Noname manuscript No.
(will be inserted by the editor)

Bright Spot Regions Segmentation and Classification for Specular Highlights Detection in Colonoscopy Videos

F. Javier Sánchez · Jorge Bernal · Cristina Sánchez-Montes · Cristina Rodríguez de Miguel · Gloria Fernández-Esparrach

Received: date / Accepted: date

Abstract A novel specular highlights detection method in colonoscopy videos is presented. The method is based on a model of appearance defining specular highlights as bright spots which are highly contrasted with respect to adjacent regions. Our approach proposes two stages; segmentation, and then classification of bright spot regions. The former defines a set of candidate regions obtained through a region growing process with local maxima as initial region seeds. This process creates a tree structure which keeps track, at each growing iteration, of the region frontier contrast; final regions provided depend on restrictions over contrast value. Non-specular regions are filtered through a classification stage performed by a linear SVM classifier using model-based features from each region. We introduce a new validation database with more than 25,000 regions along with their corresponding pixel-wise annotations. We perform a comparative study against other approaches. Results show that our method is superior to other approaches, with our segmented regions being closer to actual specular regions in the image. Finally, we also present how our methodology can also be used to obtain an accurate prediction of polyp histology.

Keywords Specular highlights · bright spot regions segmentation · region classification · colonoscopy

F. Javier Sánchez · Jorge Bernal
Computer Vision Center and Computer Science Department at Universitat Autonoma de Barcelona, Campus UAB, Edifici O, 08193, Bellaterra, Barcelona, Spain
E-mail: javier.sanchez,jorge.bernal@cvc.uab.es

Cristina Sánchez-Montes · Cristina Rodríguez de Miguel · Gloria Fernández-Esparrach
Endoscopy Unit, Gastroenterology Service, CIBERHED, IDIBAPS, Hospital Clinic, Universidad de Barcelona, 08036, Barcelona, Spain
E-mail: crsanchez,crodrigm,mgfernan@clinic.cat

1 Introduction

Colorectal cancer is the fourth cause of cancer death worldwide [11], with the early detection of its precursor lesions - polyps - being crucial to increase patient survival rate. Among the advances in endoscopy imaging [7] to allow a better scene observation, computational tools have been proposed to help clinicians during the different stages in the procedure [22,23,14]. The majority of these tools are focused on polyp characterization (detection, segmentation or classification), although some of them are focused on characterizing other elements for different applications. For instance, luminal region characterization [5] is used to help in navigation and blood vessels detection [20] could allow tracking regions of interest in the insertion and withdrawal phases. Recent works [6] also show that characterizing different elements in the scene has an impact on method performance.

In this paper, we focus in this paper on the characterization of specular highlights. They are not actual endoluminal scene structures as they appear due to image acquisition, scene illumination and the structures' reflectance properties but, as they alter the content of the scene to be analyzed by both clinicians and computational systems, their presence can affect the outcome of a colonoscopy procedure [6]. From a clinical point of view, specular highlights difficult the observation of image regions with potentially useful information for applications such as in-vivo lesion histology through visual analysis of the polyp region. Furthermore, their highly contrasted appearance may deviate clinicians' attention from actual clinical targets in the image.

As regards computational systems, the appearance of specular highlights also makes them highly attractive to computational region of interest (ROIs) detectors,

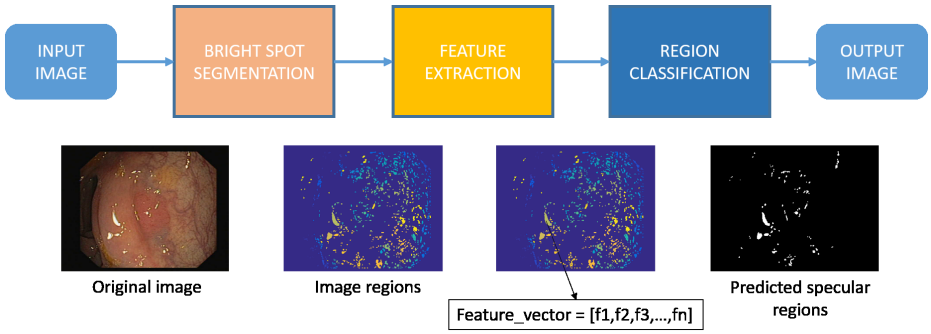
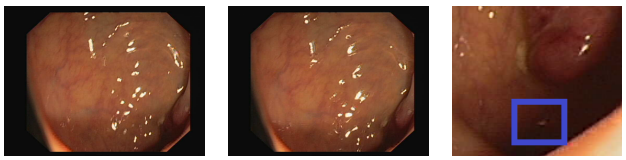


Fig. 1 Specular highlights detection pipeline.

diverting their attention from actual targets in the image. As their appearance varies in each frame, computational systems have to be aware of their presence when dealing with structure tracking in video sequence analysis. Finally, their presence over polyps might result in a loss of texture information, damaging the performance of computational lesion characterization methods. Considering all this, their detection would be a benefit for both clinicians and automatic computational tools.

There are several challenges that specular highlights detection in colonoscopy images present. First of all, it is difficult to model their appearance, as it varies greatly within an image. Besides, not all of them can be detected by intensity thresholding. As the intensity of specular highlight depends on the amount of incident light, they appear darker in in poorly illuminated areas - Fig. 2 (c) -. Secondly, slight illumination changes alter their appearance within consecutive frames, (Fig. 2 (a-b)) thus detection has to be performed separately for each frame. Consequently, efforts should be made to reduce the computational cost of the detection, in order to include it as part of a clinically useful support system working under real time constraints. Finally, the lack of public annotated databases makes it difficult to compare similar approaches, even outside colonoscopy image analysis. We associate this with the difficulty related to creating pixel-wise annotations of specular highlights.

The contributions of this paper are two-fold; the introduction of a specular highlight detection method,



(a) (b) (c)

Fig. 2 Examples of variability on specular highlights appearance. Images (a) and (b) shows variation of specular highlights appearance in number and structure among two consecutive frames. Image (c) highlights with a blue frame an example of a specular highlight within a dark region.

and the definition of a completely new hand-made pixel-wise annotated validation database. Our method is based on the definition of model a of appearance defining specular highlights as bright regions highly contrasted with respect to neighbor image areas. The first stage consists of the segmentation of these bright regions in the image. Unfortunately, other image regions - such as overexposed regions or highly illuminated structures - may fall within this category, therefore a region classification stage is performed to eliminate them from final method's output. We present our processing pipeline in Fig. 1.

Our validation database, to be publicly available upon papers' publication, contains more than 25,000 fully annotated specular highlights regions of different sizes and shapes. We use it to perform a comparative study of several general and colonoscopy-specific specular highlight detection methods. We also present an application of our methodology as part of a computational support system to predict polyp histology.

The rest of the paper is structured as follows: Section 2 describes state-of-the-art on specular highlight detection. Section 3 introduces our model of specular highlights appearance. Section 4 details bright spot regions segmentation, region classification as explained in 5. Section 6 presents the comparative study, the results of which are presented in Section 7 and discussed in Section 8. Concluding remarks are drawn in Section 9.

2 Computational approaches for specular highlights detection

Specular highlights detection has been an active research topic in the last decades. The majority of methods are based on the dichromatic reflection model (DRM), which separates highlights into diffuse and specular components. Differences among DRM-based methods are based on how this separation is performed.

Tan et al. [24] propose a distribution of specular and diffuse points in a two-dimensional maximum chromaticity-intensity space, and tackle separation of

the reflection component by identifying the maximum diffuse chromaticity which requires an additional analysis of image noise. This same space is used in the approach presented by Xu et al. [26], which states that for objects with ideal Lambertian surfaces, diffuse chromaticity is constant, but that of specular is variable. This method assumes that, in case of uniformly colored surfaces, any local diffuse chromaticity should be the same as the global one, and proposes the use of local windows to explore changes from this global value in order to detect specular highlights. We can also include in this group the work of Yang et al. [27], which estimates the maximum diffuse chromaticity values of the specular pixels by directly applying a low-pass filter to the maximum fraction of the color components of the original image. The method assumes that maximum values of diffuse chromaticity are propagated from diffuse to specular components.

Other DRM-based approaches are not based on diffuse chromaticity. Shao et al. [21] deals with specularities on non-Lambertian surfaces basing their method on defining specular highlights as regions with higher luminance in the image. Highlights are detected by applying iteratively increasing-order thresholding over luminance value. The work of Mairon et al. [16] follows a similar trend though it uses luminance and saturation from HSV color space, being specular pixels those in which the fraction between those channels is below a given threshold value. Finally the work of Yoon et al. [28] provides a specular-free two band image using simple pixel-wise real-time computations.

Angelopoulos et al. [2] method is based on Cook and Torrance reflectance model, which describes the directional distribution and spectral composition of the reflected light and its dependence on local surface geometry, surface roughness and material properties.

Specular and diffuse components separation is more difficult in colonoscopy images due to: 1) the presence of overexposed regions due to illumination, making it impossible to recover diffuse component and 2) the presence of image regions with non-linear camera response due to image post-processing for visualization purposes. Colonoscopy-specific methods thus propose alternative strategies for specular highlights detection deviating from those already mentioned.

Arnold et al. [3] propose a two-module approach: the first uses color balance adaptive thresholding to determine those pixels in the image candidate to be inside a specular region. The second deals with those remaining specular pixels presenting differences in intensity to the center of the specular region. Our previous approach [9] extends this method to tackle detection of pixels that constitute the boundary of the original specular region and whose appearance resembles a shadow of the highlight. This is done by analyzing specular highlight regions frontiers and the difference between the original image and the median of pixels not part of the original output to obtain which pixels have an intensity value marginally higher than its neighborhood. More recently, Alsaleh et al. [1] proposed the use of an adaptive threshold which considers color variation of the input image to define the threshold value from which a pixel is labelled to be part of a specular highlight. Contours of specular highlights are identified as those specular highlight pixels with strong gradients.

The main novelty of our approach from other colonoscopy approaches relies on absence of intensity thresholding contrary to the works of [3],[9] [1], as its use may end in the loss of specular regions in dark image areas. With respect to all presented approaches, our method is one of the few based on defining specular regions rather than on the analysis of individual pixels as it focuses on the contrast between highlights and their neighboring regions. Finally, our method is the first one, up to our knowledge, exploring the use of the blue channel of RGB images to differentiate between specular and diffuse component of the specular highlights, under the assumption that the latter tends to appear reddish in colonoscopy images.

Comparing presented approaches is difficult due to the lack of publicly annotated database. The only available comparison study was done for colonoscopy images domain [9] but it was a region-based analysis and the validation database was not public. Therefore, there is a need of large public datasets which, apart from pixel-wise analysis, deals with images from different domains, being this out of the scope of this paper.

3 Model of Appearance of Specular Highlights

Our specular highlights detection method is based on a model of their appearance. This model is based on how specular highlights are generated and their corresponding appearance in colonoscopy images. Specular highlights appear as a result of the reflectance of the structures that appear in the endoluminal scene as they are illuminated by the colonoscope. The colonoscope has a light source and a camera aligned in the same direction so we can model the colonoscope as a pinhole camera and a punctual illumination source placed in the same position. We also consider that the colon surface is regular and that its reflectance can be approximated by Phong's illumination model [8]. Consequently, the image acquired can be approximated as:

$$
\begin{aligned}
I &= I_a K_a + f_{att}(I_{diff} + I_{spec}) \\
I_{diff} &= I_p K_d \cos\theta \\
I_{spec} &= I_p W(\theta) \cos^n \alpha
\end{aligned}
\tag{1}
$$

where I is the light reflected by the surface towards the camera, I_a is the ambient intensity, Ka is the ambient reflection constant, f_{att} is the attenuation factor, I_{diff} is the diffuse reflection component and I_{spec} is the specular reflection component. I_{diff} is calculated as the product between incident light intensity I_p, a surface-specific diffuse reflection coefficient K_d, and the cosine of θ (angle between surface normal and incident light). I_{spec} is calculated as the product between incident light intensity I_p, the specular reflection $W(\theta)$ coefficient and the n-th power of the cosine of α (angle between the observing direction and the reflected light).

Our model defines specular highlights as bright regions fulfilling the following conditions:

1. The region contains at least a local maxima.
2. From all the regions that contain the same local intensity maximum, the final region is the one which maximizes mean gradient in its boundary.
3. Pixels belonging to each region are all above the intensity threshold which might be different for each specular region in the image.

The first condition is a direct consequence of image acquisition: as both light source and camera are at the tip of the endoscope, $cos\theta \geq \cos\alpha$. When observation of the scene is aligned to surface normal both θ and α are zero, maximizing their corresponding cosines. In this case we will have a maximum in the amount of light reflected back to the camera and, consequently, *specular highlights will contain a local maxima in the image.*

The second condition considers deviations from previous ideal situation; when the observation direction deviates from surface normal, $\alpha = 2\theta$. Considering this and that the specular component depends on the n-th power of the cosine of α, small variations in α lead to stronger decreases in specular highlight intensity. Therefore, we can deduce that *specular highlights will be delimited by strong light intensity gradients.*

Finally, and as a consequence of the big decrease in highlight intensity as surface orientation varies, we can approximate the surface generated by a specular highlight as a plane perpendicular to light and observing directions. Under this assumption, light attenuation is the same for the whole specular region and, consequently, *pixels belonging to a region are all above a same light intensity, with a different threshold for each specular region.*

4 Bright Spots Region Segmentation

4.1 Overview

Our segmentation strategy can be enclosed into region growing segmentation algorithms. In our case local maxima act as seeds defining a set of initial image regions. These regions grow by adding neighbor pixels in an intensity-ordered way so pixels closer in intensity to the region seed added first. During region growing, colliding regions are merged into a parent region. Region growing is performed until all initial regions are merged into a single one containing the whole image. Our method is based on the definition of a regions tree structure. In order to track information of the growing process, this information is used to determine which are the final regions to be provided as segmentation results.

Our approach is directly linked to the model of appearance of specular highlights. Segmentation seeds are defined as local maxima in the image and we use gradient information to select the final output regions. As we aim to obtain the region which separates better from the background, region growth is performed until the whole image is covered to avoid growth to stop at a local maxima of the contrast; in this case the region provided will be smaller than the actual specular region. Finally, our methodology is specially designed aiming at its potential inclusion in a CAD system and it is implemented using pre-computed matrices and pixel access by pointers to reduce the number of calculations.

4.2 Definitions and data structures

Our methodology is based on the concept of *region*; at the initial stage of the algorithm, regions are defined as the set of connected pixels centered at local maxima in the image and share a same region identification label. All pixels which are not part of any region are part of the *background*. We define $innerfrontier$ as the set of region pixels which have as neighbor at least one background pixel. Finally we define $externalfrontier$ of the region as the set of background pixels which have at least one pixel belonging to any region as a neighbor. Fig. 3 (b) shows a graphical example of both frontiers.

We define a $RegionInf$ structure to store information related to each region's initialization and growth. Table 1 shows the content of $RegionInf$ structure.

We use labels to identify regions during the whole segmentation process. We store this label information in $RegionInf$ but, for the sake of computational efficiency, we also define a $LabelArray$ table, indexed according to label value. This table helps to efficiently keep track of which labels are associated with a region.

Field	Explanation
PSeed	Position of the local maxima acting as seed of the region. If the region is the result of a merging process, its value will correspond to the seed of one of the children regions
Labels	List of region labels associated to the specific region
Childrens	List of regions merged into actual region
Father	Region in which actual region will be merged
CreationLevel	Luminance level the region was created at
ToUpdateLevel	Luminance level at which local contrast will be calculated
FGradSum	Sum of inner frontier pixels gradient values
FLen	Number of pixels in inner frontier
MaxFGrad	Maximum local contrast value achieved during growing process
MaxLevel	Luminance level in which current MaxFGrad value was achieved

Table 1 Fields of $RegionInf$ structure

We present each of the fields of this table in Table 3. This table allows the labelling information to be easily updated whenever two regions are merged as the values of $PRegion$ and $PNextLabel$ are changed to reflect this (see Fig. 3).

We also define a set of supporting images for computational efficiency purposes. $RegionsIm$ represents the different regions in the image. Positive values represent labels of pixels belonging to image regions; labels are different for each image region. $External frontier$ pixels have a -1 value whereas background pixels are set to 0. $NeighIm$ provides information of pixels' neighborhood at each growing iteration. Pixel value represents the number of neighbors belonging to the background so $inner frontier$ pixels have a value higher than 0.

4.3 Bright spot regions segmentation algorithm

Our segmentation algorithm consists of 4 different stages: 1) Region Initialization, 2) Region Growing, 3) Region Selection and 4) Region Marking which are detailed next. Our method needs of three inputs: 1) luminance image ($LumIm$), 2) module of gradient image ($GradIm$), and 3) minimum value of frontier contrast ($MinGrad$) a candidate region should possess. $GradIm$ should have a low value - even zero - for those pixels belonging to local maxima of $LumIm$. This condition is fulfilled in the continuous domain but not in the discrete one. To solve this, we manually set $GradIm$ to zero for image local maxima; if this is not done, local maxima may not grow as their associated inner frontier may already have a high frontier contrast. This is aggravated whenever an unsharp mask is applied in videoprocessors to enhance image visualization [7]. $MinGrad$ value is only used at region selection to prune the region tree and determine final segmented regions. Table 2 summarizes the terminology associated with the different supporting images and structures that the segmentation algorithm uses.

Term	Definition
Supporting Images	
RegionsIm	Image representing the different regions in the image. Each different pixel value represents a different image region.
NeighIm	Image representing information of pixels' neighborhood. Values for each pixel represent the number of neighbors (pixels which do not belong to any region yet).
Data Structures	
RegionInf	Data structure representing information related to each region's initialization and growth.
LabelArray	Table used to keep track of which labels are associated with a given region.
Heap	Binary max-hepa used to store ordered information about pixels belonging to external frontiers. Ordering is performed according value in $LumIm$ image.
Algorithm inputs	
LumIm	Luminance image.
GradIm	Module of the gradient applied over the input image, which is used to control region growth process.
MinGrad	Minimum value required that the frontier contrast of a given region should possess in order to be part of the output of the method.

Table 2 Summary of the terminology related to the main images and structures used in the explanation of bright spot regions segmentation algorithm.

4.3.1 Region Initialization

This first step consists of definition of the set of initial image regions as the local maxima of $LumIm$ [25]. A different label is assigned to each region in $RegionsIm$. Starting values of $FGradSum$ and $FLen$ can be calculated as $inner frontier$ and $external frontiers$ are defined - see Fig. 3 (a) -. This first step also updates $NeighIm$. Initialization stage is also used to create the regions tree which will keep track of region structure during the growing process. In this first stage all image regions, all of them independent, are connected to the root of the $LabelArray$ structure.

4.3.2 Region growing

Our region growing strategy is performed by adding neighbor $external frontier$ pixels to the actual region. Pixel addition is sorted by intensity level. We use a binary max-heap - referred to as *Heap* - to store ordered information about pixels belonging to external frontiers. Ordering is made according to luminance value. Each specific intensity level defines a

Field	Explanation
PRegion	This value points at the region the specific label belongs to
PNextLabel	This field is used to create the list of linked labels that are associated to a same region

Table 3 Fields of $LabelArray$ table

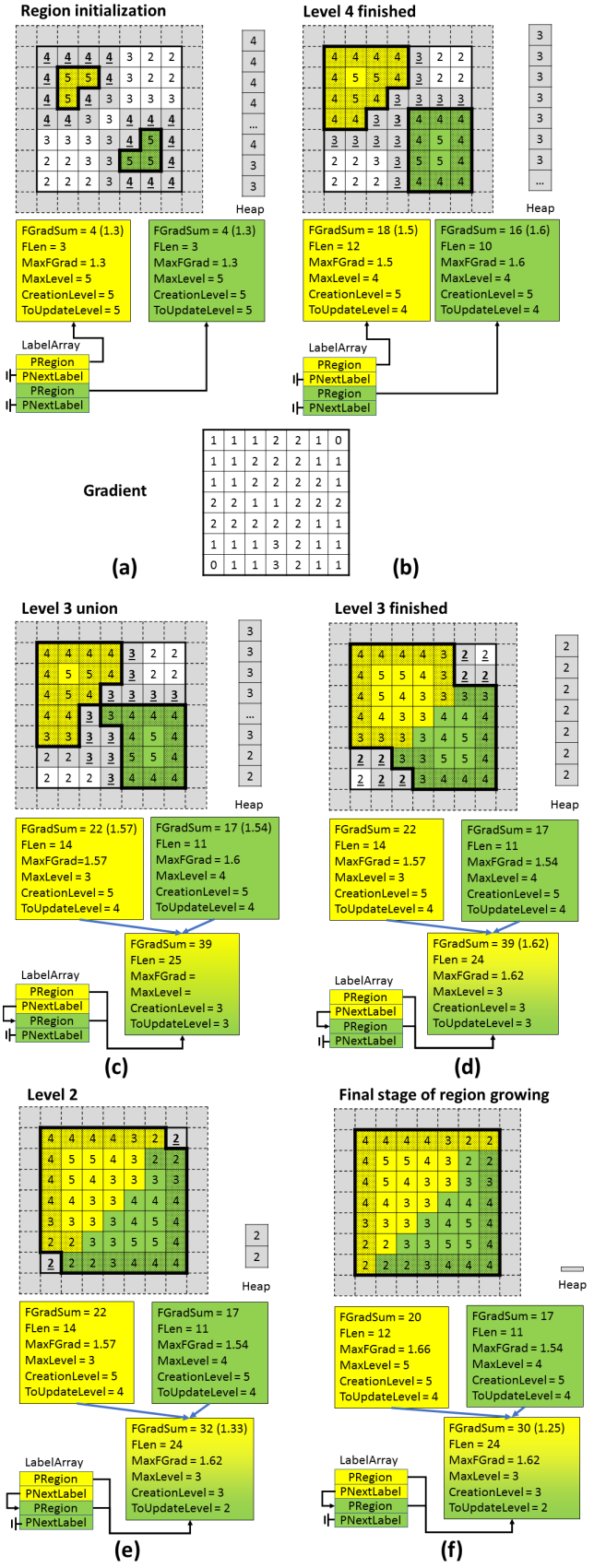


Fig. 3 Synthetic example of (a) Region initialization and (b-e) Region merging. Each region label is represented with a different color and limited by a thick contour. External frontier pixels are painted in gray, dotted pixels belong to the internal frontier. Pixels at current intensity level are underlined; those currently are also highlighted in bold.

growing iteration. This iteration ends when all pixels with $ToUpdateLevel$ intensity values have been processed. Once this happens, local contrast can be calculated. If its value surpasses current $MaxLevel$ value in the region's $RegionInf$ structure, both $MaxLevel$ and $MaxFGrad$ fields are updated. Fig. 3 (b) exemplifies this, and we can observe how $MaxFGrad$ and $MaxLevel$ are updated for the yellow region once external frontier pixels at intensity level 4 have been associated with the region.

Region growing process is performed as long as there are pixels in $Heap$. Having an empty $Heap$ means that all pixels in the image belong to any region, as shown in Fig. 3 (e). Each time a pixel p is extracted from the $Heap$, it is added to one of its neighbor regions and, consequently, corresponding value of its neighbors in $NeighIm$ will decrease by an unit. Moreover, each of the neighbors of p which earlier belonged to the background will now be part of the external frontier of regions. Consequently, their value in $RegionsIm$ is changed from 0 to -1 and they are added to the $Heap$.

In case a pixel p has more than one neighbor region, p is added to one of the regions and a new parent region is created as the result of merging all neighbor regions (see Fig. 3 (c) and (d)). Children regions' labels are combined into the parent regions's label through $LabelArray$ structure without needing to change pixel labels. In case constituting regions were created at $LumIm[p]$ level, the regions are substituted by the merged region and original regions disappear from the regions tree.

Region merging involves updating both $RegionsInf$ of constituting regions and $LabelArray$ so $PRegion$ of children regions point to the parent. The only change in $RegionsIm$ is related to assigning a region label to pixel p as shown in Fig. 3.

The result of this stage is the final regions tree of the image. The structure of this tree represents both region initialization and region growing stages as initial regions are merged until all pixels in the image are associated to a region. In this case, the remaining region covers all the image and corresponds to root of the regions tree. The full region growing algorithm is described in Algorithms 4, 3 respectively, and 2.

Contrary to some region growing approaches [18], we perform region growing until the whole image is covered in order to avoid stopping on a local maxima of the contrast. This would result on having segmented regions smaller than the ones actually appearing in the image.

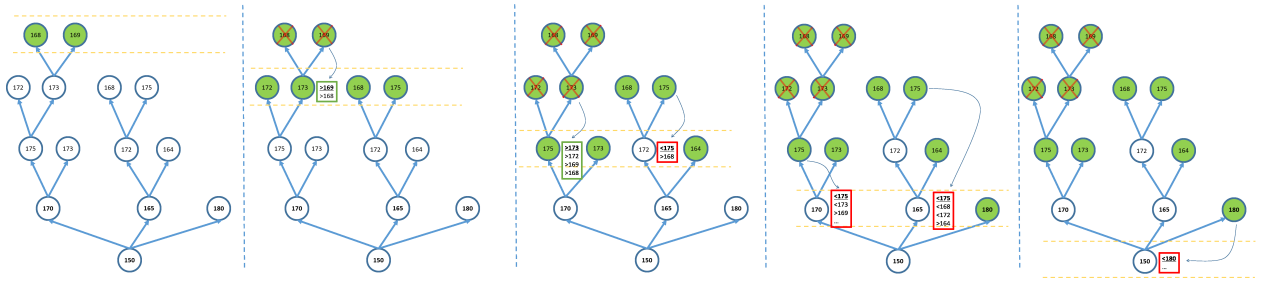


Fig. 4 Synthetic example of region selection. Each circle represents a region and its corresponding local contrast value is placed inside. Green colored regions are those pre-selected by the system, while a red cross over the green colored region represents discarded regions.

4.3.3 Region selection

Although we already have the complete regions tree, our method has still not decided what the final output regions will be. To determine this, the regions tree is recursively explored to keep only those regions which are actual candidates to be bright spot regions. These regions comply with our model of appearance and with the two following criteria: 1) all candidate regions are disjointed pixel sets and 2) its local contrast has to be the maximum achieved during its associated growing process, with this value being equal or higher than $MinGrad$.

The first criteria forces an automatic rejection of the parent region if a children region is selected as method's output. The second one fosters maximum separation between the region and the background; combined with the first one, which ensures selected regions have higher local contrast than all their children regions.

The result of this stage is a regions tree where selected regions are marked. The algorithm checks if all descendants of a region have a lower local contrast value. If this is the case, the region is marked and all its descendants are eliminated. This process starts from the leaves of the tree down to its root. We provide a pseudo-code for this step in Algorithm 5 and we show a synthetical step by step example in Fig. 4.

4.3.4 Region marking

The final step consists of labelling each pixel in the original image according to the selected regions . Considering this, possible values of pixels in the image can go from 0 - if the pixel does not belong to any of the proposed output regions - to N_{reg}, where N_{reg} corresponds to the number of regions provided as result of the previous step. The final step of our algorithm goes through the regions tree and 'paints' in the final output image each of the marked regions.

A flood-fill algorithm - [4] proposes a possible implementation - is used in order to generate full connected regions from local maxima. In our case, the seed for flood-fill algorithm for a particular region is the local maxima of one of the constituent regions. Starting from this seed, only connected pixels with luminance level equal or higher than $MaxLevel$ value of the particular region will be labelled as part of the final region. The label assigned to the pixel depends on the number of marked regions that have already been analyzed.

As can be seen, our approach searches for the regions with higher contrast with respect to the background, regardless of their intensity level and actual shape. The lack of any intensity thresholding allows us to capture bright spots appearing in darker regions of the image. The sole use of gradient information without any kind of shape information allows to focus only on contrast information as well as making it unnecessary to apply re-initialization and shape regularity constraints that are applied in other region growing approaches, especially in active contour models [18]. These methods could lead to similar segmentation results, but to perform well several parameters have to be optimized. Our approach only needs of two inputs apart from original image: $GradIm$ and $MinGrad$. The latter is only used to reduce the number of final regions provided by the method but not for defining their quality. Our approach could also be seen as similar to the concept of Maximally Stable Extremal Regions (MSER) [17,19], although our approach only provides disjointed regions as output and considers gradient information (apart from region content) in the growing process.

4.3.5 Computational efficiency of the method

Our implementation aims at computational efficiency through several design decisions. We add a 1-pixel width frame around both $RegionsIm$ and $NeighIm$ to allow for a more efficient access to pixel values by using pointers instead of coordinates, eliminating all the comparisons related to checking image margins. Displacement arrays are used to accelerate the access to the neighbours of a given pixel. These arrays store the address difference between a pixel and its neighbors.

	Feature		Feature		Feature
1	Mean gray value	9	Max. chromaticity value	17	Min. chromaticity value in boundary
2	Min. gray value	10	Mean gray value in boundary	18	Max. chromaticity value in boundary
3	Max. gray value	11	Min. gray value in boundary	19	Standard deviation of $LumIm$ inside region
4	Mean blue channel value	12	Max. gray value in boundary	20	Standard deviation of $LumIm$ in boundary
5	Min. blue channel value	13	Mean blue channel value in boundary	21	Mean $GradIm$ value in boundary
6	Max. blue channel value	14	Min. blue channel value in boundary	22	Min. $GradIm$ value in boundary
7	Mean chromaticity level	15	Max. blue channel value in boundary	23	Max. $GradIm$ value in boundary
8	Min. chromaticity level	16	Mean chromaticity value in boundary	24	Region size in pixels

Table 4 Features used for discriminating specular highlights from the total of bright regions

As regards computational complexity, our method works in parallel with each region so its complexity can be approximately related to the size of the image ($O(n)$, with n being the number of image pixels). Processing each pixel allows accessing to its neighborhood and potential updates of the region tree. All this process has a linear complexity with respect to the number of image pixels ($O(n)$). Other operations have no linear complexity, as accessing the heap has no linear complexity ($O(log(n_{ext})$, with n_{ext} being the number of external frontier pixels). $LabelArray$ updating complexity depends on the number of local maxima and on region merging progress. As can be seen, it is difficult to provide an exact calculation of the overall complexity as the number of local maxima and external frontier pixels varies on each image. However, experiments on different size images do show a certain degree of overall linear complexity. Processing a high definition (HD) image (size 1920×1080), 28 times larger than a standard definition (SD) one of 384×288, has an associated computation time of 31 times longer (0.8 seconds vs. 0.025 seconds). This puts our method in a complexity level close to other pixel-based methods with linear complexity presented in Section 2.

5 Specular highlight region classification

The output of our segmentation method is a set of bright regions but, as mentioned in Section 1, some of them may not contain a specular highlight. In order to discard them, we propose a region classification stage. This stage is also strongly linked to our model of appearance as it defines a set of model-based features to be extracted from candidate regions. These features are fed to a classifier which will then provide, as a result, the model used to generate the final output of our method.

In Table 4 we present the 24 features that we extract from each region. Features $1-6$ are based on conditions 1 and 3 of our model and explore intensity values within candidate regions. The blue channel is used under the assumption that as diffuse component tends to be reddish in colonoscopy images, specular component might be better observed in the blue channel. Features $7-9$ analyze whether color information is lost within region. As the second condition of our model defines specular highlight regions as having a strong contrast with their neighbor regions, we compute intensity and color-based features for region boundary. This boundary is defined from a 2-width pixel dilation over an area occupied by candidate region. Boundary strength is explored in Features $21-23$ in terms of $GradIm$ information. Finally we also add size information under the assumption that specular highlights tend to be small.

With respect to the choice of the classifier, we have decided to use Support Vector Machines (SVM) [13] for the sake of the simplicity of use, resource efficiency and considering potential real-time implementation of our methodology. In our case, SVM will perform a binary classification between specular highlights and the rest of bright regions. We use LIBLINEAR implementation of SVM [10] running under Matlab 2015a.

We fed the classifier with feature vectors extracted from candidate regions, in which positive examples are linked to specular highlight regions and negative ones with the rest of bright regions. Labelling of regions into positive and negative is done by comparison their degree of overlap with ground truth specular region, computed using DICE similarity score [29]. We define two parameters, pos_{thresh} and neg_{thresh} to cluster candidate regions into positive and negative examples. These values are not necessarily adjacent, for instance, we could set very restrictive values (high pos_{thresh} and low neg_{thresh}) to train the classifier with really strong examples. We perform k-fold cross validation to avoid dependence on the results on a specific dataset division.

6 Experimental setup

6.1 Validation database and performance metrics

We introduce our **CVC-ClinicSpec** database which, to the best of our knowledge, is the first one to be publicly available. **CVC-ClinicSpec** annotated specular highlights ground truth of the 612 images from CVC-Clinic database [6]. **CVC-ClinicSpec** database contains 25914 specular highlights regions of different sizes and appearances. Table 5 shows some statistics about

Size	Number of regions	Mean size and StD
1-10	13885 [57.43%]	5.31 ± 002.54
11-50	8700 [33.57%]	21.72 ± 010.07
51-100	1356 [05.28%]	69.00 ± 013.73
101-500	907 [03.50%]	187.72 ± 086.46
> 500	54 [00.20%]	795.94 ± 501.85
All	25914 [100.00%]	22.22 ± 57.53

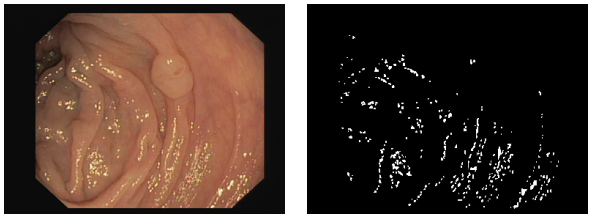
Table 5 Facts and figures of **CVC-ClinicSpec** database. StD stands for standard deviation.

these regions. More than half of them are quite small ($size < 10$ pixels). Only about 1000 regions [3.70%] could be considered as large size (> 100 pixels).

Ground truth was performed by an expert, who manually marked specular pixels in each image using a graphic tablet. The expert was trained by showing examples labelled from CVC-COLONDB database [8]. Fig. 5 shows an example of database content. Ground truth is a binary image in which white pixels correspond to specular pixels in the image.

We assessed the performance specular highlight detection methods using standard pixel-wise and region-wise performance measurements. A good performing method should not only detect all specular regions, but it also has to segment them accurately. To calculate these measurements, we assume the output of a given method to be a binary mask in which white pixels correspond to specular highlights. We compare method output with ground truth to define performance metrics.

As regards pixel-wise metrics, True Positive Pixels (TPP) are those marked by the method and present in the ground truth. False Positive Pixels (FPP) are those marked by the method but present in the ground truth. False Negative Pixels (FNP) are those absent in methods' output, but present in the ground truth. Region-wise metrics are defined similarly. It is important to mention that we only ask a given method to provide one single ground truth pixel to consider its corresponding region as detected (True Positive Regions (TPR)). We propose to use Precision, Recall and F1-score for pixel and region-wise performance metrics. We do not compute True Negative Pixels or Regions, as positive and negative classes are not balanced. Consequently, the use of related metrics (Specificity, Accuracy) would not be representative of the actual method performance.



(a) (b)

Fig. 5 Example of **CVC-ClinicSpec** database. (a) shows an original image and (b) its corresponding ground truth. White pixels in (b) correspond to specular highlights pixels.

For our experiments to reflect segmentation quality, we propose using a DICE similarity score. It is important to mention that in all cases we provide results per region. Using mean values in these kind of experiments may not provide a fair representation of the performance of a method in a way such that an isolated, badly segmented region in an image would impact negatively on the results of other better segmented ones.

6.2 BSSC parameter setting

Our bright spot regions segmentation (BSSC) method requires two input images, and one parameter. In the experiments presented in this paper, we use the gray-scale version of CVC-ClinicDB images as $LumIm$, the module of Sobel gradient over $LumIm$ as $GradIm$ and we set $MinGrad$ to 100 assuming that relevant bright spot regions should be highly contrasted with respect to their neighborhood.

As for data preparation for region classification, we decided to set pos_{thresh} to 40% and neg_{thresh} to 0% so the classifier is fed with good specular highlight regions as positive examples and strictly non-specular pixel containing regions as negative examples. Finally, we propose to use 4-fold cross-validation, as such that, for each fold, 75% of the images will be used for building the model and the remaining 25% for validation.

6.3 Comparative study

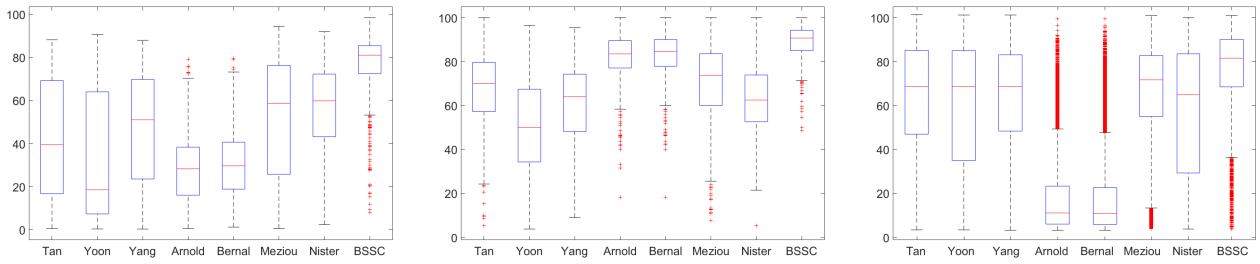
A comparative study is presented in this paper with the methods that are shown in Table 6. Inclusion criteria is based on the availability of source code to perform a fair comparison in a same database. We included an active contour model-based segmentation method in the comparison [18] and an implementation of MSER region segmentation algorithm based on [19] in order to compare our approach with other region-based approaches using the same initial seeds. For all cases we have used

Method	Domain	Information	Approach
Tan et al. [24]	General	Intensity	Pixel-based
Yoon et al. [28]	General	Intensity	Pixel-based
Yang et al. [27]	General	Intensity	Pixel-based
Arnold et al. [3]	Colonoscopy	Intensity	Pixel-based
Bernal et al. [9]	Colonoscopy	Intensity	Pixel-based
Meziou et al. [18]	General	Intensity	Region-based
Níster et al. [19]	General	Intensity	Region-based
BSSC	Colonoscopy	Intensity + Gradient	Region-based

Table 6 Methods analyzed in the comparison study.

	Pixel-wise analysis						*Region-wise analysis*					
Method	TPP	FPP	FNP	Prec	Rec	F1	TPR	FPR	FNR	Prec	Rec	F1
Tan et al.	421017	1960577	154818	17.68%	73.11%	28.47	20733	12412	5181	62.55%	80.01%	70.21
Yoon et al.	491752	6411019	84083	7.12%	85.40%	13.15	24152	49094	1762	32.97%	93.20%	48.71
Yang et al.	385796	1181934	190039	24.61%	67.00%	36.00	20876	19008	5038	52.34%	80.56%	63.45
Arnold et al.	537267	3128062	38568	14.66%	93.30%	25.34	22373	4542	3541	83.12%	86.34%	84.70
Bernal et al.	543483	2727902	32352	16.61%	94.38%	28.25	22707	4416	3207	83.72%	87.62%	85.63
Meziou et al.	405729	1152364	170106	26.04%	70.45%	38.02	16202	2776	9712	85.37%	62.52%	72.17
Níster et al.	393354	421989	182481	48.24%	68.31%	56.54	11909	1222	14005	90.69%	45.95%	61.00
BSSC	454269	173794	121566	72.33%	78.89%	75.47	22779	2273	3135	90.93%	87.90%	88.49

Table 7 Summary of specular highlight detection results over **CVC-ClinicSpec** ground truth.



(a) (b) (c)

Fig. 6 Boxplots showing distribution of (a) pixel-wise and (b) region-wise F1 scores per image obtained by each method. DICE scores per detected region are shown in (c).

the parameter setting proposed in corresponding papers. As regards MSER, for the cases in which provided regions are not disjoint between them, we opted to provide the one resulting from the intersection of all overlapped regions as final region, aiming to foster coincidence between potential output regions and to potentially reduce over-segmentation.

We propose an F1-score for region-wise and pixel-wise analysis, and DICE score for segmentation quality analysis as comparison metrics. For the F1-score, we treat each individual frame as a different study, and we compare F1-scores for each study, with this score being calculated from all the regions present in an image. As the number of detected regions may be different for each method, DICE comparison makes us deal with unpaired data, and also defining the statistical tests to be used.

In order to account for statistically significant differences in methods' performance, a Saphiro-Wilk test is performed to find out if available data follows a normal distribution. If this is the case, statistically significant differences between methods are assessed using an analysis of variance (ANOVA) over F1-score or DICE values. In case of non-normal data distribution , a non-parametric test (Kruskal-Wallis) test is used. Finally, in order to account for pair-wise statistical differences, we perform a Wilcoxon rank-sum test for paired data and Mann-Whitney U-test for unpaired data. All tests are performed using a confidence level $1-\alpha = 0.95$.

7 Experimental Results

7.1 Pixel and region-wise results

Pixel-wise performance results are shown in Table 7. A significant difference can be observed between the performance of our method and the rest of approaches, which is particularly notable in the case of Precision results. Fig. 6 (a) shows a boxplot showing the distribution of F1 scores per image obtained by each of the methods, where we can again observe how our method outperforms the rest of approaches.

Saphiro-Wilk test over F1 scores rejected the null hypothesis that data follows a normal distribution ($p-value = 0$) so we run a Kruskal-Wallis test which indicated differences between the methods analyzed, rejecting the null hypothesis ($p-value < 0.05$). Pair-wise differences were explored using a Wilcoxon rank-sum test which confirmed that our proposal shows significant differences with the rest of the methods ($p-value < 0.05$).

Table 7 shows region-wise performance of the different methods. We observe a performance gap between general and colonoscopy methods, with the main difference being related to the number of FPR, which is higher for general methods. Regarding colonoscopy methods, our approach shows the better balance between TPR, FPR and FNR, which can be better observed by differences in F1 scores. In Fig. 6 (b) a boxplot is presented, showing the distribution of F1 scores per image for each method, where we can observe the better overall performance achieved by our method.

Size	*All sizes* (25914 regions)		1 – 10 (14883 regions)		11 – 50 (8700 regions)	
Method	Detected regions	Mean DICE ± StD	Detected regions	Mean DICE ± StD	Detected regions	Mean DICE ± StD
Tan et al.	20733 [80.00%]	61.90 ± 26.80	10077 [67.70%]	62.44 ± 24.36	8331 [95.75%]	60.62 ± 27.64
Yoon et al.	24152 [93.20%]	56.17 ± 32.47	13190 [88.62%]	57.33 ± 30.38	8633 [99.23%]	55.80 ± 33.58
Yang et al.	20876 [80.55%]	62.39 ± 24.53	10279 [69.06%]	62.40 ± 23.50	8272 [95.08%]	60.76 ± 25.06
Arnold et al.	22373 [86.33%]	13.66 ± 15.33	11858 [79.67%]	06.54 ± 06.75	8218 [94.45%]	17.17 ± 14.15
Bernal et al.	22707 [87.62%]	13.50 ± 15.47	12149 [81.63%]	06.39 ± 06.67	8261 [94.95%]	16.92 ± 14.07
Meziou et al.	16202 [62.52%]	64.43 ± 22.13	5892 [39.58%]	58.98 ± 20.29	7992 [91.86%]	66.01 ± 21.55
Níster et al.	11909 [45.95%]	55.34 ± 30.26	3838 [25.78%]	30.83 ± 24.37	5862 [67.37%]	63.76 ± 26.68
BSSC	22779 [87.90%]	74.72 ± 18.57	12133 [81.52%]	69.50 ± 19.49	8434 [96.94%]	79.92 ± 15.20

Size	51 – 100 (1370 regions)		101 – 500 (907 regions)		> 500 (54 regions)	
Method	Detected regions	Mean DICE ± StD	Detected regions	Mean DICE ± StD	Detected regions	Mean DICE ± StD
Tan et al.	1365 [99.63%]	63.20 ± 31.73	906 [99.88%]	65.34 ± 32.71	54 [100.00%]	67.38 ± 28.05
Yoon et al.	1369 [99.92%]	51.71 ± 38.32	906 [99.88%]	49.96 ± 39.50	54 [100.00%]	51.29 ± 34.22
Yang et al.	1365 [99.63%]	66.84 ± 26.05	907 [100.00%]	70.08 ± 26.36	53 [98.14%]	68.63 ± 23.88
Arnold et al.	1351 [98.61%]	33.01 ± 19.35	894 [98.56%]	44.10 ± 21.38	52 [96.29%]	57.55 ± 21.83
Bernal et al.	1351 [98.61%]	33.13 ± 19.79	894 [98.56%]	45.87 ± 21.63	52 [96.29%]	62.33 ± 19.84
Meziou et al.	1359 [99.19%]	71.84 ± 24.42	905 [99.77%]	73.96 ± 25.73	54 [100.00%]	77.88 ± 24.16
Níster et al.	1294 [94.45%]	73.47 ± 20.97	868 [95.70%]	78.22 ± 16.70	47 [87.03%]	84.38 ± 10.15
BSSC	1320 [96.35%]	82.61 ± 16.22	847 [93.38%]	85.11 ± 14.80	45 [83.33%]	79.34 ± 25.94

Table 8 Summary of specular highlight region segmentation results, broken down according ground truth region size. StD stands for standard deviation.

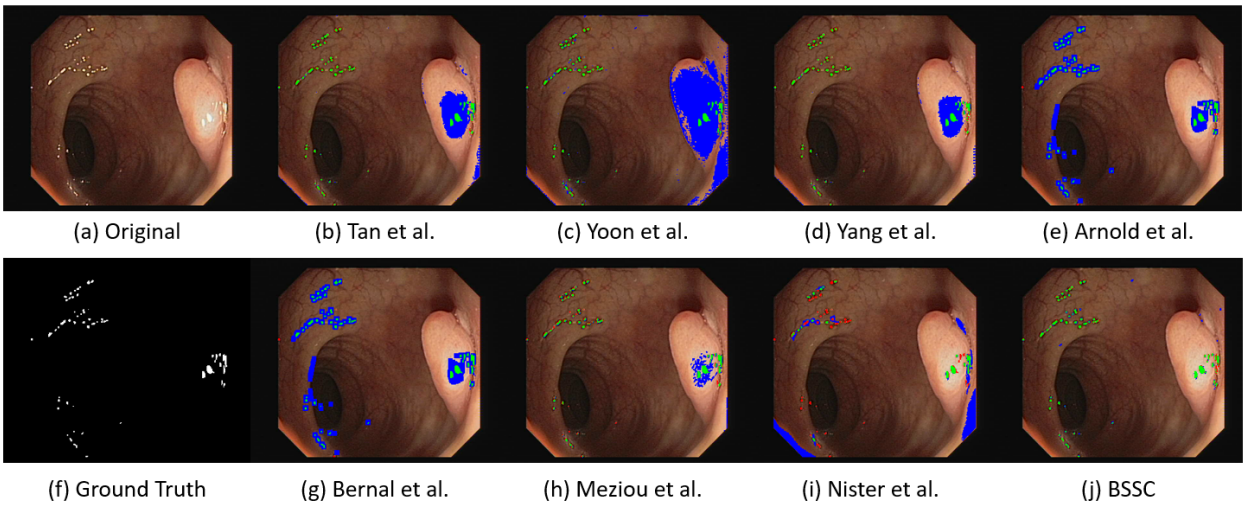


Fig. 7 Qualitative methods' comparison. (a) Original image and (d) corresponding ground truth, (b) Tan et al., (c) Yoon et al. , (d) Yang et al., (e) Arnold et al., (g) Bernal et al., (h) Meziou et al., (i) Níster et al. and (g) BSCC (our proposal). True Positive Pixels are painted in green, false positives pixels in blue and false negatives pixels in red.

Saphiro-Wilk test over F1 scores rejected the null hypothesis that data follows a normal distribution ($p-value < 0.05$). Kruskal-Wallis test also rejected the null hypothesis that all data comes from the same distribution ($p-value = 0$). Finally, pair-wise differences were explored by means of a Wilcoxon rank-sum test, which provided significant differences between BSSC and the other methods methods ($p-values < 0.05$).

7.2 Segmentation quality results

We present overall segmentation results in Table 8, and boxplot showing distribution of DICE scores per region in Fig. 6 (c). A mean and standard deviation of DICE scores is presented for each method. We also present a breakdown of the results according to region size.

Results indicate that our proposal is the one which provides higher DICE scores when all regions are considered. We can also observe from Table 8 that there is difference in performance scores according to the size of the target region. On the one hand, our proposal achieves higher DICE scores in all categories but provides lower detection rates for larger regions. On the other hand, general methods increase the detection rate for those larger regions. There are larger differences across methods happen for small regions in which detection rates and DICE scores experience a decrease. Apart from differences related to the methodologies, differences in performance for all methods can be asso-

Method	Detected regions	Mean DICE ± Std
Tan et al.	2895 [86.00%]	57.23 ± 30.07
Yoon et al.	3200 [95.06%]	50.23 ± 36.05
Yang et al.	2922 [86.80%]	60.44 ± 27.48
Arnold et al.	2926 [86.92%]	23.44 ± 19.82
BSV2013	2950 [87.64%]	23.21 ± 19.81
Meziou et al.	2765 [82.14%]	66.02 ± 25.00
Níster et al.	1920 [57.04%]	66.98 ± 26.12
BSSC	3010 [89.42%]	80.10 ± 16.07

Table 9 Summary of specular highlight region detection and segmentation results, considering only regions inside polyps.

ciated with the difficulty in manually marking specular highlights pixels, as potential small errors - of even just one pixel - in ground truth creation impact higher on the performance scores for small regions.

A statistical analysis was performed on DICE scores for all detected regions. The Saphiro-Wilk test rejected that data comes from normal distribution ($p-value < 0.05$) and posterior Kruskal-Wallis test confirmed that not all methods share a same data distribution ($p-value = 0$). Again, pair-wise differences were further explored using a Mann-Whitney U-test for unpaired data. Results indicate statistical differences between BSSC and the rest of approaches ($p-value = 0$).

All three validation experiments shows a quantitative and statistically significant difference between the performance of our methodology and both general and colonoscopy-specific approaches. In case of similar region-wise detection score, BSSC is the one that provides more accurate regions which can be observed by higher pixel-wise Precision and DICE scores.

In order to better illustrate these differences, we show a graphical comparison between compared methods in Fig. 7. In this image, we can observe how the majority of the methods are able to correctly detect the specular regions but we can also observe that some of them are impacted by other bright regions not part of a specular highlight and that, for the case of colonoscopy-specific methods, they do not accurately segment the specular region, providing a number of false positives.

Regarding the comparison between pixel-based approaches, Yang et al. is the one which offers a best compromise between number of detected regions and DICE score whereas between region-based approaches, our method outperforms the studied level sets approach, specially on pixel-wise Precision.

Taking into account, as mentioned in Section 1, that we foresee specular highlights detection as part of a bigger polyp characterization system, we consider it of interest to provide a break down of the results only considering specular highlights inside polyps. Specular highlights detection could be used to determine which image areas should not be used in later processing stages. In this context, quality of region segmentation is crucial, because if the detected region covers more pixels than the actual specular region, pixels with relevant texture information may be discarded.

Table 9 shows a summary of detection and segmentation results. The first conclusion that can be drawn, is that overall detection score improves for all methods when only specular highlights inside polyp regions are considered. With regard to the quality of the segmented regions, DICE score improves for colonoscopy-specific methods but gets worse for general methods. It is interesting to observe how our method is able to outperform similar region growing approaches. Therefore, the combined use of intensity and gradient-based features appear as a suitable solution for this problem. Again, our proposal achieves the highest DICE score appearing as the solution which would allow to keep more relevant polyp region information and therefore it is the most suitable to be included as part of a global polyp characterization computational support system.

8 Discussion

In this section we discuss some topics that have arisen after analyzing our methodology and the results of the comparison study. More precisely, we tackle the impact of the decisions regarding parameter tuning of our approach, the impact that ground truth creation may have on method performance, including the consideration of which of the regions have to be labelled and, finally, the clinical use of a specular highlight detection method by assessing its impact in two specific applications.

8.1 Impact of BSSC configuration on performance

In Section 6 we presented the specific parameter configuration of our method used in the comparison study. We set those parameter values based on our model of appearance, aiming at a strong distinction between positive and negative examples and the presence of high contrast between specular highlight regions and their neighbor regions. We analyze here the impact of each individual decision on system configuration.

As regards the output of our bright spot regions segmentation method, we decided to set $MinGrad$ to 100 aiming to eliminate those weak regions still remaining on the regions tree. This value only affects the number of positive and negative regions that are fed to the classification stage but not the quality of the segmented regions, as shown in Table 10. This happens as result of our segmentation strategy, as regions kept after the region selection stage are disjointed between them.

Even by setting $MinGrad$ to a high value, the mean size of discarded regions remains lower than 11

	Method regions			*Non-reachable ground truth regions*	
MinGrad	**Positive**	**Mean DICE**	**Negative**	**Regions**	**Mean area**
0	28027	67.77	124686	101	02.60
10	27303	69.23	523723	171	03.35
20	27237	69.36	224938	180	03.57
30	27117	69.57	116824	219	05.70
40	26987	69.77	75415	269	11.04
50	26775	70.04	54728	381	11.62
60	26518	70.32	42040	542	11.14
70	26196	70.63	33444	773	10.40
80	25787	71.00	26866	1086	10.70
90	25322	71.41	21655	1462	11.24
100	24779	71.84	17580	1914	11.05

Table 10 Impact of *MinGrad* value on the number and quality of regions provided by the bright region segmentation stage. BSCC optimal configuration sets 100 as *MinGrad* value.

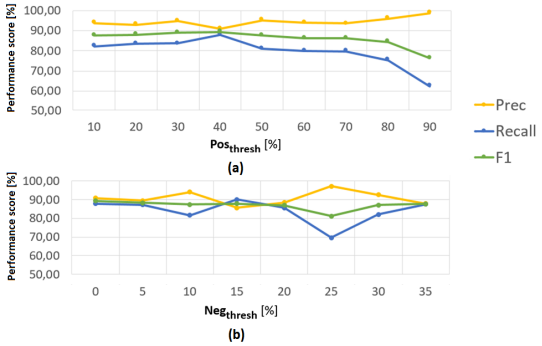


Fig. 8 Impact of pos_{thresh} and neg_{thresh} values in region-wise results of our proposal. BSCC optimal configuration sets 40 as pos_{thresh} value and 0 as neg_{thresh} value. Vertical axis represents the value for each of the three studied metrics (Precision, Recall and F1)

pixels, which implies that losing those regions would have little effect on the subsequent stages of the processing pipeline, such as textural structure recognition. Consequently, *MinGrad* values also sets the maximum performance that our method may achieve. The lower *MinGrad*, the higher the number of negative examples which could cause the introduction of some bias into the classifier. We obtain better performance scores with higher *MinGrad* values as indicated in Table 11. In this case we obtain higher Recall scores at the cost of slightly lowering Precision.

Once segmented regions are determined by *MinGrad*, we have to cluster them into positive and negative regions to train the classifier. This decision is based on an overlap coefficient over ground truth and we set $pos_{thresh} = 40\%$ and $neg_{thresh} = 0\%$, aiming to train the model with very low quality negative regions and adequate positive ones. In Fig. 8 (a) we show the results of a quantitative study on the effect on F1 score of pos_{thresh}. As expected, the lower pos_{thresh}, the higher the Recall. As pos_{thresh} goes beyond 40%, we can observe a clear decreasing trend on Recall score (and consequently, F1-score). In this case the classifier is being fed with very strong positive regions which helps in reducing the number of false positives at the cost of imposing very strong restrictions on what a specular highlight region should be.

As regards neg_{thresh}, and assuming that its value must never surpass pos_{thresh}, Fig. 8 (b) shows the results of a separate experiment in which we keep $pos_{thresh} = 40\%$ and vary neg_{thresh} from 0% to 35% in 5% steps. The results do not seem to follow any general trend, except for the F1-score, which seems to decrease as negative threshold decreases.

	Pixel-wise analysis			*Region-wise analysis*		
MinGrad	Prec	Rec	F1	Prec	Rec	F1
0	67.06%	77.35%	71.84	92.49%	83.67%	87.73
10	67.19%	77.47%	71.96	92.49%	83.67%	87.86
20	68.26%	77.69%	72.67	93.03%	83.51%	88.01
30	67.36%	78.90%	72.67	91.01%	85.49%	88.16
40	69.76%	78.99%	74.09	93.58%	84.29%	88.69
50	67.24%	79.21%	72.74	91.48%	85.53%	88.40
60	69.96%	78.02%	73.77	92.98%	83.40%	87.93
70	67.27%	79.92%	73.05	87.77%	87.70%	87.73
80	67.62%	79.45%	73.06	86.77%	87.26%	87.01
90	71.81%	78.63%	75.07	92.33%	85.53%	88.80
100	72.33%	78.89%	75.47	90.93%	87.90%	89.39

Table 11 Impact of *MinGrad* value on performance scores. BSCC optimal configuration sets 100 as *MinGrad* value.

8.2 Impact of ground truth creation

We mentioned in Section 1 that the lack of publicly available databases limits the comparison between different approaches. We associated this to the difficulty on providing accurate annotations of specular highlights and we discuss here the impact on performance of two annotation-related issues: 1) differences among observers and 2) influence of other illumination artifacts.

With respect to the former, it has to be noted that ground truth creation is a highly time-consuming and highly precision requiring task. Pixel-wise labelling of the whole database took a month and, due to this, it was not feasible to ask additional experts to label all images to account for inter-observer variability. Nevertheless, considering the importance of pixel-wise performance in terms of assessing information loss for later processing stages, differences in ground truth creation may impact performance scores, though they will affect all compared methods the same. For instance, in regions smaller than 10 pixels - prominent in our database -, an error of 1 pixel can affect Precision results in a 10%.

We asked two additional experts to label the first 100 images of our database to study the impact of annotation quality in method performance. Our experiment consisted of using one of the experts as ground truth and comparing the output provided by the rest of experts to account for differences in ground truth creation

Method	*Pixel-wise*			*Region-wise*		
	Prec	Rec	F1	Prec	Rec	F1
*Ground truth = **CVC-ClinicSpec** [3236 regions]*						
Exp1	41.26	94.37	57.42	85.83	96.26	90.75
Exp2	99.89	99.21	99.55	99.50	99.62	99.56
BSSC	62.99	76.94	69.27	95.31	85.56	90.18
Ground truth = Exp1 [2639 regions]						
Exp2	94.57	41.06	57.26	94.04	80.14	86.53
CVC-ClinicSpec	94.37	41.26	57.42	94.61	80.52	87.00
BSSC	73.93	39.48	51.47	95.79	75.18	84.24
Ground truth = Exp2 [3244 regions]						
Exp1	41.06	94.57	57.26	85.58	95.86	90.43
CVC-ClinicSpec	99.21	99.89	99.55	99.62	99.50	99.56
BSSC	62.65	77.04	69.10	94.90	85.01	89.69

Table 12 Impact of ground truth creation in pixel-wise performance results. Exp1 stands for independent expert 1, Exp2 for independent expert 2 and **CVC-ClinicSpec** stands for ground truth used in the validation framework.

and methods' performance. Table 12 shows quantitative results of this experiment. The study shows that there is a difference in the number of regions marked by each expert and whether this difference affects performance scores or not. For instance, we observe that Exp1 annotated less regions. If we validate our method against Exp1, it achieves lower region-wise Recall score as the number of detected regions is reduced. In the pixel-wise analysis, we observe great similarity between Exp2 and **CVC-ClinicSpec** and big differences of them with Exp1. Precision and Recall scores might indicate that if a given region is marked by three experts, Exp1 would include more pixels than the others.

As mentioned in Section 3, bright spot regions may include those which are not specular highlights. Among these are overexposed areas; experts did not label them as specular highlights as they do not strictly fulfill their model of appearance, especially regarding contrast between the region and all its neighbors. Nevertheless, some of the compared methods treat them as specular highlights.

In order to isolate the impact of overexposed regions labelling and for the sake of providing a fair comparison between methods and avoiding potential discrepancies on whether overexposed regions should be included or not in the ground truth, we performed an additional validation experiment excluding images with overexposed regions - 178 images in this case -. We show a summary of the results on Table 13. We can observe how performance scores increase for all metrics for the majority of the methods. The biggest increase is observed in pixel-wise Precision, which can be interpreted as methods having difficulties to accurate label highlight pixels included in overexposed regions. Fig. 9 shows examples of the output of BSSC in images with and without overexposed regions confirming our previous hypothesis.

Method	*Pixel-wise*			*Region-wise*		
	Prec	Rec	F1	Prec	Rec	F1
Tan et al.	10.06	-2.80	11.31	1.35	0.01	0.84
Yoon et al.	1.76	-2.61	2.91	3.90	-0.58	4.04
Yang et al.	16.55	-0.51	14.84	3.99	1.78	3.44
Arnold et al.	3.23	1.92	4.79	0.19	1.72	0.92
Bernal et al.	2.45	1.96	3.58	0.16	1.67	0.88
Meziou et al.	23.13	-0.70	19.65	2.61	2.69	2.72
Níster et al.	3.70	2.88	3.52	1.11	1.39	1.47
BSSC	8.30	2.14	5.36	0.28	1.71	1.02

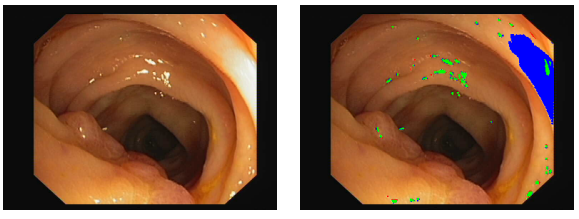
Table 13 Difference in performance associated with the absence of overexposed regions. Values indicate difference between methods' performance in the full dataset (612) against performance in a subset of images without overexposed regions (434 images).

As performance of detection methods is affected by overexposed regions, their apparition should be controlled either on clinicians' side by correcting scene lighting or on the technical side by an automatic detection of overexposed regions presence, discarding images which contains them for later processing.

8.3 Clinical applicability of specular highlights detection and bright spot regions segmentation

We mentioned in Section 1 that specular highlights detection can be part of an intelligent system for colonoscopy, playing a supporting role in lesion detection [6] or characterization tasks [12]. Regarding the former, specular highlights detection can be used in two ways. First, to indicate lesion presence, especially in large protruding polyps in which specular highlights appear larger as they reflect a higher amount of light back to the camera. Second, to eliminate potential false positive detections, as specular highlights appearance is prone to attract the attention of RoI detectors.

The positive impact of specular highlights detection in polyp detection has already been assessed [6] but, to the best of our knowledge, there is no mention about their use as part of a polyp characterization system. As mentioned in Section 1, clinicians perform visual inspection of the polyp surface to predict lesion histology. Several paradigms have been proposed but, for the case



(a) (b)

Fig. 9 Impact of overexposed reigons on method performance: (a) Original images and (b) BSSC output. True positive pixels are painted in green, false positives pixels in blue and false negatives pixels in red.

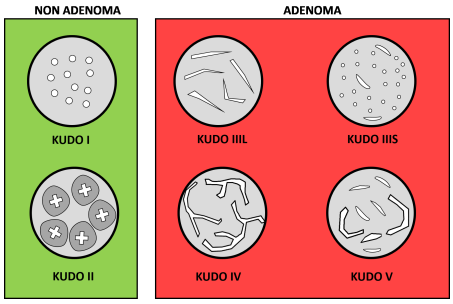


Fig. 10 Synthetical representation of the different types proposed by KUDO's pit pattern classification. Types I and II correspond to non-adenoma lesions whereas types III to V correspond to adenomatous ones.

of white light colonoscopy, KUDO's pit pattern classification [15] is one of the ones most commonly used. This classification proposes a gross classification of pit patterns into several types which have a correlation with the degree of malignancy of the lesion. Fig. 10 shows a synthetical representation of KUDO classification.

As can be observed, the identification of these patterns is based on the fact that they present a contrast with respect to their neighboring regions and they can appear and be considered as bright regions in the image. We can also see how both specular highlights and textural patterns appear as bright regions in the image, with specular highlights being more likely to attract the attention of computational methods. In this context, we foresee specular highlights detection as a way to remove those regions which could lead to providing an incorrect automatic prediction of polyp histology.

Fig. 11 shows a potential computational diagnosis error caused by the apparition of specular highlights. If our system only relied on the shape of the bright regions, a high presence of tubular patterns would be assigned to the polyp, thus indicating an adenomatous histology when, in fact, there are no prominent tubular structures on its surface, and the polyp is actually a non-adenomatous one. Therefore, although our bright spot regions segmentation could be used to extract those textural patterns (textons), we also have to consider firstly removing those which belong to specular highlights.

We present here the results of a preliminary study of the potential of using our bright spot regions segmentation to both detect specular highlights and characterize remaining textons, as part of a CAD system to obtain an automatic prediction of polyp histology. Our proposed pipeline has two stages: specular highlights detection and texton characterization from bright spot regions segmentation. It should be noted that we remove specular highlights first and then we again apply bright spot regions segmentation to the image without specular highlights. In the second time our segmentation approach is applied, we use as $LumIm$ the image resulting after applying a top-hat filter to the original image. Consequently, $GradIm$ corresponds to the gradient of the top-hat filtered image. Finally, it is important to mention that, contrary to what we do for polyp detection, we do not apply any kind of inpainting [9] to restore image regions below the highlights as in this case it would create new textural elements which could affect methods' performance. We just ignore the output within regions covered by them.

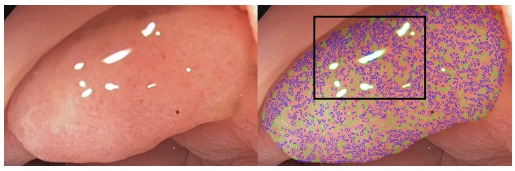


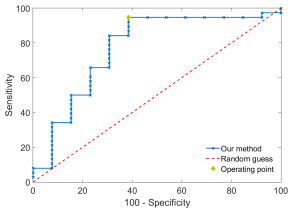
Fig. 11 Potential impact of specular highlights over polyp region on in-vivo polyp histology. The image shows how both textural patterns appear as bright regions in the image, with specular highlights being more salient.

Once specular highlights are removed, we characterize bright regions using the following tubularity formula $Tub = \frac{Area^2}{d}$, where $Area$ equals the number of pixels of the region and d the sum of the Euclidean distance from each region pixel to the region contour. Tubularity is designed to obtain low values for circular shapes and high values for tubular regions of the same area. We use the mean of all tubularity values for an image to classify it into two classes: Adenoma (malignant) and Non-Adenoma (benign). As shown in Fig. 10, the main difference between these groups is related to the appearance of tubular structures in adenomas. Consequently, the use of tubulary metrics over good segmented textons has the potential to obtain a first distinction between these two histological groups.

We compare the histology predicted by our system and the actual histology obtained after lesion removal. A total of 51 HD images showing a different polyp were collected in the Hospital Clinic, Barcelona, Spain. These images were selected to show as much variability in polyp appearance as possible. Experts provided annotations of the polyp regions using a GUI interface. Once the lesion was removed, histological analysis was performed and the images were classified into the two mentioned groups: Non Adenoma and Adenoma.

We present the ROC curve in Fig. 12 (a) and a confusion matrix in Fig. 12 (b) to show the performance of the system. An optimal threshold value of $Tub = 13.14$ to separate the two classes, adenoma vs. non adenoma, was selected from the operating point of the ROC curve.

Results show that bright spot regions segmentation and the use of tubularity metric have potential to obtain an accurate prediction of polyp histology. Our method is able to provide an accurate 'diagnosis' in 44 out of



(a)

n=51	Predicted: Adenoma	Predicted: Non-adenoma
Actual: Adenoma	36	2
Actual: Non-adenoma	5	8

(b)

Fig. 12 Results of our automatic histology prediction system. (a) Receiver Operating Curve (Area Under Curve = 0.77) (b) Confusion matrix.

51 (86.3%) images. If we used this prediction to determine whether the lesion has to be removed or not, we can observe that only in 2 cases (3.9%) using our method would incorrectly suggest leaving the lesion in the colon. Finally our approach achieves a 94.7% of accuracy when dealing with adenomatous polyps as it provides 36 correct predictions out of a total of 38 adenomatous images. This accuracy level decreases when non-adenomatous polyps are considered with 8 (61.5%) correct detections out of 13 images. Finally, as regards the potential use of our technology by clinicians, full processing of a single HD image takes 2.7 seconds, making its use feasible in the examination rooms as real-time constraints are not compulsory for this application.

9 Conclusions and Future work

In this paper, we have presented a methodology for specular highlights detection in the context of colonoscopy image analysis. Our proposal is based on an appearance model defining specular highlights as highly contrasted regions containing a local image maxima. Our method consists of two stages, bright spot regions segmentation and region classification. The former provides a shape and intensity-independent image partition into a set of disjointed regions that present the highest level possible of contrast achieved during their individual growth process. Segmented regions are described using model-related features and are fed to a binary classifier to discard non-specular regions.

We validate our method in a new database, to be publicly available, and we compare its performance against general and colonoscopy-specific detection methods. The results show that our approach outperforms all methods, especially regarding pixel-wise Precision. Regions provided by our method are closer to the actual specular regions and their removal would lead to discard less potentially relevant image information.

Our methodology has been tested successfully in a specific domain of application, but future work should investigate the potential of applying our methodology - especially bright spot regions segmentation - in more general application domains. We have shown a preliminary study in which specular highlights detection is included as part of a global computational support system, but this study might be extended with new cases and covering other clinical applications. Finally, an in-depth analysis of the validation experiments suggest that future comparative studies could be improved to tackle some of the topics raised up in the discussion, such as dealing with the presence of overexposed regions or considering ground truth created by several observers to account for inter-observer variability.

ACKNOWLEDGEMENTS

This work was supported by the Spanish Government through the funded project iVENDIS (DPI2015-65286-R), by the FSEED and by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya, 2014-SGR-1470 and 2014-SGR-135.

References

1. Alsaleh, S.M., Aviles, A.I., Sobrevilla, P., Casals, A., Hahn, J.K.: Adaptive segmentation and mask-specific sobolev inpainting of specular highlights for endoscopic images. In: Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the, pp. 1196–1199. IEEE (2016)
2. Angelopoulou, E.: Specular highlight detection based on the fresnel reflection coefficient. In: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pp. 1–8. IEEE (2007)
3. Arnold, M., Ghosh, A., Ameling, S., Lacey, G.: Automatic segmentation and inpainting of specular highlights for endoscopic imaging. Journal on Image and Video Processing **2010**, 9 (2010)
4. Asundi, A., Wensen, Z.: Fast phase-unwrapping algorithm based on a gray-scale mask and flood fill. Applied optics **37**(23), 5416–5420 (1998)
5. Bernal, J., Gil, D., Sánchez, C., Sánchez, F.J.: Discarding non informative regions for efficient colonoscopy image analysis. In: International Workshop on Computer-Assisted and Robotic Endoscopy, pp. 1–10. Springer (2014)
6. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Medical Imaging and Graphics **43**, 99–111 (2015)
7. Bernal, J., Sánchez, F.J., Rodríguez de Miguel, C., Fernández-Esparrach, G.: Screening for Colorectal Cancer with Colonoscopy, vol. 1, chap. Building up the Future of Colonoscopy - A Synergy between Clinicians and Computer Scientists, pp. 109–141. InTech (2015)
8. Bernal, J., Sánchez, J., Vilarino, F.: Towards automatic polyp detection with a polyp appearance model. Pattern Recognition **45**(9), 3166–3182 (2012)

9. Bernal, J., Sánchez, J., Vilarino, F.: Impact of image preprocessing methods on polyp localization in colonoscopy frames. In: Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, pp. 7350–7354. IEEE (2013)
10. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: Liblinear: A library for large linear classification. The Journal of Machine Learning Research **9**, 1871–1874 (2008)
11. Fernández-Esparrach, G., Bernal, J., López-Cerón, M., Córdova, H., Sánchez-Montes, C., Rodríguez, d.M.C., Sánchez, F.: Exploring the clinical potential of an automatic colonic polyp detection method based on the creation of energy maps. Endoscopy (2016)
12. Hafner, M., Brunauer, L., Payer, H., Resch, R., Gangl, A., Uhl, A., Wrba, F., Vécsei, A.: Computer-aided classification of zoom-endoscopical images using fourier filters. IEEE Transactions on Information Technology in Biomedicine **14**(4), 958–970 (2010)
13. Hearst, M.A., Dumais, S.T., Osman, E., Platt, J., Scholkopf, B.: Support vector machines. Intelligent Systems and their Applications, IEEE **13**(4), 18–28 (1998)
14. Iakovidis, D.K., Koulaouzidis, A.: Software for enhanced video capsule endoscopy: challenges for essential progress. Nature Reviews Gastroenterology & Hepatology **12**(3), 172–186 (2015)
15. Kudo, S.E., Wakamura, K., Ikehara, N., Mori, Y., Inoue, H., Hamatani, S.: Diagnosis of colorectal lesions with a novel endocytoscopic classification–a pilot study. Endoscopy **43**(10), 869–875 (2011)
16. Linker, R., Kelman, E.: Apple detection in nighttime tree images using the geometry of light patches around highlights. Computers and Electronics in Agriculture **114**, 154–162 (2015)
17. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and vision computing **22**(10), 761–767 (2004)
18. Meziou, L., Histace, A., Precioso, F.: Alpha-divergence maximization for statistical region-based active contour segmentation with non-parametric pdf estimations. In: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, pp. 861–864. IEEE (2012)
19. Nistér, D., Stewénius, H.: Linear time maximally stable extremal regions. Computer Vision–ECCV 2008 pp. 183–196 (2008)
20. Núñez, J.M., Bernal, J., Ferrer, M., Vilariño, F.: Impact of keypoint detection on graph-based characterization of blood vessels in colonoscopy videos. In: International Workshop on Computer-Assisted and Robotic Endoscopy, pp. 22–33. Springer (2014)
21. Shao, F., Jiang, G., Yu, M., Ho, Y.S.: Highlight-detection-based color correction method for multiview images. ETRI journal **31**(4), 448–450 (2009)
22. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. International Journal of Computer Assisted Radiology and Surgery **9**(2), 283–293 (2014)
23. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. IEEE transactions on medical imaging **35**(2), 630–644 (2016)
24. Tan, R., Ikeuchi, K.: Separating reflection components of textured surfaces using a single image. Digitally Archiving Cultural Objects pp. 353–384 (2008)
25. Vincent, L.: Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. IEEE transactions on image processing **2**(2), 176–201 (1993)
26. Xu, S.C., Ye, X., Wu, Y., Zhang, S.: Highlight detection and removal based on chromaticity. In: Image Analysis and Recognition, pp. 199–206. Springer (2005)
27. Yang, Q., Wang, S., Ahuja, N.: Real-time specular highlight removal using bilateral filtering. Computer Vision–ECCV 2010 pp. 87–100 (2010)
28. Yoon, K., Kweon, I.: Correspondence search in the presence of specular highlights using specular-free two-band images. Computer Vision–ACCV 2006 pp. 761–770 (2006)
29. Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M., Kaus, M.R., Haker, S.J., Wells, W.M., Jolesz, F.A., Kikinis, R.: Statistical validation of image segmentation quality based on a spatial overlap index 1: Scientific reports. Academic radiology **11**(2), 178–189 (2004)

Appendix: Bright spot regions segmentation algorithms

Algorithm 1 Initialization algorithm

- *Calculation of LMax, local maxima of LumIm.*
- *Addition of a 1-pixel width frame around RegionsIm with* -1 *pixel value.*

for $nr = 1 \rightarrow \#(LMax)$ **do**
  $RegionsIm[LMax[nr]] = nr$;
  $LabelArray[nr].PRegion = newRegionInf$;
end for
for pixel $p \in RegionsIm$ **do**
  $NeighIm[p] = \#\{q|q$ is connected neighbor of p $\&RegionsIm[q] == 0| - 1\}$;
  $LabelArray[nr].PRegion = newRegionInf$;
  if $RegionsIm[p] \geq 0$ **then**
    if $NeighIm[p] \geq 0$ **then**
      $r = LabelArray[RegionsIm[p]].PRegion$;
      $r.Seed = p$;
      $r.FGradSum = r.FGradSum + GradIm[p]$;
      $r.FLen = r.FLen + 1$;
    end if
  else if $NeighIm[p] \leq MaxNeighbors$ **then**
    $RegionsIm[p] = -1$;
    $Heap.Push(p)$;
  end if
end for
- *Addition of a 1-pixel width frame around NeighIm;*

for $nr = 1 \rightarrow \#(LMax)$ **do**
  $r = LabelArray[nr].PRegion$;
  $r.CreateLevel = LumIm[r.Seed]$;
  $r.ToUpdateLevel = r.CreateLevel$;
end for

Algorithm 2 Region growing algorithm

```
while isNotEmpty(Heap) do
    p = Heap.Pop();
    level = LumIm[p];
    FR = NULL;
end while
for {q ∈ RegionsIm | q is neighbor of p} do
    NeighIm[q] = NeighIm[q] − 1;
    if RegionsIm[q] == 0 then
        RegionsIm[q] = −1;
        Heap.Push(q);
    else if RegionsIm[q] ≥ 0 then
        SR = LabelArray[RegionsIm[q]].PRegion
        if NeighIm[q] == 0 then
            SR.FLen = SR.FLen − 1;
            SR.FGradSum = SR.FGradSum+GradIm[q];
            if FR == NULL then
                FR = SR;
                UpdateMaxContrast(FR, level);
                - Add Pixel q to FR;
                RegionsIm[p] = RegionsIm[q];
                if NeighIm[p] > 0 then
                    FR.Len = FR.Len + 1;
                    FR.FGradSum = FR.FGradSum +
GradIm[p];
                end if
            else if FR! = SR then
                UpdateMaxContrast(SR);
                JoinRegions(FR, SR);
            end if
        end if
    end if
end for
```

Algorithm 3 JoinRegions algorithm

```
if FR.CreationLevel! = level then
    if SR.CreationLevel! = level then
        NewReg = new RegionInf;
        NewReg.Childrens.Push(FR);
        NewReg.MaxFGrad = 0;
        NewReg.Labels = FR.Labels;
        for l ∈ NewReg.Labels do
            LabelsArray[l].PRegion = NewReg;
        end for
        FR = NewReg;
    else
        tmp = FR;
        FR = SR;
        FR = temp;
    end if
end if
FR.FLen = FR.FLen + SR.FLen;
FR.FGradSum = FR.FGradSum + SR.FGradSum;
for l ∈ SR.Labels do
    LabelsArray[l].PRegion = FR;
end for
FR.Concat(SR.Labels);
if FR.CreationLevel == SR.CreationLevel then
    FR.Children.Concat(SR.Children);
    Delete SR;
else
    FR.Children.Push(SR);
end if
```

Algorithm 4 UpdateMaxContrast algorithm

```
if R.ToUpdateLevel! = level then
    FGrad = R.FGradSum/R.FLen;
    if FGrad > R.MaxFGrad then
        R.MaxLevel = R.ToUpdateLevel;
        R.MaxFGrad = FGrad;
    end if
    R.ToUpdateLevel = level;
end if
```

Algorithm 5 Region Selection algorithm

```
for child ∈ pRegion.Children do
    MaxG = max(MaxG, SelectRegions(child, MinGrad));
end for
if MaxG ≤ pRegion.MaxFGrad then
    DeletepRegions.Children
    if pRegion.MaxFGrad ≥ MinGrad then
        pRegion.Mark = true;
    end if
    return pRegion.MaxFGrad;
else
    eturn MaxG;
end if
```