
This is the **accepted version** of the journal article:

Bernal del Nozal, Jorge [et al.]. «WM-DOVA maps for accurate polyp highlighting in colonoscopy : Validation vs. saliency maps from physicians». *Computerized Medical Imaging and Graphics*, Vol. 43 (July 2015), p. 99-111 DOI 10.1016/j.compmedimag.2015.02.007

This version is available at <https://ddd.uab.cat/record/326494>

under the terms of the  license.

WM-DOVA Maps for Accurate Polyp Highlighting in Colonoscopy: Validation vs. Saliency Maps from Physicians

Jorge Bernal^{a,1}, F. Javier Sánchez^a, Gloria Fernández-Esparrach^b, Debora Gil^a, Cristina Rodríguez^b, Fernando Vilariño^a

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

^bEndoscopy Unit, Gastroenterology Service, CIBERHED, IDIBAPS, Hospital Clinic, Universidad de Barcelona, Barcelona, Spain

Abstract

We introduce in this paper a novel polyp localization method for colonoscopy videos. Our method is based on a model of appearance for polyps which defines polyp boundaries in terms of valley information. We propose the integration of valley information in a robust way fostering complete, concave and continuous boundaries typically associated to polyps. This integration is done by using a window of radial sectors which accumulate valley information to create WM-DOVA¹ energy maps related with the likelihood of polyp presence. We perform a double validation of our maps, which include the introduction of two new databases, including the first, up to our knowledge, fully annotated database with clinical metadata associated. First we assess that the highest value corresponds with the location of the polyp in the image. Second, we show that WM-DOVA energy maps can be comparable with saliency maps obtained from physicians' fixations obtained via an eye-tracker. Finally, we prove that our method outperforms state-of-the-art computational saliency results. Our method shows good performance, particularly for small polyps which are reported to be the main sources of polyp miss-rate, which indicates the potential applicability of our method in clinical practice.

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

Keywords: Polyp localization, Energy Maps, Colonoscopy, Saliency, Valley detection

¹ WM-DOVA: Window Median Depth of Valleys Accumulation

1. INTRODUCTION

1.1. Introduction to Intelligent Systems for Colonoscopy

Colon cancer is nowadays the fourth cause of cancer death worldwide. In 2012, an estimated number of 746.298 cases were diagnosed with colorectal cancer (CRC) and a total number of 373.631 died of the disease [1]. CRC's survival rate depends on the stage it is detected on, going from rates higher than 95% in early stages to rates lower than 35% in latter ones [2], hence the necessity of an early colon screening.

Clinical guidelines based on scientific evidence recommend screening with colonoscopy which allows the identification and removal of polyps, which are the first step in the sequence 'adenoma'–'high grade dysplasia'–'carcinoma', leading to the development of CRC [3, 4]. Conventional colonoscopy is currently the most precise procedure for early CRC detection and prevention, demonstrating several prospective studies with a large number of cases that conventional colonoscopy with polypectomy reduces CRC's incidence by 40–90% [5, 6]. However, colonoscopy presents some drawbacks being polyp miss-rate the most relevant -reported to be as high as 22% [7]-, resulting in a lack of total effectiveness of colonoscopy [8]. The miss-rate increases significantly in smaller sized polyps (2% for adenomas ≥ 10 mm versus 26% for adenomas < 5 mm) [9] and this is meaningful in clinical practice because the prevalence of high grade dysplasia among patients with adenomas is related to the polyp size, with figures that vary from 50% in adenomas larger than 10 mm to 16% in adenomas between 6 and 9 mm [10].

Other alternatives proposed for CRC screening are colon capsule endoscopy and CT colonography but they present a lower accuracy for small polyps detection and do not allow to obtain histological confirmation. These data reinforce the need to improve the detection rate of polyps at colonoscopy in order to be able to remove all small adenomas.

In this sense, there has been an increasing effort during last years on exploring the use of intelligent systems to provide additional information to the output of colonoscopy [11], including the task of automatic polyp localization. The contributions of this paper go in this direction and they are two-fold: 1) The introduction of Window Median Depth of Valley Accumulation (WM-DOVA) energy maps as a tool for automatic polyp localization in colonoscopy images, and 2) the definition of the, up to our knowledge, first validation framework over a fully annotated database and over visual saliency maps created from physicians' fixations obtained via an eye-tracker device and state-of-the-art computational saliency models. Two new databases are introduced: 1) *CVC-ClinicDB* which is the only fully annotated database with clinical metadata associated and 2) *CVC-EyeTrackerDB* which contains annotations of polyps throughout all the frames in the sequence and information about physicians' fixations when observing these sequences.

Our polyp localization method takes as input a colonoscopy frame, and it offers as output image an energy map in which high energy concentration is related with high likelihood of polyp presence. Our method is based on a model of appearance for polyps, which characterizes polyp boundaries by using valley information. The energy maps obtained by our method are validated in two ways: 1) We check whether the maximum of the maps coincides with the polyp location in the image, and 2) we assess whether their energy distribution has correspondence with visual saliency models from physicians fixations and state-of-the-art saliency models. As our method is based on valley information associated to polyp boundaries, our model also considers the impact of different elements of the scene with valley information associated. In the next subsection we make a review of existing methods on the characterization of different elements of the endoluminal scene.

1.2. Related Work on Endoluminal Scene Characterization

The different elements that appear on the endoluminal scene are: 1) Lumen; 2) Wrinkles and Folds; 3) Blood vessels; 4) Polyps; 5) Fecal content, and 6) Specular highlights [11]. We focus our review on available methods for polyp localization, but also on two other elements with valley information associated: specular highlights and blood vessels.

1.2.1. Polyps

The majority of the works related to polyp characterization are based on the calculation of a certain feature descriptor over the image to support the final decision about polyp presence. Existing approaches are divided in two groups: 1) Shape-based; and 2) Texture and color-based. Shape-based approaches aim to search in colonoscopy frames those specific shapes that polyps commonly have. This group can be divided into two categories: detection by curvature analysis and by ellipse fitting. The first subgroup cluster a series of methods which observe the curvature value of the boundaries observed in the image, which are obtained generally using low-level image processing operations such as gradient, edges or valley information. Methods belonging to this category include the works of [12], [13] or [14]. The second group of methods assumes that polyps tend to have elliptical shape. Examples belonging to this group can be found in the work of [15] or at [16]. Finally some other works like the one proposed in [17] combine both curvature analysis and shape fitting to detect polyps in the image whereas other methods [18, 19] base their localization of polyps in the pure analysis of low-level features of the image such as valleys or edges. The use of texture and color-based approaches in polyp characterization has been gaining interest during the last years. Some works are based on the use of wavelet descriptors -such as [20] - paying attention on the detail and approximation coefficients of the wavelet transform. Other

alternatives include the use of local binary patterns [21] or co-occurrence matrices [22]. Finally an interesting approach is the use of several texture and color MPEG-7 descriptors to build up polyp detection methods [23]. One big drawback of this group of methods is that they tend to need of an exhaustive training and they are very sensitive to parameter tuning.

We propose in this paper our Window Median Depth of Valleys Accumulation method (WM-DOVA), which can be enclosed into the shape-based group. Although WM-DOVA shares similarities with some available methods in the use of low-level image processing [12], our method is the only one which also searches for concavity of boundaries and considers other elements of the scene such as blood vessels and specular highlights. WM-DOVA also admits variable polyp shapes, unlike other methods [15, 16] which are specifically focused on elliptical shapes. In addition, WM-DOVA also integrates color information for discarding those valleys originated by blood vessels.

One big problem when comparing different approaches is the lack of a public fully annotated database [18]. Nevertheless none of the published methods consider clinical impact of polyp localization, since the databases used in the experiments do not convey any kind of clinical information regarding polyp size or type. In order to solve this we contribute in this paper with the introduction of the ²*CVC-ClinicDB* database, built in collaboration with Hospital Clinic of Barcelona, Spain.

1.2.2. Specular highlights

The output of a valley detector gives response both inside the specular highlight and in a neighborhood of pixels around it, impacting the performance of algorithms relying on valley information such as the one we propose. In order to mitigate their impact two different operations must be performed: 1) Detection of specular highlights, and 2) inpainting, defined as the substitution of the specular highlight area to avoid the apparition of valley information.

We find several related works in the literature regarding specular highlights detection, such as the work of [24] which assign sudden color changes to highlights, or the works of [25, 26, 27], which exploit the definition of specular highlights as a combination of diffuse and specular reflection. Other methods offer a combined solution for detection and inpainting, although some of them [25, 26] are not suitable for our case since there is no non-overexposed channel from which to extract information needed for inpainting. Finally, there are two works that present a solution for both detection and inpainting in colonoscopy images [28, 19] with high reported accuracy against general methods.

² CVC-ClinicDB is currently available at http://www.polyp2015.com/wp/?page_id=141 and at <http://mv.cvc.uab.es>

1.2.3. Blood vessels

The majority of the literature regarding the presence of blood vessels is related to their segmentation and consequently different methods have been proposed [29] in non endoscopy domains. In our particular case we are not concerned on blood vessels segmentation but on their contribution in terms of valley information, and taking this into account, we opted to study if their impact changed according to the color space we were working with, as a way to mitigate their presence. Current experimental results [19] show that by using the channel B of the standard RGB image we both enhance polyp-related valleys and mitigate blood vessels valleys.

2. Windows Median DOVA Energy Maps based on a Model of Appearance for Polyps

We present in this section WM-DOVA energy maps which are based on an extended model of appearance for polyps.

2.1. Model of appearance for polyps

In order to build up our model we consider both the physical model of a polyp along with how colonoscopy images are acquired. In our model we consider polyps as abnormalities that appear as protruding surfaces on the colon wall. Polyps create a sloping surface in their union with the colon wall.

This surface generates a contour around the polyp which is meant to delimit the polyp from the colon wall. The contour can be approximated as a closed and continue curve with smooth curvature profile. Taking this into account, we can approximate polyps as spheres -with some degree of deformation embedded on the colon wall.

Regarding image acquisition, colonoscopes have a light source and a camera, both attached and pointing in the same direction towards the colon walls. This can be approximated as illuminating a tubular surface with an axial light. Thus, surfaces with their normals oriented in the same direction as the pair light-camera appear brighter as they reflect the incident light towards the camera. Conversely, as we deviate from this situation, the amount of light reflected decreases [18].

Given this model, we consider three different scenarios: 1) Optimal scenario (with respect to model fitting), with a complete zenithal view of the polyp -see Figure 1 (a)-; 2) Semi lateral view of the polyp -see Figure 1 (b)- and 3) Lateral view of the polyp -Figure 1 (c)-.

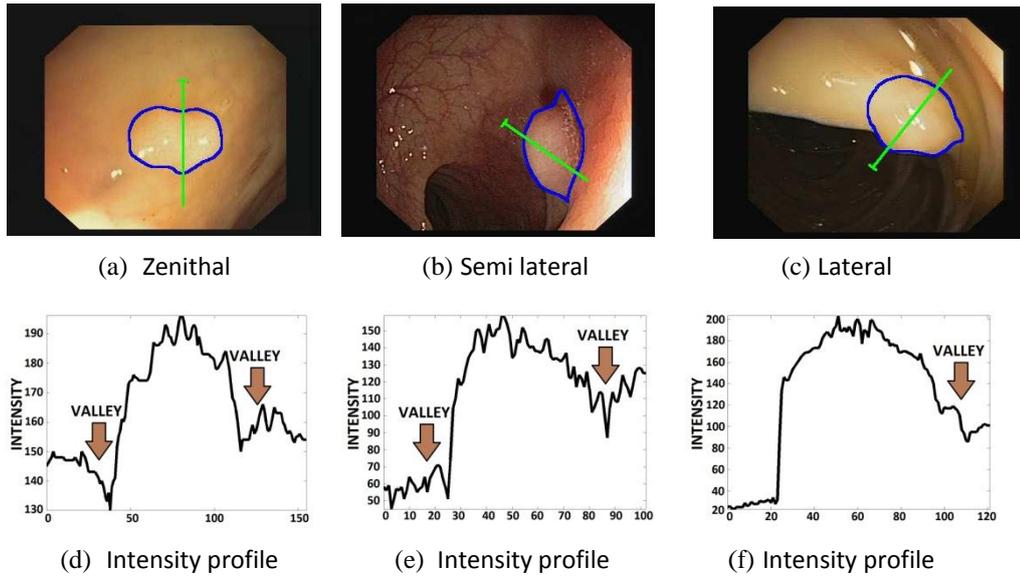


Figure 1: Examples of the three possible scenarios that need to be considered when defining a model of appearance for polyps: (a-d) Zenithal view; (b-e) Semi lateral view; (c-f) Lateral view. A green line is superimposed in the original images to highlight the area of the image to be studied.

In zenithal view, the normal at the polyp center shares orientation with normals on the colon wall so polyp surface will appear with high intensity value due to maximal reflection. Conversely, polyp frontier appears dark since its normal deviates from the wall's normal. If we move from the center of the polyp towards its boundary, the corresponding intensity profile will constitute a valley, as can be observed in Figure 1 (b). Consequently, polyps in zenithal views appear as surrounded by intensity valleys.

As we deviate from the optimal scenario, the orientation of the pair light-camera diverts from the orientation of the normal on the colon wall -semi lateral view, Figure 1 (c)-. The worst possible scenario will happen when the camera looks laterally at polyps while looking in parallel to the colon axis -lateral view, Figure 1 (e)-. Even in semi lateral and lateral views valley information is still present. In semi lateral views valleys are present in both extremes of the polyp profile as long as some part of the colon wall can still be observed. In lateral views, the valley behind the polyp is substituted by a steep contour which separates the lumen from the polyp -Figure 1 (f)-.

In conclusion, our model implies that polyps are surrounded completely by valleys in zenithal views and partially when they are viewed laterally. Valleys surrounding the polyp will constitute a closed curve which can be partial or completely closed. This curve will appear as concave from inside the polyp. The intensity and strength of the valley depends greatly on the direction of the illumination and on how protruding is the polyp. Considering this, according to Paris classification criteria [30] our model fits completely to polyp appearances of types Is, IIa, IIb, and Ip polyps in zenithal view, and partially to polyps in lateral and semi lateral views.

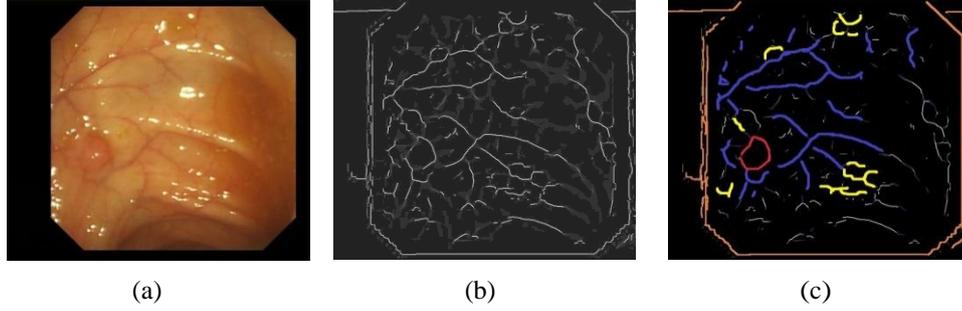


Figure 2: Valley information sources: (a) Original image; (b) Valley image; (c) Manually- marked valley image. Marked valleys are from polyps (red), blood vessels (blue), specular highlights (yellow) and black mask (orange).

Unfortunately polyps are not the only source of valley information (see Figure 2). Both specular highlights and blood vessels convey valley information that should be treated properly in order not to have an impact in polyp localization results. Image preprocessing is performed to mitigate impact of non-polyp valleys. Preprocessing is also used here to eliminate some image artifacts that are caused by the image acquisition system. We apply four different preprocessing operations: 1) Image deinterlacing to remove artificial valleys created by time difference in the acquisition of the odd and even lines of a frame; 2) Specular highlights detection and inpainting; 3) Black mask inpainting, and 4) Blood vessels mitigation by using channel B of sRGB image. The interested reader can find more details about these operations in [19].

Once image preprocessing operations are applied to the colonoscopy frame, we need to obtain the valley information which will be used to characterize polyp boundaries. Consequently, the use of a valley detector is key in our localization scheme. There are several approaches for valley detection in the literature but we opt to use a geometrical valley detector [31] which adjusts better to the structures we are searching in the image. This valley detector succeeds at localizing the valleys in the image but it does not give enough information of the depth of the valley, i.e., the strength of its contribution.

To cope with this, we developed the Depth of Valleys image (DV image) [18] which combines valley information -from the valley detector [31] - with gradient information -via morphological gradient-. The DV is achieved through the pixel-wise multiplication of the ridges/valleys detector and morphological gradient in the following way:

$$DV = V(\sigma_d, \sigma_i)MG = V(\sigma_d, \sigma_i) (I \oplus C_{\sigma_i} - I \ominus C_{\sigma_i}) \quad (1)$$

, where V stands for the output of valley detector, I for the original input image and C for the disk structuring element used in the dilation (\oplus) and erosion (\ominus) morphological operations. Formulated in this way, DV image has higher values in the points that constitute the relevant valleys of the image. It must be noticed that, in order to calculate the morphological gradient, the ridge and valley extractor needs two parameters -differentiation σ_d and integration scales σ_i - that must be set in correspondence to the size of the structuring

element (s_d) [31]. More precisely, σ_i should span the same size as s_d to both work in the same scale.

2.2. Generation of WM-DOVA Energy Maps

The basic idea of our localization method is that, as our model defines polyps as surrounded by intensity valleys, any given pixel in the image inside a valley structure is prone to be part of the polyp. WM-DOVA energy maps are created to integrate valley information in a way such polyp boundaries are fostered. High energy areas of WM-DOVA energy maps are meant to be related with areas with high likelihood of polyp presence. In order to favor polyp boundaries against other structures in the image, WMDOVA applies four different constraints to boundary candidates, namely: Completeness, robustness, continuity and concavity.

1) Completeness: Our model states that polyp boundaries tend to be complete, that is, a pixel inside the polyp should be surrounded by valleys in the majority of directions. In order to favor complete boundaries from those which only have contribution in a few directions we use the following approach: we define a ring of radial sectors centered on each pixel in the image and we accumulate DV values for all the sectors. As our DV image has been defined to enhance deep valleys, we only accumulate the maximum of DV image under each sector. In this way we are fostering boundaries composed by pixels with high DV value which have contribution for a high number of directions. We can see an example of this in Figure 3 (a-b), where we can observe that the maximum of accumulation is higher for a more complete boundary than for the incomplete one.

2) Robustness against spurious responses with high DV value: Our system should discard contributions from boundaries composed by pixels with high DV value concentrated in a few directions against contributions from boundaries composed by pixels with lower DV value comprising the majority of directions. To achieve this, we apply a median operation to the final value to be accumulated. By doing this, we penalize those undesired boundaries such as the one shown on the right part of Figure 3 (c-d). We can observe in Figure 3 (d) that the accumulation is higher for the circular (complete) boundary, although it is composed by pixels with lower DV value.

3) Continuity: WM-DOVA energy maps should favor continuous boundaries against discontinuous. This is achieved by the use of an *angular median* operator Med . The angular median calculates the median of the distances from the maximum of DV under each sector to the center of the ring of sectors, for a given number of sectors $2w_s + 1$. This operator favors the contributions of consecutive sectors having the maximum of DV at a similar distance to the center. We can see an example of how this is fostered by observing Figure 3 (e). We can observe how the synthetic boundary on the right

is composed by two incomplete boundaries which do not have a maximum of DV at the same distance to the center. By applying the angular median, the maximum of accumulation for this boundary will be very low, as the median of distances to the center between two adjacent parts of the two boundaries will be between both of them, hence mitigating their contribution to the final accumulation value, as it can be seen in Figure 3 (f).

4) Concavity: The final condition that our model imposes is that boundaries should be locally concave from the point of view of a pixel inside the polyp, which implies that if we take a window of sectors comprising several directions, the distance from the boundary candidate for a given direction to the median of distances under the window should be small. We use the median of the distances to the center as exposed before but in this case we calculate a pondering factor which will modify the contribution of maximum of DV under each sector. This pondering factor will depend on the distance from the maximum under the current sector to the median of the distances for the considered window of sectors. By applying this pondering factor we favor boundaries such as the ones shown in the left part of Figure 3 (g), as in this case all the pixels constituting the boundary have a more similar distance to the center than those in the right part of the Figure. WM-DOVA energy maps will be consequently more focused on concave boundaries than to those locally convex, as can be observed in Figure 3 (h).

By imposing the four mentioned constraints we help also to mitigate the impact of other elements of the endoluminal scene. The definition of the window of radial sectors helps to mitigate response of blood vessels or folds as its contribution in terms of DV image will happen only in a few directions and without concave profile. The use of the median in the accumulation has also effect in the removal of the contribution of spurious valleys that can appear due to non-homogeneous illumination.

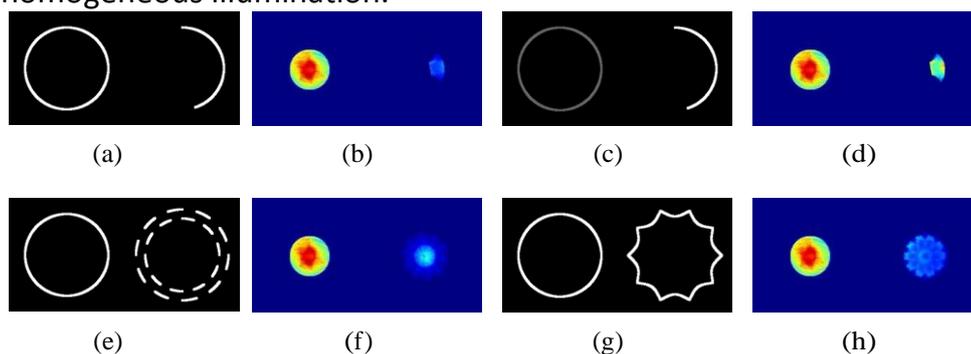


Figure 3: Example of use of WM-DOVA energy maps to foster: (a-b) Completeness in boundaries; (c-d) Sensitivity to noise; (e-f) Continuity and (g-h) Concavity. Image (a,c,e,g) show original synthetic images whereas images (b,d,f,h) represent WM-DOVA energy maps. Pixels with high DV value are marked in white in the original image. Hotter areas in the map correspond with zones with high accumulation value.

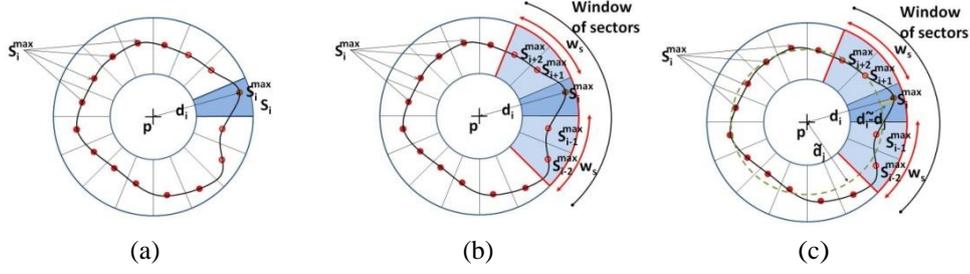


Figure 4: WM-DOVA algorithm: (a) Calculation of S_i^{max} under S_i ; (b) Definition of window of sectors and calculation of S_n^{max} of sectors under window; (c) Obtention of distances d_n from S_n^{max} to center p and calculation of \hat{S}_i^{max} .

As mentioned above, WM-DOVA defines a window of $2w_s + 1$ sectors which integrates the information of neighboring ring sectors fostering in this way concavity and continuity. To ease the comprehension of the WM-DOVA algorithm we will support on the graphical example shown in Figure 4, in which the black contour is a binary approximation of the DV output for a polyp contour. For each pixel p in the image, the WM-DOVA algorithm consists of 4 different steps:

1. Calculation for each sector S_i of the following: a) $S_i^{max} = \max_{S_i}(DV)$ as the maximum of DV under sector S_i ; b) $p_{S_i^{max}} = \text{argmax}_{S_i}(DV)$ as the position of the maximum of DV under sector S_i ; and c) $d_i = \text{dist}(p, p_{S_i^{max}})$ as the distance between current pixel p and $p_{S_i^{max}}$.
2. Definition of a window of width $2w_s + 1$ sectors centered on S_i (Fig. 4 (b)).
3. Calculation of the contribution of S_i^{max} in the final accumulation value. To do so we first obtain $\tilde{d}_i = \text{Med}(d_j)$, with $j \in [i-w_s, i+w_s]$, as the angular median of the distances d_j to p for the window of sectors centered on S_i (Fig. 4 (c)).

Then we calculate a pondering factor γ_i which is calculated as follows:

$$\gamma_i = \frac{1}{1 + \frac{|d_i - \tilde{d}_i|}{\tilde{d}_i}} \quad (2)$$

, with $\gamma_i \in [0, 1]$. γ_i is a mitigation coefficient which penalizes isolated high DV values which break the continuity of the contour. On the one hand, if the distance $p_{S_i^{max}}$ from the current pixel to S_i^{max} is similar to \tilde{d}_i , there will be no mitigation. On the other hand, in cases like the one shown in Fig. 4 (c) where $p_{S_i^{max}}$ is substantially higher than \tilde{d}_i , the corresponding S_i^{max} value will be mitigated, as γ_i will be low. The final contribution of S_i^{max} to the final accumulation value is calculated as follows:

$$\hat{S}_i^{max} = \gamma_i S_i^{max} \quad (3)$$

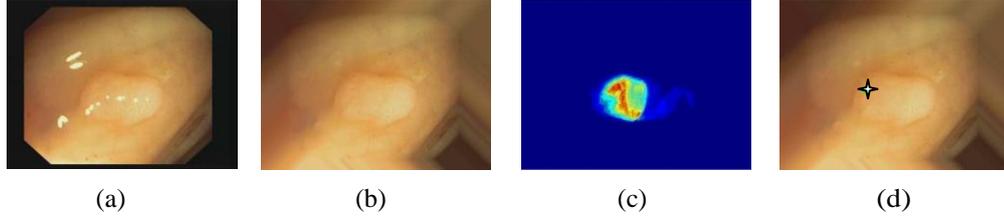


Figure 5: WM-DOVA energy maps generation: (a) Original image; (b) Fully preprocessed image; (c) WM-DOVA energy map and (d) Output image with maximum of WM-DOVA energy map superimposed and marked as a star. Energy maps are shown as heat maps where hotter areas correspond to areas with high likelihood of polyp presence.

WM-DOVA also favors concavity since for the case of locally circular contours the line that joins the maxima of DV under the sectors of the window tends to coincide with the boundary (particularly true for polyp contours).

4. Calculation of the final accumulation value for pixel p as: $Acc_p = \text{Med}_i(\widehat{S}_i^{max})$, with $i \in [1, n_s]$.

WM-DOVA needs of three different parameters: 1) Minimum radii of the sectors ($radmin$); 2) Maximum radii of the sectors ($radmax$), and 3) Width of the windows of sectors (w_s). The number of sectors n_s has been set to 180 aiming to cover all the needed different directions. Regarding the scale at which WM-DOVA works, we want to stress that parameters $radmin$ and $radmax$ are not actually defining a particular scale but a range of scales. This allows the method to detect both small and large polyps. Should $radmin$ be smaller, directional information would be lost due to the smaller number of pixels in the lower part of the sector.

As WM-DOVA has been built according to our model of appearance of the polyps, the maxima of the WM-DOVA energy map should fall inside the polyp. Our polyp localization processing scheme therefore consists of preprocessing the input image by applying the different operations explained in Section 2.1, and then DV image is calculated. Finally WM-DOVA energy maps are generated. We can see a complete example of the different steps in Figure 5.

3. RESULTS

We present in this section a double validation of WM-DOVA energy maps. The first validation is focused on assessing the performance of WM-DOVA as part of an automatic polyp localization method. WM-DOVA energy maps are tested on two different databases, including a new one with clinical metadata associated and results are compared with previous available methods. The second validation aims to assess whether the WM-DOVA energy maps

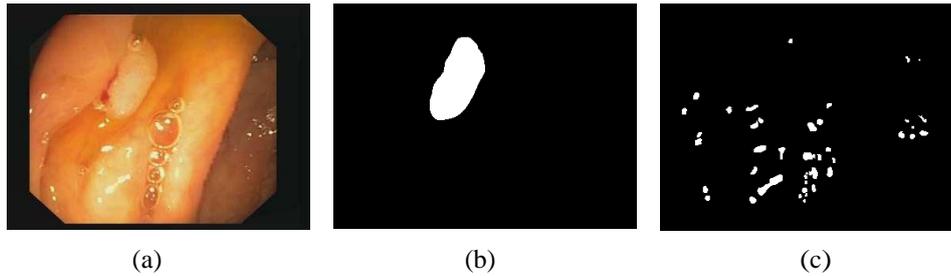


Figure 6: An example of the content of CVC-ClinicDB database: (a) Original image; (b) Polyp mask; (c) Specular highlights mask.

distribution has correspondence with visual saliency models from physicians fixations and state-of-the-art saliency models.

3.1. Polyp localization results

3.1.1. Databases used in the experiment

In order to validate the outcome of our polyp localization method we need a ground truth defining the position of the polyp in the image. Up to our knowledge there is only one public fully annotated database, (*CVC-ColonDB*) [18, 32]. Unfortunately *CVC-ColonDB* did not contain any kind of clinical metadata.

We introduce in this paper the *CVC-ClinicDB* database built in collaboration with Hospital Clinic of Barcelona, Spain. *CVC-ClinicDB* has been generated from 23 different video studies from standard colonoscopy interventions with white light. For each study all the sequences containing a polyp were extracted, providing a total number of 31 sequences of 31 different polyps. We rejected all the frames with extremely high patient preparation or bad visualization quality due to image blurring. In this way, 31 frame sequences with an average number of 25 frames were obtained; paying particular attention that for each sequence we obtain as many different polyp appearances as possible. *CVC-ClinicDB* database comprises 612 polyp images of size 576×768 .

In addition to the frames, a ground truth was created by the experts by manually defining a mask on the region covered by the polyp. To assess the impact of image preprocessing in polyp localization results, experts also provided ground truth for specular highlights. We can see an example of the frames of our database and their corresponding ground truths in Figure 6.

Clinicians also provided the following clinical metadata associated to each polyp:

- Polyp size in mm: polyps are clustered in three categories according to [7]: a) Diminutive (≤ 5 mm); b) Small ($6 \text{ size} \leq 9$ mm); c) Large (> 10 mm).
- Polyp classification according to Paris criteria [30]: 6 groups are defined: Ip (protruded, pedunculated), Is (protruded, sessile), IIa (superficial, elevated), IIb (flat), IIc (superficial shallow, depressed) and III (excavated).

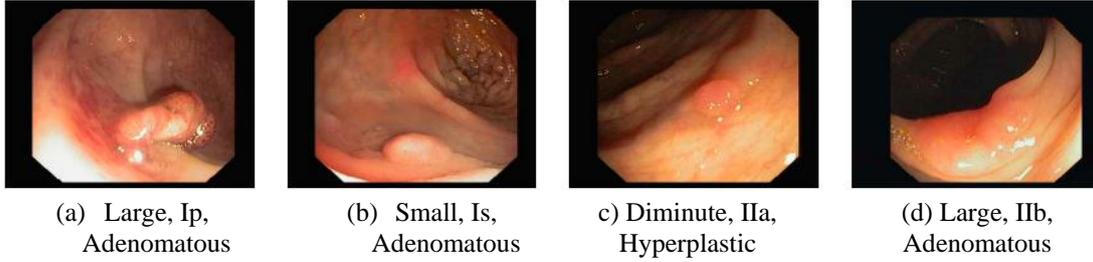


Figure 7: Example of polyps present in CVC-ClinicDB database along with their categorization.

Polyps of categories IIc and III have a low prevalence and they were not found in the interventions recorded when creating the database.

- Histological type of polyp after biopsy: in this case polyps have been subdivided into two groups: Adenomatous -488 images (79.74%)- and Hyperplastic -124 images (20.26%)-.

We show in Figure 7 an example of some polyps from the database categorized according to the metadata provided.

3.1.2. Metrics and parameters of the experiment

The following performance metrics are defined for this experiment:

1. Accuracy: The accuracy of the maximum of WM-DOVA to localize the polyp is calculated as $Acc = \frac{TP}{TP+FP}$. For each frame, we have a True Positive -TP- when the maximum of WM-DOVA energy map is inside the polyp mask. If the maximum falls outside the mask we have a False Positive -FP-.
2. Concentration Ratio: -CR- as the ratio energy of the image that falls under the polyp mask. CR is defined as: $100 \frac{E_{polyp}}{E_{frame}}$, where E_{polyp} corresponds to the total energy under the polyp and E_{frame} corresponds to the total energy of the image. A high CR value will correspond to a map focused on the polyp whereas a low CR value will denote a sparser energy map.

Our sampling size (612 for the case of *CVC-ClinicDB*) warrants that accuracy in localization is estimated with a variability under $\frac{1}{\sqrt{N}} = \frac{1}{\sqrt{612}} = 4\%$. A variability of less than 10% is not considered as relevant in clinical trials [33].

WM-DOVA parameters were set to $radmin = 10$ and $radmax = 120$ considering the minimum and maximum radii of the polyps -in pixels- that clinicians observe during intervention time. $radmin = 10$ implies a circumference of 60 pixels and $radmax = 120$ implies a circumference of 720 pixels. Being $n_s = 180$ sectors, we are already sharing 1 pixel between 3 sectors in the lower part and we are using 4 pixels per sector in the upper part of the sector. If $radmin$ becomes smaller the meaning of directions is lost. This ranges polyps from 0.3%

Method	Acc in CVC-ClinicDB	Acc in CVC-ColonDB
EF-DOVA	43.30% (265 images)	46.00% (138 images)
VO-DOVA	45.59% (279 images)	55.44% (166 images)
SA-DOVA	50.65% (310 images)	61.66% (185 images)
MSA-DOVA	57.03% (349 images)	67.66% (203 images)
WM-DOVA	70.26% (430 images)	72.33% (265 images)

Table 1: Polyp localization results in CVC_ClinicDB database (612 images and CVC-ColonDB (300 images).

of the image area to 41% of the image area. In this sense, our method actually could fail in highlighting polyps occupying more than 40% of the image size, which is a clinically irrelevant case. Finally, the optimal value for w_s was tuned on a validation set of 30 images from the same videos but not belonging to the final database with a final value of $w_s = 18$, being the total window size of 37 sectors comprising 74° .

3.1.3. Experimental Results

We present in Table 1 a comparison of our WM-DOVA polyp localization method against previous available methods using valley information, which represent all the different categories of shape-based methods which were exposed in Section 1.2. For instance EF-DOVA applies ellipse-fitting to boundaries whereas VO-DOVA exploits curvature of boundaries.

Regarding CVC_ClinicDB, we can see in Table 1 how WM-DOVA offers the best performance with an accuracy of 70.26% -430 images-, outperforming the rest of approaches. We can see that WM-DOVA improves results of EF-DOVA (which uses ellipse fitting) in 165 images which confirms our original hypothesis that constraining our method to a certain shape could lead to good particular results for those polyps that fit this shape, but a bad overall result. The joint use of completeness, robustness, continuity and concavity lead to an improvement over the rest of the approaches, surpassing its closest competitor MSA-DOVA in 80 images. We can observe the same trend for *CVC-ColonDB* although in this case the performance of all the methods increases. Performance differences are related to being *CVC-ClinicDB* a more complete and difficult database with appearances that does not exist in *CVC-ColonDB* as it almost triples the number of different polyps it contains (31 against 12).

Size category	# of Frames in CVC-ClinicDB	Correct localization
Diminutive (size ≤ 5 mm)	268 (43.79%)	172 (64.17%)
Small ($6 < \text{size} \leq 9$ mm)	162 (26.47%)	136 (83.95%)
Large (size ≥ 10 mm)	182 (29.73%)	122 (67.03%)

Table 2: Polyp localization results according to polyp size.

Paris Classification	# of Frames in CVC-ClinicDB	Correct localization
Ip	174 (28.43%)	119 (68.39%)
Is	208 (33.98%)	136 (65.38%)
Ila	168 (27.45%)	121 (72.02%)
Ilb	62 (10.13%)	54 (87.09%)

Table 3: Polyp localization results according to Paris Classification.

We present a breakdown of polyp localization results using WM-DOVA as energy map in Table 2 and Table 3. As it can be seen from Table 2, our localization method works better for polyps smaller than 10 mm, which are the largest group in our database and also, as mentioned in Section 1.1, those cited by clinicians as the main cause for the miss-rate. Regarding Paris classification we can see in Table 3 that our method seems to be particularly well suited for Ilb and Ila polyp types, which is expected considering our model of appearance for polyps. Concerning the histology of the polyp after biopsy, WM-DOVA is more successful -72.54% against 61.29%- when locating adenomatous polyps, which comprise the larger number of frames in our database.

One important thing to mention is that WM-DOVA is able to locate the polyp present in all the 31 sequences of 31 different polyps that *CVC-ClinicDB* comprises, showing the robustness of our method even considering the great variability of polyp appearances observed in the database.

For the case of the concentration ratio, we present in Figure 8 boxplots that show, for each algorithm, the distribution of the ratios of energy inside the polyp mask. We can observe that there is indeed a significant difference in performance between WM-DOVA and the rest of the methods. More precisely, the mean CR value inside the polyp mask is 89.28% for WMDOVA, followed by MSA-DOVA with a 27.01%, SA-DOVA with a 22.61%, VO-DOVA with a 19.55% and EF-DOVA with a 16.09%. Consequently, the maximum of WM-DOVA energy maps does not only localize correctly the polyp in more images but WM-DOVA energy maps also concentrate their energy more densely inside the polyp mask.

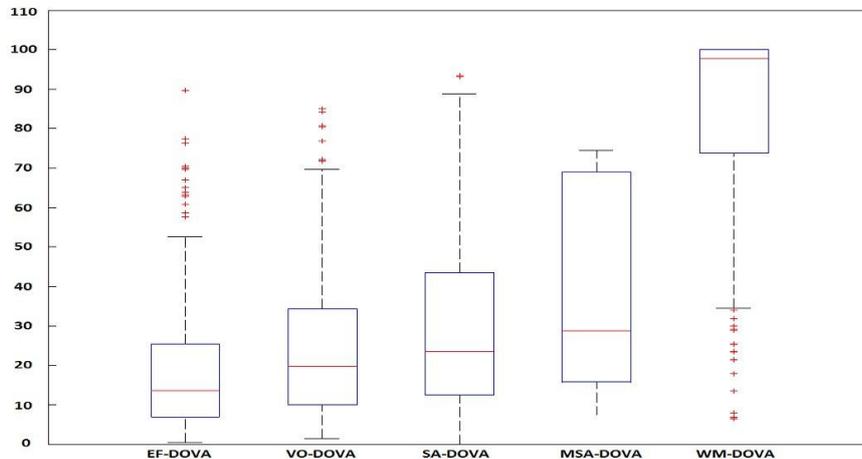


Figure 8: Comparison of the percentage of energy inside the polyp mask among existing DOVA energy maps

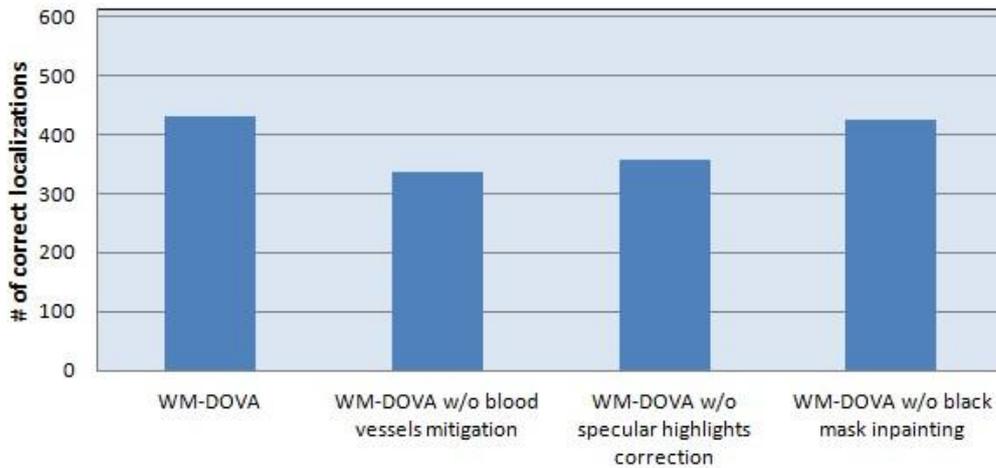


Figure 9: Impact of image preprocessing on polyp localization results. Total number of images = 612.

As mentioned in Section 2.1, we tackle in our image preprocessing stage the impact of other elements of the endoluminal scene in terms of valley information. We present a summary of the results in Figure 9. We can see in this case that blood vessels mitigation shows to have a relevant impact on polyp localization results, obtaining a difference in terms of successful polyp localization in 93 images (more than a 15% difference). We also improve accuracy results in polyp localization in 72 images (11.76%) by tackling specular highlights correction. Finally black mask inpainting has no impact on the results which can be related to the fact that physicians inherently tend to place polyps away from the borders during the colonoscopy intervention. Image preprocessing has been proven to have an impact in the overall performance of our localization method. We show in Figure 10 an example of the impact of image preprocessing on polyp localization.

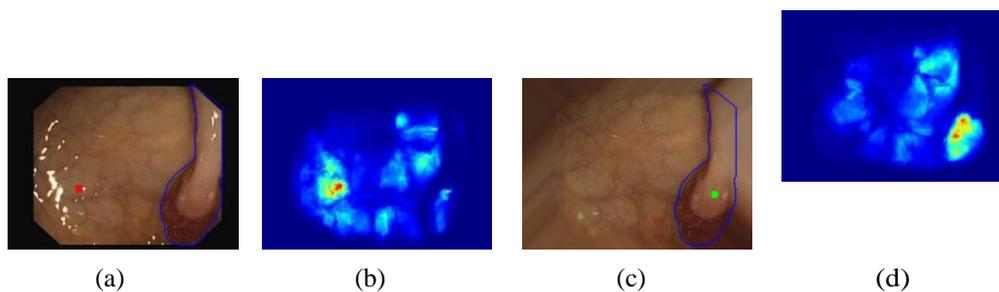


Figure 10: Impact of image preprocessing on polyp localization results: (a) Original image; (b) WM-DOVA energy map; (c) Preprocessed image; (d) Corresponding WM-DOVA energy map; Maxima of WM-DOVA are painted as a green square when they fall inside the polyp mask and as a red square if they fall outside. The polyp contour is painted in blue.

3.2. Experimental results on validation of WM-DOVA energy maps as saliency models

The aim of this section is to validate the potential of WM-DOVA energy maps as saliency models. In order to assess the applicability of WM-DOVA in clinical environments, we will compare its performance when searching for polyps against physicians' visual attention models which are obtained by integrating fixations gathered by means of an eye-tracker device. In this case, if the maxima of WM-DOVA maps also correspond with the areas that drive the attention of the physicians during the visual search for polyps, we can assess that our system not only localizes polyps in an accurate way but also identifies the same potential regions for polyp presence by performing -seeing- in the same way as a physicians do. In this context, WM-DOVA can also be seen as a saliency model considering that polyps are meant to be the most salient part of a colonoscopy frame. Considering this we also compare WM-DOVA with state-of-the-art computational saliency methods aiming to confirm the hypothesis that our method is able to capture better the saliency of a colonoscopy frame in the context of polyp localization.

3.2.1. Use of task-driven saliency maps to integrate physicians' fixations

WM-DOVA energy maps correspond with a bottom-up approach that highlights image regions that are associated to a high likelihood of polyp presence. Unlike general methods for saliency detection -with different degrees of bottom-up [34, 35, 36, 37, 38, 39] or top-down implementation [40, 41] - our approach appears especially tuned to the model of appearance associated for polyps [18].

Saliency maps modeling the physicians' visual attention were obtained from the fixations acquired by eye-tracking the physicians during the screening of colonoscopy sequences under the task of finding polyps. Maps generated in this way do not only take into account the low level saliency of the images but also involve high-level cognitive factors related with the task, that turn into relevant certain areas of the image during the screening process [42].

The work presented in [43] shows how fixations can be integrated into task-driven saliency maps. In this case, the saliency map is represented by a fixation density map in which the measures obtained by different subjects are integrated into the density map to create an average subject. The fixation density map is then created from a set of discrete points (x_i^f, y_i^f) , $i = 1, \dots, N$, where N is the total number of fixations in a frame and (x_i^f, y_i^f) is the location of the i -th fixation point. Those fixations correspond with the centers of Gaussian functions that generate the fixation density map $s(x, y)$:

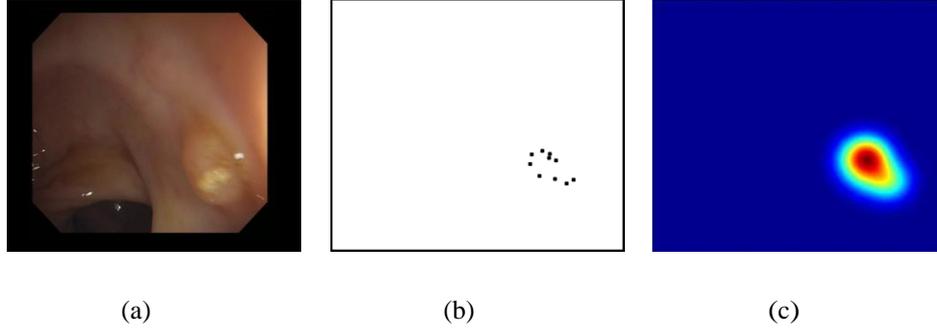


Figure 11: (a) Original image; (b) Physicians' fixations; (c) Physicians' saliency map

$$s(x, y) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2\pi\sigma_s^2} \exp\left(-\frac{((x-x_i^f)^2 + (y-y_i^f)^2)}{2\sigma_s^2}\right), \quad (4)$$

where σ_s is the standard deviation of the Gaussian function, determined according to the visual angle accuracy. A pixel in a densely populated fixation area is more attractive to the attention than a pixel in a sparsely populated area. An example on physicians-related task-driven saliency maps is show in Figure 11 (b) and (c).

3.2.2. Databases used in the experiment

As our objective is to compare computational saliency models –including our proposal WM-DOVA- against physicians' task driven saliency maps, we need to define a ground truth related to physicians' attention model which was built in the following way: a series of experiments were run in association with 22 physicians from Beaumont Hospital and Saint Vincent's Hospital in Dublin, Ireland. The experiments consisted of showing, to each physician, 11 different videos and, by using an eye tracking device, acquiring the gaze position for each frame. We show in Table 4 a summary of the number of interventions done by the different physicians with different degrees of expertise. Comparison studies between subjects of different level of expertise (experts vs. novices) have been carried out in several domains such as threat assessment [44] and on the analysis of fixations during a laparoscopic operation [45]. The task of labeling an individual to the expert or novices cluster depends on the specific application. In our case, the threshold number of interventions used to divide physicians into

Physician	1	2	3	4	5	6	7	8	9	10	11
Procedures	50-100	2000	5	0	600	800	100+	400-500	0	2000	600
Physician	12	13	14	15	16	17	18	19	20	21	22
Procedures	20000	80	1250	200	8	200	20	80	10	500	50

Table 4: Number of interventions performed by each physician.

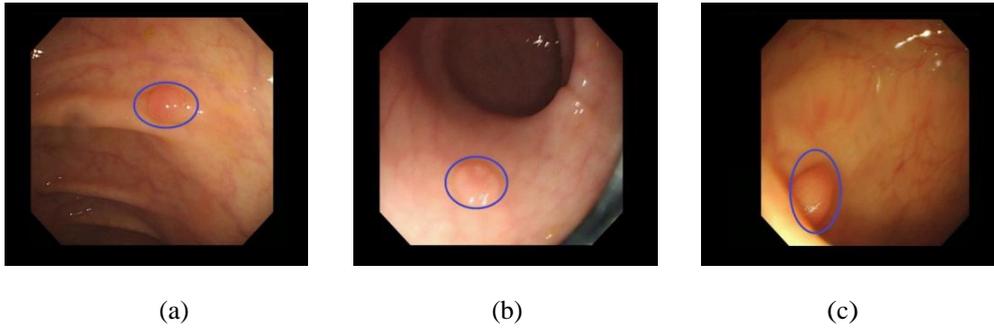


Figure 12: Examples of content of CVC-EyeTrackerDB.

experts or novices was set to 100, as suggested in [46] and we use these clusters in the comparison with WM-DOVA and the rest of computational saliency models. Using these two clusters, an *average expert* and *average novice* task-driven saliency map was generated by integrating experts and novices fixations as explained in the previous section.

In order to carry out the two different analyses proposed we have created *CVC-EyeTrackerDB* database from the above mentioned 11 different videos [47]. *CVC-EyeTrackerDB* comprises all the frames with a polyp from each video and two different kinds of ground truth: one is composed by the fixations of all the physicians that observed each particular frame of the video whereas the second is an estimation of the polyp mask. In this case, due to the huge amount of frames available (around ten thousand), experts did not provide exact polyp masks but an elliptical approximation which aims at covering as much of the polyp as possible. We can see some examples of the ground truth superimposed to original frames in Figure 12.

3.2.3. Metrics of the experiment and statistics

To compare the different saliency models we need to define two different concepts that will be used to calculate performance metrics. We denote as *Polyp Fixation Frame (PFF)* the frame in which the maximum of the saliency map falls under the polyp mask. Consequently, the *First Polyp Fixation Frame (FPFF)* is defined as the first PFF in a given sequence. Considering this we define the following metrics:

- *Reaction Time (RT)*: difference in number of frames between the first frame of the sequence with a polyp and FPFF. We would like to note that the lower RT value is, the better diagnostic yield we have.
- *Dwell Time (DT)*: percentage of PFF out of all the frames with a polyp in a sequence. In this case better diagnostic yield is associated to high DT values.
- *Mean Concentration Ratio (MCR)*: CR corresponds to the ratio energy of the image that falls under the polyp mask, as defined in Section 3.1.2. In this case we define MCR as the mean CR value for all the frames of a

given sequence with a polyp. A high MCR value will correspond to energy maps focused on the polyp throughout all the sequence.

Statistic significant differences across methods are assessed using an analysis of variance to detect differences in each of the proposed metrics. Since all metrics are related to time and other quantitative always positive, we do not expect that they follow a normal distribution. Non-normality in data is accounted using non-parametric Kruskal-Wallis (KW) test [48].

For each metric and test, the individuals are given by their value in the 11 videos and ANOVA factors are the different saliency methods, namely: 1) *average expert* ; 2) *average novice*; 3) WM-DOVA; 4) Bruce and Tsotsos (BT) [34]; 5) Itti-Koch (IK) [35]; 6) Graph Based Visual Saliency (GBVS) [36]; 7) Seo et al. [37]; 8) SIM method developed by Murray et al. [38] and 9) SUN method by Zhang et al. [39].

All tests have been done at a confidence level $1 - \alpha = 0.95$. Using such confidence level and this experimental design, tests have an approximate power of 80% to detect differences of approximately 0.45 units of variance. Such magnitudes are appropriate for an exploratory analysis like the ones presented in [33].

3.2.4. Experimental Results

Figure 13 (a) shows boxplots for the distribution of the RT values achieved by each of the *average expert*, *average novice*, WM-DOVA and computational saliency maps. A qualitative visual inspection of the boxplot indicates that the *average expert* has the lowest RT, followed by the *average novice*, WMDOVA, SIM and Zhang. All these methods seem to present comparable times whereas boxplots for the remaining ones are clearly larger. The KW test detects significant differences across RT values ($p - \text{value} < 10^{-6}$) that are explored using the multicomparison test shown in Figure 13 (b). The best ranked group corresponds to the blue horizontal line, comparable groups to grey lines and groups with statistically significant different performances to red lines. This test indicates that there are no significant differences among the *average expert*, *average novice* and WM-DOVA.

Figure 13 (c) shows boxplots for the distribution of DT values. Visual analysis of boxplots indicates that the *average expert*, *average novice* and WM - DOVA have a similar distribution in contrast to the lower range of the remaining saliency methods. It is also worth noticing that DT values distribution are very similar for the *average expert* and the *average novice*, and differences could be attributable to differences in RT. This follows from the fact that, once the polyp is found, clinicians keep focusing their attention inside the polyp for the rest of the sequence. Again, a KW test detects significant differences across DT values ($p\text{-value} < 10^{-6}$), which are explored in the multicomparison test results plotted in Figure 13 (d). Such difference is due to GBVS and Seo being

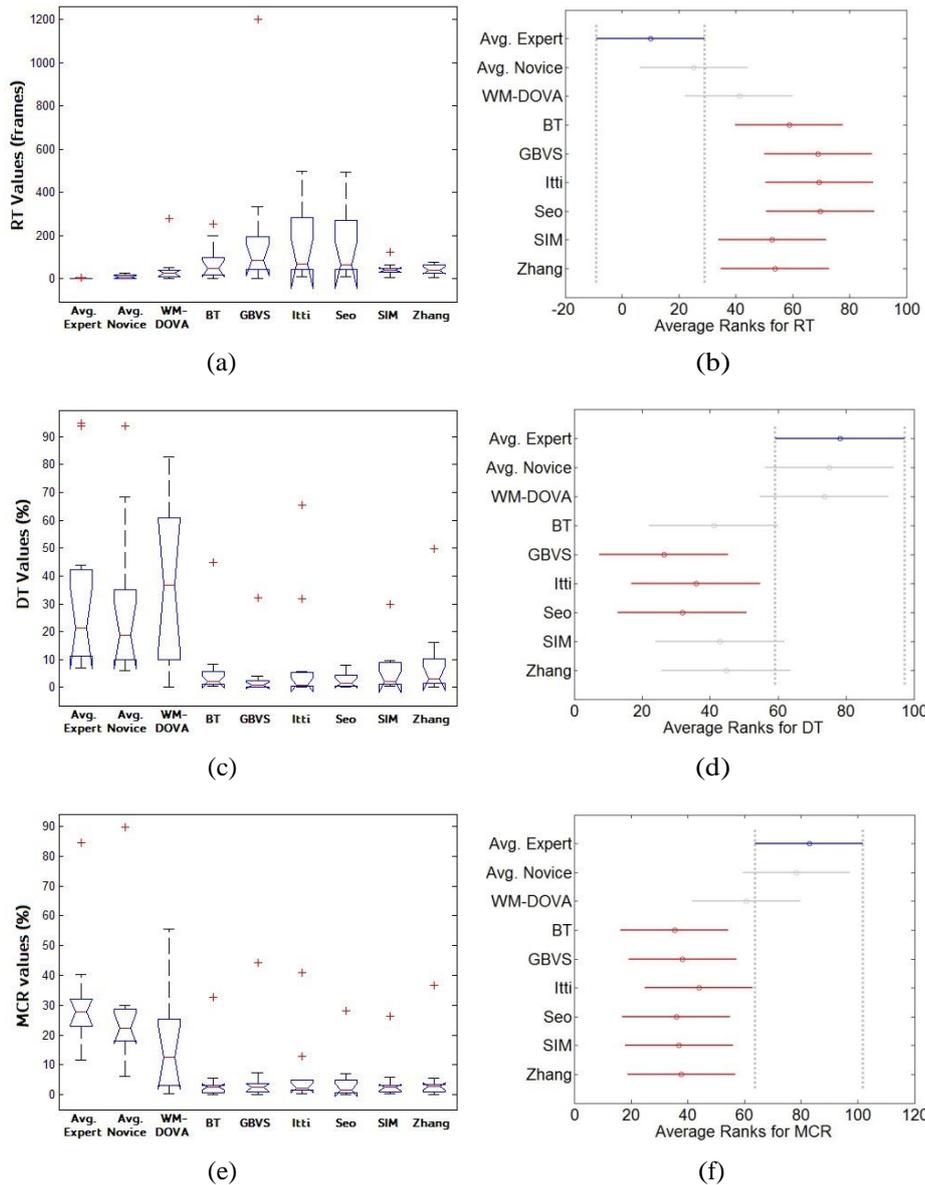


Figure 13: Boxplots showing distribution of: (a) Reaction Time (RT) values; (b) Dwell Time (DT) values and (c) Mean Concentration Ratio (MCR) values. Results of multi-comparison analysis for: (b) RT values; (d) DT values and (f) MCR values. For each of the multicomparison tests, best ranked group is represented with a blue horizontal line, comparable methods in grey and statistically significant different methods in red.

significantly worse than the rest of the methods. Among the remaining groups, it is worth noticing that the *average expert*, *average novice* and WM-DOVA are better than BT, SIM and Zhang.

Boxplots showing the distribution of MCR values are shown in Figure 13 (e). A qualitative analysis again shows that the *average expert* presents the higher MCR value, followed by the *average novice*, WM-DOVA and the

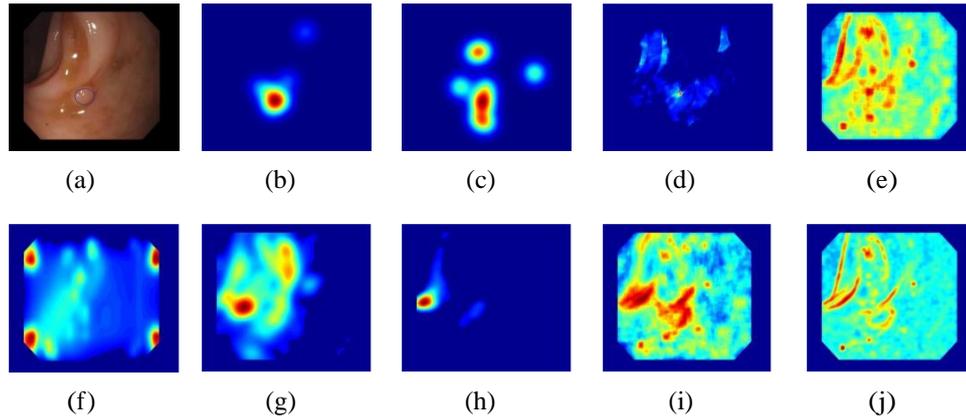


Figure 14: Comparison of saliency maps represented as heat maps: (a) Original image with polyp mask superimposed; (b) Avg. Expert; (c) Avg. Novice; (d) WM-DOVA energy map; (e) Bruce and Tsotsos; (f) GBVS; (g) Itti-Koch; (h) Seo; (i) SIM; (j) SUN. High saliency areas correspond to hot regions in the image.

remaining saliency methods. Our method seems to perform better than other computational saliency methods, having MCR values distribution closer to the one achieved by *average expert* and the *average novice*. Trends hinted by observing boxplots are confirmed by the non-parametric ANOVA test which detects significant differences among all methods (p - value $< 10^{-5}$). The multicomparison plot -Figure 13 (f)- shows that all saliency methods apart from WM-DOVA have MCR ranks significantly lower than the *average expert* and the *average novice* -which both present very similar MCR ranks-. Concerning WM-DOVA, its MCR is not significantly different than the one achieved by clinicians. Finally we present a qualitative comparison of all methods in Figure 14.

The *average expert* saliency map is clearly focused inside the polyp. The *average novice* presents a sparser saliency map where the energy maximal values still fall inside the polyp. WM-DOVA is the best candidate among computational saliency maps as it has the most polyp-focused energy maps. The remaining computational saliency methods present sparser energy maps which do not seem to highlight the polyp region against other parts of the image.

4. DISCUSSION

Colonoscopy is still considered nowadays as the gold standard for colon screening although it presents some drawbacks, being polyp miss-rate the most relevant of them. In this sense, we presented in this paper a novel polyp localization method along with a new validation framework which includes the creation of the first, up to our knowledge, fully annotated database with clinical metadata associated. Our polyp localization method is based on a general model of appearance for polyps, including the characterization of specular highlights and blood vessels. WM-DOVA is a parametric method that adapts easily to all polyp sizes and appearances, not favoring any predetermined shape like other methods in the literature. We analyze in this section WM-DOVA

performance, indicating whether it has potential to be included in clinical practice along with indicating future improvements and research lines.

4.1. Conclusions: Clinical applicability of WM-DOVA energy maps

Our polyp localization method using WM-DOVA energy maps outperforms current available methods over both our publicly available fully annotated database and *CVC-ClinicDB*. Moreover, our method succeeds locating the polyp in all the 31 different sequences with 31 different polyps, showing the robustness of our method even considering the great variability of polyp appearances observed in *CVC-ClinicDB*.

The analysis of localization results using clinical metadata shows that WM-DOVA is able to localize polyps of all sizes, with a higher sensitivity for the smallest ones (size $\leq 10\text{mm}$). Regarding Paris classification, WMDOVA has the higher localization accuracy better for flat polyps which, by their geometry and appearance in videos, are the ones having a higher chance to be missed. Concerning this we observe that there is not a consensus about polyp miss-rate causes. We would like to note that another important source of polyp miss-rate in current colonoscopy procedures is associated to folds that might hide polyps. It is clear that no computer vision method could be able to locate a polyp when it does not appear in the image and improving the scope field of view is an active area of endoscopy technology development. In fact, several manufacturers are working on innovative proposals such as Fuse endoscope developed by Endochoice [49, 50], which provides 330° of visualization as a way to make the clinician have a more complete view of the endoluminal scene. In this scenario, our method could play a key role to aid the clinician to locate polyps in the multiple views provided by Fuse endoscope.

Concerning the validation of WM-DOVA as saliency map, our pilot trial study using 11 cases shows that it is the best candidate to be used in clinical practice as part of an intelligent system for colonoscopy. Our statistical analysis does not detect any significant differences among WM-DOVA and clinicians in all the metrics considered. Our method clearly outperforms state-of-the-art bottom-up saliency models assessing that our model is more suitable for detecting polyps as the most salient region of the image when analyzing a full sequence of frames with a polyp.

Regarding computational cost of our proposal, it is worth mentioning that the results presented in this paper have been obtained with a PC with an Intel Core i7 3930K twelve-core processor with 8 GB of RAM memory. In order to implement the different algorithms proposed we have used Matlab scripts and compiled functions to incorporate the GPL libraries of the CrossVisions package [51]. Current implementation of WM-DOVA energy maps takes 10.54 seconds to fully process an image. WM-DOVA algorithm is naturally parallelizable and it can be integrated into GPU architectures by image partitioning and individual pixel assignation to core. These optimizations will surely lead to a decrease in

processing time therefore enhancing the applicability of our method in clinical practice.

4.2. Analysis of sources of error in polyp localization using WM-DOVA

Experimental results show a good performance of WM-DOVA as part of polyp localization method, being able to localize the polyp in all the sequences regardless of polyp type or size. Nevertheless there are a few cases in which WM-DOVA fails when localizing the polyp in the image. After a deep analysis of the results, we have identified three different sources of error, namely: 1) Contribution of other elements of the endoluminal scene in terms of valley information -Figure 15 (a-d)-; 2) Impact of patient preparation - Figure 15 (b-e)- and 3) Presence of lateral polyps which deviate from our model of appearance - Figure 15 (c-f)-.

First of all, we have to consider that polyps, blood vessels and specular highlights are not the only source of valley information in colonoscopy frames: other elements, especially folds or the lumen, also convey valley information and their characterization could potentially lead to an improvement of polyp localization results.

Secondly, we have to consider the impact of patient preparation: there are some cases with extremely bad patient preparation in which neither clinicians nor WM-DOVA are not able to locate the polyp. Concerning this, our method is able to cope with bad patient preparation although we have to consider that in cases with extremely bad patient preparation clinicians tend to repeat the operation and, as clinicians performance will do, WM-DOVA performance will benefit from having a cleaner endoluminal scene.

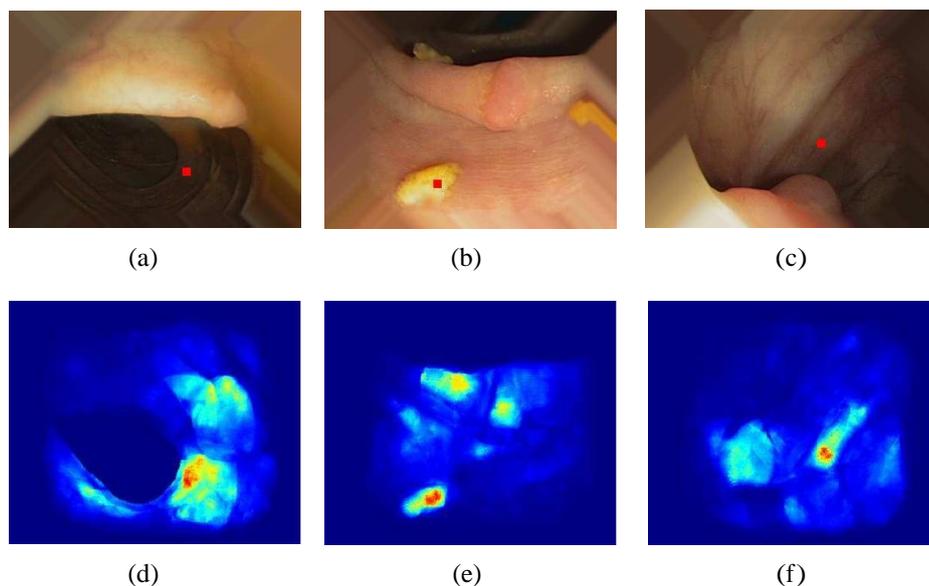


Figure 15: Errors in polyp localization results: (a,b,c) Polyp localization by means of maximum of WM-DOVA; (d,e,f) Corresponding WM-DOVA energy maps. Maximum of WM-DOVA is marked as a green square inside polyp mask and as a red square outside the polyp mask. Hotter areas of the energy maps correspond with areas with high likelihood of polyp presence.

Finally there is another source of localization errors related to lateral polyps as in this case we do not have complete boundary information, which does not help in the accumulation process, as depicted in model explanation in Section 2.1.

4.3. Future work

The analysis of the results also indicates some areas of improvement. Future work in this area should involve the complete characterization of other elements of the endoluminal scene rich in valley information such as folds or the lumen. Currently, WM-DOVA does not apply any kind of spatial or temporal coherence, contrary to what clinicians inherently do. This implies that WM-DOVA does not take into account valuable information from previous frames which could enrich the analysis of a full colonoscopy video. This fact and the presence of some images with bad visualization quality impacts accuracy in localization results when WM-DOVA is applied over a full sequence.

Shape, color and texture provide useful information for characterization of polyp. We plan to include the use of texture and color contrast between regions in order to complement our method as part of our future work. Finally, a deeper analysis of the comparison between WM-DOVA saliency maps and physicians' should also let us learn better which are the attractive regions in the image in order to identify the computer vision features that better describe them.

ACKNOWLEDGEMENTS

This work was supported in part the Spanish Government through the founded projects 'COLON-QA' (TIN2009 – 10435) and 'FISIOLOGICA' (TIN2012 – 33116).

REFERENCES

- [1] J. Ferlay and I. Soerjomataram et al., "Globocan 2012 v1.0, cancer incidence and mortality worldwide: Iarc cancerbase no. 11." 2013.
- [2] A. Tresca, "The Stages of Colon and Rectal Cancer", *New York Times (About.com)*, p. 1, 2010.
- [3] J. Kerr and P. Day et al., "Systematic review of the effectiveness of population screening for colorectal cancer", *Journal of the New Zealand Medical Association*, vol. 120, no. 1258, 2007
- [4] E. Quintero and A. Castells et al., "Colonoscopy versus fecal immunochemical testing in colorectal cancer screening", *New Engl. J. Med*, vol. 366, no. 8, pp. 697–706, 2012.

- [5] A. D. Muller and A. Sonnenberg, "Prevention of colorectal cancer by flexible endoscopy and polypectomy: a case-control study of 32.702 veterans", *Ann Intern Med*, vol. 123, no. 12, pp. 904–910, 1995.
- [6] S.J. Winawer and A.G. Zauber et al., "Prevention of colorectal cancer by colonoscopic polypectomy", *New Engl. J. Med*, vol. 329, no. 27, pp. 1977–1981, 1993.
- [7] A. Leufkens, M. van Oijen, F. Vleggaar, and P. Siersema, "Factors influencing the miss rate of polyps in a back-to-back colonoscopy study", *Endoscopy*, vol. 44, no. 05, pp. 470–475, 2012.
- [8] H. Singh, D. Turner, L. Xue, L. E. Targownik, and C. N. Bernstein, "Risk of developing colorectal cancer following a negative colonoscopy examination: evidence for a 10-year interval between colonoscopies", *Jama*, vol. 295, no. 20, pp. 2366–2373, 2006.
- [9] J.C. van Rijn and J.B. Reitsma et al., "Polyp miss rate determined by tandem colonoscopy: a systematic review", *Am J. Gastroenterol*, vol. 101, no. 2, pp. 343–350, 2006.
- [10] J.-F. Bretagne, S. Manfredi, C. Piette, S. Hamonic, G. Durand, and F. Riou, "Yield of high-grade dysplasia based on polyp size detected at colonoscopy: a series of 2295 examinations following a positive fecal occult blood test in a population-based study", *Diseases of the Colon & Rectum*, vol. 53, no. 3, pp. 339–345, 2010.
- [11] J. Bernal, F. Vilariño, and F. J. Sánchez, "Towards Intelligent Systems for Colonoscopy", *Colonoscopy*, In-Tech, 2011.
- [12] S. Krishnan, X. Yang, K. Chan, S. Kumar and P. Goh, "Intestinal abnormality detection from endoscopic images", *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, vol. 2, pp. 895–898, 1998
- [13] Y. Iwahori and T. Shinohara et al., "Automatic polyp detection in endoscope images using a hessian filter", in *Proceedings of MVA2013*, (Kyoto, Japan), pp. 21–24, (in press), May 2013.
- [14] J. Bernal, F.J. Sánchez, and F. Vilariño, "Integration of valley orientation distribution for polyp region identification in colonoscopy", *Abdominal Imaging. Computational and Clinical Applications*, pp. 76–83, 2012.
- [15] J. Kang and R. Doraiswami, "Real-time image processing system for endoscopic applications", in *IEEE Canadian Conference on Electrical and Computer Engineering*, vol. 3, pp. 1469 – 1472 vol.3, May 2003.
- [16] J. Bernal, F. J. Sánchez, and F. Vilariño, "Depth of valleys accumulation algorithm for object detection," in *CCIA*, pp. 71–80, 2011.
- [17] S. Hwang, J. Oh, and W. et al. Tavanapong, "Polyp detection in colonoscopy video using elliptical shape feature", in *Proceedings of ICIP 2007*, vol. 2, pp. II-465, IEEE, 2007.
- [18] J. Bernal, F. J. Sánchez, and F. Vilariño, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognition*, 45(9), pp. 3166-3182, 2012.
- [19] J. Bernal, F.J. Sánchez, and F. Vilariño, "Impact of image preprocessing methods on polyp localization in colonoscopy frames", in *Proceedings of the 35th IEEE EMBC*, (Osaka, Japan), (in press), July 2013.
- [20] S. Karkanis, D. Iakovidis, and D. Maroulis et al., "Computer-aided tumor detection in endoscopic video using color wavelet features", *IEEE T. Inf. Technol. B*, vol. 7, no. 3, pp. 141–152, 2003.

- [21] J. Bernal, J. Sánchez, and F. Vilariño, “Feature Detectors and Feature Descriptors: Where We Are Now”, *Tech. Rep. 154, Computer Vision Center*, September 2010.
- [22] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilariño, “Texture based polyp detection in colonoscopy”, *Bildverarbeitung für die Medizin*, pp. 346–350, 2009.
- [23] M. Coimbra and J. Cunha, “MPEG-7 visual descriptors and contributions for automated feature extraction in capsule endoscopy,” *IEEE T Circ Syst Vid*, vol. 16, no. 5, pp. 628–637, 2006
- [24] G. J. Klunker, S. A. Shafer, and T. Kanade, “Using a color reflection model to separate highlights from object color”, in *Proc. ICCV*, pp. 145–150, IEEE, 1987.
- [25] R. Tan and K. Ikeuchi, “Separating reflection components of textured surfaces using a single image”, *Digitally Archiving Cultural Objects*, pp. 353–384, 2008.
- [26] Q. Yang, S. Wang, and N. Ahuja, “Real-time specular highlight removal using bilateral filtering”, *ECCV 2010*, pp. 87–100, 2010.
- [27] Ch. Saint-Pierre and J. Boisvert and G. Grimard et al., “Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images”, *Mach Vision Appl*, vol. 22, pp. 171–180, 2011.
- [28] M. Arnold, A. Ghosh, S. Ameling, and G. Lacey, “Automatic segmentation and inpainting of specular highlights for endoscopic imaging”, *Journal on Image and Video Processing*, vol. 2010, p. 9, 2010.
- [29] A. Mendonca and A. Campilho, “Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction”, *IEEE T. Med. Imaging*, vol. 25, no. 9, pp. 1200–1213, 2006.
- [30] H. Inoue and H. Kashida et al., “The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to december 1, 2002”, *Gastrointest Endosc*, vol. 58, no. 6 Suppl, pp. S3–43, 2003.
- [31] A.M. López and F. Lumbreras et al., “Evaluation of methods for ridge and valley detection”, *IEEE T Pattern Anal*, vol. 21, no. 4, pp. 327–335, 1999.
- [32] Machine Vision Group, “Cvc-databasecolon: A database for assessment of polyp detection,” 2012. [Online; accessed 24-July-2012].
- [33] S. A. Julious, “Sample sizes for clinical trials”, CRC Press, 2009.
- [34] N. Bruce and J. Tsotsos, “Saliency based on information maximization”, *NIPS*, vol. 18, p. 155, 2006
- [35] L. Itti and C. Koch, “A saliency-based search mechanism for overt and covert shifts of visual attention”, *Vision research*, vol. 40, no. 10, pp. 1489–1506, 2000.
- [36] J. Harel, C. Koch, P. Perona, et al., “Graph-based visual saliency”, *NIPS*, vol. 19, p. 545, 2007.
- [37] H. J. Seo and P. Milanfar, “Static and space-time visual saliency detection by self-resemblance”, *Journal of vision*, vol. 9, no. 12, p. 15, 2009.
- [38] N. Murray, M. Vanrell, X. Otazu, and C. A. Parraga, “Saliency estimation using a non-parametric low-level vision model”, in *CVPR 2011*, pp. 433–440, IEEE, 2011.
- [39] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, “Sun: A Bayesian framework for saliency using natural statistics”, *Journal of vision*, vol. 8, no. 7, p. 32, 2008.
- [40] D. Gao and N. Vasconcelos, “Discriminant saliency for visual recognition from cluttered scenes”, in *Advances in Neural Information Processing Systems*, pp. 481–488, 2004.
- [41] A. Torralba, A. Oliva, M. S. Castelhana, and J. M. Henderson, “Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search”, *Psychological review*, vol. 113, no. 4, p. 766, 2006.

- [42] A. Borji, D. Sihite, and L. Itti, “Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study,” *IEEE T. Image Process*, 2012.
- [43] H. Chen, W. Lee, T. Huang, and S. Yeh, “Learning-based prediction of visual attention for video signals”, *IEEE T. Image Process*, vol. 20, no. 99, pp. 1–1, 2011.
- [44] C.M. Mann and H.J. Godwin et al., “Rapidly imparting the skills of experts to novice participants in threat assessment tasks”, *Book of Abstracts of ECEM 2013*, vol. 6, no. 3, p. 245, 2013.
- [45] R. S.A. Khan and G. Tien et al., “Analysis of eye gaze: Do novice surgeons look at the same location as expert surgeons during a laparoscopic operation?”, *Surgical endoscopy*, vol. 26, no. 12, pp. 3536–3540, 2012.
- [46] R. Barton, “Accrediting competence in colonoscopy: validity and reliability of the UK joint advisory group/nhs bowel cancer screening programme accreditation assessment”, *Gastrointestinal Endoscopy*, vol. 67, no. 5, pp. AB77–AB77, 2008.
- [47] J. Bernal, F. Vilariño, F. J. Sánchez, M. Arnold, A. Ghosh, and G. Lacey, “Experts vs. novices: Applying eye-tracking methodologies in colonoscopy video screening for polyp search”, in *Proceedings of ETRA 2014*, (Florida, EEUU), March 2014.
- [48] E. Theodorsson-Norheim, “Kruskal-Wallis test: Basic computer program to perform nonparametric one-way analysis of variance and multiple comparisons on ranks of several independent samples”, *Computer methods and programs in biomedicine*, vol. 23, no. 1, pp. 57–62, 1986.
- [49] I. M. Gralnek, P. D. Siersema, Z. Halpern, O. Segol, A. Melhem, A. Suissa, E. Santo, A. Sloyer, J. Fenster, L. M. Moons, et al., “Standard forward-viewing colonoscopy versus full-spectrum endoscopy: an international, multicentre, randomized, tandem colonoscopy trial”, *The lancet oncology*, vol. 15, no. 3, pp. 353–360, 2014.
- [50] Endochoice, “Fuse full spectrum endoscope”, 2014.
- [51] F. J. S´anchez, “Proyecto CrossVisions”, 2006.