

# Discarding Non Informative Regions for Efficient Colonoscopy Image Analysis

Jorge Bernal, Debora Gil, Carles Sánchez, and F. Javier Sánchez

Computer Vision Centre and Computer Science Department, Campus Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain  
{Jorge.Bernal, Debora.Gil, Carles.Sanchez, Javier.Sanchez.Pujadas}@uab.cat  
<http://www.cvc.uab.es>

**Abstract.** The diagnostic yield of colon cancer screening using colonoscopy could improve using intelligent systems. The large amount of data provided by high definition equipments contains frames with large non-informative regions. Non-informative regions have such a low visual quality that even physicians can not properly identify structures. Thus, identification of such regions is an important step for an efficient and accurate processing. We present a strategy for discarding non-informative regions in colonoscopy frames based on a model of appearance of such regions. Three different methods are proposed to characterize accurately the boundary between informative and non-informative regions. Preliminary results shows that there is a statistically significant difference between each of the methods as some of them are more strict when deciding which part of the image is informative and others regarding which is the non-informative region.

**Keywords:** Non informative region, Colonoscopy, Segmentation, Likelihood Maps

## 1 Introduction

Colon cancer is the fourth most common cause of cancer death worldwide and its survival rate depends on the stage it is detected on, going from rates higher than 95% in the first stages to rates lower than 35% in later stages [1]. Although colonoscopy is the gold standard screening modality for early cancer detection, a polyp miss-rate, especially for the smallest ones, around 6% [2] limits its diagnosis yield.

The high definition videos provided by recent equipments allow a more accurate detection of the smallest polyps [3] at the cost of an increase of computational time if all available information is processed. Such increase in computational time is definitely a flaw for an effective deployment in clinical environments as part of an intelligent system [4]. During an exploration, the navigation along the flexible colon anatomy produces many non-informative frames in videos and non-informative regions in valid images. Efficient identification of such non-informative data would substantially speed-up a further processing of colonoscopy explorations.

Most of the existing works on non-informative data identification focus on frame detection. The work of [5] addresses the identification of non-informative frames -with low quality or without any useful clinical information- by analyzing the energy of the

detail coefficients of the wavelet decomposition of a given image, which is used as the input to the classification system. The work of [6] defines a measure *-isolated pixel ratio* (IPR)- to classify frames into informative, ambiguous and non-informative. The IPR measure is calculated from the edges of the image, being an isolated pixel one that is not connected to any other edge pixel. Some authors [7] analyze image content to discern between frames that correspond to either a diagnostic or a therapeutic operation. The identification of the operation shots is based on the detection of diagnostic or therapeutic instruments, mapping in this case the detection of instrumental to the problem of detecting the cables of these instruments.

Concerning non-informative region -NIR- identification, the only work addressing the topic is, up to our knowledge, the content-based approach presented in [8, 9]. On the grounds that dark regions do not contain data valid for a further image processing, images were split into darker and brighter regions using watershed segmentation. Although efficient, a main concern was that intensity does not suffice for describing the endoluminal scene [10]. Moreover, the cited method included into its definition of NIR region the black borders of the image which has a clear impact in its performance.

In this paper we propose to extend the geometric-appearance models of the lumen introduced in [10] to segment NIR regions. We generate a one-parametric family of likelihood maps which minimal curve progressively approaches the non-informative region. For each such a likelihood map we define a non-informative cost function which minimum selects the parameter that best splits images. NIR region boundaries are extracted using three different operators that are evaluated using non-parametric Analysis of Variance (ANOVA) to determine the most suitable ones in terms of area overlap scores.

The structure of the paper is as follows: we explain our non-informative region segmentation method in Section 2. Experimental results are exposed in Section 3. We close this paper with the conclusions and future work in Section 4.

## 2 A Strategy for Discarding Non Informative Regions in Colonoscopy

Non informative regions are areas of such a low visual quality that neither physicians nor computer vision methods would be able to discern anything inside them. Non informative regions include lumen and other dark parts of the image generated by protruding objects which decrease the reflection towards the camera of structures below them. Therefore they correspond to dark region of images whose center is the hub of image gradients. Following [10], we characterize dark areas of the image using our Dark Region Identification (DRI) given by convolving the image,  $I = I(x, y)$ , with an isotropic gaussian kernel of standard deviation  $\sigma$ :

$$DRI := g_{\sigma} * I = \frac{1}{(2\pi)\sigma^2} e^{-\left(\frac{x^2}{2\sigma^2} + \frac{y^2}{2\sigma^2}\right)} * I(x, y)$$

Meanwhile, image gradient centralness is formulated using a Directed Gradient Accumulation (DGA) image given by [10]:

$$DGA(x, y) := \sum_{(x_0, y_0)} \chi_{\gamma_{\nabla I(x_0, y_0)}}(x, y)$$

for the image  $\chi_{\gamma_{\nabla I(x_0, y_0)}}(x, y)$  defined for each image pixel  $(x_0, y_0)$  as the mask associated to its gradient line,  $\gamma_{\nabla I(x_0, y_0)}(\lambda) = (x_0, y_0) + \lambda \nabla I(x_0, y_0)$  and  $\lambda$  as the free parameter of the gradient line equation.

Non informative region pixels will have low DRI and high DGA values. We build up a feature space given by (DRI, DGA) which should discriminate pixels from non informative region from others belonging to informative endoluminal structures. The semi-supervised strategy described in [10], classifies the space (DRI, DGA) into informative and non-informative classes by means of a linear classifier:

$$LK_0 = (DRI - DRI_0)V_{DRI} + (DGA - DGA_0)V_{DGA} = 0 \quad (1)$$

for  $(V_{DRI}, V_{DGA})$  the normal to the line defined by the classifier to separate the two categories and which passes through  $(DRI_0, DGA_0)$ . The linear classifier (1) defines a 1-parametric family of likelihood maps depending on the intercept, namely  $l$ :

$$\begin{aligned} LK_l &:= |DRI \cdot V_{DRI} + DGA \cdot V_{DGA} - (DRI_0 \cdot V_{DRI} + DGA_0 \cdot V_{DGA})| = \\ &= |DRI \cdot V_{DRI} + DGA \cdot V_{DGA} - l| \end{aligned} \quad (2)$$

for  $|\cdot|$  the absolute value. The values  $LK_l$  can be interpreted as the distance (in the feature space) to the set of pixels that define the border (given by  $LK_l = 0$ ) between non-informative and informative regions. This set of pixels correspond to a local minima of  $LK_l$  which appears as an energy valley in Lumen Energy Map -LEM- images. It follows that the boundary, which we note by  $\gamma_{LK_l}$  separating non-informative and informative regions could be extracted using a suitable valley detector.

In order that the boundary curve properly encloses non-informative regions, a main requirement is that the valley operator yields closed regular curves. Closed contours are required for the dynamic selection of the optimal  $l$  value, while regularity is a must for an accurate region segmentation. Under these considerations we have tested 3 region valley boundary detection methods:

1. **GSM2**. This valley operator presented in [11] produces complete valleys by combining steerable filters [12] with operators based on level sets geometry [13]. The geometric operator [13] is computed as the divergence of the maximum eigenvector of the structure tensor,  $\vec{V} = (P, Q)$ , reoriented along the image gradient:

$$NRM := \text{div}(\vec{V}) = \partial_x P + \partial_y Q \quad (3)$$

where NRM stands for Normalized Ridge Map and  $P, Q$  stand for the components of the structure tensor  $\vec{V}$ . The above operator assigns positive values to ridge pixels and negative values to valley ones. A main advantage is that  $NRM \in [-N, N]$  for  $N$  the dimension of the volume, so that it is possible to set a threshold common

to any volume for detecting significant valleys. A main limitation is that it could produce fragmented curves. In [11] this was overcome by further convolving NMR with a bank of steerable filters defined by 2nd derivatives of (oriented) anisotropic Gaussian kernels  $g_{\sigma}^{\theta}$  with standard deviation  $\sigma_y$ :

$$\partial_y^2 g_{\sigma}^{\theta} = (\tilde{y}^2 / \sigma_y^4 - 1 / \sigma_y^2) g_{\sigma}^{\theta} \quad (4)$$

for  $(\tilde{x}, \tilde{y})$  the coordinates given by a rotation of angle  $\theta$  that transform the y-axis into the unitary vector  $(\cos(\theta), \sin(\theta))$  and the scales set to  $\sigma_x = 2 * \sigma_y$ . The maximum response for a sampling of the angulation,  $\theta_i$  given by  $\theta_i = \{\frac{i\pi}{N}, i = 1..N\}$ , defines the Geometric Steerable Map (GSM2) as:

$$GSM2 := \max_i \left( \partial_y^2 g_{\sigma}^{\theta_i} * NRM \right) \quad (5)$$

2. **Depth of Valleys Accumulation-based segmentation (DOVA).** Depth of Valleys Accumulation -DOVA- energy map are linked with likelihood of polyp presence in colonoscopy images [14]. These maps are built from a model of appearance for polyps which describe their boundaries in terms of valley information. In order to generate these maps the authors use a ring of radial sectors to accumulate contributions of pixels with high valley information. We propose here to use the same rationale but in this case taking advantage of the fact that we already know which is the lumen center - $c^{max}$ - and we have *LEM* maps. The application of DOVA maps here consists of calculating valley information from the *LEM* maps to obtain a first approximation of NIR region boundary which is later refined using the ring of radial sectors. The steps are:
  - (a) Definition of a ring of  $ns$  radial sectors centred in  $c^{max}$ .
  - (b) Calculation of the valley image *VLEM* from *LEM* maps using valley detection method proposed in [13].
  - (c) Calculation of the position of the maximum of *VLEM* image under each sector  $S_i$  of the ring as  $p_i^{max} \in S_i \mid \forall k \in S_i, VLEM(c_i^{max}) \geq VLEM(q)$ , with  $i \in [1, ns]$ . By directly joining the positions of  $p_i^{max}$  we can obtain a first segmentation of the NIR region. As this segmentation may present some spikes we propose the use of median filtering of  $p_i^{max}$  in the polar domain to correct these irregularities in order to have similar distances from consecutive sectors maximums to  $c^{max}$ .
3. **Watershed with markers (WSM).** Watershed segmentation considers a grayscale image as a topographic surface and achieves the segmentation by a process of 'filling' of catchment basins from local minimums. Providing markers helps the algorithm to define the catchment basins that must be considered in the process of segmentation [15]. For our specific applications we will use  $c^{max}$  as the internal marker, placing the external marker in a padding masks surrounding the whole image.

For all the methods above, the non informative region was identified as the one containing the center of the lumen. Such point can be computed from the linear classifier (1) using the semi-supervised strategy described in [10].

We observe that the  $LK_l$  map best separating non-informative and informative regions should split image pixels into the darkest and brightest ones. Darkest pixels should

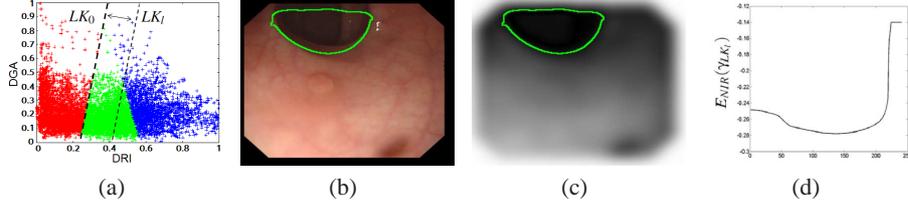


Fig. 1: Complete example of non informative region identification: (a) Colonoscopy feature space; (b) Original image with boundary of NIR superimposed; (c) LEM map with boundary superimposed; (d) Selection of the optimal threshold as the minimal value of function  $E_{NIR}(\gamma_{LK_l})$ .

all lie in the interior of  $LK_l$  boundary,  $\gamma_{LK_l}$ , while the image region outside  $\gamma_{LK_l}$  should present a significantly brighter intensity level. We will note pixels belonging to the interior of  $\gamma_{LK_l}$  by  $NIR$  and by  $NIR^c$  pixels outside  $\gamma_{LK_l}$ . The difference between internal and external intensities can be measured using the following cost function:

$$E_{NIR}(\gamma_{LK_l}) := \frac{1}{|NIR|} \sum_{(x,y) \in NIR} I(x,y) - \frac{1}{|NIR^c|} \sum_{(x,y) \in NIR^c} I(x,y) \quad (6)$$

for  $|NIR|$  denoting the number of pixels in the NIR region and  $|NIR^c|$  the number of pixels outside NIR region. The function  $E_{NIR}(\gamma_{LK_l})$  attains a minimal value for the optimal intercept, namely  $\tilde{l}$  that best separates  $NIR$  and  $NIR^c$ . This optimal value can be efficiently obtained by exhaustive search of all possible intercept values or using any gradient descent method.

We show in Figure 1 a complete example of our non-informative region identification method.

### 3 Experiments

In order to validate the performance of our non informative region identification method, we have used the same database presented in [10] taken from 15 different sequences with a polyp from colonoscopy interventions. We have selected those colonoscopy frames having non-informative regions including lumen and low visibility regions. The final dataset used in our experiments contains 100 frames. One expert provided a mask labelling non-informative regions excluding the black borders which surround natively colonoscopy frames. We show some examples of frames of our database along with non-informative masks in Figure 2.

Assessment of the proposed methods was quantified using the Annotated Area Covered (AAC) and the Dice Similarity Coefficient (DICE) given by: [16].

$$AAC = 100 \cdot \frac{T_{NIR}}{GT_{NIR}} \quad DICE = 100 \cdot \frac{T_{NIR}}{AUT_{NIR}}; \quad (7)$$

where  $T_{NIR}$  stands for the number of pixels correctly labelled as non informative region,  $GT_{NIR}$  for the number of annotated pixels and  $AUT_{NIR}$  for the number of pixels detected

using the methodology explained. Both measures are complementary, as the former calculates the amount of annotated  $NIR$  area while the latter complements it with the amount of  $NIR^c$  information that is kept in the region.

In order to explore differences across the three methods for region extraction proposed in Section 2, we have used a non-parametric analysis of variance given by the Kruskal-Wallis one-way analysis of variance by ranks [17]. The Kruskal-Wallis test has been done for the DICE and AAC scores obtained by each region extractor method, which define the anova groups. Quality scores have been summarized using confidence intervals for their average values [18]. The Kruskal-Wallis test and average score confidence intervals have been computed at a significance level  $\alpha = 0.05$ .

For both scores, the Kruskal-Wallis test shows a significant difference on the performance of methods ( $p - val < 10^{-4}$ ). Figure 3 shows, for each score, a comparison of the average ranks (the lower, the better) of each method with Tukey-Kramer correction for the multi-comparison. The intervals shown in the graph are computed so that to a very close approximation, two estimates being compared are significantly different if their intervals are disjoint, and are not significantly different if their intervals overlap [19]. We have highlighted in blue the best ranked methods and in red the worse ones. For AAC, *GSM2* is significantly the best ranked method compared to the ranks of *DOVA* and *WSM*. Although there is no significance difference between the latter (their rank intervals overlap), *DOVA* has a worse average rank than *WSM*. This trend reverts in the case of *DICE* as *GSM2* rank is significantly worse than *DOVA* and *WSM*. As before, there are no significant differences between *DOVA* and *WSM*, but *DOVA* has a better average rank.

Table 1 reports AAC and DICE confidence intervals for the 3 methods. Results indicate that *GSM2* is the most strict one in terms of the amount of selected valid information (with AAC around 85% and DICE around 65%) and it might discard areas that could be considered valid for a further inspection. On the other hand, *DOVA* is the most permissive in terms of information discarding (with DICE around 85% but AAC around 65%) and might include some non-informative areas. Finally, *WSM* achieves a

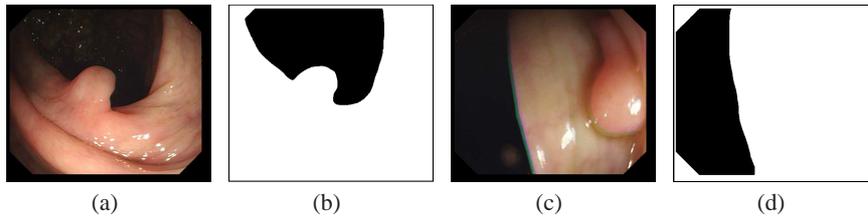


Fig. 2: Examples of the content of the dataset used in the experiments: (a,c) Original images; (b,d) Non-informative region masks provided by manual annotation by the expert.

	<i>GSM2</i>	<i>DOVA</i>	<i>WSM</i>
AAC	[81.3620%,87.7628%]	[61.4614%, 73.6962%]	[62.9300%,78.9571%]
DICE	[58.7298%,73.4484%]	[77.3452%,89.3385%]	[71.3674%,85.2215%]

Table 1: Confidence Intervals for average AAC and DICE

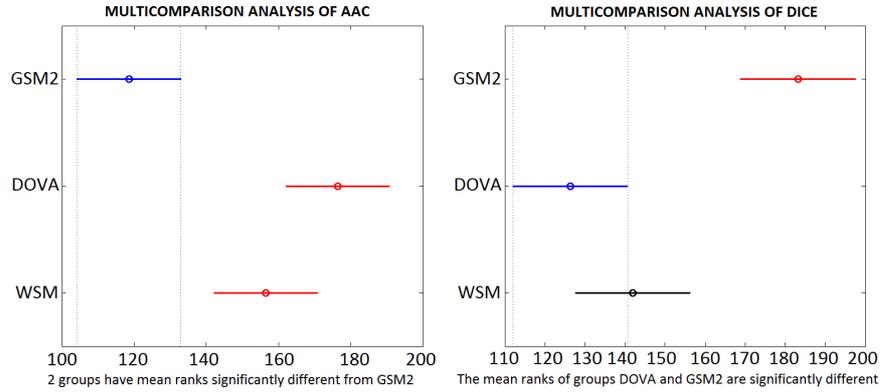


Fig. 3: Multicomparison Analysis for the two Quality Scores. Horizontal axis represent the average ranks (the lower, the better).

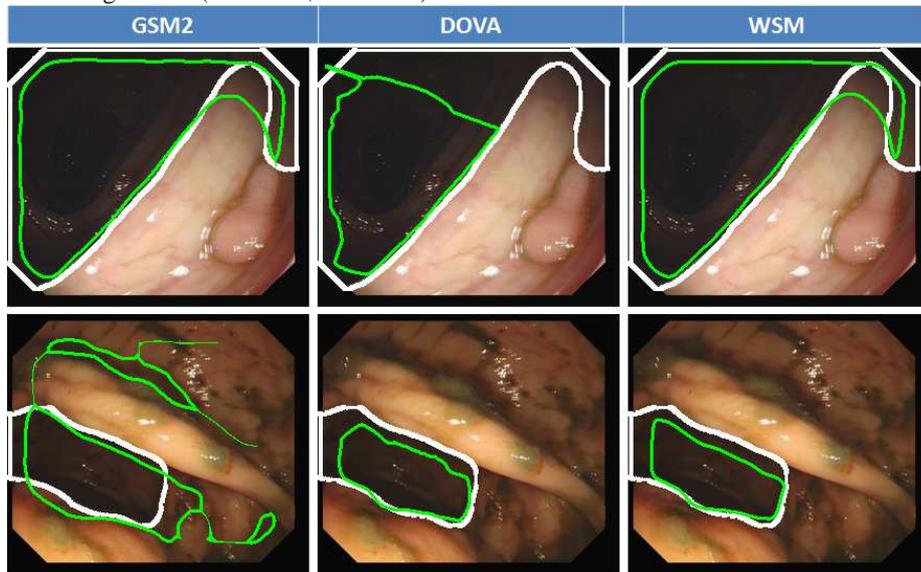


Fig. 4: Qualitative examples of NIR region identification using as the three proposed methods -GSM2, DOVA and WSM- as boundary detector.

compromise between AAC and DICE, with both indexes round 75%. Figure 4 shows some qualitative examples of NIR region identification using the 3 methods for delimiting their boundary. Manual boundaries are shown in white and automatic ones in green. The first row shows an example of *DOVA* under-segmentation but accurate *GSM2* and *WSM* segmentation of the non-informative region, which includes the lumen and a shadow. The second row shows the opposite behavior with *DOVA* and *WSM* providing a more accurate segmentation in contrast to a larger non-informative *GSM2* region.

Finally and regarding computation time, all 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 develop the different algorithms we have used Matlab scripts and compiled functions to incorporate the GIPL libraries of the CrossVisions package [20]. Obtaining LEM energy maps from a single image takes 3.87 seconds in a single core. The computation time different algorithms for final NIR region calculation from LEM maps is as follows: 3.82 seconds for GSM2, 4.52 seconds for DOVA and 0.02 seconds for WSM. The direct computation of NIR regions using the method proposed in [8, 9] takes 80.46 seconds. All the proposed algorithms are naturally parallelizable and they can be integrated into GPU architectures by image partitioning and individual pixel assignation to core.

## 4 Conclusions and Future Work

This paper addresses identification of non-informative regions, NIR, in colonoscopy frames which should be discarded at later stages of clinical support algorithms. An automatic discard of NIR saves a computational time that allows a more accurate processing of valid parts of the image. Aside from computational time savings, discarding frames with large non-informative regions could also be used to automatically create summaries of colonoscopy videos [4], omitting those non-informative frames.

In this work, three different alternatives for NIR segmentation, *GSM2*, *DOVA* and *WSM*, have been presented and evaluated according to AAC and DICE score. The methods can be ranked according to the amount of valid information discarded from the most strict *GSM2* to the most permissive *DOVA* and *WSM* presenting the best compromise with average scores over 75%. This already represents a huge improvement of previous results [8, 9] which achieved average DICE and AAC scores around 50%. Although very promising, our results have room for improvement. First, visual identification of non-informative regions is a difficult task presenting a significant variability within observers. In order to account for it, images will be annotated twice. Second, the proposed feature space works in the gray intensity domain, which usually discards larger areas that include information valid in the color space. This could be overcome by working in a 3D color space and it is currently under research.

## ACKNOWLEDGMENTS

This work was supported by a research grant from Universitat Autònoma de Barcelona 471-01- 2/2010 and by Spanish projects *TIN2009* – 10435, *TIN2009* – 13618 and *TIN2012* – 33116.

## References

- [1] R. Siegel, C. DeSantis, and A. Jemal, “Colorectal cancer statistics, 2014,” *CA: a cancer journal for clinicians*, vol. 64, no. 2, pp. 104–117, 2014.
- [2] B. Bressler, L. Paszat, Z. Chen, D. Rothwell, C. Vinden, and L. Rabeneck, “Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis,” *Gastroenterology*, vol. 132, no. 1, pp. 96–102, 2007.

- [3] P. Pickhardt, P. Nugent, P. Mysliwiec, J. Choi, and W. Schindler, "Location of adenomas missed by optical colonoscopy," *Annals of internal medicine*, vol. 141, no. 5, pp. 352–359, 2004.
- [4] J. Bernal, F. Vilariño, and J. Sánchez, *Colonoscopy Book 1: Towards Intelligent Systems for Colonoscopy*. In-Tech, 2011.
- [5] M. Arnold, A. Ghosh, G. Lacey, S. Patchett, and H. Mulcahy, "Indistinct frame detection in colonoscopy videos," in *2009 13th International Machine Vision and Image Processing Conference*, pp. 47–52, 2009.
- [6] J. Oh, S. Hwang, W. Tavanapong, P. de Groen, and J. Wong, "Blurry-frame detection and shot segmentation in colonoscopy videos," in *Proceedings of SPIE*, vol. 5307, p. 531, 2003.
- [7] Y. Cao, D. Liu, W. Tavanapong, J. Wong, J. Oh, and P. de Groen, "Computer-aided detection of diagnostic and therapeutic operations in colonoscopy videos," *Biomedical Engineering, IEEE Transactions on*, vol. 54, no. 7, pp. 1268–1279, 2007.
- [8] J. Bernal, J. Sánchez, and F. Vilariño, "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, November 2010.
- [9] J. Bernal, J. Sánchez, and F. Vilarino, "Towards automatic polyp detection with a polyp appearance model," *Pattern Recognition*, 2012.
- [10] C. Sánchez, J. Bernal, D. Gil, and F. J. Sánchez, "On-line lumen centre detection in gastrointestinal and respiratory endoscopy," in *Clinical Image-Based Procedures. Translational Research in Medical Imaging*, pp. 31–38, Springer, 2014.
- [11] S. Vera, D. Gil, A. Borràs, and M. A. Linguraru, Marius G. and González Ballester, "Geometric steerable medial maps," *Machine Visions and Applications*, in press.
- [12] W. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.
- [13] López, A.M. and Lumbreras, F. et al., "Evaluation of methods for ridge and valley detection," *IEEE T Pattern Anal*, vol. 21, no. 4, pp. 327–335, 1999.
- [14] J. Bernal, 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.
- [15] X. Zhang, F. Jia, S. Luo, G. Liu, and Q. Hu, "A marker-based watershed method for x-ray image segmentation," *Computer methods and programs in biomedicine*, 2014.
- [16] F. Riaz, M. Ribeiro, and M. Coimbra, "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, September 2009.
- [17] K. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American Statistical Association*, vol. 47(260), p. 583â621, 1952.
- [18] J. L. Myers and A. D. Well, *Research Design and Statistical Analysis 2Ed*. New Jersey: Lawrence Erlbaum Associates, 2009.
- [19] Y. Hochberg and A. Tamhane, *Multiple Comparison Procedures*. Wiley, 1987.
- [20] F. J. Sánchez, "Proyecto crossvisions." 2006.