camera generates a video signal from the interior of the human body [2] This signal is stored in a video file to faithfully document the findings and endoscopic interventions and are treated as an important part of the medical records for further analysis [3].

Colonoscopy, which is a type of endoscopy, is used as a great tool for the detection of colorectal cancer, a disease considered one of the leading causes of death in several countries [4]. Although colonoscopy has contributed to a significant decrease in deaths related to colorectal cancer, the analysis of the videos resulting from this medical examination presents some drawbacks, lots parts of the resulting video are out of focus due to the rapidity with which it advances the camera, intense reflections of light, gastric mucosa of the colon and some residues adhere to the camera, then blurred frames - called "non-informative frames" - are created which do not show significant information [5][6][7].

To classify medical diagnostic video frames into informative and non-informative, various efforts have been focused on exploiting digital imaging techniques. Specifically, lots research have been developed in recent years to improve the visualization of endoscopic examinations [8] [9] and other recent researches suggest methods for the identification of informative frames in resultant videos from an examination of endoscopy; in the research Informative frame classification for endoscopy videos [5] with the purpose of reducing the number of images seen by a physician and analyzed by a CAD system, a technique (edge-based and clustering-based) is proposed to classify the frames into informative and non-informative. However, because intensive specular reflections reduce the accuracy of the classification we also propose a specular reflection detection technique, and use the detected specular reflection information to increase the accuracy of informative frame classification; Obtaining an accuracy of up to 95% for the identification of informative frames.

In this research two methods were implemented for the automated extraction of informative frames from video-recorded colonoscopies by processing and sorting them using and edge detection technique and a motion detection technique using dense optical flow [10]. The former extracts and counts the contours of each frame while the latter obtains the motion percentage by analyzing the transition from one frame to the next. The K-Means algorithm is used to sort frames into informative and non-informative.

2. Related Works

Several previous works have been carried out to extract informative frames from medical diagnosis video sequence. For instance, in Leszcuk and Mariusz's [11] work a Discrete Cosine Transform (DCT) algorithm was implemented to exclude non-informative frames and ensure the inclusion of informative ones to document areas or reveal endobronchial pathological lesions during bronchoscopic examination. DCT was used to detect edges based on the fact that non-informative frames show lower spectra in the frequency domain compared with frames that do have many edges. The application developed allocated for video sequence selection and inclusion/exclusion frame parameter definitions until the user is satisfied.

Hwang *et al* [12] developed a method based on the Discrete Fourier Transform (DFT) for evaluating the colonoscopy quality and to exclude non-informative frames. The rationale behind their work was that the frequency spectra of the informative frames and non-informative show different patterns; for example, non-informative frames have clear object information except for four strong edges at the corners of an image running approximately ± 45 grades, while informative frames have a lot of clear edges in their spectrum and not show prominent components along the ± 45 grades. They obtained seven texture characteristics for each frame: entropy, contrast, correlation, homogeneity, the dissimilarity, the second angular momentum and energy. The sorting stage into informative and non-informative frames was carried using the K-MEANS algorithm in a two-level form using the seven texture characteristics as the criteria.

3. Methodology

Two methods are proposed for extracting video informative frames. The first method is based on the number of identifiable contours, i.e non-informative frames have few contours compared to informative frames. The second method is based on motion detection using dense optical flow followed by the K-Means algorithm which takes advantage of the lower spectra in non-informative frames compared to informative ones.

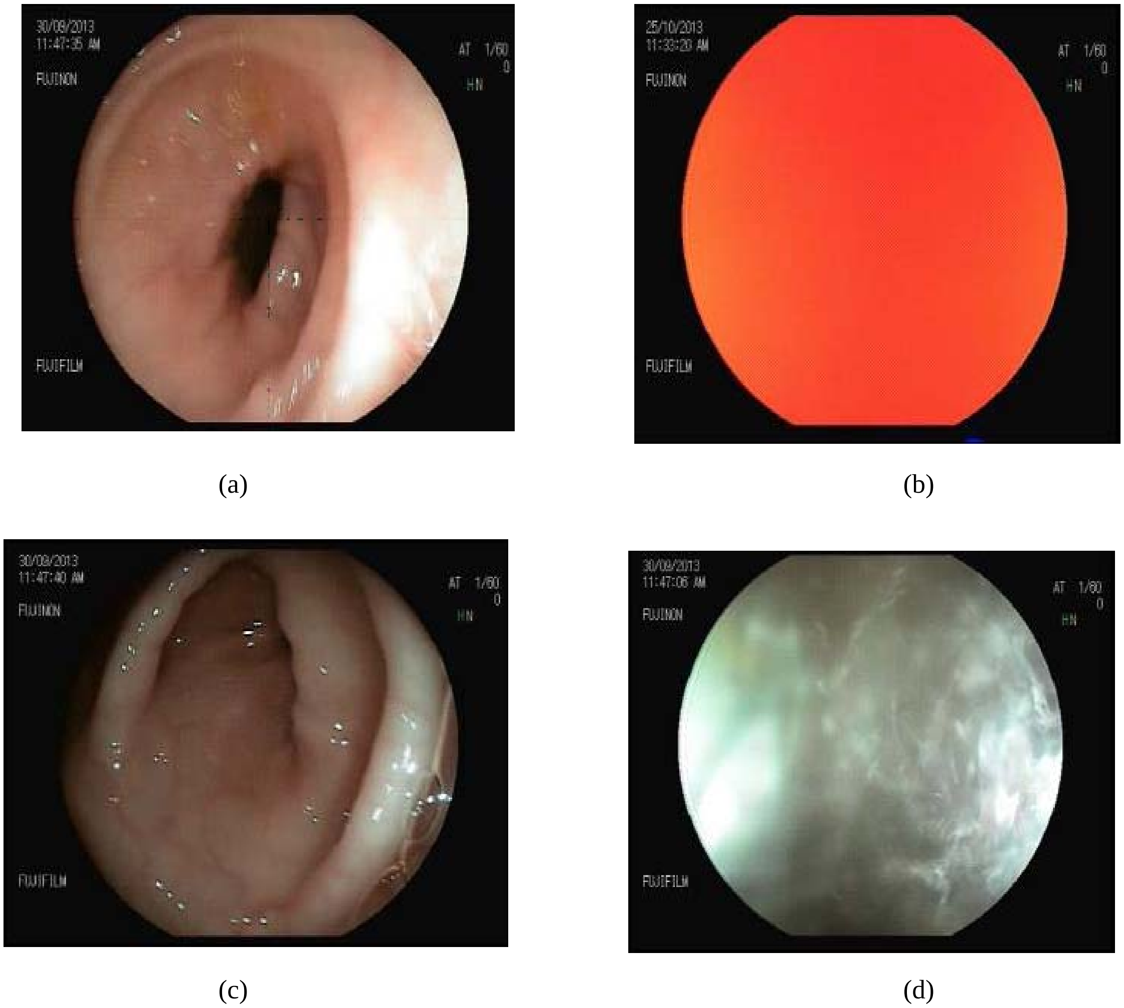

Figure 1. Illustration of (a, c) informative frame, and (b, d) non-informative frame

3.1. Extraction of Informative Frames Based on Number of Identifiable Contours

The proposed method is variant of Leszcuk and Mariusz's [11] scheme which includes a different edge detection algorithm [13] The number of contours detected determines if a frame is informative or non-informative. If the frame has few contours it is a frame

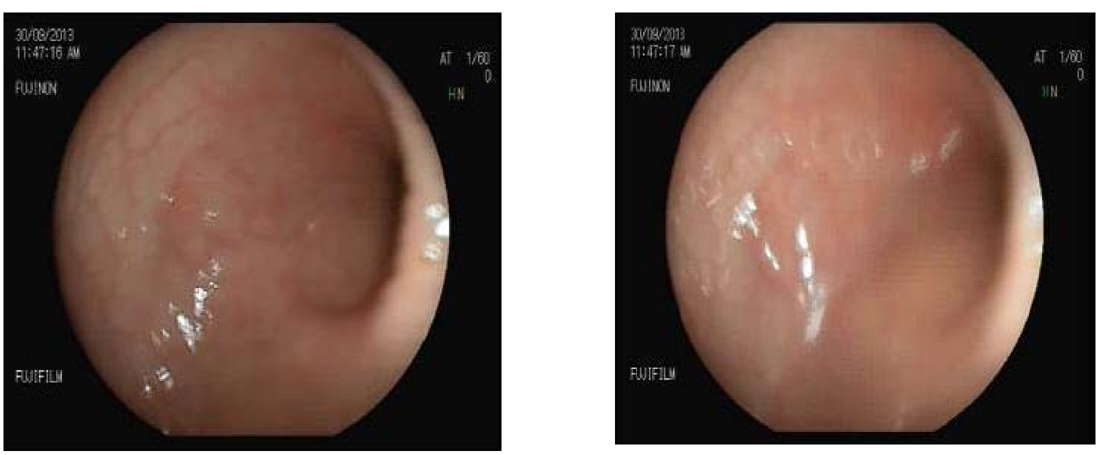

Figure 2. Illustration of consecutive frames in RGB

with no significant information because it is unfocused due to a rapidly changing camera positioning or due to the clashing of the camera lens with the mucosa walls.

3.1.1 Pre- Processing

Two consecutive frames are initially processed after the video sequence is input (*Fig. 2*), iterating until all frames have been processed.

RGB frames (Fig. 2) are change into grayscale (Fig. 3) using the following equation:

$$y = 0.299R + 0.587G + 0.114B \qquad (1)$$

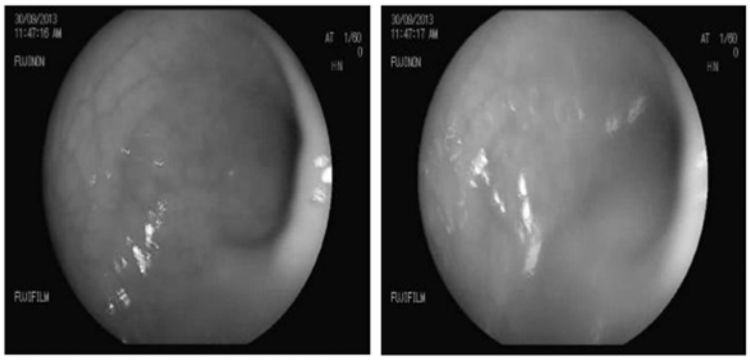

Figure 3. Consecutive frames converted to grayscale

3.1.2 Absolute Difference Operation

Using the two-new grayscale frames an absolute difference between the two frames is created. Given the grayscale frames $I_i(x,y)$ and $I_{i+1}(x,y)$, the frame difference is obtained $I_{diff}(x,y)$, and a binary image is created such that:

$$I_{diff}(x,y) = \begin{cases} |I_i(x,y) - I_{i+1}(x,y)|, & si|I_i(x,y) - I_{i+1}(x,y)| \leq 255 \\ 255, & si|I_i(x,y) - I_{i+1}(x,y)| > 255 \end{cases} \qquad (2)$$

3.1.3 Filter of Non-informative Frames

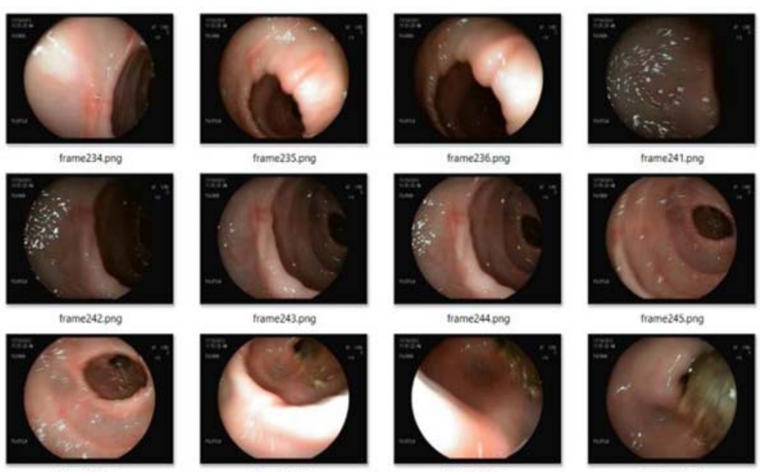

Figure 5. Frames classified as informative

To determine if a frame is informative or not, the number of contours is calculated in each frame. After using a conditional structure are classified into informative (Fig. 5) or and non-informative (Fig. 6) by the number of contours that have. If a lot of contours is found the frame is considered as informative. If a relatively small number of contours is found, then the frame is non-informative.

.

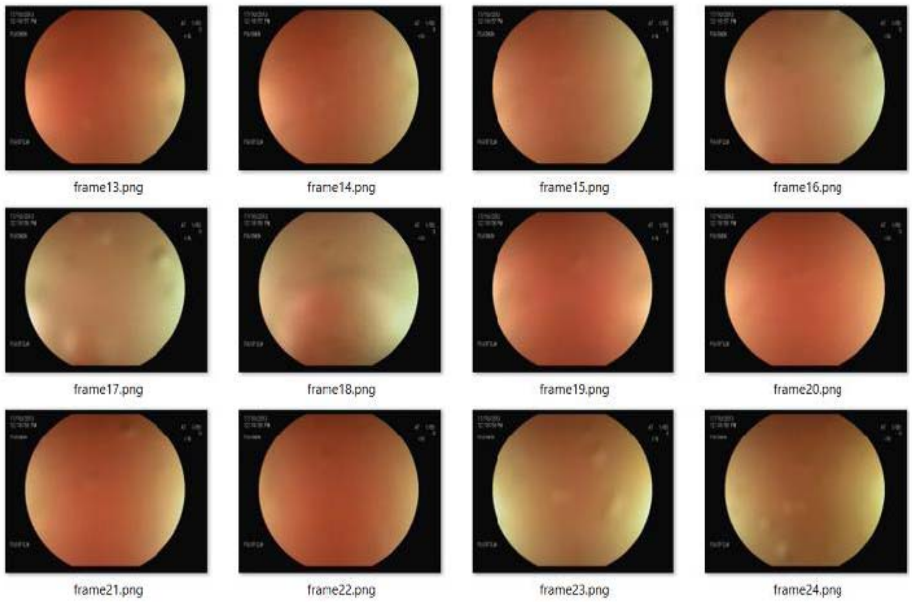

Figure 6. Frames classified as non-informative

3.2 Extraction of Informative Frame used Dense Optical Flow

This method is based on the motion detected between frames using the dense optical flow[10][14]. If consecutive frames have little or no movement set the camera has bumped into the colonic mucosa or some intestinal residue and therefore have no meaningful information. If these frames show regular motion, it is said that we have informative frames. Additionally, if high motion is detected it implies that the captured frames are blurred due to a high camera speed leading again to non-informative frames.

3.2.1 Initial Process

As in the previous method, frames are converted from RGB to greyscale.

3.2.2 Motion Percentage of Each Frame

To obtain the motion, be $I(x, y, t)$ the pixel (x, y) intensity in frame at time t and suppose that there is a translation in (v_x, v_y) such that:

$$I(x + v_x, y + v_y, t + 1) = I(x, y, t). \tag{3}$$

It was used the development of first order Taylor series:

$$\frac{\partial I}{\partial t}(x, y, t) + \frac{\partial I}{\partial x}(x, y, t)v_x + \frac{\partial I}{\partial y}(x, y, t)v_y = 0. \tag{4}$$

Two matrices are obtained use dense optical flow depending on whether frame motion is detected in the "*x*" coordinate, $A(x, y, t)$, or in the "*y*" coordinate, $B(x, y, t)$. A third matrix, $C(x, y, t)$, is obtained such that $\{0,1\} \in C(x, y, t)$

$$C(x, y, t) = A(x, y, t) * B(x, y, t). \tag{5}$$

Once "*C*" is obtained, the percentage of pixels, "*P*", representing motion, i.e. the percentage of elements with value 1 in $C(x, y, t)$, is calculated as:

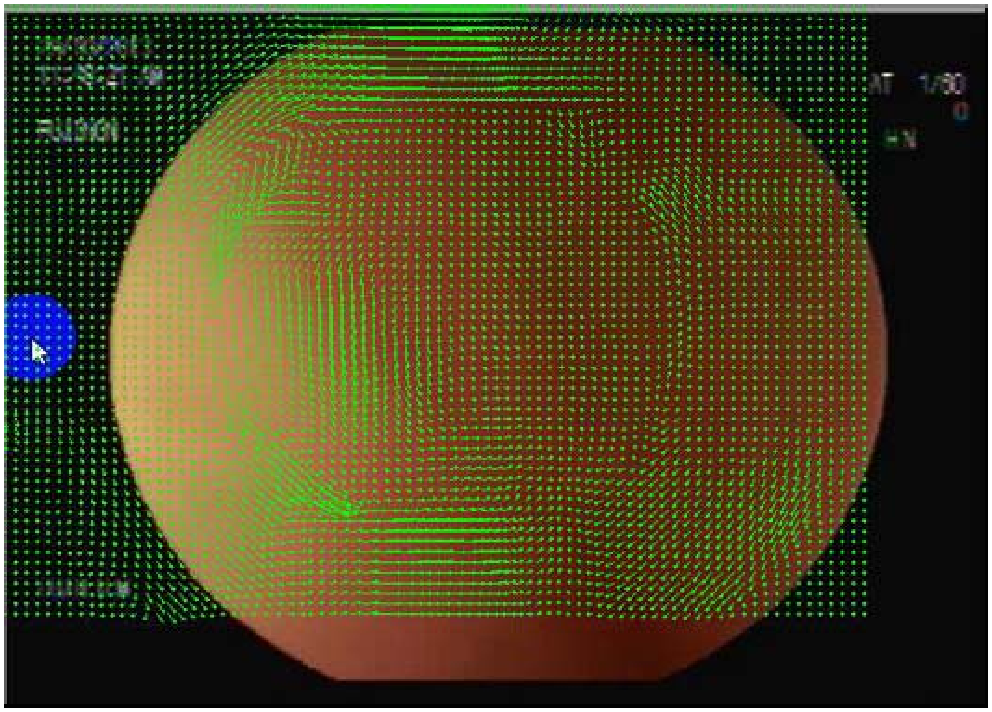

Figure 7. Dense optical flow calculation of a frame

$$P = \left(\left(\sum_{1}^{x*y} C_{(x_i, y_i, t)} \right) / (x * y) \right) * 100. \tag{6}$$

The percentage of motion (P) that exists in the transition from one frame (F) to another is shown in Table 1.

3.2.3 Frame Group

The percentage of motion obtained is used as grouping variable along with the K-means algorithm. An initial stage for classifying frames is to set the number groups to 3 ($\boldsymbol{m}1^{(1)}, \boldsymbol{m}2^{(1)}, \boldsymbol{m}3^{(1)}$): large motion, mean motion and little motion; initially randomly it sets a centroid for each group. The procedures continue with two iterating steps:

Assignment Step: Performed first assigning percent value of each frame movement (x^{p}) to the group ($S_i^{(t)}$).

$$S_i^{(t)} = \{x_p : ||x_p - m_i^{(t)}|| \le ||x_p - m_j^{(t)}|| \forall 1 \le j \le k\} \tag{7}$$

Update Step: The new centroids are computed as the centroids of the observations in the group.

$$\frac{\partial I}{\partial t}(x, y, t) + \frac{\partial I}{\partial x}(x, y, t) v_x + \frac{\partial I}{\partial y}(x, y, t) v_y = 0. \tag{8}$$

These two steps are repeated until there is no change in the frame assignment to one of the three defined groups.

After frame grouping is finished three groups (G) are considered for which 0 indicates frames with large motion, 1 indicates medium motion frame, and 2 indicates little motion. Table 2 shows the firsts 124 frames processed assigned to its corresponding group.

F	P	F	P	F	P	F	P
1	2.52	32	53.17	63	0.52	94	45.63
2	0.00	33	35.04	64	0.00	95	52.38
3	4.73	34	4.13	65	2.13	96	50.13
4	0.00	35	2.21	66	1.60	97	55.48
5	4.42	36	0.52	67	0.50	98	67.54
6	6.38	37	0.31	68	1.79	99	44.04
7	3.48	38	0.00	69	0.94	100	34.96
8	7.77	39	2.35	70	0.88	101	53.48
9	37.58	40	8.94	71	0.00	102	43.48
10	6.31	41	30.73	72	2.25	103	38.71
11	13.52	42	41.17	73	2.27	104	62.67
12	20.92	43	48.42	74	6.77	105	59.06
13	41.81	44	8.29	75	12.42	106	67.75
14	11.08	45	18.71	76	20.42	107	58.85
15	7.31	46	21.08	77	36.52	108	64.21
16	14.81	47	18.94	78	0.08	109	62.50
17	17.58	48	16.56	79	3.98	110	65.31
18	5.27	49	34.33	80	21.96	111	79.04
19	30.94	50	60.96	81	6.67	112	68.54
20	42.52	51	71.81	82	12.10	113	62.13
21	37.31	52	13.81	83	6.71	114	33.42
22	50.19	53	19.33	84	6.75	115	19.63
23	61.44	54	11.81	85	3.00	116	22.65
24	70.48	55	5.02	86	6.79	117	14.71
25	69.58	56	9.35	87	9.71	118	17.90
26	55.98	57	3.88	88	29.13	119	21.42
27	68.77	58	3.27	89	48.48	120	22.50
28	72.06	59	3.17	90	46.92	121	5.77
29	75.29	60	0.79	91	24.35	122	7.48
30	25.77	61	1.31	92	47.42	123	4.92
31	42.46	62	0.00	93	51.19	124	4.10

Table 1. First 124 frames of a video sample and its corresponding percentage of motion

F	G	F	G	F	G	F	G	F	G	F	G
1	2	22	0	43	1	64	2	85	2	106	0
2	2	23	0	44	2	65	2	86	2	107	0
3	2	24	0	45	2	66	2	87	2	108	0
4	2	25	0	46	2	67	2	88	1	109	0
5	2	26	0	47	2	68	2	89	1	110	0
6	2	27	0	48	2	69	2	90	1	111	0
7	2	28	0	49	1	70	2	91	2	112	0
8	2	29	0	50	0	71	2	92	1	113	0
9	1	30	1	51	0	72	2	93	0	114	1
10	2	31	1	52	2	73	2	94	1	115	2
11	2	32	0	53	2	74	2	95	0	116	2
12	2	33	1	54	2	75	2	96	0	117	2
13	1	34	2	55	2	76	2	97	0	118	2
14	2	35	2	56	2	77	1	98	0	119	2
15	2	36	2	57	2	78	2	99	1	120	2
16	2	37	2	58	2	79	2	100	1	121	2
17	2	38	2	59	2	80	2	101	0	122	2
18	2	39	2	60	2	81	2	102	1	123	2
19	1	40	2	61	2	82	2	103	1	124	2
20	1	41	1	62	2	83	2	104	0		
21	1	42	1	63	2	84	2	105	0		

Table 2. First 124 frames of a video sample labeled with a corresponding group

In Figure 8 some informative frames samples (Group 1) are shown. Non-informative frames are shown in Figure 9 (Group 2, little motion) and in Figure 10 (Group 0, large motion).

4. Experiments and Results

To evaluate the methods proposed is used a database composed of 20 videos of colonoscopy, was evaluate 309 frames per video, with a total of 12360 frames in each test performed.

The proposed methods were tested for comparison using the following metrics:

4.1 Average Processing Time

This is the average time it takes to process a video sequence while extracting informative frames.

The identifiable contours method is 83.5% faster compared with the dense optical flow method.

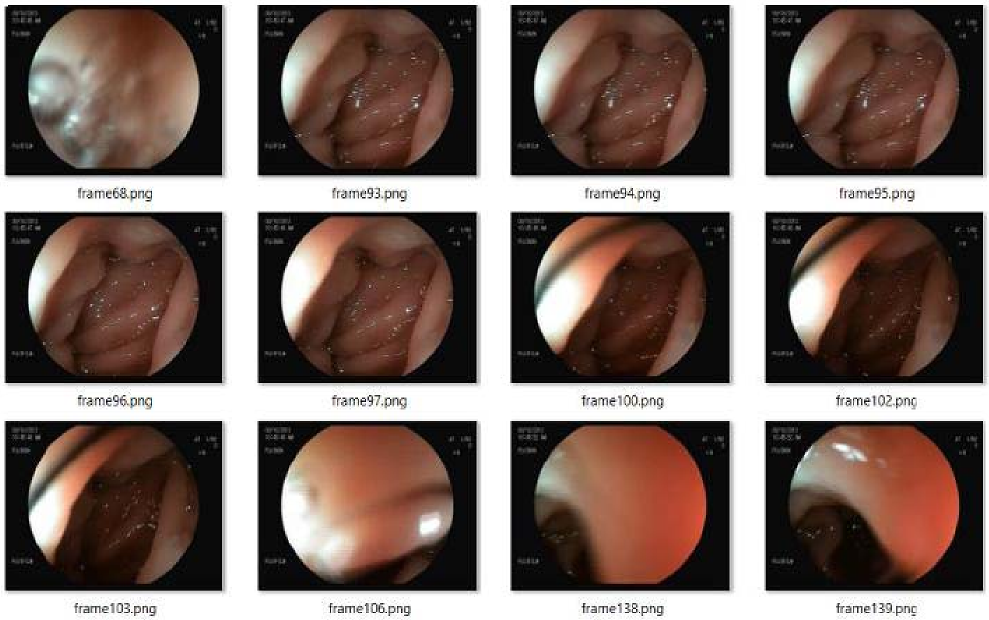

Figure 8. Group 1, informative frames

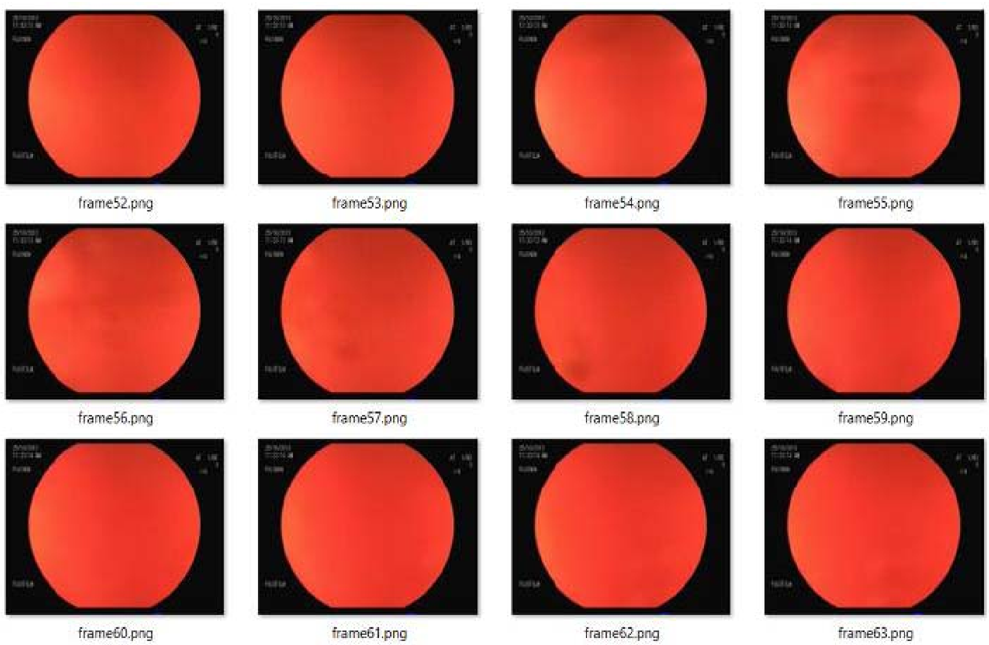

Figure 9. Group 2, Non-informative frames

Average processing time	
Methods	**Time (sec)**
Identifiable contours method	2,099
Dense optical flow method	3,853

Table 3. Average processing time of each method

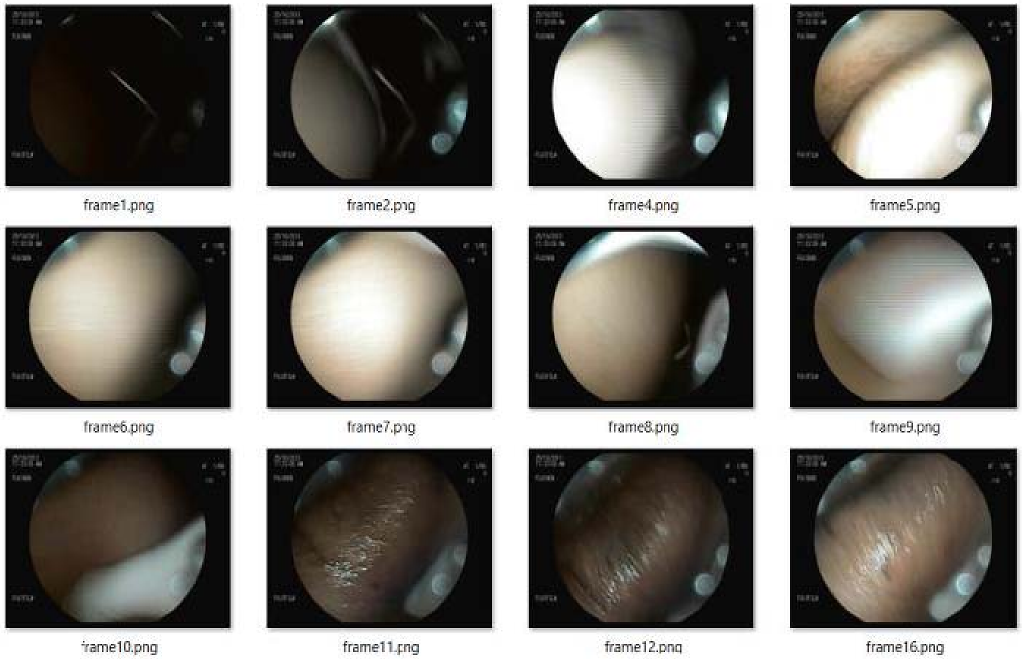

Figure 10. Group 0, non-informative frames

with the dense optical flow method.

4.2 Percentage Of Informative Frames Correctly Classified

This metric is defined by the percentage of informative frames correctly classified and is given by:

$$P = \frac{|\#\mathrm{T}success \cap \#\mathrm{T}samples|}{|\#\mathrm{T}samples|} \qquad (9)$$

Where, $\#T_{samples}$ is the total number of frames processed and with the help of a specialist physician is manually determined $\#T_{success}$ the number of frames successfully extracted.

Table 4 shows that the identifiable contours method is 68.49% accurate for informative frame extraction while the dense optical flow method reaches only 60.10% in this metric.

Methods	Average Correct	Average Incorrect	% Accuracy	% Error
Identifiable contours method	97,40	43,85	68,49%	31,51%
Dense optical flow method	60,25	36,20	60,10%	39,90%

Table 4. Average accuracy of each method to classify information frames

4.3 Percentage of Non-Informative Frames Correctly Classified

Table 5 is built using an equation similar to equation (9), in this table it shows that the method based on the amount contour obtains greater accuracy, an average of 82.85%, for the classification of not informative frames; while the method based on dense optical flow method obtained a percentage of 79.58% accuracy for classifying non-informative frames.

Methods	Average Correct	Average Incorrec t	% Accuracy	% Error
Identifiable contours method	139,55	28,20	82,85%	17,15%
Dense optical flow method	170,70	41,85	79,58%	20,42%

Table 5. Average accuracy of each method to classify non-informative frames

4.4 Time Reduction

The percentage of time reduction obtained in each video sequence is calculated using the following equation:

$$R = \frac{\text{TimeFinalInformative}}{\text{TimeVideoInitial}} x\ 100 \tag{10}$$

Table 6 shows the percentage of reduction of each video after extracting the non-informative frames, the contours based classification method obtained a reduction average of 45.71%, while the method based on the movement obtained a percentage of 31.21% reduction in the duration of the resulting videos.

Methods	Reduction
Identifiable contours method	45,71%
Dense optical flow method	31,21%

Table 6. Average reduction of the resultant videos including only the informative frames

4.5 Overall Accuracy

To obtain a general precision, the percentage obtained by each method to correctly classify informative frames and non-informative frames has been averaged. The results of the final accuracy of each method shown in Table 7. As it can be inferred from Table 7, contours based classification method is more accurate (76.7%) than the dense optical flow based classification method (74.7%).

Method	Accuracy
Identifiable contours method	76.7%
Dense optical flow method	74.7%

Table 7. Average of each method to classify frames in informative and non-informative

5. Conclusion

In this paper, a method based on edge detection - using your identifiable contours - and other method based on motion estimation - using dense optical flow -was proposed to classify colonoscopy frame into two groups: Informative and Non-Informative.

The contours based classification method is able to correctly extract frames without relevant information to colonoscopy diagnosis, this method may be used to significantly reduce duration of videos – frames classified as Non-Informative are excluded from the colonoscopy video – before being analyzed by medical specialist.

After the metrics tests, it was determined that out of the two proposed methods the most efficient one – fewer resource consumption and less processing time - is the identifiable contours method, which obtained an accuracy of 76.7% compared to the dense optical flow, 74.7%, in extracting informative frames from colonoscopy videos.

6. Future Research

In the near future, the use of the algorithm DCT will be pursued to extract the clustering characteristics in order to ensure that absolutely all informative frames needed for medical diagnosis are included in the final processed video.

References

[1] National Library of Medicine. (2017). Diagnóstico por imágenes, [Online]. Available: https://medlineplus.gov/spanish/diagnosticimaging.html. [Accessed: 26-May-2017].

[2] Stanek, S. R., Tavanapong, W., Wong, J., Oh, J. H., de Groen, P. C. (2012). Automatic real-time detection of endoscopic procedures using temporal features, *Comput. Methods Programs Biomed.*, 108 (2) 524–535, 2012.

[3] Tang, S. J., Raju, G. (2015). Endoscopic photography and image documentation, *Gastrointest. Endosc.*, 82 (5) 925–931.

[4] Areia, M. (2014). One day of upper gastrointestinal endoscopy in a southern European country, *GE J. Port. Gastrenterologia*, 21 (3) 97–101.

[5] Oh, J., Hwang, S., Lee, J., Tavanapong, W., Wong, J., de Groen, P. C. (2007). Informative frame classification for endoscopy video, *Med. Image Anal.*, 11 (2) 110–127.

[6] Li, C., Ben Hamza, A., Bouguila, N., Wang, X., Ming, F., Xiao, G. (2002). Online redundant image elimination and its application to wireless capsule endoscopy, *Signal, Image Video Process.*, 8 (8) 1497–1506.

[7] Wang, Y., Tavanapong, W.,Wong, J., Oh, J. H., de Groen, P. C. (2015). Polyp-Alert: Near real-time feedback during colonoscopy, *Comput. Methods Programs Biomed.*, 120 (3) 164–179.

[8] Parot *et al.* V. (2013). Photometric stereo endoscopy, *J. Biomed. Opt.*, 18 (7) 76017.

[9] Manivannan, S., Wang, R., Trucco, E., Hood, A. (2013). Automatic Normal-Abnormal video frame Clasification for Colonospopy, *CVIP Comput. Vis. Image Process. Gr. Sch. Comput. , Univ. Dundee, UK Leeds Inst. Mol. Med.*, p. 644–647, 2013.

[10] Fortun, D., Bouthemy, P., Kervrann, C. (2015). Optical flow modeling and computation: A survey, *Comput. Vis. Image Underst.*, 134, p 1–21.

[11] Leszczuk, M. I., Duplaga, M. (2011). Algorithm for Video Summarization of Bronchoscopy Procedures, *Biomed. Eng. Online*, 10 (1) 110.

[12] Hwang, S., *et al.* (2005). Automatic measurement of quality metrics for colonoscopy videos, *In:* Proceedings 13th Annu. ACM Int. Conf. Multimed. - Multimed. '05, p. 912, 2005.

[13] Rosten, E., Drummond, T. (2006). Machine Learning for High Speed Corner Detection, *Comput. Vis. — ECCV 2006*, 1, p 430–443.

[14] Barron, J. L., Thacker, N. A. (2005).Tutorial: Computing 2D and 3D optical flow, *Imaging Sci. Biomed. Eng. Div. Med. Sch. Univ. Manchester*, 2004, p. 1–12.