Towards Real-Time Polyp Detection in Colonoscopy Videos: Adapting Still Frame-Based Methodologies for Video Sequences Analysis

Quentin Angermann[1], Jorge Bernal[2(✉)], Cristina Sánchez-Montes[3],
Maroua Hammami[1], Gloria Fernández-Esparrach[3], Xavier Dray[1,4],
Olivier Romain[1], F. Javier Sánchez[2], and Aymeric Histace[1]

[1] ETIS Lab, ENSEA, University of Cergy-Pontoise, CNRS, Cergy, France
{quentin.angermann,maroua.hammami,
olivier.romain,aymeric.histace}@ensea.fr
[2] Computer Vision Center, Universitat Autonoma de Barcelona, Barcelona, Spain
{jorge.bernal,javier.sanchez}@cvc.uab.cat
[3] Digestive Endoscopy Unit, Hospital Clinic de Barcelona, Barcelona, Spain
{crsanchez,mgfernan}@clinic.cat
[4] St. Antoine Hospital, APHP, Paris, France
xavier.dray@aphp.fr

Abstract. Colorectal cancer is the second cause of cancer death in United States: precursor lesions (polyps) detection is key for patient survival. Though colonoscopy is the gold standard screening tool, some polyps are still missed. Several computational systems have been proposed but none of them are used in the clinical room mainly due to computational constraints. Besides, most of them are built over still frame databases, decreasing their performance on video analysis due to the lack of output stability and not coping with associated variability on image quality and polyp appearance. We propose a strategy to adapt these methods to video analysis by adding a spatio-temporal stability module and studying a combination of features to capture polyp appearance variability. We validate our strategy, incorporated on a real-time detection method, on a public video database. Resulting method detects all polyps under real time constraints, increasing its performance due to our adaptation strategy.

Keywords: Polyp detection · Colonoscopy · Real time · Spatio temporal coherence

1 Introduction

Colorectal cancer (CRC) is the second leading cause of cancer death in United States, causing about 49,190 deaths during 2016 [1]. CRC's early diagnose is crucial for patient's survival, as precursor lesions (known as polyps) may degenerate into cancer over time. Several techniques have been proposed for lesion screening,

M.J. Cardoso et al. (Eds.): CARE/CLIP 2017, LNCS 10550, pp. 29–41, 2017.
DOI: 10.1007/978-3-319-67543-5_3

such as Wireless Capsule Endoscopy (WCE) or Virtual Colonoscopy (VC) but colonoscopy is still considered as the gold standard tool as it can detect lesions of any size (contrary to VC) and it allows lesion detection and removal during the same procedure (contrary to WCE). Nevertheless, colonoscopy has its own drawbacks being the most relevant of them polyp miss-rate, reported to be up to 22% for the case of small size or flat polyps [10].

Three types of approaches have been tackled to overcome these drawbacks: (1) improvement of endoscopic devices (magnification endoscopes [6]), (2) the development of new imaging technologies such as virtual chromoendoscopy [7,12] and (3) the proposal of computational support systems for colonoscopy aiming to support clinicians during/after the procedure.

Regarding computational systems, several efforts have already tackled automatic polyp detection in colonoscopy videos, ranging from classical hand-crafted shape-based methods [4] to pure machine learning approaches [2,8]. Recently, trending techniques such as deep convolutional networks have been also proposed [13,14] and a comparison between a large number of them was presented in [5] in the context of a global polyp detection challenge.

Despite the large number of approaches, none of them, to the best of our knowledge, are currently used in the exploration room due to: (1) not meeting real-time constraints, (2) not being tested on full length colonoscopy procedures and (3) being developed using still frame data (as fully public annotated video databases are not available). Regarding the latter, development over still frame data present the following problems associated to video analysis: absence of temporal coherence in method output and lack of adaption to higher variability in structures appearance (polyps and other elements) and image quality.

We present in this paper a methodology to adapt existing still-frame based polyp detection methods to video analysis. Our strategy consists of the addition of a spatio-temporal coherence module to stabilize methods output and the combination of different feature types to capture polyp appearance variability throughout a video. We integrate our strategy over an real-time polyp detection method [2]; the whole methodology is validated over a fully publicly annotated video database [3]. This validation is performed using a set of performance metrics chosen to fully represent method performance.

The structure of the rest of this paper is as follows. Section 2 introduces the adaptation strategy as well as the reference polyp detection method. In Sect. 3 we detail the experimental setup, results of which are shown in Sect. 4. We discuss in-depth the performance of the proposed methodology in Sect. 5. We finally the main conclusions of this study are drawn in Sect. 6.

2 Method

2.1 Reference Real-Time Still Frame-Based Polyp Detection Method

As explained in Sect. 1, we will use as reference method the one proposed in [2] which offers a good tradeoff between performance and associated processing time (0.039 ms, meeting real-time constrains over 25 fps videos). This active learning

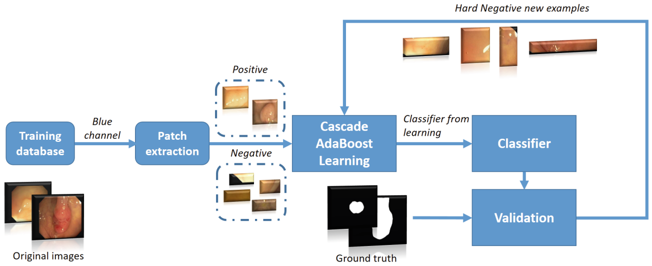


Fig. 1. Still-frame processing pipeline

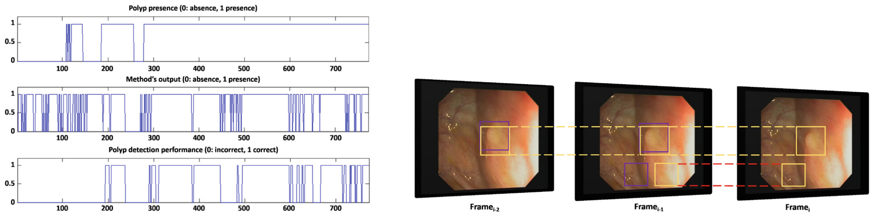
methodology consists of two different stages: (i) a Cascade AdaBoost learning step for the computation of a classifier, and (ii) a strengthening strategy based on active learning principle using Hard Negative examples [16]. Active learning is used to reinforce the classification performance by adding new negative examples produced by the initial classifier to the learning database.

This initial classifier is trained using six patches from each image of the training database (CVC-ClinicDB [4]): one positive patch covering completely the polyp and five negative ones without any polyp content. We use the Cascade Adaboost strategy (10 stages, with for each of them a targeted true positive rate of 99% and a false positive rate of 50%) to obtain this initial classifier, which is tested as a polyp detector function on each of the images of the complete dataset. As a result, the classifier provides a set of regions of interest (RoIs) where it predicts polyp presence. We compare prediction results over ground truth; all RoIs that do not contain a polyp are fed into the learning process as hard negative training patches so a new Cascade Adaboost classifier is created. An overview of the full processing training/learning scheme is shown Fig. 1. This process is repeated several times to obtain an optimal performance level. The interested reader can find a full description of the methodology at [2].

2.2 Combination of Feature Types

The use of texture-based descriptors (Local Binary Patterns) was proposed in [2] due the polyp appeared different enough from its surroundings due to the good selection of polyp shots from the corresponding videos. Unfortunately, in full video analysis, the number of false alarms grow due to variations in image quality and polyp appearance and due to the presence of other endoluminal scene elements which can deviate detectors' attention from the polyp.

The reference method allows an straightforward aggregation of other features to complement LBP though it is important to consider the potential impact in computational time of these new features. We propose to combine LBP with Haar features [11] because of the following two reasons: first, they can be fastly computed by using the usual "integral image strategy". Second, they can offer complementary information to LBP in a way such if LBP are more sensitive to the gradient information inside an image, Haar, by computing contrast/homogeneity parameter, can be related to geometrical local properties of a given RoI.



(a) (b)

Fig. 2. Spatio-temporal coherence module: (a) Example of spatio temporal instability in the output of still frame polyp detection methods when applied on full sequences. (b) Graphical explanation of the proposed solution: Green boxes represent the output in the current frame, blue boxes represents outputs in the previous frames. Green dashed lines connect similar RoIs in consecutive frames (kept in method output) whereas red ones represent unconnected RoIs between consecutive frames (removed in the output). (Color figure online)

2.3 Spatio-Temporal Coherence Module

One big drawback of the use of still-frame based methods for video analysis is that, by default, they do not consider information of previous frames to determine the output of the current ones. Due to this, a given method can show a performance like the one shown in Fig. 2(a) where we can observe that the method is not able to provide a stable output between consecutive frames.

To mitigate this, we propose the decision tree shown in Fig. 3. It is important to mention that to calculate the initial output for a given frame we first perform intra-frame block fusion to only provide as output candidates those RoIs where more individual outputs have been provided by the classifier. Once this is done, when calculating the final output for a given frame, the system considers also the RoIs provided by the classifier in the two previous frames in a way such if RoIs from the previous frame overlap with RoIs provided for the actual frame, these RoIs are kept to generate the final output. If it is not the case, those RoIs without spatio-temporal overlap are not included in the final output.

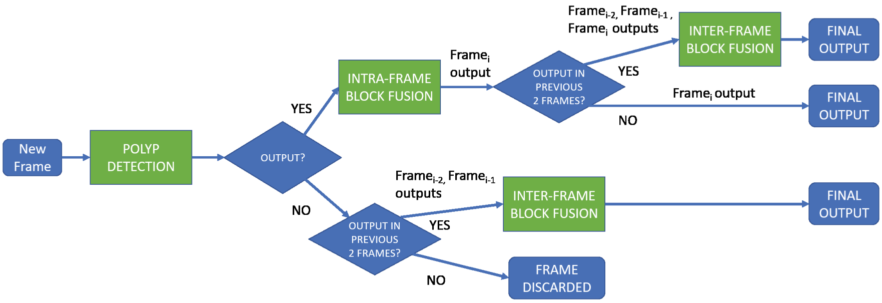


Fig. 3. Decision tree implemented to warrantee spatio-temporal coherence in method output

Table 1. Statistics of CVC-ClinicVideoDB database. PF stands for polyp frames, NPF for non-polyp frames and Paris represents morphology of the polyp according to Paris classification (0-Is for sessile polyps, 0-Ip for pedunculated polyps and 0-IIa for flat-elevated polyps).

Video	PF	NPF	Paris	Video	PF	NPF	Paris	Video	PF	NPF	Paris
1	386	112	0-Is	7	338	103	0-Is	13	620	4	0-Is
2	597	176	0-Is	8	405	44	0-IIa	14	2015	45	0-Is
3	819	153	0-Is	9	532	19	0-Ip	15	360	215	0-Is
4	350	40	0-Is	10	762	78	0-IIa	16	366	5	0-Is
5	412	78	0-Is	11	370	130	0-Is	17	651	146	0-IIa
6	522	335	0-Ip	12	261	124	0-IIa	18	259	122	0-Ip

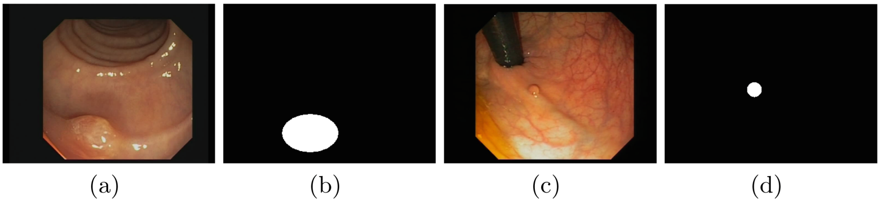


Fig. 4. Examples of (a, c) original image and (b, d) associated ground truth.

3 Experimental Setup

3.1 Validation Database

We validate our complete methodology over the first fully publicly available video annotated database (CVC-ClinicVideoDB) database, which comprises 18 different standard definition video sequences all showing a polyp. These sequences have been recorded using OLYMPUS QF190 endoscopes and Exera III videograbber. CVC-ClinicVideoDB contains 10924 frames of size 768 × 576, of which 9221 contain a polyp. Table 1 shows statistics of each of the videos of CVC-ClinicVideoDB, including Paris morphology [9] of the different polyps. Ground truth for each frame corresponds to a binary image in which white pixels correspond to polyp pixels in the image (images without polyps do not have any white pixels). CVC-ClinicVideoDB ground truth consists of an ellipse approximating polyp boundary. We show some examples of original images and their corresponding ground truth in Fig. 4.

3.2 Performance Metrics

Before defining the different metrics used to assess method performance, it is worth to mention that the output of the method for a particular frame consist of a series of bounding boxes representing the different RoIs provided by the classifier.

Following guidelines depicted in [5], we will use as first indication of correct detection (True Positive, TP) if the centroid of the RoI falls within the polyp mask. As in this first version of the database we provide ellipses for a weak labelling of the polyp, we have incorporated two additional criteria to determine a TP: (1) having pixel-wise precision within the RoI is higher than 50% (to cover the case of very big polyps against an small RoI) or (2) having a small distance to the centroid of the RoI to the border of the ground truth mask or a pixel-wise recall higher than 50% (to cover the case of small polyps enclosed within a large ground truth area). It is important to mention that we will only account one TP per polyp region in the image, no matter how many RoIs detect it. In case a polyp in a frame is not detected, we have a False Negative (FN) - we can have as many FNs as polyps in the image -. RoIs without overlap with a polyp region are accounted as False Positives (FP) - there can be more than one FP per image - and, finally, the absence of RoIs in a frame without a polyp is defined as a True Negative (TN).

From these definitions, we can calculate the following aggregation metrics: (1) Precision ($Prec = 100 * \frac{TP}{TP+FP}$), (2) Recall ($Rec = 100 * \frac{TP}{TP+FN}$) and (3) F1-score ($F1 = \frac{2*Prec*Rec}{Prec+Rec}$).

We also calculate the following metrics to account for clinical usability:

- Polyp Detection Rate (PDR) checks whether a method is able to detect the polyp at least once in a sequence, following guidelines depicted in [15].
- Mean Processing Time per frame (MPT). Considering videos are recorded at 25 fps, 40 milliseconds is the maximum time processing of a new frame can take to avoid delaying the intervention. MPT includes both frame processing time as well as displaying the results on the monitor.
- Mean Number of False Positives per frame (MNFP).
- Reaction Time (RT) represents the delay (in frames and seconds considering a frame rate of 25fps) between first appearance of the polyp in the sequence and the first correct detection provided by the method [4].

4 Results

4.1 Quantitative Results

We present quantitative results in Table 2, broken down by the different aspects we wanted to test in the study (impact of adaptation strategy, computational efficiency). Before introducing a breakdown of the results, it is important to mention that, as the methodology over which we have incorporated our adaptation strategy incorporates strengthening stages, we will distinguish each strengthening iteration with a cardinal index starting by 0 in a way such classifier Ni will refer to a classifier computed with i strengthening steps.

The first important result to be extracted from Table 2 is that the methodology is able to detect all different polyps in the different sequences at least in one frame, using the same definition proposed in [15]. The basic configuration of the system, as presented in [2], achieves the smallest reaction time.

Table 2. Overall performance results.

Method	PDR	MPT	MNFP	Prec	Rec	F1	RT
Impact of the type of feature descriptor used							
LBPN0	100%	140ms	3.5	12.42%	54.65%	20.24%	7.2 [0.3 sec]
HaarN0	100%	24ms	1.4	23.29%	46.82%	31.10%	17.5 [0.7 sec]
Impact of spatio-temporal coherence (STC)							
LBPN0 noSTC	100%	140ms	3.5	12.42%	54.65%	20.24%	7.2 [0.3 sec]
LBPN0	100%	140ms	1.9	16.25%	41.25%	23.31%	35.0 [1.4 sec]
HaarN0 noSTC	100%	24ms	1.4	23.29%	46.82%	31.10%	17.5 [0.7 sec]
HaarN0	100%	36ms	0.9	27.02%	39.61%	32.12%	38.3 [1.5 sec]
Impact of network strengthening							
LBPN0	100%	140ms	1.9	16.25%	41.25%	23.31%	35.0 [1.4 sec]
LBPN1	100%	160ms	1.1	27.11%	46.02%	34.12%	43.7 [1.7 sec]
LBPN2	100%	162ms	0.7	29.88%	34.96%	32.22%	45.9 [1.8 sec]
HaarN0	100%	36ms	0.9	27.02%	39.61%	32.12%	38.3 [1.5 sec]
HaarN1	100%	21ms	0.6	39.14%	42.56%	40.78%	27.3 [1.1 sec]
Impact of feature aggregation)							
LBPN2	100%	162ms	0.7	29.88%	34.96%	32.22%	45.9 [1.8 sec]
HaarN1	100%	21ms	0.6	39.14%	42.56%	40.78%	27.3 [1.1 sec]
Aggregation	100%	185ms	1.1	30.39%	52.40%	38.47%	15.0 [0.6 sec]

With respect of *the type of features used*, we can observe a positive difference associated to the use of Haar features which leads to a great reduction in the number of false positives while keeping real-time constraints and a similar recall (higher F1-score). LBP offers a slower processing time (140 ms per image) and an excessive number of false alarms (around 3.5 FP per image), which makes its use not compatible with a clinical use.

We broke down the results according to polyp morphology, under the assumption that Haar features should perform better for those types in which the contour can be clearly observed. We present results of this side experiment in Table 3. On the one hand, we can observe how LBP achieves a higher F1-score for flat polyps (higher recall for a similar precision); we associate Haar's worse performance to the lack of strong contours. In this case, LBP takes advantage of the difference in pattern between polyp and mucosa. On the other hand, for peduncular polyps in which their contours are clearly recognizable, we can observe a clearly superior performance of Haar in all performance metrics, especially with respect to RT (difference of more than 3.5 s with respect to LBP).

The use of our *spatio-temporal coherence* module results on an improvement in the overall performance for both descriptors, decreasing in a significant way the average number of FPs per image (lower than one for Haar descriptor). We can also notice that, for both descriptors, the average detection latency is now more than a second. We associate this to false positives damaging posterior good detections. Only Haar presents a MCT compatible with real time constraints

Table 3. Impact of Paris morphology on overall performance results. $N1$ classifiers are used for both for LBP and Haar features, as well as spatio-temporal coherence.

Method	MNFP	Prec	Rec	F1	RT
0-Is (sessile,11 polyps)					
LBPN1	1.3	23.93%	40.84%	30.18%	40.6 [1.6 sec]
HaarN1	0.6	38.01%	41.32%	39.59%	22.3 [0.9 sec]
Aggregation	1.2	27.93%	48.18%	35.36%	21 [0.8 sec]
0-Ip (peduncular,3 polyps)					
LBPN1	1.0	31.4%	51.10%	38.90%	89.0 [3.6 sec]
HaarN1	0.5	50.46%	57.50%	53.75%	4.0 [0.1 sec]
Aggregation	1.1	40.28%	64.73%	49.66%	4.0 [0.1 sec]
0-IIa (flat, 4 polyps)					
LBPN1	1.1	34.62%	59.35%	43.73%	18.0 [0.7 sec]
HaarN1	0.6	35.14%	37.08%	36.08%	58.5 [2.3 sec]
Aggregation	0.8	32.32%	58.10%	41.54%	7.0 [0.3 sec]

(36 ms). Considering the overall positive impact of spatio temporal coherence, in the following it will be applied for all experiments.

Though clearly more specific to the reference methodology used, Table 2 shows the benefit of the strengthening strategy for both descriptors. The overall performance is improved though, for the case of LBP, the mean computation time remains incompatible with a clinical use, and the detection latency is not far from 2 s for $LBPN2$ classifier. Haar descriptors definitely appear here more compatible with a daily routine use since for $HaarN1$ the mean latency is only of 1.1 s but with 14 videos (on the overall 18) presenting with an average RT lower than 0.4 s; the mean computation time is only of 21 ms with a max value of only 25 ms for video 14 and, finally, the overall performance levels obtained are the best from all the experiments presented in this paper in terms of trade-off between true and false alarms.

One of the reasons of studying the use of different type of features was to observe whether the *combination of several feature types* could lead to an overall performance improvement. Our experiments yield an interesting result: the combination of LBP and Haar classifiers leads to a significant increase of the TP detection rate since the Recall reaches its highest value considering the all set of experiments achieved in this section. Results indicate that LBP and Haar can detect different kind of polyp (RoIs) in a complementary way. Nevertheless, from a clinical applicability perspective, even if the mean RT is only of 0.7 s when combining both classifiers, as expected, the mean processing time per frame is constrained by LBP classifier performance which is of an average value of more than 185 ms. Finally, we can observe in Table 3 how the combination of feature descriptors help to improve recall scores and to reduce computation time regardless polyp morphology.

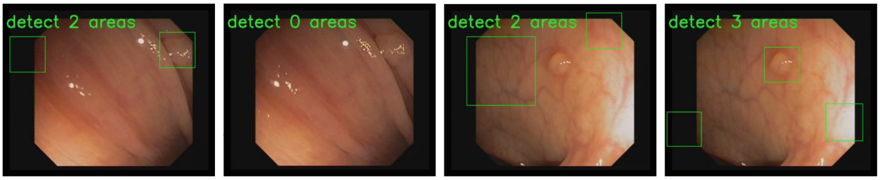
The first conclusion that we can extract from the analysis of the results presented is that our proposed adaptation strategy does improve the performance of still-frame based methods when dealing with video analysis. The use of spatio-temporal coherence leads to a reduction in the number of false positive alarms whereas the combination of different types of features lead to an increase in the number of polyp frames correctly detected. It is important to mention that some of these improvements come at the cost of losing real-time capabilities and efforts should be made to improve the computational cost of some of the proposed improvements (such as the combination of LBP and Haar features).

5 Discussion

5.1 Impact of Adaptation Strategy on Method's Performance

With respect of the specific feature descriptor used, we observed better performance related to the use of Haar features. We associate this to the fact that in video sequences, differences in texture between mucosa and the polyp become less relevant and, in this case, the presence of strong boundaries delimiting the different structures such as polyps in the image may appear more useful than texture analysis. Nevertheless, it has to be taken into account LBP's offers best performance for the case of flat polyps, which are those recurrently mentioned by clinicians as one of the main causes of polyp miss-rate. If the decision on the descriptor to use depends on real time constraints, Haar is the way to go but, as Fig. 5(a–d) shows, the combination of both descriptors might increase overall performance.

With respect to spatio-temporal coherence module, its inclusion has led to a reduction in the number of false alarms but it has also lead to a decrease in performance scores on Recall or RT. We associate this decrease to isolated correct detections not kept through consecutive frames therefore leading to miss the polyp in the whole subsequence of frames. In this case efforts should be made to clearly identify the polyp target to be tracked in order to only mitigate false alarms and not those correct ones. Consequently, efforts should be put on



(a) (b) (c) (d)

Fig. 5. Differences in performance associated to the specific feature descriptor used: (a, c) show the output of Haar descriptor whereas (b, d) show the output achieved using LBP as descriptor.

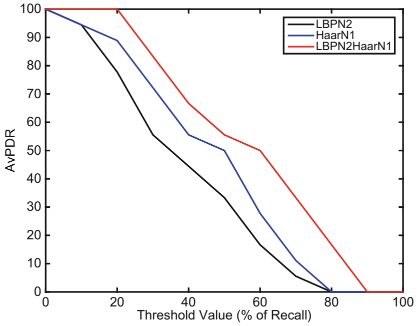


Fig. 6. Evolution of the AvPDR metric with respect to the threshold value applied to the Recall on each video.

identifying and tackling appropriately the source of those false alarms which might involve, as some authors propose [4], considering the impact of other elements of the endoluminal scene.

5.2 Frame-Based Analysis vs. Clinical Applicability

We have presented in Sect. 3 two sets of metrics to represent the performance of a given method. It is clear that clinicians will be mainly interested on whether the computational system is able to detect the polyp once it appears in the image. As the polyp is detected, their attention will deviate to other areas in the image. Considering this, a good performing method could be one that only detects the polyp in one frame being this frame the first in which the polyp appears in the sequence. As this kind of system does not warrantee good performance under exploratory conditions, frame-based and clinical applicability metrics should be combined to represent actual method performance.

To solve this, we propose to combine the most clinically relevant metric (PDR) with Recall into a new metric representing both whether the method is able to detect the polyp and that this detection occurs in a relevant number of frames. We define the Average Polyp Detection Rate (AvPDR) to checks whether a method is able to detect the polyp in a set of sequences with respect to a minimum value for Recall. We calculate AvPDR in the following way: for each video we set individual (IndPDR) score to 1 if Recall score for the particular video surpasses Rec_{thres} value. Final AvPDR score for the whole dataset will be calculated as the mean of individual InDPRs. To illustrate this, Fig. 6 shows the evolution of the AvPDR for different values of the Recall and for the three last computed classifiers LBPN2, HaarN1 and aggregation of both.

As it can be seen, the AvPDR brings very interesting insights on the capacity of a given method to detect the polyp with a given minimum Recall. In our case, the aggregation of LBPN2 and HaarN1 classifiers makes possible to systematically detect the polyp in all videos with a minimum Recall of 20%.

5.3 Analysis of Methods' Performance in the Context of the State-of-the-Art

As mentioned in Sect. 1, there are many available polyp detection methods in the literature, some of them already showing quite good performance as it can be observed in [5]. The main objective of our work was not to develop the best polyp detection method but to show how still frame-based methodologies databases could still be valid for full sequences analysis.

Due to the lack of publicly available annotated video databases, we can only compare global performance scores of different methods even if they have not been tested under the same conditions. In this sense, our approach obtains similar performances in PDR and Reaction time than those achieved by the best methods presented in [5]. As mentioned before, we are not worried here about frame-based performance (though it has to be improved for sure) but on whether the system can be of actual clinical use hence the focus on real time performance. We also believe that, once public video databases become more available, methods performance (especially machine learning ones) will benefit from being trained on them as they will cover a wide variety of polyp appearances.

Finally and to assess actual clinical applicability of a given method on the exploration room, we believe efforts should also be made on incorporating full realistic interventions as part of the databases in a way such once the polyp is found the clinician progresses through the colon without the need of observing the polyp in different views, typical from still frame database creation protocols,

6 Conclusions

We have presented in this paper a study on how to adapt still frame based polyp detection methodologies to full sequences analysis. Our adaptation strategy involves the addition of a spatio-temporal coherence module and the combination of feature descriptors. We have tested the impact (in performance and computational efficiency) of this adaptation strategy implementing them over an already existing real time polyp detection method trained on still frame based databases. We validate the complete methodology over a newly published video database of 18 sequences using a set of clinical and technical performance metrics.

The main conclusion extracted from this study is that the addition of a spatio-temporal coherence module and the combination of feature descriptors lead to an overall improvement on method performance over full sequences; once these modules are applied over the reference method, the proposed methodology is able to detect all different polyps in at least one frame in the sequence.

It has to be noted that the best performing configuration is not ready for clinical use due to not meeting real time constraints; efforts should be made to increase the computational efficiency of the different modules proposed. Apart from this, we also foresee the following areas of improvement: (i) add an image preprocessing stage to mitigate the impact of other elements of the endoluminal scene (which can impact when the first correct detection occurs), (ii) incorporate computationally efficient camera motion tracking methods to improve spatio-temporal coherence and (iii) study the possibility of incorporating additional feature descriptors to improve overall performance. Moreover, our method should be trained over video sequences in order to capture better the great variability of polyp appearance within a same sequence.

Acknowledgements. The authors acknowledge the support of the following agencies for research funding: Spanish government through funded project iVENDIS (DPI2015-65286-R), SATT IdFInnov (France) through the Project Smart Videocolonoscopy under Grant 186, Catalan government through SGR projects 2014-SGR-1470 and 2014-SGR-135, CERCA Programme/Generalitat de Catalunya and FSEED.

References

1. ACS2016: Key statistics for colorectal cancer. online (2016)
2. Angermann, Q., Histace, A., Romain, O.: Active learning for real time detection of polyps in videocolonoscopy. Procedia Comput. Sci. **90**, 182–187 (2016)
3. Bernal, J., Histace, A.: Gastrointestinal Image Analysis (GIANA) sub-challenge. online (2017)
4. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., et al.: Wm-dova maps for accurate polyp highlighting in colonoscopy: validation vs. saliency maps from physicians. Comput. Med. Imaging Graph. **43**, 99–111 (2015)
5. Bernal, J., Tajbakhsh, N., et al.: Comparative validation of polyp detection methods in video colonoscopy: Results from the MICCAI 2015 endoscopic vision challenge. IEEE Trans. Med. Imaging **36**(6), 1231–1249 (2017). doi:10.1109/TMI.2017.2664042
6. Bruno, M.: Magnification endoscopy, high resolution endoscopy, and chromoscopy; towards a better optical diagnosis. Gut **52**(suppl. 4), iv7–iv11 (2003)
7. Coriat, R., Chryssostalis, A., Zeitoun, J., et al.: Computed virtual chromoendoscopy system (FICE): a new tool for upper endoscopy? Gastroentérologie clinique et biologique **32**(4), 363–369 (2008)
8. Gross, S., Stehle, T., Behrens, A., et al.: A comparison of blood vessel features and local binary patterns for colorectal polyp classification. In: SPIE Medical Imaging, p. 72,602Q. International Society for Optics and Photonics (2009)
9. Inoue, H., Kashida, H., Kudo, et al.: The paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to december 1, 2002. Gastrointest Endosc **58**(6 Suppl.), S3–S43 (2003)
10. Leufkens, A., van Oijen, M., Vleggaar, F., Siersema, P.: Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. Endoscopy **44**(05), 470–475 (2012)
11. Lienhart, R., Maydt, J.: An extended set of haar-like features for rapid object detection. In: Proceedings of the 2002 International Conference on Image Processing, vol. 1, p. I-900. IEEE (2002)
12. Machida, H., Sano, Y., Hamamoto, Y., et al.: Narrow-band imaging in the diagnosis of colorectal mucosal lesions: a pilot study. Endoscopy **36**(12), 1094–1098 (2004)
13. Park, S.Y., Sargent, D.: Colonoscopic polyp detection using convolutional neural networks. In: SPIE Medical Imaging, pp. 978, 528–978, 528. International Society for Optics and Photonics (2016)

14. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 79–83. IEEE (2015)
15. Wang, Y., Tavanapong, W., Wong, J., et al.: Polyp-alert: Near real-time feedback during colonoscopy. Comput. Methods Programs Biomed. **120**(3), 164–179 (2015)
16. Wang, Z., Song, Y., Zhang, C.: Efficient active learning with boosting. In: Proceedings of the SIAM International Conference on Data Mining Society for Industrial and Applied Mathematics, pp. 1232–1243. Society for Industrial and Applied Mathematics (2009)