

Towards Automatic Polyp Detection with a Polyp Appearance Model

J. Bernal^{a,*}, J. Sánchez^a, F. Vilariño^a

^aComputer Vision Center and Computer Science Department, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

Abstract

This work aims at the automatic polyp detection by using a model of polyp appearance in the context of the analysis of colonoscopy videos. Our method consists of three stages: region segmentation, region description and region classification. The performance of our region segmentation method guarantees that if a polyp is present in the image, it will be exclusively and totally contained in a single region. The output of the algorithm also defines which regions can be considered as non-informative. We define as our region descriptor the novel Sector Accumulation-Depth of Valleys Accumulation (SA-DOVA), which provides a necessary but not sufficient condition for the polyp presence. Finally, we classify our segmented regions according to the maximal values of the SA-DOVA descriptor. Our preliminary classification results are promising, especially when classifying those parts of the image that do not contain a polyp inside.

Keywords: Colonoscopy, Polyp Detection, Region Segmentation, SA-DOVA descriptor

1. Introduction

Colon cancer's survival rate depends on the stage in which it is detected, decreasing from rates higher than 95% in the first stages to rates lower than 35% in stages IV and V [1]; hence the importance of detecting it on its early stages by using screening techniques, such as colonoscopy [2].

Although colonoscopy is considered nowadays as the gold standard for colon screening there are still open challenges to overcome, such as the reduction of the miss-rate [3]. During the last decades there is a trend that consists of developing intelligent systems for medical applications. Intelligent systems are currently being used to assist in other medical interventions. For instance, there are systems that can interpret medical data automatically, such as KARDIO [4], which was developed to interpret electrocardiograms. It is possible to find many examples of intelligent systems built to assist in cancer detection. The interested reader can consult some works in the field of breast cancer detection [5] or prostate cancer detection [6].

Our objective is to add significant value to the colonoscopy procedure by using methods based on computer vision or artificial intelligence. In the case of colonoscopy, there is a number of possible areas where an intelligent system can potentially help [7]. It is possible to think of an intelligent system that can assist in the diagnosis procedure, by highlighting parts of the colon that are likely to contain lesions or polyps as the physician progresses the instrumental through the patient. There is also a potential in the use of the information extracted from the analysis of a colonoscopy video in order to build up systems that can provide an objective assessment of the physician's skills.

By doing so, training programs could be developed without the cost that a real intervention has. Another possible area of application could be to provide a whole description of what appears on the scene for automatic reporting. Finally, another possible domain of application could be the extension of the information that the colonoscopy intervention provides, by leading to a development of patient-specific models.

In our case, in the context of developing intelligent systems for colonoscopy, the main objective is to define a robust model of polyp appearance. This model can be potentially used to indicate which regions in the image are more likely to contain a polyp inside, which can be useful for several of the applications mentioned before. Particularly, our whole processing scheme is built on the fact that intensity valleys appear to surround polyps as the light of the colonoscope and the camera are in the same direction. For this reason, we propose to valley and ridge information as cue to detect polyps. In our case, our method will not try to fit a certain model into the images but it will look for several appearance cues which will guide it.

In this paper we present our work on polyp detection, which extends the works on the depth of valleys image [8]. Our detection method consists of three stages: region segmentation, region description and region classification.

We present our first contribution, the region segmentation stage, in which an input colonoscopy image is segmented into a minimum number of informative regions, one of these regions containing the polyp in a complete way. Since all the non-informative regions are rejected, the size of the problem is reduced largely. The concept of informative and non-informative regions are used here in the context of assuring that no polyp is inside the given region and, therefore, there will be no need for further processing [9]. These results can be used later to classify the informative regions into polyp- vs. non-polyp-containing candidates.

*Corresponding author

Email address: jbernal@cvc.uab.es (J. Bernal)

URL: <http://www.cvc.uab.es/> (J. Bernal)

Our second contribution consists of the introduction of the Sector Accumulation-Depth of Valleys Accumulation (SA-DOVA), which aims to find which points on the image are interior to objects, which are meant to be delimited by points with high value in the depth of valleys image. Finally, we classify the segmented regions into polyp-containing vs. the opposite, according to their maximum values of the DOVA descriptor.

The structure of the paper is as follows: in Section 2 we introduce previous approaches on polyp detection in colonoscopy videos. We present in Section 3 the theoretical model on which we base our polyp detection method, which is presented in Section 4. In Section 5 we show our experimental setup along with polyp detection results. In Section 6 we discuss in depth the performance of each stage. Finally we finish this paper in Section 7 with the main conclusions extracted from our approach and our proposals for future work.

2. Related work

The main objective of the colonoscopy procedures is to check the status of the colon of the patient, with the aim to find possible lesions and cancer polyps on it. The general appearance of the polyps has been covered widely by medical bibliographic sources [10]. However, there is a great variability in polyp appearance in colonoscopy videos, since there are some challenges that hinder polyp detection, namely: 1) the non-uniform appearance of polyps (see Figure 1 (a-d)); 2) their shape, flat or peduncular (Figure 1 (a) and (d)); 3) the effects of image acquisition, such as changes in pose, blurring, occlusions, specular highlights (Figure 1 (e-g)), and 4) the high similarity between the tissues inside and outside the polyp, which disables the possibility of relying only on texture or color cues (Figure 1 (h)), just to mention a few.

In the case of polyp detection, the great majority of the research approaches that have been carried out is based on the analysis of features detected in the image. In the context of image processing, features can be defined as singular visual traits, associated to the visual primitives that constitute an object, such as edges, corners or lines, among others. The usual procedure is to use *feature detection* methods to locate the potential ROIs of the image and then describe them using one or many *feature descriptors*. Feature descriptors are commonly divided into [11]: shape, texture, color and motion. In the case of polyp detection, the bibliography can be divided into two categories: polyp detection by means of shape descriptors and polyp detection by means of texture (and color) descriptors.

- *Shape descriptors*: In general, polyps present two different shapes: flat (Figure 1 (a)) or peduncular (Figure 1 (d)), but the way polyps appear in images makes difficult a simple polyp detection by means of finding their general shape directly. Since pose, size and appearance of polyps vary largely, many of the approaches try to find polyps by means of detecting parts of the image that may indicate their presence. Shape-based approaches can also be divided into two groups: detection by ellipse-fitting and

detection by means of curvature. In the former, the difficulty relies on fitting ellipses in partially complete contours [12]. In the latter, the use of curvature-based measures is used to aid polyp detection, as can be seen in [13], where several measures are defined to analyze the local shapes in the colon walls. However the authors point out one of the big problems when using contour-based information, the incompleteness of the boundaries detected, which has an impact in the performance. Other approaches [14] use curvature to measure the protrudedness of a candidate object by means of second principal curvature flow.

- *Texture and color descriptors*: One option referred in the field of texture descriptors for polyp detection is the use of wavelets, either using features extracted from both approximating and detail coefficients [15] or by combining wavelets with other types of feature descriptors, such as local binary patterns or co-occurrence matrices [16], [17]. The characterization of the local texture information for a given pixel and its neighborhood has been shown to provide a different solution in this field, since statistics extracted from all texture units over the image can reveal global texture aspects [18].
- *MPEG-7 descriptors*: MPEG-7 shape and texture descriptors have also been used for polyp detection [19]. A region-based shape descriptor using moments is presented. Related to texture descriptors, *homogeneous texture* and *local edge histogram* are introduced and, in the sub-field of color descriptors, methods such as *dominant color*, *scalable color* or *color structure* are presented (see [11] for a further explanation of them).

In Table 1 we can see a summary of the main characteristics of recent approaches for polyp detection, some of which have been introduced above. Unfortunately, it is very difficult to make a comparison between them, due to the fact that there is no common local database where to test the different methods. Therefore, we could only compare different methods in terms of certain measures, such as precision or recall, always taking into account that these measures will have been calculated for a specific dataset. Our approach could be enclosed in the subgroup of Shape descriptors, although in this case we aim at a more general solution, based on a theoretical model of a polyp that is valid for both flat and peduncular polyps.

3. A theoretical model for polyp appearance

3.1. Illumination

In order to define model of appearance for polyps in colonoscopy videos, we need both an a priori model about the polyp and a model of the illumination. For the sake of simplicity let us consider a polyp as a semi-spherical shape protruding from the colon wall plane. We will also consider that the polyp surface is regular and that its reflectance can be approximated by the Phong's illumination model [22]. The colonoscope itself is modeled by a pinhole camera and a punctual illumination source placed in the same position. Figure 2 (a) shows a

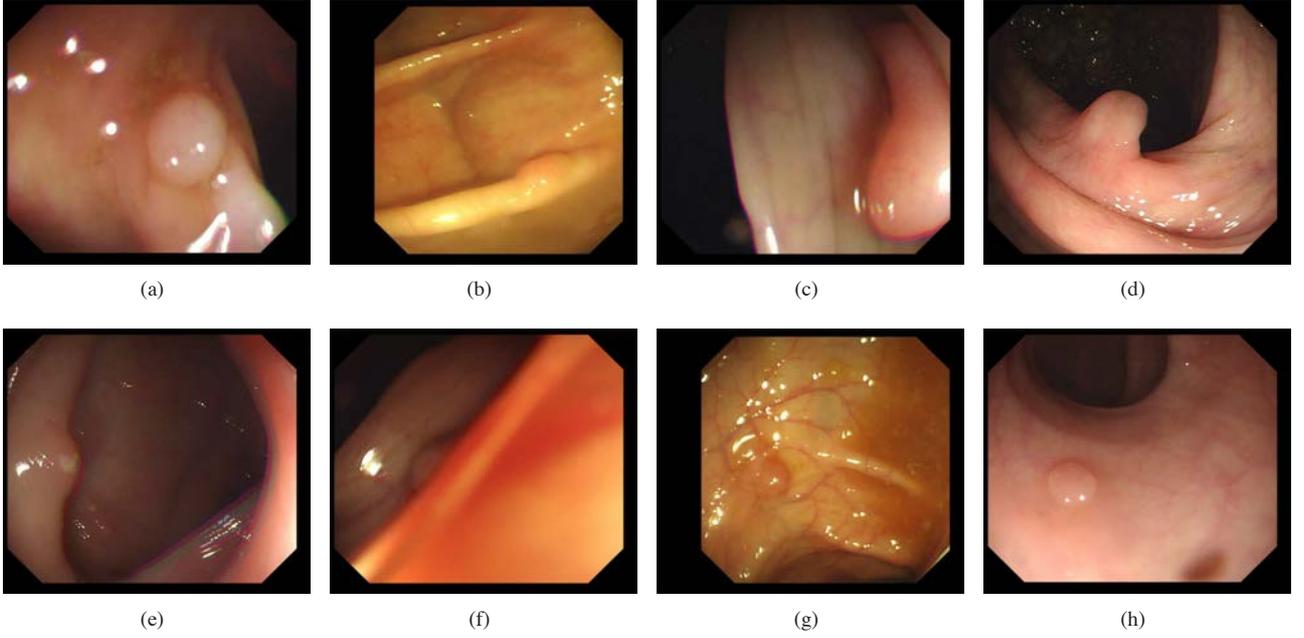


Figure 1: Challenges in polyp detection: (a-d) non uniform appearance; (e) partial (lateral) views; (f) blurred images; (g) specular highlights, and (h) uniform texture and color inside and outside the polyp.

Author	Group	Method	Classification	Datasets	Bib
Krishnan et al.	Shape	Edge detection to extract contours. Curvature	-	2 normal and 4 abnormal images	[20]
Hwang et al.	Shape	Ellipse fitting. Curvature	-	27 polyp shots	[12]
van Wijk et al.	Shape	Amount of protrudness. Curvature	5 measures (MaxIntChange, LongAxis, ShortAxis, MinHU, MaxHU), distances	84 studies, 168 scans, 108 polyp shots	[14]
Dhandra et al.	Shape	Segmentation of color images followed by watersheds	-	50 normal and 50 abnormal images	[21]
Zhu et al.	Shape	Curvature-based shape measures	By segmentation results	1 phantom image and 1 colon patient	[13]
Coimbra et al.	Shape and Texture	MPEG-7 descriptors: (Shape) Region-based shape descriptor (texture) homogeneous texture, local edge histogram (color) dominant color, scalable color, color structure.	Mean of descriptor values for each event group	Blood (832 images from 17 events), ulcers (400 images from 21 events), polyps (899 images from 31 events), and normal (55000 images from an entire uneventful exam) images	[19]
Karkanis et al.	Texture	Wavelets applied in different color spaces (CWC features)	LDA	5 different videos	[17]
Li et al.	Texture	Division of image in patches. Mean and standard deviation of DWT coefficients	Support Vector Classifiers	46 colonoscopic images with multiple categories of abnormal regions and 12 normal	[15]
Tjoa et al.	Texture	Measures extracted from texture spectra in chromatic and achromatic domain. Texture units	Helped by PCA, Nearest Neighbor	12 normal and 54 abnormal images	[18]
Ameling et al.	Texture	Local binary patterns (also in opponent color space) and grey-level co-occurrence matrices.	SVM	4 videos	[16]

Table 1: Summary of texture descriptor-based methods for polyp detection.

schematic representation of this scenario. Under such assumptions, the image provided following these approaches is calculated as:

$$I = I_a * K_a + f_{att} * I_p * [K_d * \cos \theta + W(\theta) * \cos^n \alpha], \quad (1)$$

where I is the light reflected by the surface towards the camera, I_a is the ambient intensity, K_a is the ambient reflection constant, f_{att} is the attenuation factor, I_p is the punctual intensity, K_d is the diffuse reflection coefficient, θ the angle between the surface normal and the illumination source, $W(\theta)$ is the fraction

of light reflected in a specular way, α is the angle between the surface normal and the camera, and n modulates the decay in the specular reflection. This model is implemented with a set of arbitrary values for the sake of visualization in Figure 2 (b) and (c), which show a rendering of a synthetic polyp for tilt angles of $\alpha = 0^\circ$ and $\alpha = 60^\circ$. In this scenario, sharp edges, gradual shading and specular reflections are created (these same elements can be visualized in the real example of Figure 2 (d)).

3.2. Polyp characterization

The characterization of the polyp is obtained through the shadings, related to valleys in the intensity image. Hence, the

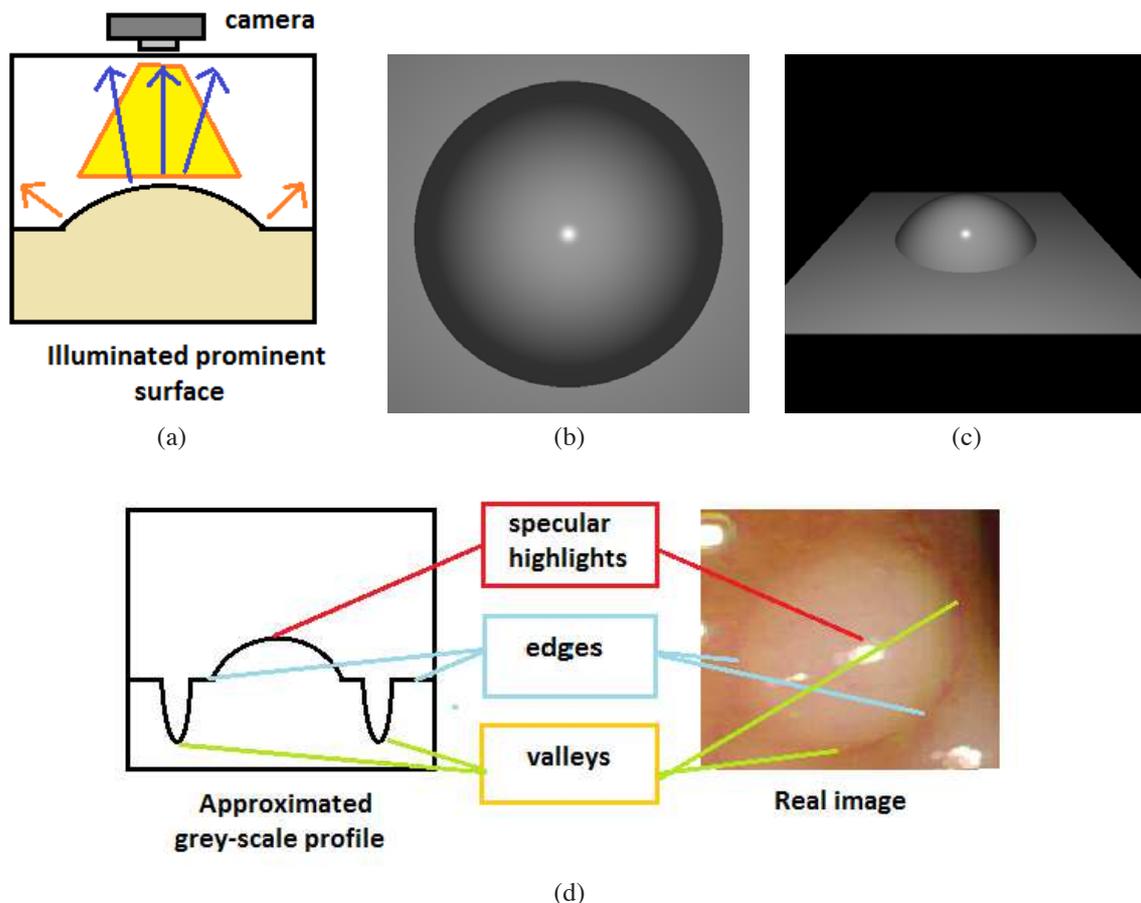


Figure 2: Model of appearance and illumination of polyps: (a) Scheme of an illuminated prominent surface; (b) and (c) synthetic model rendering for 0 and 60 degrees. (d) a real example.

detection of polyps must be linked to the identification of valleys. However, due to the different points of view of the colonoscope, and the potential presence of occlusions, it is not guaranteed that the polyp is fully surrounded by strong valleys in all directions. Our proposal deals with these situations by integrating the valley information as an accumulation operator in the following way:

$$Acc(x) = \int_{\alpha=0}^{\alpha=2\pi} \max_{r \in [R_{min}, R_{max}]} V(x + r * (\cos(\alpha), \sin(\alpha))) d\alpha, \quad (2)$$

where V is the valley image, x represents the coordinates of pixel in the image, R_{min} and R_{max} define the area in which the valleys are searched, and α allows the angular integration of the maximum values in all directions. Acc will be large when the structure presents strong valleys in all directions. However, the presence of multiple valleys in the same angle will not affect the value of Acc . Because of that, weak valleys coming from thin vessels or wrinkles will not be computed at a given angular orientation, and only the strong valleys associated to the polyp would tend to provide high outputs. It is important to notice that the ring structure defined by R_{min} and R_{max} provides invariance to rotation. Defined in this way, this approach will perform well

in circular or elliptical patterns, but the Acc operator will not be linked to any particular shape, as long as a substantial amount of (strong) valleys reside under the ring. This property makes Acc is robust to occlusions and potential lack of information for a number of directions. Acc can be digitally implemented in an efficient way as the sum of the valley maxima that are found in the sectors by following the SA-DOVA approach described in Section 4.2.

3.3. Depth of valleys image

As mentioned above, we base our whole processing scheme on the evidence of valley information surrounding polyps. The idea behind the *depth of valleys image* is to complement the information that a valley detector provides with the morphological gradient in order to achieve a method that enhances both valley and contour information. The rationale of this approach is that in certain type of views -in general, in lateral views- we do not have a whole valley surrounding the polyp, but still non-connected edges are available in these cases.

Ridges and valleys in n -dimensional images are commonly identified as loci of minimum gradient magnitude along the relief's level curves [23]. If $|\lambda_1| \geq \dots \geq |\lambda_d|$ are the eigenvalues of $\nabla \nabla L$ and v_1, \dots, v_d their corresponding eigenvectors, then a nD

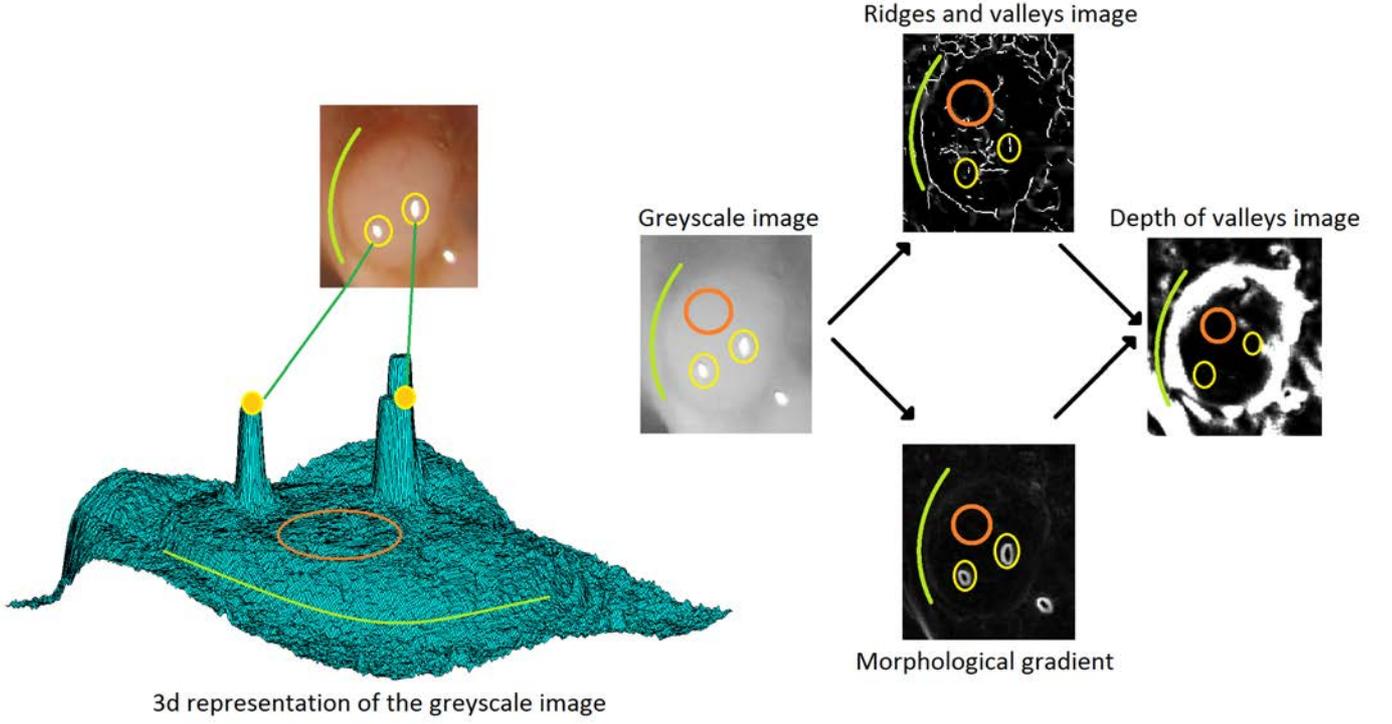


Figure 3: Example of the calculation of a depth of valleys image.

crease ($1 \leq n \leq d$) is characterized as:

$$\forall i \in \mathbf{I}_{d-n} \quad \nabla L \cdot v_i = 0, \quad (3)$$

where L is the scale-space representation of the image intensity, obtained by convolving the image with a gaussian function and I is the image. Considering this, we can state that if $\lambda_i < 0$ we have a ridge, and if $\lambda_i > 0$, a valley.

In 2D, ridges/valleys can be also identified as positive maxima/negative minima of the curvature of the relief's level curves. Maxima are connected from one level to the next therefore constituting a subset of the vertex curves. As shown in Figure 3, the output of a ridges and valley detector applied in one of our target images informs us about where valleys are in the image but not about their intensity. The results of these detectors are also affected by the presence of reflections or artifacts in the image.

In order to obtain information about the intensity of the valleys the morphological gradient can be used. The morphological gradient is defined as the difference between the dilation and the erosion of a given image [24] and it gives as output an image where each pixel value indicates the contrast in intensity in the close neighborhood of that pixel.

As shown in Figure 3, the depth of valleys image (normalized in grey level for a proper visualization) is achieved through the pixel-wise multiplication of the ridges/valleys detector and morphological gradient. In the points where we have a valley (marked as a green line) and the morphological gradient value is high the result is high value of the depth of valleys image.

Conversely, in the points where there exist a valley but the morphological gradient is low (or vice versa) the depth of valley will not result in a maximal value. The orange line surrounds an area with no local features, and the yellow circles surround the frontal reflections with a typical saturation pattern. The mathematical representation for this definition of the depth of valleys image is as follows:

$$\forall i, j \in \mathbf{I} \quad DV(i, j) = V(i, j) \cdot MG(i, j), \quad (4)$$

where DV stands for the depth of valleys image, V for the output of the ridges and valleys detector and MG for the morphological gradient image, both normalized to unit. Formulated in this way, the depth of valleys image has higher values in the points that constitute the relevant valleys of the image and lower to points inside the object surrounded by the valley.

It must be noticed that in order to calculate the morphological gradient, the ridge and valley extractor needs two parameters to be set in correspondence to the size of the structural element (sd) [23]. These parameters are the differentiation scale σ_d and the integration scale σ_i . In our case, the structural element is a disk. More precisely, σ_i should span the same size as sd in order to work in the same scale. If this does not happen, maximal points of the ridge and valley extractor could be located in places where the morphological gradient is not maximal, and therefore the desirable properties of the resulting depth of valleys image would be lost.

4. Methodology

4.1. Region segmentation

The general scheme of the segmentation algorithm consists of 4 different stages which will be described next (see Figure 4 for a graphical scheme).

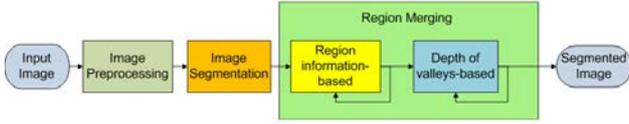


Figure 4: Scheme of the Segmentation Algorithm.

1. **Image Preprocessing:** Before applying any segmentation algorithm there are some operations that should be done: 1) converting the image into grey-scale; 2) de-interleaving (as our images come from a high definition interleaved video source); 3) correction of the specular highlights, and 4) inverting the grey-scale image.

To correct the specular highlights we do the following: 1) Detect the location of the specular highlights by using an intensity threshold (Figure 5 (b)); 2) starting from the boundary points of the specular highlights, we substitute their original pixel value by the mean of the non-specular highlights neighbor pixels (Figure 5 (c)), and 3) smooth the area where the correction has been applied (Figure 5 (d)). This correction of the specular highlights is crucial because their appearance, motivated as a result of having frontal illumination, can modify the performance of certain algorithms, for instance the valleys detector.

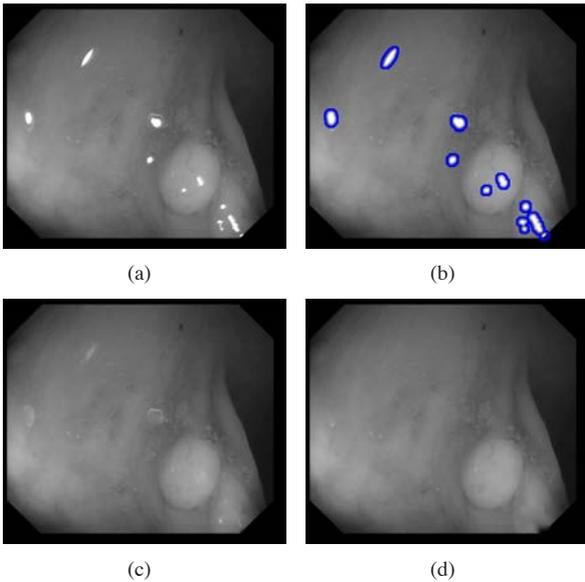


Figure 5: Correction of specular highlights: (a) Original image; (b) location of the specular highlights; (c) correction of the specular highlights, and (d) final corrected image.

2. **Image Segmentation:** We apply watersheds to the gradient image. The rationale of this choice is underpinned by the empirical evidence that the boundaries between the regions obtained in such a way are closer to the boundaries that separate the different structures [9] (an example of this can be seen in Figure 6).

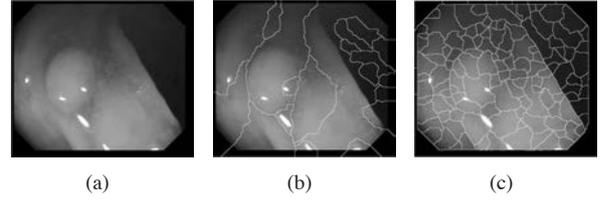


Figure 6: Use of gradient information: (a) Original Image; (b) segmentation using grey level image, and (c) segmentation using gradient image .

3. **Region Merging:**

The output of the first segmentation will yield a high number of regions that should be merged in order to fulfill our objective of providing the minimum number of relevant regions.

a) Region information-based: We first calculate the neighborhood map of the image and identify the frontier pixels between each pair of regions and then categorize the regions and frontiers, in terms of the amount of information that they contain. For instance, a low information region will have a very high mean (or very low) grey level and very low standard deviation of this grey level. We will only merge regions with the same kind of information separated by weak frontiers. In this case, in order to consider a frontier as weak we propose a frontier weakness measure as defined in Equation 5.

$$FrontierWeakness = \alpha * gradient + \beta * median \quad (5)$$

This measure combines the information of the mean gradient at the frontier pixels and the strength of the frontiers (weighted by alpha), taking into account the frontiers that are kept after applying a two consecutive order-increasing median filtering (weighted by beta), in order to reach a balanced compromise between both components [7]. The filtering also boosts the removing of regions created by veins. We merge and categorize regions until their number is stabilized or there are no weak frontiers left.

b) Depth of valleys-based: In this point we use information from the depth of valleys images (see Figure 7), in order to merge regions with frontiers constituted by pixels with low value in the DV image. We merge regions until there are no weak frontiers according to a threshold on the DV value. This new parameter allows the rejection of non-desired structures such as vessels or wrinkles, which have a marginal DV value inherent to them. Finally, the region merging ends when the number of regions is stabilized.

4.2. Region description

Once introduced the depth of valleys concept we present an algorithm that performs object identification from a depth of

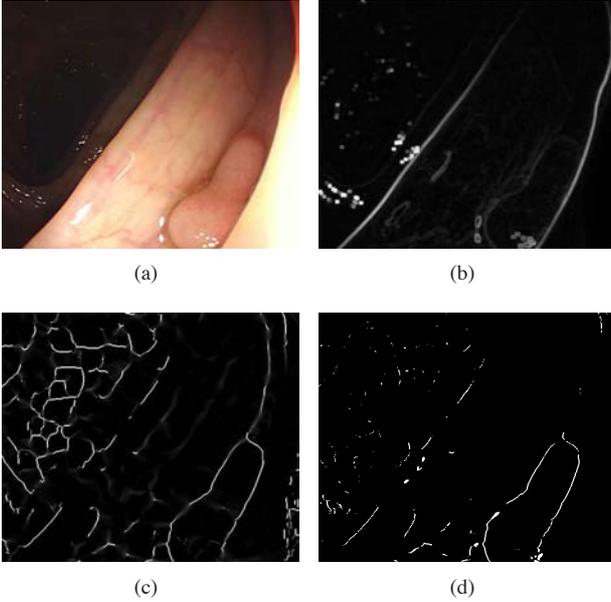


Figure 7: Creation of the depth of valleys image: (a) Original image; (b) morphological gradient image; (c) valleys image, and (d) depth of valleys image.

valleys image. The basis of the descriptor is that any point inside an object should be enclosed by a boundary, which will be constituted by points with high value in the DV image. Therefore, the goal of our Sector Accumulation-Depth Of Valleys Accumulation descriptor (SA-DOVA) is to decide which points in the image are more likely to be inside an object (in our case, a polyp) by accumulating the depth of valleys information from the boundaries that surround it. SA-DOVA uses the information that the DV image provides to guide the accumulation process but, in this case, we are not searching for elliptical boundaries but for more general shapes. SA-DOVA algorithm consists of the following steps:

1. *Choice of the starting point:* We accumulate information from the DV image for every pixel position.
2. *Obtaining the array of sectors:* In order to accumulate values from the DV image, an array of sectors centered in each pixel position is built. This structure has 3 associated parameters, the impact of which will be focus of a deep analysis in the results section. These parameters are: 1) The number of sectors, 2) the minimum radius, and 3) the maximum radius. The circular shape of the detector provides invariance to rotation, which is mandatory in the case of polyp detection. The number of segments is linked to the definition of structures to be detected, and minimum and maximum radii are associated to an implicit scale factor in which the detector provides optimal results.
3. *Accumulation:* Accumulation process in SA-DOVA is done in order to implement the operator defined in Equation 2. For each sector of the array we will accumulate, in each pixel position, the value of the maxima of the DV image that falls under it. This process can be performed in an efficient way for all the pixels in the image by approximating the acquisition of the maxima to a dilation opera-

tion, using each sector as the structural element, and then adding up all the contributions from each partial dilation.

The way this algorithm works can be better understood by following a graphical example, as shown in Figure 8. We start with the original image and we calculate the DV image (see Figure 8 (b)). Once this is done, we start with the accumulation method. The sectors that guide the accumulation process are previously calculated (in order to save computation time) so we can directly apply them to calculate the accumulation image. In Figure 8 (c) we can see in yellow the area that each different sector covers (in this case, in order to enhance the understanding of the algorithm, we only plotted one third of the sectors). We can also see how, for each sector, there are some pixels that are painted in blue. This represents the value of the DV image that will be accumulated in each point. We can see in Figure 8 (d) our resulting accumulation image, where brighter areas correspond to higher values in the accumulation image. Finally, our region classification stage is based on the maxima of the SA-DOVA descriptor to each segmented region.

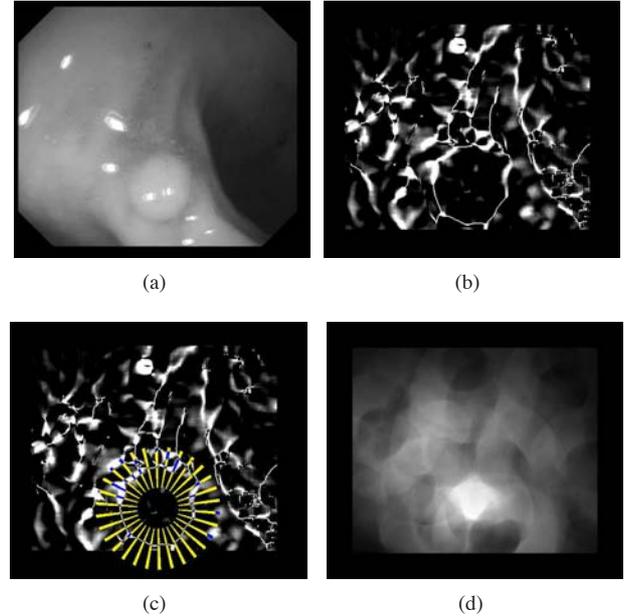


Figure 8: Explanation of SA-DOVA algorithm: (a) Original image; (b) depth of valleys image; (c) sector accumulation, and (d) final accumulated image.

5. Experimental results

5.1. Database

We built a database from our data in order to test the performance of our segmentation and description methods. In Table 2 we present the key data of our database: length, number of frames and polyp shape (flat or peduncular).

We were provided with 15 random cases, in which the experts (physicians) annotated all the sequences showing polyps, and a random sample of 20 frames per sequence was obtained, with frame size of 500×574 pixels. The central portion of the images was cropped in order to reject the non functional

Video	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Length (mm:sec)	1:02	1:30	1:07	1:05	1:15	1:09	0:54	2:44	1:05	2:36	0:37	1:21	0:36	0:43	0:35
Frames	1570	2252	1679	1648	1885	1773	1361	4121	1639	3912	947	2038	923	1097	879
Shape	Flat	Flat	Flat	Peduncular	Flat	Peduncular	Peduncular	Flat	Peduncular						

Table 2: Database description.

black borders. The experts guaranteed that all these 20 frames showed a significantly different point of view within the scene by rejecting similar frames. As this work is focused on testing the validity of our model of polyp appearance, our database consists only of frames which contains a polyp. In Figure 9 the reader can observe the great variability between the different types of polyp appearance along the different videos. This allows us to maximize the variability of the images used, while not jeopardizing any bias.

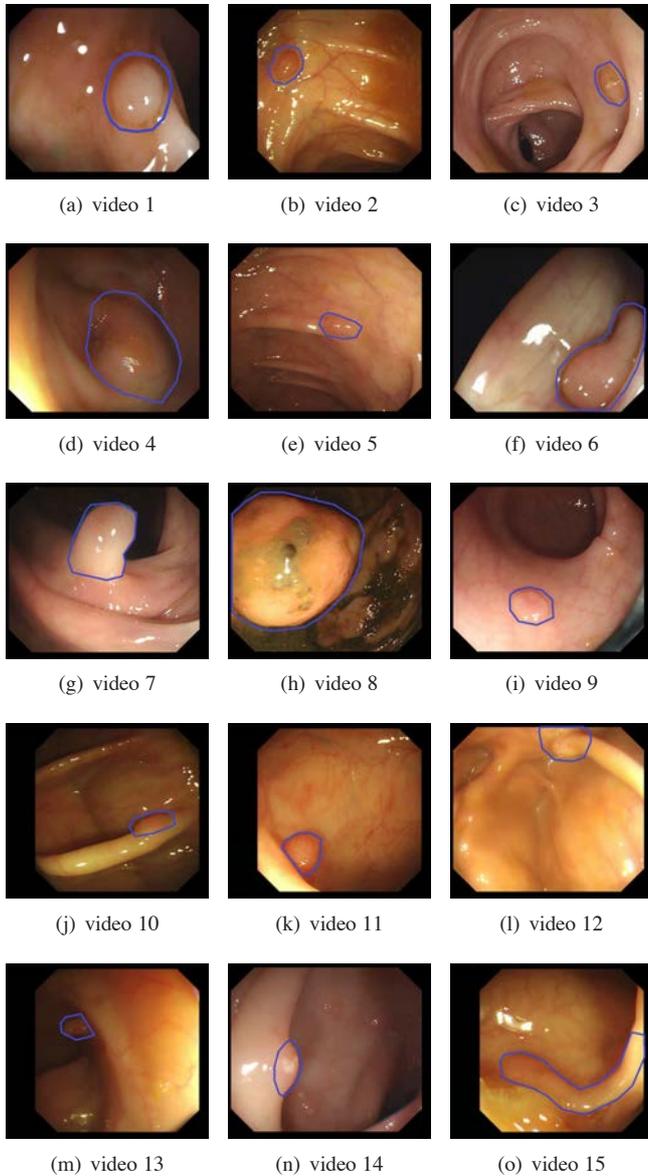


Figure 9: Examples of the type of polyp present in each colonoscopy video. Polyps are surrounded by a blue contour.

As it will be mentioned in the classification results, up to our knowledge, there exists no common database currently that can be used to test the performance of the different methods. Thus, each author has to build up their own database. This can hinder any comparative analysis because, although we can rely on common measures such as precision or recall, the data will be obtained using images from different databases. Taking this into account, our aim was to create a database, in our case consisting of 300 different images, which can cover as many types of polyp appearances as possible. Our database is available in our webs site through the following link: <http://mv.cvc.uab.es/projects/colon-qa/cvccolondb> [25].

5.2. Segmentation results

5.2.1. Experimental setup

In order to test the performance of our segmentation algorithm we run a series of experiments on the database images. We evaluate the performance of our method by using two different measures: Annotated Area Covered (AAC) (Eq. ?? and Dice Similarity Coefficient (DICE) (Eq. ?? [26]. Both measures are complementary as the former calculates the amount of annotated polyp area while the latter complements it with the amount of non-polyp information that is kept in the region. We can see in Figure 10 two different segmentation results that can help to understand the decision about using these two concrete measures. The segmentation results for the first image (Figure 10 (a-c)) is be very good in AAC but bad in terms of DICE, since the region that contains the polyp also contains lots of non-polyp information. In the other hand, segmentation results for the second image (Figure 10 (d-f)) may be a little worse in terms of accuracy but the final polyp region will contain more polyp than non-polyp information.

We compare our final segmentation results with related techniques such as *normalized cuts* [27] and *turbo pixels* [28]. To do so we set the number of final regions that we have to obtain to the minimum number of final regions that we obtained with our method that gives us best results in terms of AAC and DICE. The objective of these experiments is to check the incidence that the depth of valleys threshold has in the number of final regions and in the performance of the segmentation method.

5.2.2. Experimental Results

For this experiment we used threshold values from 0.55 to 0.8, considering that the maxima of the depth of valleys image is normalized to 1. As can be seen in Table 3 there is a clear relation between the number of final regions and the depth of valleys threshold: the higher the threshold is, the fewer the final regions are. The reason behind these results is that, by increasing the threshold value, more boundaries are considered as weak, and therefore we have more candidates to be merged.

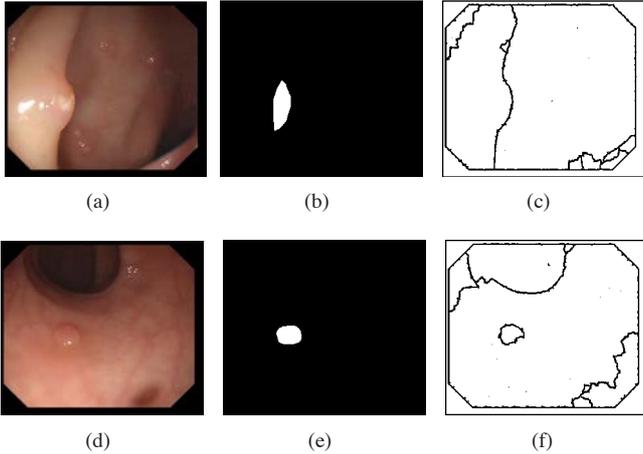


Figure 10: Examples of AAC and DICE results: (a) and (d) Original image; (b) and (e) polyp masks (the polyp is shown in white), and (c) and (f) segmentation results.

Depth of valleys threshold	0.55	0.6	0.65	0.7	0.75	0.8
Mean number of regions	29	26	23	20	18	15
Maximum number of regions	96	84	74	68	66	58
Minimum number of regions	4	3	2	2	2	2

Table 3: Number of final regions related to the depth of valleys threshold.

In Table 4 we show results for polyp region detection comparing the performance of our method, with the performance achieved by normalized cuts and turbo pixels (with the same number of final regions). In this case we only considered 3 depth of valleys threshold values: 0.6, 0.7 and 0.8. We get better results than normalized cuts and turbo pixels in terms of AAC and DICE. This means that our regions which contain polyps have more polyp information than the ones that normalized cuts and turbo pixels, and they better define the real polyp region. The reason why the performances of these two methods are worse than the one achieved by our method is that both methods are aimed to offer as output a certain number of regions that are more or less uniform in size while in our case, the final regions can be of any size.

It may seem that segmentation results improve, if we only take into account the AAC measure, as we increase the threshold value. But, as we mentioned before, this has an effect in a decrease in the number of final regions (with the possible loss of some contours). We can see from Table 4 that this is indeed the case, because AAC results do improve at the expense of decreasing DICE results. Hence, we are getting a lower number of regions, that are bigger and contain more non-polyp information than the regions obtained with a lower threshold value.

We can also see in Figure 11 the effect of the depth of valleys threshold on the performance of each different method. Our method performs equal or better than the two alternatives presented for each of the threshold values presented, although DICE results are not good for any of the methods when the threshold value is set to a higher value.

In Figure 12 we can see examples of each method’s output. It can be seen that the images segmented with our method (see Figure 12 (c)) fit better the polyp mask (that was segmented by

experts). It can be also seen that the regions that our method provides can have a different size whereas the regions that both normalized cuts and turbo pixels generate are more uniform in size. We can see also that our results could be improved by merging some little regions that appear inside the polyp region (Figure 12 (c), third row).

5.3. SA-DOVA descriptors results

The purpose of this subsection is to analyze the performance of SA-DOVA description method once it is applied to our database images. SA-DOVA algorithm needs five different parameter values:

1. *Differentiation sigma*: Parameter of the ridges and valleys detector [23] used to create the DV image.
2. *Integration sigma*: Parameter of the ridges and valleys detector. It has to be equal to the size of the structural element used to calculate the morphological gradient.
3. *Radius 1*: Minor radius of the sector.
4. *Radius 2*: Major radius of the sector.
5. *Number of sectors*: Number of sectors used to guide the accumulation method.

We fixed 3 possible values for each parameter, which are shown in Table 5. In this case we have, for each image, 243 combination of parameters. We group nearby high accumulation points via morphological dilation in order to constitute accumulation blobs. We present in Figure 13 a summary of the True Positives (TP), False Positives (FP) and False Negatives (FN) results, as there is no point on using here the True Negative (TN) measure since the number of non-polyp pixels outnumbers the number of polyp pixels (in this case a TP means that our accumulation blob falls into the polyp mask, and it has to be noticed that we may have more than one accumulation blob per image).

The best results for each measure can be consulted at Table 6. By a general observation of the results obtained by applying SA-DOVA accumulation method, we observe that our best precision results are above 54% and our best recall results are above 72%. The reason for the low precision results is that we have a relatively high number of FP, which is not bad for our purpose. We prefer, taking into account the scope of our method, to have a certain number of FP rather than having a high number of FN, which will indicate that we are losing a great number of polyps.

We can see how this method works in Figure 13. An example of good performance of SA-DOVA can be seen in Figure 13 (a), where the method places all the maxima of accumulation inside the polyp region and also places the maxima of all the accumulation points in the center of the polyp. It is true that SA-DOVA does not perform well in some cases, as can be seen in Figure 13 (b) where non-polyp structures of the image such as wrinkles or blood vessels makes the accumulation system perform in an undesired way. In general, the system shows a good balance of TP and FP results and SA-DOVA performs well, in terms of discriminative power, for a great majority of the images.

Measure / Method	Ours	NCuts	TurboP	Ours	NCuts	TurboP	Ours	NCuts	TurboP
Threshold Value	0.6			0.7			0.8		
AAC	61.91%	63.66%	63.65%	70.29%	69.06%	69.2%	75.79%	70.86%	71.98%
DICE	55.33%	44.97%	45.98%	44.6%	37.75%	38.04%	36.44%	34.01%	34.66%

Table 4: Comparison between the results obtained by our method, Normalized Cuts and TurboPixels with respect to the depth of valleys threshold.

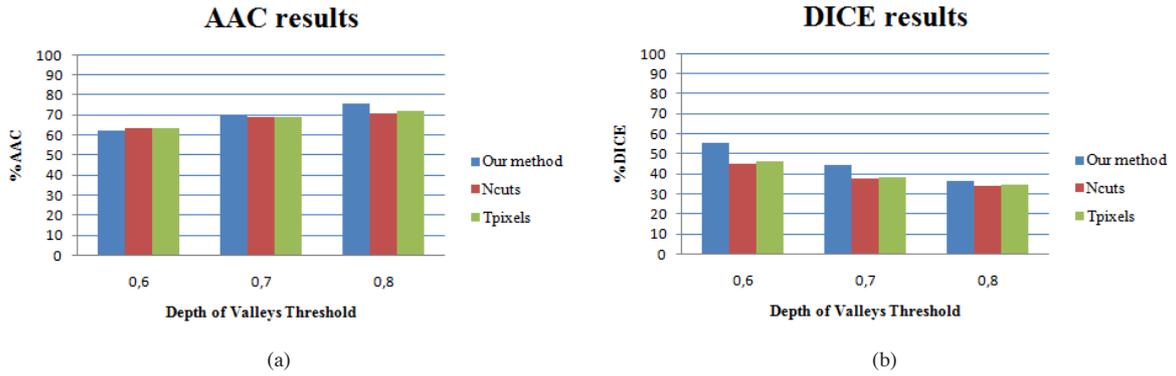


Figure 11: Comparison of segmentation results: (a) AAC results, and (b) DICE results.

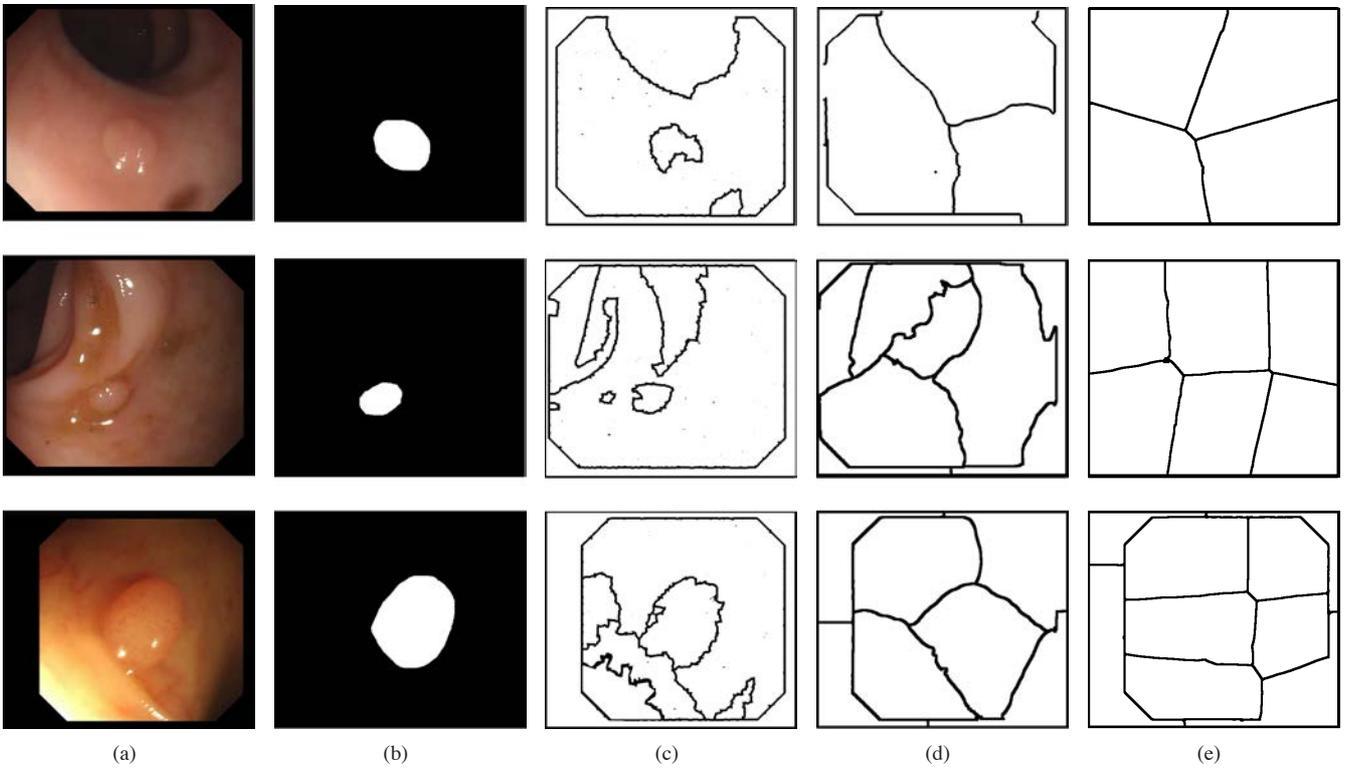
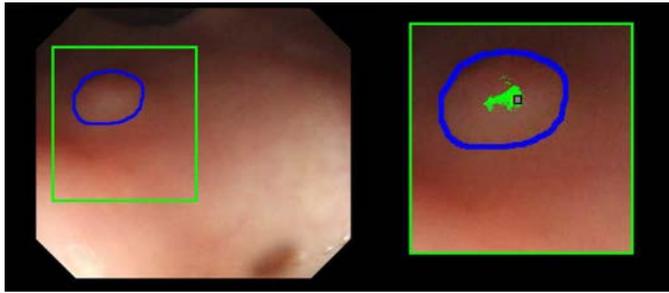
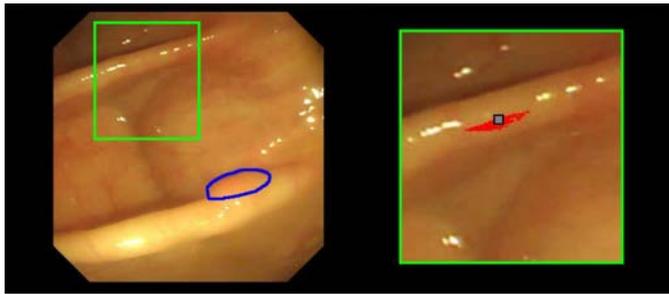


Figure 12: Comparison of segmentation results: (a) Original images; (b) polyp masks; (c) our method's output; (d) normalized cuts' output, and (e) turbo pixels' output.



(a)



(b)

Figure 13: Successful (a) and unsuccessful (b) SA-DOVA results: Right image zooms on the ROI showing TP in green, FP in red and the position of the maxima of accumulation marked by a blue square. The blue line bounds the polyp.

In order to measure, as a first approximation, the discriminative power of SA-DOVA we performed the following experiment: we counted, for each combination of parameters, in how many images we place the maxima of the accumulation method inside the labeled polyp region. Results of this experiment are shown in Figure 14. By doing so, we can assess if our method is suitable for polyp detection and, therefore, applicable later in the classification stage.

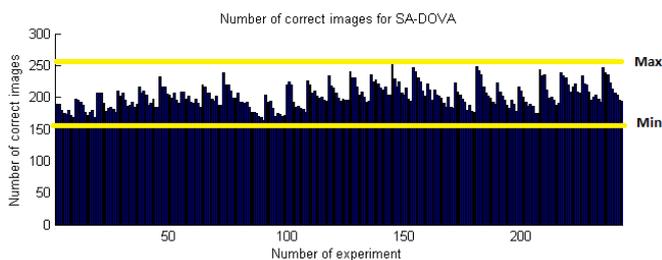


Figure 14: SA-DOVA results related to the placement of the maxima of accumulation.

As can be seen in Figure 14, SA-DOVA, in the worst of the experiments, places the maxima in 164 images and in the best experiment, in 252 out of all the 300 images of our database. While the performance of SA-DOVA in classification purposes will be discussed in the next subsection, the preliminary results shows that where SA-DOVA places the maxima of accumulation, there is a great chance that is inside the polyp.

Parameter	Values
Differentiation sigma	[2, 4, 8]
Integration sigma	[4, 8, 10]
Rad 1	[30, 40, 50]
Rad 2	[80, 100, 120]
Number of sectors	[60, 120, 180]

Table 5: Parameters of SA-DOVA.

Measure	Value	Diff. σ	Int. σ	Min r	Max r	Sectors
TP	215	4	4	30	120	120
FP	241	2	10	50	80	60
FN	85	4	4	30	120	120
Prec	47.14%	4	4	30	120	60
Recall	71.66%	4	10	50	80	60

Table 6: Best results for SA-DOVA.

Finally, in order to end this comparison, it is important to know how much processing time SA-DOVA takes which is shown in Table 7. SA-DOVA takes 19 seconds to calculate the accumulation values for each image. This can be achieved because SA-DOVA algorithm can be applied to the whole image at once as we are using the dilation operation to approximate the maxima accumulation. The SA-DOVA method has an inherent computational complexity $O(n)$ with the number of sectors and its parallelization is straightforward when using the efficient dilation proposed for the structural elements. These results were obtained by a PC with an Intel-i7-930 quad-core processor and 12 GB of memory, using Matlab scripts for the software implementations.

Number of experiments	72900
Mean processing time per experiment (mins)	76.95
Mean processing time per image (secs)	19

Table 7: Processing time results.

5.4. Polyp detection results

The region segmentation and description stages of our processing scheme have been detailed in previous subsections. In this subsection we present our polyp detection results, obtained by applying a classification criteria by means of our SA-DOVA descriptor, which is applied to every segmented region. Before showing our detection results, we will introduce the concepts that we will later use to explain and analyze our classification results.

5.4.1. Experimental setup

In order to test the performance of our whole processing scheme, we will apply our whole method to all the 300 images from our database. In this case, we will use in the region segmentation the combination of parameters that gave better results in terms of both AAC and DICE. We will show results of applying SA-DOVA to those segmented regions.

In our images, non-polyp pixels outnumber polyp pixels and, therefore, non-polyp regions outnumber polyp regions. In this case, the number of TP and FN in the $[0, 300]$ range whereas the number of FP and TN will be higher. In order to show how

our method performs in a region-based scenario or in a polyp detection scenario, we will consider two different sets of results:

1. Region-based results: For each image, we will classify each region into polyp containing or not, according to the value of the maximum of the DOVA descriptor for this region. In order to measure the performance of our method we will count, for each image, the number of TP, FP, TN and FN regions.
2. Frame-based results: For each image we will have only two regions: the one that the system predicted the polyp to be inside (according to the value of the DOVA descriptor) and the region where the system did not predict polyp. In this case all TP, FP, TN and FN values will be in the range of [0, 300] as there is only one polyp region in each one of our database images.

In order to measure the performance of our detection method we will use the following measures: 1) Precision (6); 2) recall (7); 3) accuracy (8); 4) specificity (9); 5) fallout (10); 6) F2measure (11) and 7) Area Under Curve (AUC).

$$Precision = \frac{TP}{TP + FP}; \quad (6)$$

$$Recall = \frac{TP}{TP + FN}; \quad (7)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}; \quad (8)$$

$$Specificity = \frac{TN}{TN + FP}; \quad (9)$$

$$Fallout = \frac{FP}{FP + TN}; \quad (10)$$

$$F2measure = 5 \cdot \frac{TP}{5 \cdot TP + 4 \cdot FN + FP}; \quad (11)$$

In our context, precision will measure how good we are at defining what is a polyp region, whereas recall will denote if some relevant results were missed -in this case, if the system missed polyp regions-. For our concrete case precision and recall appear to be good indicators of the performance of our system and we will use a complementary measure, F2measure, that combines their values into an only measure [29]. F2measure can be relevant in our case, as it enhances the recall value, which is crucial for our interests, i.e., it is more important for our objective that we get a very low number of false negatives than the number of false positives. Finally we will also show performance results by means of ROC curves and Precision-Recall curves.

5.4.2. Experimental results

We present in the central column of Table 8 the global results for the best combination of parameters. We get high values for both precision, recall and F2measure. That is, our global processing scheme performs well when detecting both what is a polyp and what is not a polyp in the image. We get a very low number of FN, which, as it will be later exposed, are due to the

Measure	Region-based	Frame-based
Precision	89%	88%
Recall	89%	89%
Accuracy	97%	89%
Specificity	98%	88%
F2measure	0.89	0.89
TP	265	265
FP	31	35
TN	2262	269
FN	35	31

Table 8: Comparison between results obtained by using the two defined criteria: Region-based and Frame-based.

presence of elements in the scene that can alter the performance of our depth of valleys approximation.

In order to complete this analysis, we plot in Figure 15 and in Figure 16 several graphs to find out which of the five parameters have more relevance, showing that in some cases the variation of the parameters values does not affect significantly the results. A number of interesting conclusions can be extracted from this analysis:

- TP results are better when both differentiation and integration sigma are equal, and improve as the size of both radius increases.
- FP results are better as both radius increases and appear to improve as the differentiation sigma is bigger.
- TN improve as the differentiation sigma increases and are more stable as the integration sigma is increased. TN results are better the less the difference between radius values.
- FN results improve as differentiation sigma and both radius decrease their values.
- Precision results increases as the differentiation sigma increases. Recall results also improves as differentiation sigma increases. This results on a slight increase in the F2measure value for middle values of both differentiation and integration sigmas.
- Accuracy and specificity results appear to vary little, due to the great difference in the number of TP and FN compared to the number of FP and TN.

Comparison of SA-DOVA results with different classification setups

The second column of Table 8 we shows the results for the frame-based approach. The main difference between both methods relies on the fact that with the second criteria we have a lower number of true negatives. This is because, for each image, we will only have two regions so there will be only one TP or FP and only one TN or FN per image. In general the performance results are very similar and, as it can be expected, the number of TN has greatly decreased.

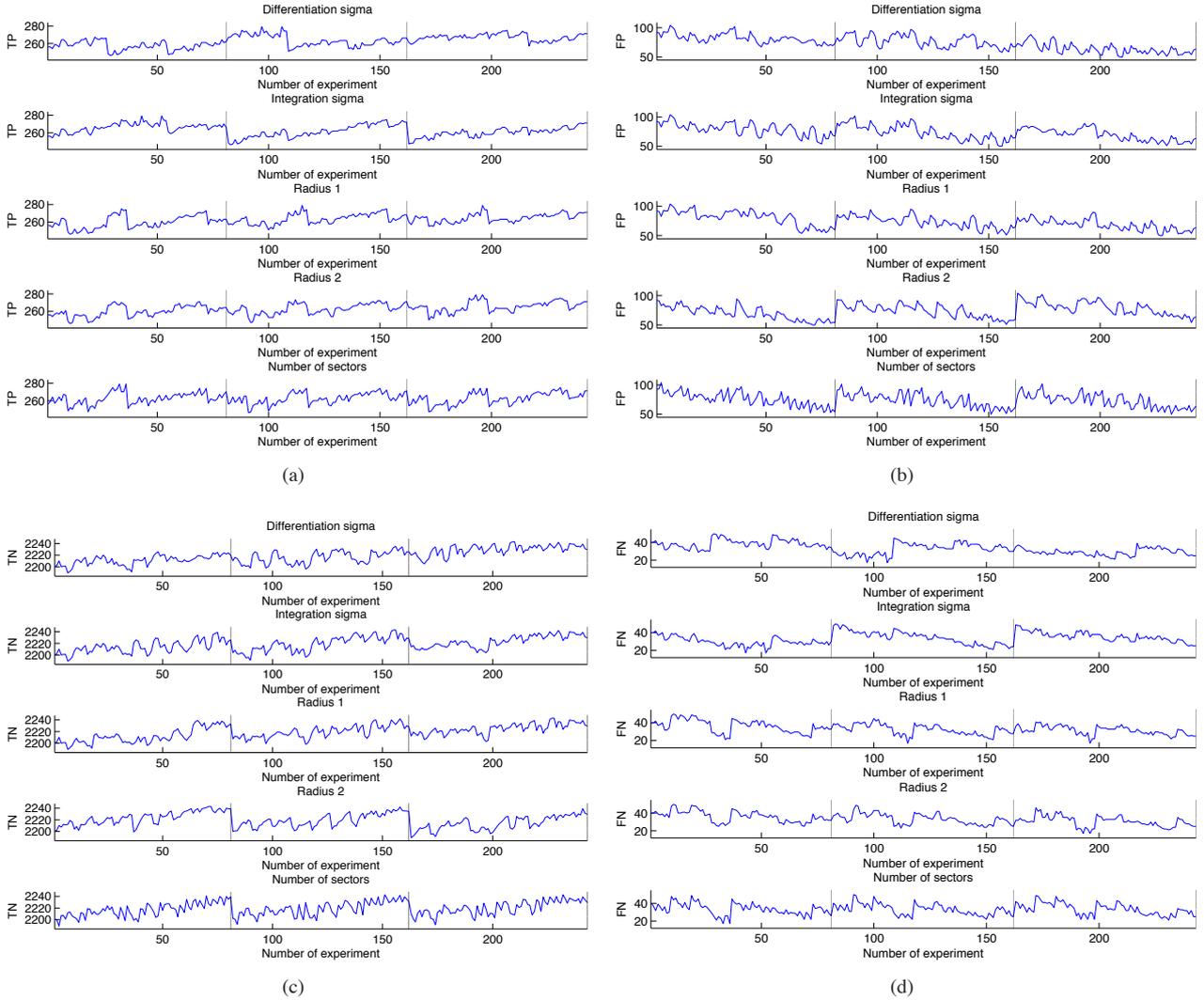


Figure 15: SA-DOVA results related to the parameters values. Red vertical lines bound the experiments for each of the 3 possible values of the studied parameter, while changing the remaining four parameters. (a) TP; (b) FP; (c) TN and (d) FN.

We have also shown detection results by using a binary classification criteria based on the maxima of the DOVA descriptor, i.e., the threshold value used to determine if a region contains a polyp or not is the maxima of the SA-DOVA descriptor for the concrete image. But in order to further study the performance of our detection scheme we set up experiments where this threshold value can vary from 0 to the maxima of the DOVA descriptor. We show results for several different parameter value combinations and for the two before mentioned criteria, are shown in form of ROC and Precision-Recall curves [30] in Figure 17.

One good measure to assess the performance of a certain method based on the study of ROC or Precision-Recall curves is the Area Under the Curve (AUC). AUC is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. As explained in [30] there is a relationship between both ROC and Precision-Recall curves. ROC curves are commonly used to present results for binary decision problems in machine learning. However, when dealing with highly skewed datasets,

Precision-Recall (PR) curves give a more informative picture of an algorithm's performance. In both cases, although the methods used to calculate the AUC are different, an AUC measure nearer to 1 indicates a good test whereas values near to 0.5 represents a worthless test ([31]).

We show in Table 10 results for both AUC for ROC and Precision-Recall curves for the top four combinations of parameters (presented in Table 9) in terms of performance for the sake of clarity. We can see from the results of the table that with the first region-based classification criteria we obtain very good results in terms of AUC for ROC curve, and good results for the area under the Precision-Recall curve. Both AUC measures are worse for the second classification criteria as, in this case, results are affected by a reduction in the number of TN.

If we do an analysis of the results we can see that we are in a good direction, at least in terms of the identification of non-polyp-containing regions. It is true that we do not identify correctly where is the polyp in some images, but there are some reasons to explain this fact. First, segmentation is not al-

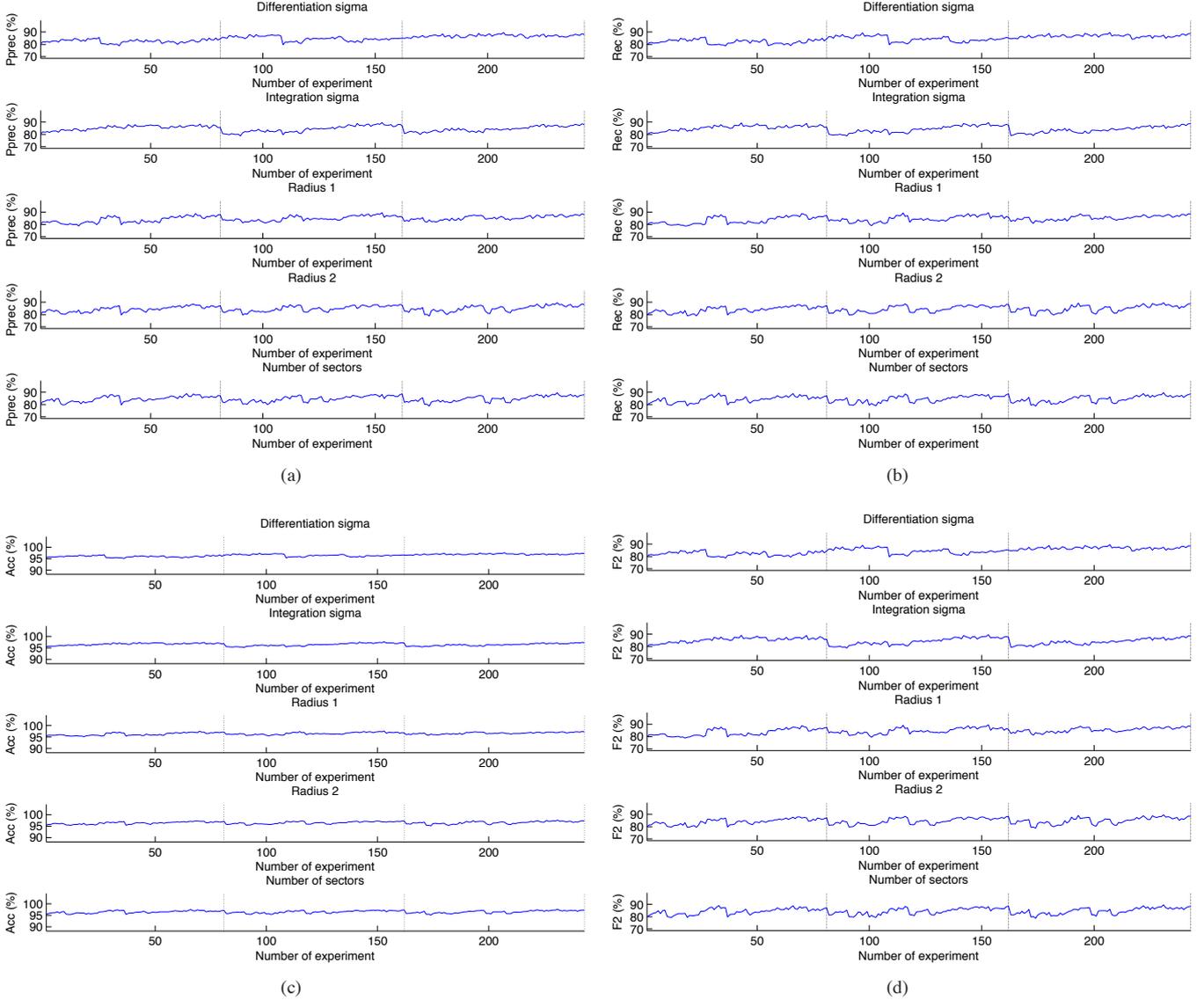


Figure 16: SA-DOVA results related to the parameters values. Red vertical lines bound the experiments for each of the 3 possible values of the studied parameter, while changing the remaining four parameters. (a) Precision; (b) recall; (c) accuracy, and (d) F2measure.

Combination	Diff. σ	Int. σ	Min r	Max r	Sectors
1	8	8	40	80	180
2	8	4	50	80	180
3	2	10	30	120	60
4	2	8	30	100	60

Table 9: Combinations of parameters used in the experiments.

Combination	ROC C1	ROC C2	PR C1	PR C2
1	0.98	0.76	0.77	0.45
2	0.98	0.74	0.75	0.44
3	0.96	0.73	0.65	0.40
4	0.96	0.72	0.63	0.38

Table 10: Comparison between AUC results obtained by using the two defined criteria: Region-based and Frame-based.

ways perfect and classification is affected by its results. We can place the maxima in the real polyp region but the segmentation of the image may not be accurate so we will denote polyp regions incorrectly as non-polyp regions. Second, there are some problems that should be tackled in order to improve SA-DOVA descriptor. SA-DOVA appears to perform well, specially when determining what is not a polyp, but it is affected by the apparition of non-desired structures as wrinkles or blood vessels which affect the DV image, where SA-DOVA descriptor takes the information for the accumulation method.

To conclude with the results section, we show in Figure 18 some real polyp detection results in images from our database. We can see from the results that our whole processing scheme can lead to accurate polyp detection results. For some images we do not label as polyp-containing the whole polyp region (second row from Figure 18) but, in this case, this happens due

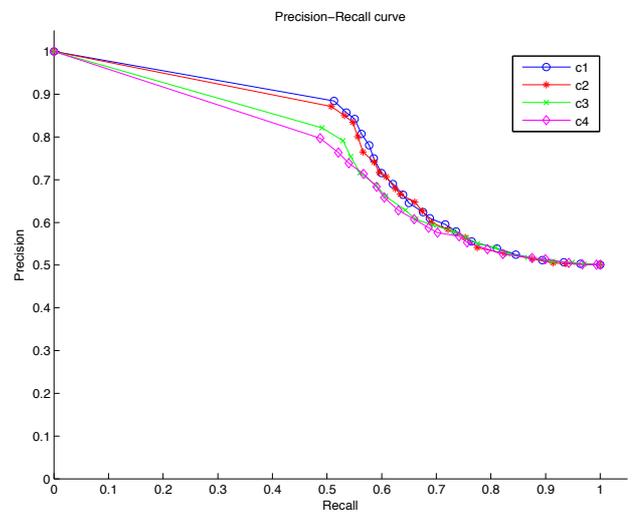
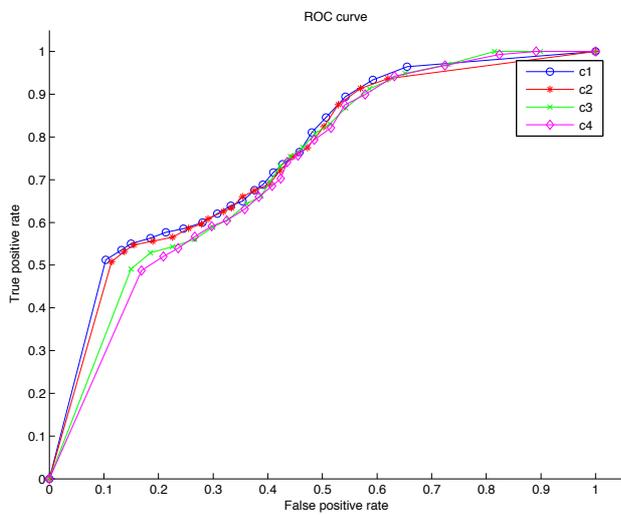
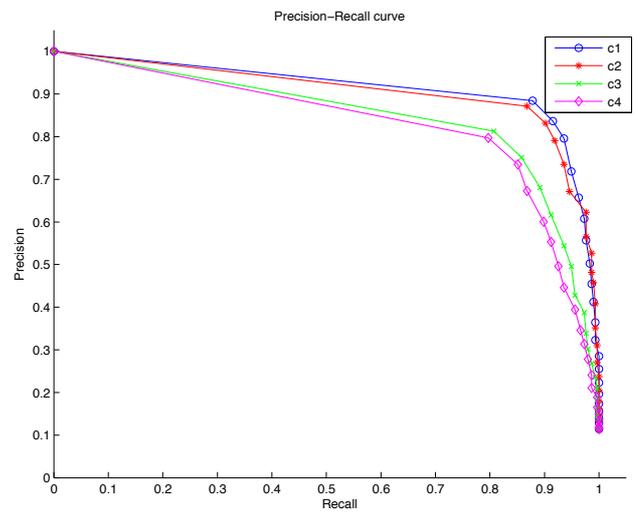
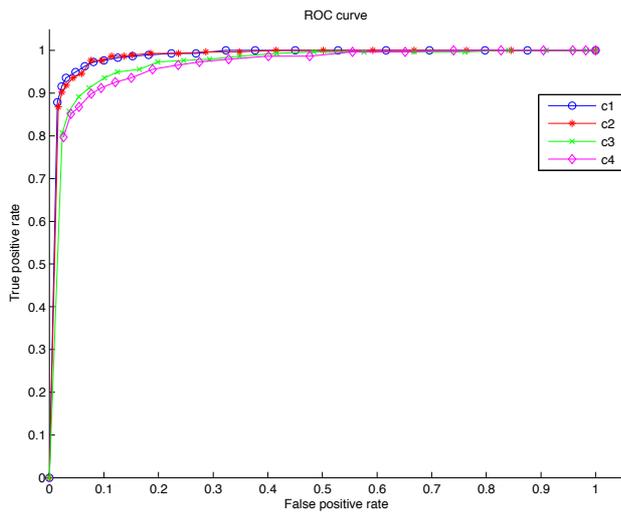


Figure 17: Region-based experiments: (a) ROC curve; (b) Precision-Recall curve. Frame-based experiments: (c) ROC curve; (d) Precision-Recall curve.

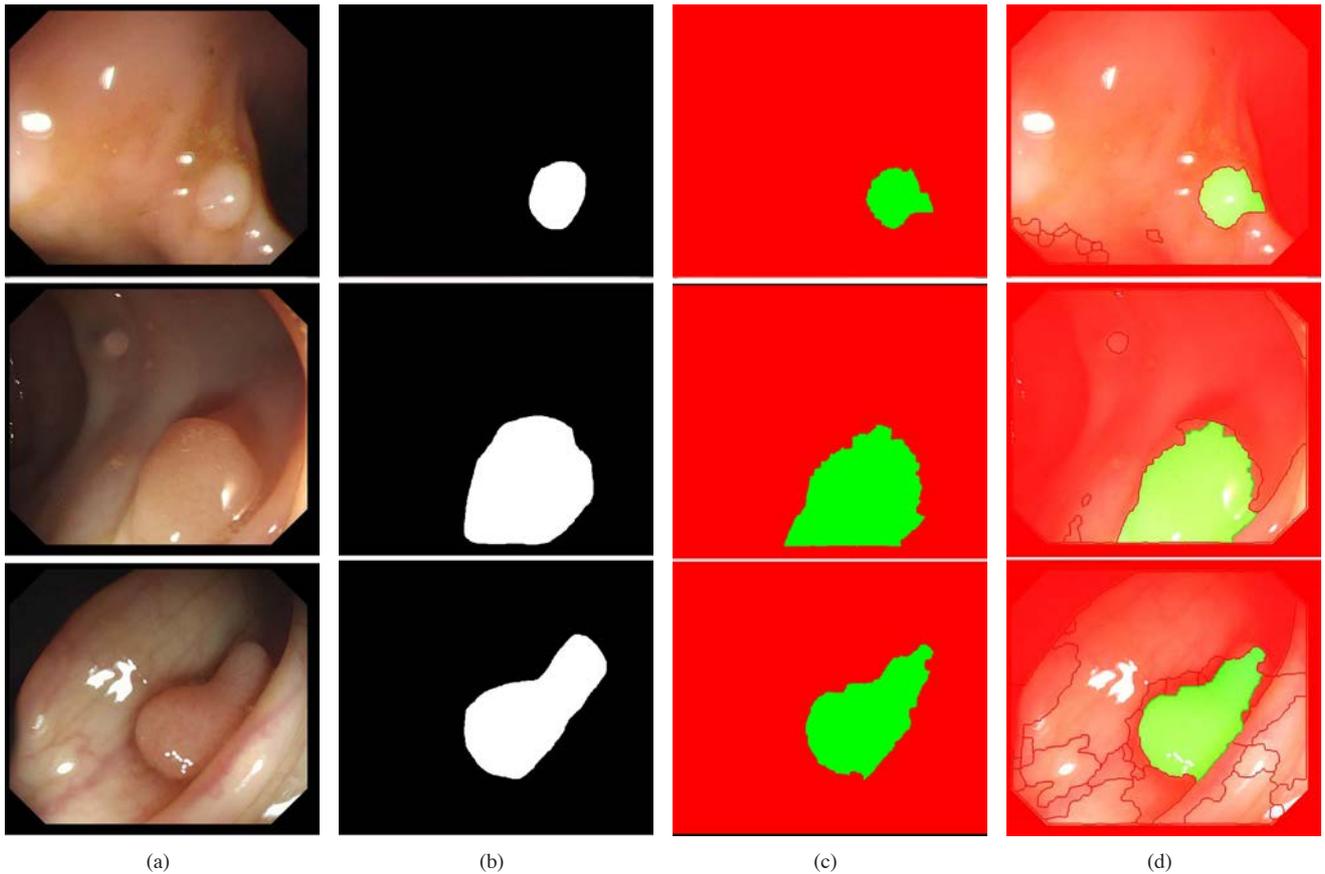


Figure 18: Polyp detection results: (a) Original images; (b) polyp masks; (c) region classification, and (d) final classified image. Areas with greenish color equal to regions classified as polyp containing. Areas with reddish color correspond to regions classified as non-polyp containing

to an imperfect region segmentation.

6. Discussion

In this section we discuss the performance of each stage of the processing scheme and we sketch how the performance of each of them could be improved to obtain better global detection results.

The region segmentation stage is based on the previously defined model of polyp appearance. The experimental results showed that segmentation results depend on the threshold that we apply to the DV image in a way such the higher the threshold, the lower the number of final regions. But, in this case, as the threshold is increased, the performance is damaged, as we can lose, for some images, some of the boundaries of the polyp region if we apply a high threshold value, as it can be seen in Figure 19 (b) and (c).

As there are some elements on the endoluminal scene that also have high response in the DV image, such as folds or blood vessels, in order to obtain a better segmentation of the polyp region we should be able to detect which responses of the detector are produced by these elements. In a different approach, another possibility could be to use results from the SA-DOVA descriptor, such as its global maximum, to guide the segmentation process. We can think of a watershed marker-controlled

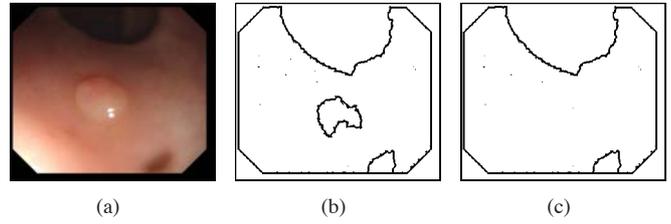


Figure 19: Problems in segmentation: (a) Original image; (b) segmentation result with depth of valleys threshold 0.7, and (c) Segmentation result with depth of valleys threshold 0.8.

segmentation system, similar to the ones that have been proposed recently in the literature [32].

Our region descriptor, SA-DOVA does not perform well for a reduced number of images. If we want to improve SA-DOVA results, apart from the elimination of the effect of the before mentioned structures such as folds or the lumen (see Figure 20), we could think of using additional information from the source of the DV image such as the orientation of the valleys combined with the orientation of the sectors in order to filter undesired accumulations. The problem in this case comes when the polyp does not appear fully on the image and, in this case, we could even get worse results. Another improvement can come by the reduction of the number of parameters, in order to obtain a more



Figure 20: Erroneous location of SA-DOVA maxima: (a) Influence of structures such as wrinkles or vessels; (b) erroneous location of the maxima of accumulation due the influence of the lumen.

robust descriptor with less degrees of freedom. One possibility could be to calculate automatically the minor and major radius, based on the distance between the different edges and valleys that appear on the image. Another alternative could be to develop an accumulation method that takes into account the approximation that polyps have an elliptical shape, although we believe that this could be more suitable only for certain cases.

Finally, we showed promising detection results, especially for the case of non-polyp region identification, obtaining a low number of FN. It is important to mention that classification results are affected by the performance of both previous stages so we can see in Figure 21 how, in some cases, a good description result can be affected, in terms of the final classified image, by the results of the segmentation stage. In both examples we identify correctly the polyp region by means of the value of the SA-DOVA descriptor, but the final output result is worse than expected due the segmentation results.

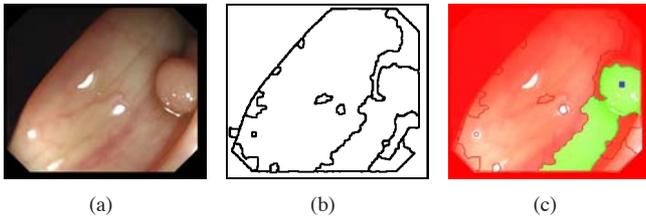


Figure 21: Problems associated to classification: (a) Original images; (b) segmentation output; (c) final classified image. Polyp-labeled regions are shown in green, non-polyp labeled regions are shown in red and the maxima of accumulation is marked by a blue square

It may seem logical that classification results will improve as both segmentation and description stages improve their results. We should add that our polyp detection system is completely automatic, that is, there is no learning in our system. Therefore, one possible improvement to the whole system could be to apply machine learning techniques so, instead of relying on maxima of accumulation for each image, we could arrive to a threshold value to be used for all the images.

As mentioned in Section 2, it is very difficult to do a proper comparison between different approaches of polyp detection in colonoscopy video being the main reason the lack of a public database to test the different algorithms and because of that we find difficult to compare different approaches just by sim-

ply comparing precision or recall data. For instance, we could compare our work to the one by [12], because it also uses the approximation of elliptical shape of polyps, but in this case the authors only look for one certain type of polyps (flat polyps) and only test their method on 27 different images. We can also use the work of [13] in order to compare our results. In this case the authors perform their test on virtual colonoscopy images, which have the advantage of knowing precisely the sizes of the objects that appear on the image, which does not happen in our case. We can also compare our global results with the ones achieved by texture-based approaches, such as the work of [19]. In this case the authors apply different methods on a larger database but they offer problem-specific measures along with accuracy results, which we find difficult to compare with ours as the type of images and experiments are very different.

7. Conclusions and Future Work

In this paper we presented a polyp detection scheme based on a model of polyp appearance in the context of the analysis of colonoscopy images. This model is built on the appearance of valleys surrounding polyps as the light of the colonoscope and the camera are in the same direction, causing the apparition of shadows around prominent surfaces such as polyps. We also presented the concept of depth of valleys image, which combines the information of a ridges and valleys detector with the morphological gradient. This approximation seems to be suitable for the images we are working with, as it enhances both valleys and edges that constitute the boundaries of the structures that appear on the images.

Our processing scheme consists of three stages, namely: region segmentation, region description and region classification. Region segmentation is built considering some of the challenges that colonoscopy images present such as the presence of specular reflections and its performance results show promising results for a great number of the images, outperforming segmentation results obtained by state-of-the-art techniques.

We presented our region descriptor, the SA-DOVA accumulation algorithm, built on the approximation that points interior to objects (in this case, polyps) will be surrounded by boundaries constituted by pixels with high value in the depth of valleys image.

In order to assess the performance of our detection system, we built a database which covers as many types of polyp appearance as possible. To classify the final segmented regions we use the maxima of the SA-DOVA descriptor on each region, determining as polyp containing those regions where the value of the descriptor is maximum. Our current classification results are promising, specially when determining which are the non-polyp regions in the image.

In order to improve our results, we plan to implement in the near future methods that can help to mitigate the effects that the apparition of certain structures, such as the lumen, specular highlights, wrinkles or blood vessels have in our detection method. Some other possible future lines could consist of applying machine learning methods in order to build more a robust classification systems.

Finally, we showed how our polyp detection method works in images that contain polyps but we also plan to extend and test our performance method in order to work in a whole colonoscopy video. Our final objective is to build a tool that can identify which regions of the image the physician should pay special attention to and, for the moment, we are concentrated in defining a robust model of polyp appearance that can be integrated in this tool, which can cover as many types of polyp appearance as possible.

ACKNOWLEDGEMENTS

The authors would like to acknowledge David Vázquez, Dr. Antonio López, Dr. Debora Gil and the GV2 group from Trinity College for their contribution in the discussions during the writing of this paper. This work was supported in part by a research grant from Universitat Autònoma de Barcelona 471-01-3/08, by the Spanish Government through the founded project "COLON-QA" (TIN2009-10435) and by research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018).

REFERENCES

References

- [1] Tresca, A., The Stages of Colon and Rectal Cancer, New York Times (About.com) (2010) 1.
- [2] Hassinger, J.P., Holubar, S.D. et al., Effectiveness of a Multimedia-Based Educational Intervention for Improving Colon Cancer Literacy in Screening Colonoscopy Patients, *Diseases of the Colon & Rectum* 53 (2010) 1301.
- [3] N. Segnan, J. Patnick, L. von Karsa, European guidelines for quality assurance in colorectal cancer screening and diagnosis, Luxembourg: Publications Office of the European Union, 2011.
- [4] I. Bratko, I. Mozetič, N. Lavrač, KARDIO: A study in deep and qualitative knowledge for expert systems, MIT Press (1990).
- [5] J. Wei, H. Chan, C. Zhou, Y. Wu, B. Sahiner, L. Hadjiiski, M. Roubidoux, M. Helvie, Computer-aided detection of breast masses: Four-view strategy for screening mammography, *Medical Physics* 38 (2011) 1867.
- [6] S. Viswanath, D. Palumbo, J. Chappelow, P. Patel, B. Bloch, N. Rofsky, R. Lenkinski, E. Genega, A. Madabhushi, Empirical evaluation of bias field correction algorithms for computer-aided detection of prostate cancer on T2w MRI, in: *Proceedings of SPIE*, volume 7963, p. 28.
- [7] Bernal, J., Sánchez, J., and Vilariño, F., *Colonoscopy Book 1: Towards Intelligent Systems for Colonoscopy*, In-Tech, 2011.
- [8] Bernal, J., Sánchez, J., and Vilariño, F., A Region Segmentation Method for Colonoscopy Images Using a Model of Polyp Appearance, in: *Proceedings of the IbPRIA 2011*, pp. 134–142.
- [9] Bernal, J., Sánchez, J., Vilariño, F., Reduction of Pattern Search Area in Colonoscopy Images by Merging Non-Informative Regions, in: *Proceedings of the XXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, Madrid, Spain, pp. 88–96.
- [10] T. Byers, B. Levin, D. Rothenberger, G. D. Dodd, R. A. Smith, F. T. A. C. S. Detection, T. A. G. on Colorectal Cancer), American cancer society guidelines for screening and surveillance for early detection of colorectal polyps and cancer: Update 1997, *CA: A Cancer Journal for Clinicians* 47 (1997) 154–160.
- [11] Bernal, J. and Vilariño, F. and Sánchez, J., Feature Detectors and Feature Descriptors: Where We Are Now, Technical Report 154, Computer Vision Center & Computer Science Department Universitat Autònoma de Barcelona, 2010.
- [12] Hwang, S. and Oh, J.H. and Tavanapong, W. et al., Polyp detection in colonoscopy video using elliptical shape feature, in: *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 2, IEEE, pp. II-465.
- [13] H. Zhu, Y. Fan, Z. Liang, Improved curvature estimation for shape analysis in computer-aided detection of colonic polyps, Beijing, China (2010) 19.
- [14] C. van Wijk, V. van Ravesteijn, F. Vos, L. van Vliet, Detection and segmentation of colonic polyps on implicit isosurfaces by second principal curvature flow, *Medical Imaging, IEEE Transactions on* 29 (2010) 688–698.
- [15] P. Li, K. Chan, S. Krishnan, Learning a multi-size patch-based hybrid kernel machine ensemble for abnormal region detection in colonoscopic images, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (2005).
- [16] S. Ameling, S. Wirth, D. Paulus, G. Lacey, F. Vilariño, Texture-based polyp detection in colonoscopy, *Bildverarbeitung für die Medizin 2009* (2009) 346–350.
- [17] Karkanis, S.A. and Iakovidis, D.K. and Maroulis, D.E. et al., Computer-aided tumor detection in endoscopic video using color wavelet features, *Information Technology in Biomedicine, IEEE Transactions on* 7 (2003) 141–152.
- [18] M. Tjoa, S. Krishnan, Feature extraction for the analysis of colon status from the endoscopic images, *BioMedical Engineering OnLine* 2 (2003) 1–17.
- [19] M. Coimbra, J. Cunha, MPEG-7 visual descriptors and contributions for automated feature extraction in capsule endoscopy, *Circuits and Systems for Video Technology, IEEE Transactions on* 16 (2006) 628–637.
- [20] S. Krishnan, X. Yang, K. Chan, S. Kumar, P. Goh, Intestinal abnormality detection from endoscopic images, in: *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 2, pp. 895–898.
- [21] B. Dhandra, R. Hegadi, M. Hangarge, V. Malemath, Analysis of abnormality in endoscopic images using combined hsi color space and watershed segmentation, in: *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4, pp. 695–698.
- [22] B. T. Phong, Illumination for computer generated pictures, *Communications of ACM* 18 (1975) 311–317.
- [23] López, A.M., Lumbreras, F. et al., Evaluation of methods for ridge and valley detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (1999) 327–335.
- [24] Serra, J., *Image Analysis and Mathematical Morphology*, Academic Press, Inc. Orlando, FL, USA, 1983.
- [25] J. Bernal, J. Sánchez, F. Vilariño, Cvc-colondb: A database for assessment of polyp detection, *Database*, 2012.
- [26] Riaz, F., Ribeiro, M.D., Coimbra, M.T., Quantitative comparison of segmentation methods for in-body images, in: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 5785–5788.
- [27] J. Shi, J. Malik, Normalized cuts and image segmentation, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22 (2000) 888–905.
- [28] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson, K. Siddiqi, Turbopixels: Fast superpixels using geometric flows, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31 (2009) 2290–2297.
- [29] M. Sokolova, N. Japkowicz, S. Szpakowicz, Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation, *AI 2006: Advances in Artificial Intelligence* (2006) 1015–1021.
- [30] J. Davis, M. Goadrich, The relationship between Precision-Recall and ROC curves, in: *Proceedings of the 23rd international conference on Machine learning, ACM*, pp. 233–240.
- [31] T. Tape, *Interpreting diagnostic tests*, University of Nebraska Medical Center, 2011.
- [32] S. Xu, H. Liu, E. Song, Marker-controlled watershed for lesion segmentation in mammograms, *Journal of Digital Imaging* (2011) 1–10.