Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 12–40

COMPUTER-ASSISTED BLEEDING DETECTION IN WIRELESS CAPSULE ENDOSCOPY IMAGES

ISABEL N. FIGUEIREDO, SUNIL KUMAR AND PEDRO N. FIGUEIREDO

Abstract: Wireless capsule endoscopy (WCE) has revolutionized the diagnosis and treatment of gastrointestinal tract, especially the small intestine where traditional endoscopies can not reach. However, this new technology leads to the inspection of a large number of images, which is a time consuming process and also too hard by naked eyes for doctors. In this article we propose a new computerized method for bleeding detection in WCE images. We employ the second component of CIE Lab color space together with appropriate segmentation and enhancement techniques, involving an adaptive anisotropic diffusion (alike Perona Malik diffusion). As a result of this procedure, it is possible to devise four functions to discriminate between bleeding and normal regions, in WCE images. Experimental results on several medical datasets show that the new algorithm achieves a very good rate of success and promising performance for bleeding detection.

Keywords: anisotropic diffusion, segmentation, bleeding detection, wireless capsule endoscopy.
AMS Subject Classification (2010): 68U10,65D18,65K10,35K55.

1. Introduction

Bleeding in gastrointestinal (GI) tract can result from a number of conditions, including vascular lesions (angiodysplasia), small bowel tumors, coeliac disease and Crohn's disease (which may be suspected because of other symptoms). The best way to diagnose bleeding in GI tract is to directly view the GI tract. Wireless capsule endoscopy (WCE) is a noninvasive procedure designed to help doctors to see all the areas of the GI tract, especially the small intestine, where traditional invasive endoscopies such as colonoscopy and gastroscopy cannot reach. The WCE was first invented by *Given Imaging* in 2000 [13] and the first device got approval from the US food and drug administration (FDA) in 2001. Later, some other companies like *Olympus* and *IntroMedic* have also developed a similar device. During the procedure, patients swallow a very small capsule (with the size and shape of a vitamin

Received October 29, 2012.
This work was partially supported by the research project UTAustin/MAT/0009/2008 of the UT Austin | Portugal Program (http://www.utaustinportugal.org/) and by CMUC and FCT (Portugal), through European program COMPETE/FEDER..

pill), consisting of a miniaturized camera, a light source and a wireless circuit for the acquisition and transmission of signals. As the capsule moves through the GI tract, due to peristalsis, images are transmitted to a data recorder, worn on a belt outside the body. After about 8h, the WCE battery lifetime, the stored images are transferred to a computer workstation for offline viewing.

Although WCE has shown significant improvement over traditional endoscopies, there is still some room for improvement. One problem with this new technology is that it generates approximately $50,000$ images per examination for one patient, which makes it hard for medical experts to examine. Furthermore, bleeding regions or abnormalities may be missed because of their size or distribution, or of visual fatigue. This opens a door for computer-aided decision support systems, to assist the clinicians in the analysis of the WCE images, by reducing the time required to reach the diagnosis and thus the cost of the procedure, making it a more efficient and affordable technique.

The manufacturer *Given Imaging* provides a software tool to detect some diseases. But its sensitivity and specificity were reported to be very low (respectively, 21.5% and 41.8% in [10]; see also [20] for a related paper). Recently, some efforts to automatically inspect WCE images have been made, and among several papers we refer in particular to [1, 6, 7, 8, 14, 16, 19, 18, 21, 17]. In [1] the authors propose a method for detecting informative frames from WCE videos. In [6] the usefulness of a particular multimedia content descriptor is measured, for detection of a variety of events, such as bleeding, ulcers and polyps. In [7] six color features in the hue/saturation/intensity (HSI) color space are proposed, to discriminate between normal and bleeding images. In [14] it is compared the performance of color and texture features in WCE image classification: the color features are extracted from local color histograms of the components of the hue/saturation/value (HSV) color space and texture features are obtained by wavelet decomposition, on each color component. In [16] a new idea of chrominance moment is proposed, as the color part of color texture feature, which makes use of Tchebichef polynomials and illumination invariant of the HSI color space; moreover, a combination of this with uniform local binary pattern is used to discriminate between normal and bleeding regions in WCE images. In [19] a computer aided intelligent system based on probabilistic neural network is suggested: color texture features of bleeding regions, from red/green/blue (RGB) and HSI color spaces, are extracted, and a probabilistic neural network classifier, to recognize the

Vision and Pattern Recognition (CVPR '08), pages 1–8, Los Alamitos, Calif., USA, June 2008.

Isabel N. Figueiredo
CMUC, Department of Mathematics, University of Coimbra, 3001-501 Coimbra, Portugal
E-mail address: isabelf@mat.uc.pt
URL: http://www.mat.uc.pt/∼isabelf

Sunil Kumar
CMUC, Department of Mathematics, University of Coimbra, 3001-501 Coimbra, Portugal
E-mail address: skumar@mat.uc.pt

Pedro N. Figueiredo
Faculty of Medicine, University of Coimbra, 3004-504 Coimbra, and Department of Gastroenterology, University Hospital of Coimbra, 3000-075 Coimbra, Portugal
E-mail address: pnf11@sapo.pt

bleeding, is built. In [18], it is proposed two color vector similarity coefficients for measuring the color similarity degree in RGB color space; based on which the authors describe an algorithm, for bleeding detection in WCE images.

In this paper we propose a new methodology/algorithm for bleeding detection in WCE images. We first apply a CIE Lab [23] transform to the WCE image. Interestingly, but not surprisingly, we have observed that the second component of the transformed WCE image, somewhat highlights the bleeding region and has also the property of somewhat separating mucosa from uninformative parts (like bubbles, trash, lumen and so on). Based on these facts our methodology relies on enhancement techniques, for evidencing the bleeding, and segmentation, for discarding uninformative regions. For improving the contrast between the bleeding lesion and its surrounding normal mucosa region, we perform a smoothing enhancement of the second component, of the CIE Lab transformed WCE image. We compute an extended structure tensor [3, 26] and furthermore, inspired on [15], we apply an adaptive anisotropic diffusion (alike Perona Malik diffusion [22]) to an appropriate transform of this extended structure tensor (more precisely, to the Frobenius norm of the Hessian of this extended structure tensor). For separating mucosa part from uninformative regions, in the WCE image, we apply an automatic two-phase segmentation (based on the well known Chan and Vese model [2, 5]) to the second component, of the CIE Lab transformed WCE image. Finally, by combining the enhancement and segmentation we devise four appropriate functions, to serve as bleeding detector candidates. These are proven to correctly locate the bleeding region and to differentiate between bleeding and normal image frames, with high rates of success in several medical datasets.

After this introduction the rest of the paper is organized in the following way. In section 2 we describe the enhancement, the segmentation, and the definition of the four bleeding detector functions. The results of these operations are illustrated in several figures through this section. Then, we summarize the main steps of our proposed algorithm in section 3. We also report, in section 4, on the numerical results achieved on there different medical datasets, consisting of 700, 1200 and 2087 images, respectively. The ROC (receiving operating characteristics) curves are exhibited, for the different four bleeding detector functions, proposed in the paper. Finally, a section with the conclusions and outlook finishes the paper.

2. Methodology

First of all we fix some notations that are used in the paper. Let $I : \Omega \subset R^2 \to R^3$ be a vectorial function representing an image, with 3 channels (in this paper the WCE images are RGB images, thus the 3 channels are the red, green and blue channels). The set Ω is an open subset of R^2, representing the image (or pixel) domain. We denote by (x, y) any arbitrary point in Ω.

For any scalar, smooth enough, function u defined on Ω, we denote the L^1 norm of u, by $\|u\|_{L^1(\Omega)}$ (which is the integral of the Euclidean norm of the gradient of $\|\nabla u\|$, over the domain Ω), and the L^2 norm of u, by $\|u\|_{L^2(\Omega)}$ (which is the integral of u^2, over the domain Ω). Finally the L^∞ norm of u, is denoted by $\|u\|_{L^\infty(\Omega)}$ (which is the essential supremum of u in Ω). Furthermore, we indicate by $|.|$ the absolute value (modulus) of any real number, and by $\|.\|$ the Euclidean norm.

2.1. Choice of the color channel. Colors are one of the most important visual primitives, which facilitate various visual tasks in the real life. The use of color in computer vision is motivated by two important factors. First, color is a powerful descriptor that often simplifies object identification from an image. Second, humans can distinguish between thousands of color shades and intensities, compared to about only a few dozen shades of gray level. CIE Lab (the color space specified by the International Commission on Illumination, see [23] and also [12] for other color spaces) is the most complete color space used conventionally to describe all the colors visible to the human eye. The three components L, a and b represent, respectively, lightness of the color, its position between red and green, and between yellow and blue.

We have observed that, for an WCE image I, containing a bleeding region, the second component of the transformed of I in the CIE Lab space color space, hereafter denoted by I^a (also referred in the paper as the *a-channel of I*), is the best channel for localizing bleeding regions. This is illustrated for instance in Figure 1: the sub-figures in column (a) display several WCE frames containing bleeding regions (which are the small red spots) and the sub-figures in column (b) show the corresponding I^a channels.

We remark that, since many WCE images suffer from illumination variations, due to specific imaging circumstances, such as motion of the camera and limited range of illuminations in digestive tract, it is necessary to compensate for nonuniform illumination. Therefore, before computing the CIE

[7] L. Cui, C. Hu, Y. Zou, and M. Q.-H. Meng. Bleeding detetction in wireless capsule endoscopy images by support vector classifier. In *Proceedings of the 2010 IEEE Conference on Information and Automation*, pages 1746–1751, Harbin, China, June 2010.

[8] J. P. Silva Cunha, M. Coimbra, P. Campos, and J. M. Soares. Automated topographic segmentation and transit time estimation in endoscopic capsule exams. *IEEE Transactions on Medical Imaging*, 27:19–27, 2008.

[9] T. Fawcett. An introduction to ROC analysis. *Pattern Recogn. Lett.*, 27(8):861–874, June 2006.

[10] R.D. Francis. Sensitivity and specificity of the red blood identification (rbis) in video capsule endoscopy. In *3rd INt. Conf. Capsule Endoscopy*, Miami, FL, USA, Feb 2004.

[11] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer-Assisted Intervention*, pages 130–137, Cambridge, MA, USA, 1998.

[12] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., NJ, USA, 2006.

[13] G. Idan, G. Meron, and A. Glukhovsky. Wireless capsule endoscopy. *Nature*, 405:417–417, 2000.

[14] P.C. Khun, Z. Zhang, L.Z. yang, L.Liyuan, and L. Jiang. Feature selection and classification for wireless capsule endoscopic frames. In *International Conference on Biomedical and Pharmaceutical Engineering*, Singapore, December 2009.

[15] B. Li and M. Q.-H. Meng. Wireless capsule endoscopy images enhancement via adaptive contrast diffusion. *J. Vis. Commun. Image R.*, 23:222–228, 2012.

[16] B. Li and M. Q.-H-Meng. Computer-aided detection of bleeding regions for capsule endoscopy images. *IEEE Transactions on Biomedical Engineering*, 56:1032–1039, 2009.

[17] M. Liedlgruber and A. Uhl. Computer-aided decision support systems for endoscopy in the gastrointestinal tract: a review. *IEEE Reviews in Biomedical Engineering*, 4:73–88, 2011.

[18] G. Pan, F. Xu, and J. Chen. A novel algorithm for color similarity measurement and the application for bleeding detection in WCE. *I.J. Image, Graphics and Signal Processing*, 5:1–7, 2011.

[19] G. Pan, G. Yan, X. Qiu, and J. Cui. Bleeding detection in wireless capsule endoscopy based on probabilistic neural network. *J. Med. Syst.*, 35:1477–1484, 2011.

[20] S. C. Park, H. J. Chun, E. S. Kim, B. Keum, Y. S. Seo, Y. S. Kim, Y. T. Jeen, H. S. Lee, S. H. Um, C. D. Kim, and H. S. Ryu. Sensitivity of the suspected blood indicator: An experimental study. *World J. Gastroenterolgy*, 18(31):4169–4174, 2012.

[21] B. Penna, T. Tilloy, M. Grangettoz, E. Magli, and G. Olmo. A technique for blood detection in wireless capsule endoscopy images. In *17th European Signal Processing Conference (EUSIPCO 2009)*, pages 1864–1868, 2009.

[22] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.

[23] J. Schanda. *Colorimetry: Understanding the CIE system*. Wiley & Sons, NJ, USA, 2007.

[24] The Mathworks, Inc. *http://www.matlab.com*.

[25] F. Vilariño, P. Spyridonos, O. Pujol, J. Vitrià, and P. Radeva. Automatic detection of intestinal juices in wireless capsule video endoscopy. In *18th International Conference on Pattern Recognition (ICPR 2006)*, pages 719–722, 2006.

[26] X.-F. Wang, D.-S. Huang, and H. Xu. An efficient local Chan-Vese model for image segmentation. *Pattern Recognition*, 43:603–618, 2010.

[27] Y. Zheng, J. Yu, S. B. Kang, S. Lin, and C. Kambhamettu. Single-image vignetting correction using radial gradient symmetry. In *Proceedings of the 26th IEEE Conference on Computer*

Finally, we observe that the implementation of all the functions and mathematical models, described in the paper, is done in MATLAB® R2012a [24], on a 64bit Windows7 machine, with Intel Core2 Quad processor running at 2.66 GHz, and 4.00Gb RAM. The execution time for one frame, on an average, is 11 seconds.

5. Conclusions

We propose a new computer-assisted bleeding detector, to discriminate between bleeding and non bleeding regions, in small bowel images, obtained *in vivo* with wireless endoscopy capsules. Four functions were devised to serve as bleeding detectors. The methodology is tested on three data sets prepared by clinicians. The results demonstrate that all the detectors achieve a good rate of success and evidence promising performance for bleeding detection. In particular, these detectors are normalized functions, relying on the second derivatives of a particular modification of the second component of the CIE Lab transform of the original image. This modification involves a smoothing-enhancement of the bleeding region, with appropriate techniques, and also segmentation, for discarding uninformative parts in the image. The analysis of the performance of this algorithm is done by taking images individually. We think it might be better if the images are analyzed in a sequence. In effect, by examining the video sequence, created by the wireless capsule, there is a greater chance that a bleeding region appearing in a frame, also appears in the previous and following frames. This would improve the sensitivity of the algorithm.

References

[1] M.K. Bashar, T. Kitasaka, Y. Suenaga, Y. Mekada, and K. Mori. Automatic detection of informative frames from wireless capsule endoscopy images. *Medical Image Analysis*, 14:449–470, 2010.

[2] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *J Math. Imaging Vis.*, 28:151–167, 2007.

[3] T. Brox, J. Weickert, B. Burgeth, and P. Mrázek. Nonlinear structure tensors. *Image and Vision Computing*, 24:41–55, 2006.

[4] F. Catte, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 29:182–193, 1992.

[5] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, 2001.

[6] M.T. Coimbra and J.P.S Cunha. MPEG-7 visual descriptors-contributions for automated feature extraction in capsule endoscopy. *IEEE Transactions on Circuits and Systems for Video Technology*, 16:628–637, 2006.

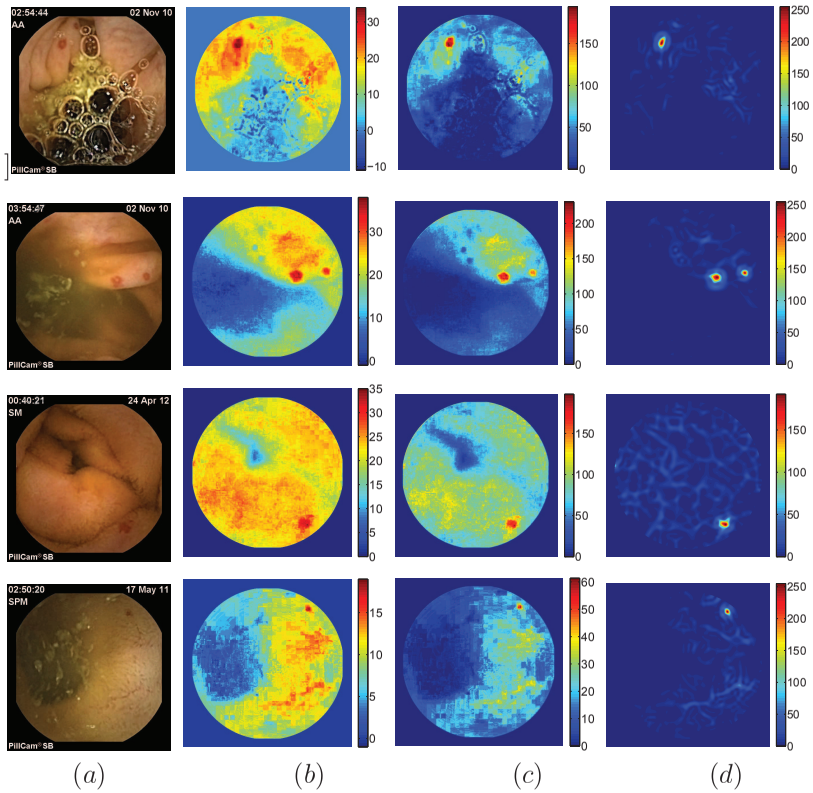


FIGURE 1. Original frames with bleeding regions in (a), channel I^a in (b), generalized structure tensor S^a in (c) and in (d) the diffused Frobenius norm $\widehat{F_s^a}$.

Lab transform of the original WCE image I, we first apply to I the automatic illumination correction scheme described in [27].

2.2. Contrast smoothing enhancement. For further enhancing the contrast between the bleeding region and the rest of the image, in I^a (the *a-channel of* I), we proceed in two steps. First, we modify I^a, by using a generalized structure tensor of I^a, hereafter denoted by S^a. Secondly, we

apply a Perona Malik anisotropic diffusion model to the Frobenious norm of the Hessian matrix of S^a.

First step : Definition of S^a. The classical structure tensor and its generalizations are important concepts in image processing and computer vision, providing several image information as for example, orientation, structure and texture (see [3, 26] and the references therein). The classical structure tensor of the scalar function I^a is a matrix field obtained by Gaussian smoothing the tensor product of the gradient of I^a. The generalized structure tensor of I^a involves in addition I^a itself. Their definitions are, respectively

$$\begin{pmatrix} G_\tau * (I_x^a)^2 & G_\tau * I_x^a\, I_y^a \\ G_\tau * I_x^a\, I_y^a & G_\tau * (I_y^a)^2 \end{pmatrix} \text{ and } \begin{pmatrix} G_\tau * (I_x^a)^2 & G_\tau * I_x^a\, I_y^a & G_\tau * I_x^a\, I^a \\ G_\tau * I_x^a\, I_y^a & G_\tau * (I_y^a)^2 & G_\tau * I_y^a\, I^a \\ G_\tau * I_x^a\, I^a & G_\tau * I_y^a\, I^a & G_\tau * (I^a)^2 \end{pmatrix}, \tag{1}$$

where $*$ is the convolution operator, the subscripts x and y represent first order partial derivatives with respect to x and y, and G_τ is a two-dimensional Gaussian function with variance τ defined by

$$G_\tau(x,y) := \frac{1}{2\pi\tau^2} \exp\Big(-\frac{x^2+y^2}{2\pi\tau^2}\Big) \tag{2}$$

This generalized (or extended) structure tensor yields six different feature channels for each scale τ. Then S^a, mentioned above, is the average of these channels

$$S^a = \Big(G_\tau * \big((I_x^a)^2 + I_x^a\, I_y^a + (I_y^a)^2 + I_x^a\, I^a + I_y^a\, I^a + (I^a)^2\big)\Big)/6. \tag{3}$$

A measure for the strength of the intensity variation of S^a is the Frobenius norm of the Hessian matrix of S^a, which can be expressed using both its eigenvalues (it is the sum of the squares of the eigenvalues). Based on this fact, we define the Gaussian smoothed Hessian of S^a, at a point (x,y) and at a scale s, by

$$H_s(x,y) = \begin{pmatrix} G_s * S_{xx}^a & G_s * S_{xy}^a \\ G_s * S_{xy}^a & G_s * S_{yy}^a \end{pmatrix}, \tag{4}$$

where again G_s is another two-dimensional Gaussian function with variance s (see (2)) and the subscripts xx, xy and yy represent second order partial derivatives with respect to x and y. Thus, denoting by $\lambda_s^1(x,y)$ and $\lambda_s^2(x,y)$ the two eigenvalues of the Hessian matrix $H_s(x,y)$, at point (x,y) and scale s, the Frobenius norm of $H_s(.,.)$, for scale s, denoted hereafter by F_s^a, is defined by

$$F_s^a = (\lambda_s^1)^2 + (\lambda_s^2)^2. \tag{5}$$

The accuracy, defined by

$$\text{accuracy} = \frac{\text{number of TN} + \text{number of TP}}{\text{number of TN} + \text{number of FP} + \text{number of TP} + \text{number of FN}}.$$

is another measure, used to assess the overall performance of the algorithm, that quantifies the actual true values, and reflects sensitivity and specificity in relation to each other.

Tables 3 and 4 display, for each data set, and for far (= 1- specificity) $\leq 10\%$, the best sensitivity (sens.), and corresponding accuracy (acc.) of each bleeding detector function, as well as the corresponding thresholds (thres.).

TABLE 3. Best sensitivity of B_1 and B_2, for far $\leq 10\%$

Datasets	B_1				B_2			
	sens.	far	acc.	thres.	sens.	far	acc.	thres.
D1	89.3	9.5	90.4	8.1072e+04	92.9	8.9	91.1	3.1468e+05
D2	79	9.4	90	6.6166e+04	79	9.7	89.7	2.4330e+05
D3	75.2	9.9	87.4	7.4588e+04	76.3	9.6	87.9	2.5913e+05
D4	75.7	9.5	88.8	7.4023e+04	76.8	9.5	88.9	2.6357e+05

TABLE 4. Best sensitivity of B_3 and B_4, for far $\leq 10\%$

Datasets	B_3				B_4			
	sens.	far	acc.	thres.	sens.	far	acc.	thres.
D1	92.9	7.3	92.7	5.1176e+05	92.9	8.9	91.1	7.7735e+04
D2	82.3	10	89.5	4.7411e+05	75.8	9.7	88.6	6.3941e+04
D3	62.1	10	84.9	4.2679e+05	77.7	10	87.8	7.1801e+04
D4	63.9	10	86.9	4.4871e+05	78.1	9.7	88.9	7.0794e+04

The graphs of the ROC curves, in Figure 6, and also Tables 3 and 4 reveal that the performance of the detectors B_1, B_2, B_3 and B_4 is somewhat similar, though the best sensitivity results are achieved with B_4, for three datasets, and, for the datasets D3 and D4, the worse performance is with the detector B_3. We should emphasize that dataset D1 contains bleeding regions that are very pronounced, as opposed to D2 and D3, where the bleeding regions are in general blurred, not so obvious to human vision perception. This difference is well reflected in the results, where the highest sensitivity and lowest far is reached for the dataset D1, with the bleeding detector B_3.

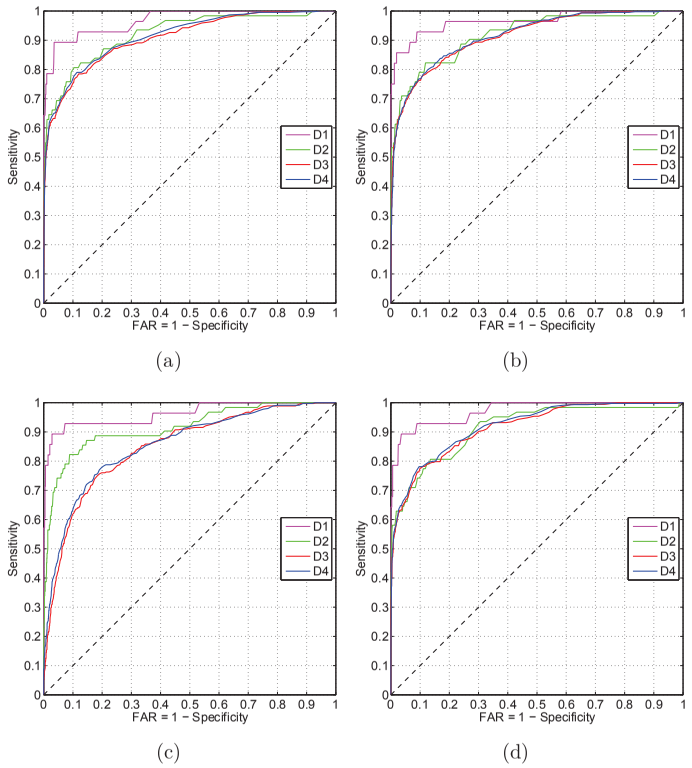


FIGURE 6. ROC curves for the four datasets (D1, D2, D3, D4) with the bleeding detectors: B_1 in (a), B_2 in (b), B_3 in (c) and B_4 (d).

the ROC curve, representing 100% sensitivity (no false negatives) and 100% specificity (no false positives). A completely random guess would give points along the diagonal line. In addition, the diagonal divides the ROC space: points above the diagonal represent good classification results, and points below the diagonal bad results.

Figure 6 represents the ROC curves for the four functions B_1, B_2, B_3, and B_4, and for all the four data sets.

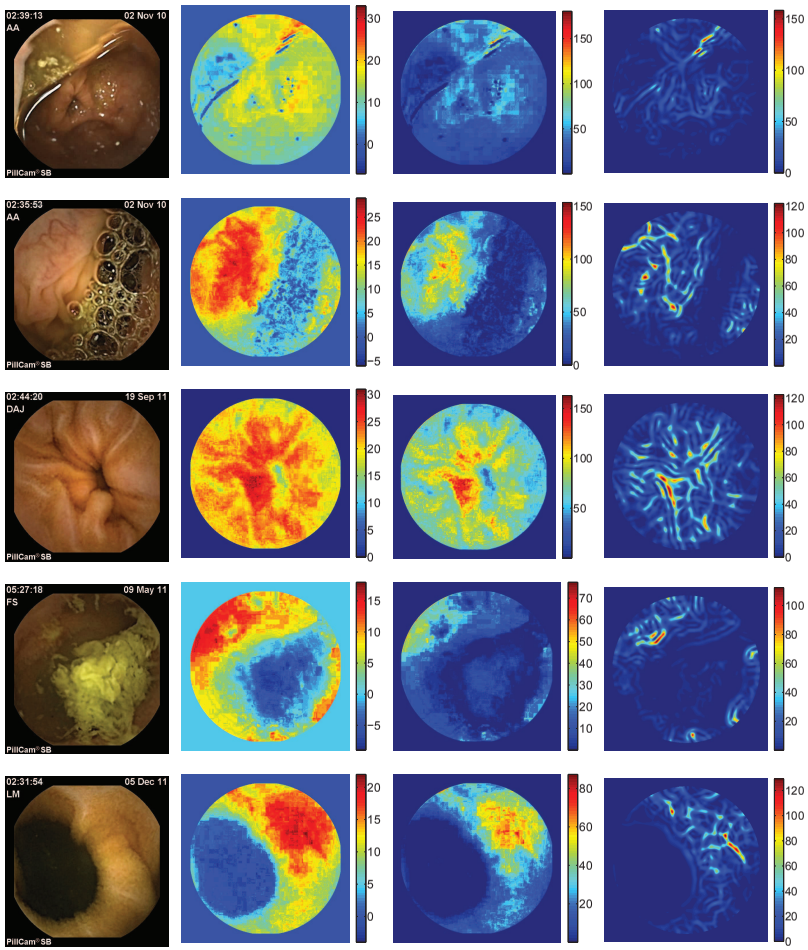


FIGURE 2. Original frames without bleeding and displaying liquid, bubbles, normal mucosa, green trash and lumen in (a). Corresponding channel I^a in (b), generalized structure tensor S^a in (c) and the diffused Frobenius norm $\widehat{F_s^a}$ in (d).

Second step : Anisotropic diffusion of F_s^a. Inspired on [15], we then apply a Perona Malik alike adaptive anisotropic diffusion model to F_s^a, that is, we solve the following time-spatio equation

$$\begin{aligned} &\frac{\partial u}{\partial t} = \operatorname{div}\left(g(u)\nabla u\right), \\ &u(0) = F_s^a \end{aligned} \tag{6}$$

where ∇ and div denote respectively, the gradient and divergence operators. The initial condition for the unknown u is F_s^a, at time $t = 0$, and g is the diffusion function defined by

$$g(u) = \frac{1}{1 + (|u|/\alpha)^2} \quad \text{with} \quad \alpha = \frac{1}{\sqrt{F_s^a}}. \tag{7}$$

Here α is a local adaptive parameter, depending on F_s^a, that determines the amount of the diffusion according to the value of F_s^a at each point (a large α leads to smoothness, while a small α originates sharpness). In the original Perona Malik model [22], $\|\nabla u\|$ is used instead of u, in the diffusion function g above, and the parameter α is then used to control the diffusion, by smoothing the image along but not across the edges. The experimental results obtained by Perona-Malik model were visually impressive, edges remain stable over a very long time; however, the model had some serious theoretical problems and to overcome them, in [4] it is proposed to replace $g(||\nabla u||)$ by $g(||G_\tau * \nabla u||)$, where G_τ is a Gaussian function (see (1)). In WCE images, edges are unobtrusive, but, we are interested in enhancing the images, since they lack contrast, that's why the definition of g in (7) is appropriate in this context.

The solution u of (6), *i.e.* the diffused Frobenious norm of the generalized structure tensor of I^a, hereafter denoted by $\widehat{F_s^a}$, is then normalized as follows:

$$\widehat{F_s^a} = \frac{u - u_{min}}{u_{max} - u_{min}} \times 255, \tag{8}$$

where u_{min} and u_{max} are, respectively, the minimum and maximum of u.

The Figures 1 and 2 show several bleeding and non-bleeding frames, and corresponding I^a and S^a channels. They also illustrate, in column (d), the contrast enhancement results obtained with the Perona Malik alike anisotropic diffusion model applied to the Frobenius norm F_s^a. As we can see the bleeding regions are singled out.

by definition, far=1-specificity is the false alarm rate), for a particular decision threshold. The threshold is a numeric value defined by the function B_i, that produces a discrete (binary) classifier : if $B_i(F)$, the value of the frame F by B_i, is above the threshold the classifier produces a positive result (P), else a negative (N) result. On the other hand, for a fixed threshold, the sensitivity measures the percentage of bleeding frames correctly detected by the algorithm, and similarly, the specificity is the percentage of non-bleeding frames correctly identified as so by the algorithm. The exact definitions for the sensitivity and specificity are

$$\text{sensitivity} = \frac{\text{number of TP}}{\text{number of TP+number of FN}} \quad \text{and} \quad \text{specificity} = \frac{\text{number of TN}}{\text{number of TN+number of FP}}. \tag{20}$$

Here TP, FN, FP and TN stand for, respectively, the following four categories of frames:

- TP: is a true positive frame, that is, a frame containing a bleeding region, correctly detected by the algorithm, *i.e.* classified by the algorithm as positive (P).
- FN: is a false negative frame, that is, a frame containing a bleeding region, incorrectly classified as negative (N) by the algorithm.
- FP: is a false positive frame, that is, a frame that does not contain a bleeding region, and incorrectly classified as positive (P) by the algorithm.
- TN: is a true negative frame, that is, a frame not containing a bleeding region, and correctly identifed by the algorithm, *i.e.* classified by the algorithm as negative (N).

The following Table 2 summarizes these notions, combining the different classes of frames and their relation with the classifier induced by the bleeding detector function, for a given threshold.

TABLE 2. Classifier defined by B_i for threshold T

Frame F	Bleeding	Non-Bleeding
$B_i(F) > T \rightarrow$ Positive	True Positive (TP)	False Positive (FP)
$B_i(F) \leq T \rightarrow$ Negative	False Negative (FN)	True Negative (TN)

We also note that the closer to 1 the values for both the sensitivity and specificity of a scheme, the better its performance. In terms of ROC curve, and in our context, this statement means the best possible bleeding detector method would yield a point in the upper left corner, in coordinate (0,1) of

(5) For removing uninformative parts (such as bubbles, trash, liquid, lumen, and so on), we perform a two-phase segmentation over the geometric (or piece-wise smooth) part of I^a, the a-channel component of I in the CIE Lab color space.
(6) Finally, we compute the bleeding detectors B_1, B_2, B_3 and B_4, using the diffused Frobenius norm of the Hessian of the generalized structure tensor of I^a. Moreover, applying the segmentation mask obtained in the previous step, we confine these computations to the mucosa part of the frame.

4. Numerical Results

Three data sets (hereafter denoted by D1, D2 and D3) were prepared by medical experts for evaluating the performance of the proposed algorithm. In order to make these datasets representative, the frames were collected from ten different patient WCE video segments. These videos were recorded at the Department of Gastroenterology, in the University Hospital of Coimbra (Portugal) and obtained with the capsule *PillCam SB*, manufactured by *Given Imaging*, Yoqneam, Israel.

During the process of collecting the datasets all the frames were labeled, independently, by two experts, as bleeding or non-bleeding frames. The result of this classification is displayed on Table 1. Moreover, we also consider a fourth dataset (D4), which is the combination of all the three datasets together.

TABLE 1. Expert Classification of the Datasets

Frames	Bleeding	Non-Bleeding	Total
D1	28	672	700
D2	62	1138	1200
D3	367	1720	2087
D4	457	3530	3987

For evaluating the performance of our algorithm, for each bleeding detector, we use receiver operating characteristics (ROC), or ROC curve [9]. In our case, the ROC curves considered are the graphical plots of the sensitivity versus the specificity, for the different bleeding detector functions defined before. More precisely, for each bleeding detector B_i, we define a ROC curve, such that each point on this curve represents the pair sensitivity/far (where,

In these Figures 1 and 2, for the computation of the diffused $\widehat{F_s^a}$, we have used $s = 10$, for the Gaussian function G_s in (4), and pre-specified the maximum number of iterations to halt diffusion in (6) (only 5 iterations, with a time-step $\delta t = 1/7$). These values are also used in all the numerical experiments of section 4.

2.3. Segmentation. To remove uninformative parts, such as bubbles, trash, lumen regions and so on (we refer to [1], for a very detailed classification on uninformative frame parts on WCE images, and also to [25] for a related paper), and which can interfere with the bleeding detection, we perform a two-phase segmentation on the scalar image I^a, aiming at separating colonic mucosa from uninformative parts. In effect, we have observed that I^a has the property of distinguishing mucosa, form other regions in the image. More precisely, we first perform a decomposition of the scalar image I^a into geometric (or piece-wise smooth) part and texture part, using the modified Rudin, Osher and Fatemi (ROF) model described in [2] (section 2.3) and then segment this geometric part of I^a. The reason for using the geometric part of I^a instead of I^a directly, is for achieving a better segmentation.

The segmentation scheme, we have applied, relies on a reformulation of the Chan and Vese variational model (see [2, 5]), and can be briefly summarized in the following steps:

- Step 1- Start with initial guesses c_1 and c_2 (representing the averages of I^a in a two-region image partition).
- Step 2- For each pair (c_1, c_2) solve the minimization problem

$$\min_{u,v} \left\{ TV_g(u) + \frac{1}{2\theta}\|u - v\|^2_{L^2(\Omega)} + \int_\Omega \Big(\lambda\, r(I^a, c_1, c_2)\, v + \alpha\, \nu(v) \Big)\, dx\, dy \right\} \quad (9)$$

with respect to u and v, separately: u represents the two-phase segmentation and v is an auxiliary unknown. The segmentation curve, which divides the image into two disjoint parts, is a level set of u , $\{(x, y) \in \Omega :\ u(x, y) = \mu\}$, where in general $\mu = 0.5$ (but μ can be any number between 0 and 1, without changing the segmentation result, because u is very close to a binary function).

The term $TV_g(u) := \int_\Omega g(x, y)|\nabla u|\, dx\, dy$ is the total variation norm of the function u, weighted by a positive function g. Moreover, the fitting term is $r(I^a, c_1, c_2)(x, y) := \big(c_1 - I^a(x, y)\big)^2 - \big(c_2 - I^a(x, y)\big)^2$, and $\alpha\, \nu(v)$ is a term resulting from a reformulation of the model as a convex

unconstrained minimization problem (see theorem 3, in [2]). Finally, $\theta > 0$ is a fixed small parameter and $\lambda > 0$ a constant parameter weighting the fitting term.

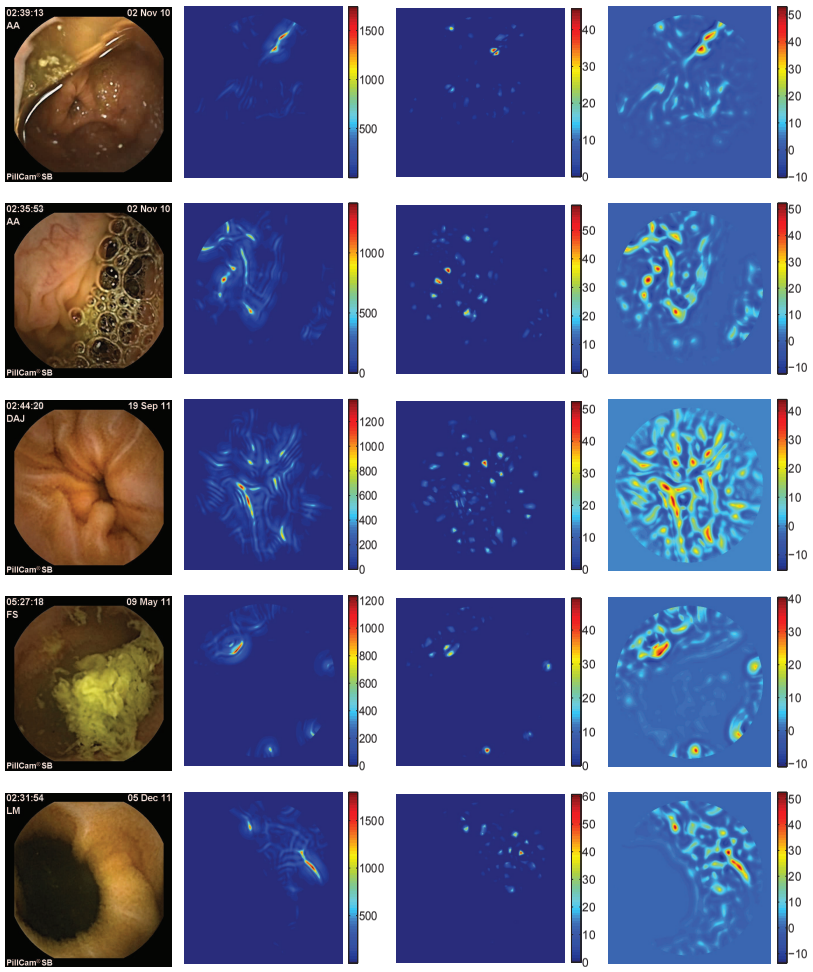
This Step 2 is solved by using a dual formulation of the TV-norm, which leads to a fast algorithm.

- Step 3- Update the averages by $c_1 = \widetilde{u_{in}}$ and $c_2 = \widetilde{u_{out}}$, which are the averages of u (u is the solution obtained in Step 2) in the two following sub-regions of the image (or pixel) domain: $\{(x, y) \in \Omega : \ u(x, y) \geq 0.5\}$ and $\{(x, y) \in \Omega : \ u(x, y) < 0.5\}$, respectively. Then return to Step 2.
- Step 4- Stop when a stopping criterium, imposed to u and v is verified.

Figure 3 illustrates the performance of the segmentation procedure. Subfigures 3(a) correspond to the original images, 3(b) show the segmentation masks, and 3(c) display the segmentation curves overlapping the original images. Here, and also for all the numerical experiments reported in section 4, the values chosen for the parameters involved in the definition of (9), are those used in [2], with $g(x, y)$ the following edge indicator function $g(\nabla u) = \frac{1}{1+\beta\|\nabla u\|^2}$, and $\beta = 10^{-4}$.

2.4. Bleeding detector functions. As a result of the contrast enhancement described before, the bleeding regions appear as isolated brighter patches or blobs outstanding to their surroundings (see Figure 1). We have chosen four normalized functions for bleeding detectors. For defining them, let first F be an arbitrary frame, and Ω_F the region, in the image domain, corresponding to the segmented mucosa part of that frame, which means, $\Omega_F = \Omega \cap \Omega_{seg}(F)$, with $\Omega_{seg}(F)$ the sub-domain corresponding to the mucosa part in the two-phase segmentation of frame F. We also consider some predefined scales, s_{min} and s_{max} , corresponding to the minimum and maximum scales at which the bleeding regions are expected to be found (while performing the Gaussian smoothing of the second derivatives, described before in section 2.2). In particular we also denoted, in the sequel, by s_1 a predefined scale.

Moreover, let λ_s^1 and λ_s^2 be the two eigenvalues of the Hessian matrix of the diffused Frobenius norm $\widehat{F_s^a}$, for a scale s. In particular, we recall that at a point belonging to a blob region, these two eigenvalues have the same sign (the sign is an indicator of the brightness/darkness of the blob: if positive it is a bright blob on a dark background, and if negative it is a dark blob on a



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

FIGURE 5. (a) Original image. (b) Function $\widehat{B_2}$. (c) Function $\widehat{B_3}$. (d) Function L.

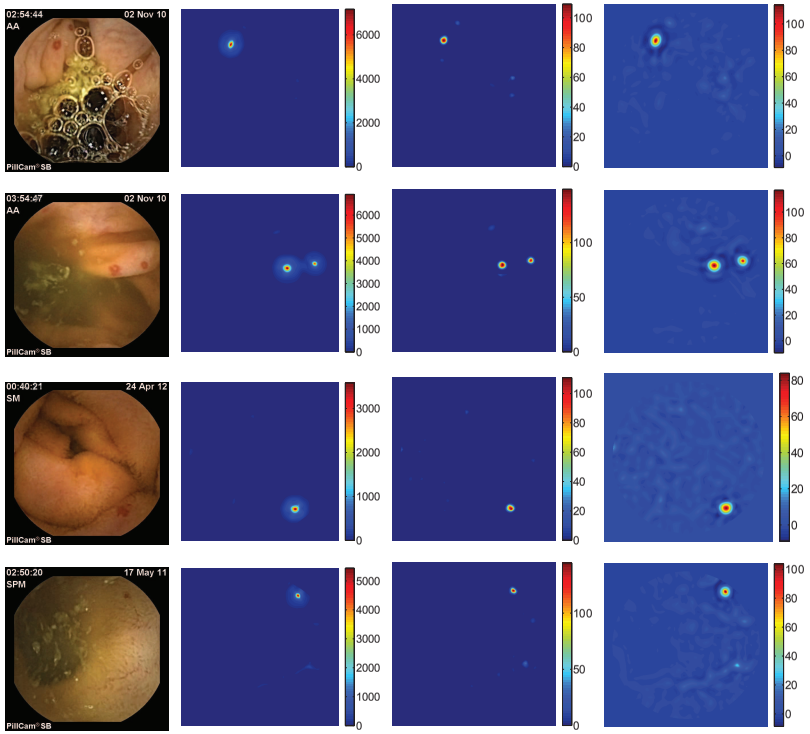


FIGURE 4. (a) Original image. (b) Function $\widehat{B_2}$. (c) Function $\widehat{B_3}$. (d) Function L.

(2) To correct the illumination and to improve the quality of the images, we apply the automatic illumination correction scheme of [27].
(3) Then we apply the CIE Lab color space to the illumination corrected image, and consider its a-channel component (denote before by I^a).
(4) We then proceed by smoothing and enhancing I^a, by using the generalized structure tensor S^a, and applying a Perona Malik alike anisotropic and adaptive diffusion model to the Frobenius norm of the Hessian of S^a.

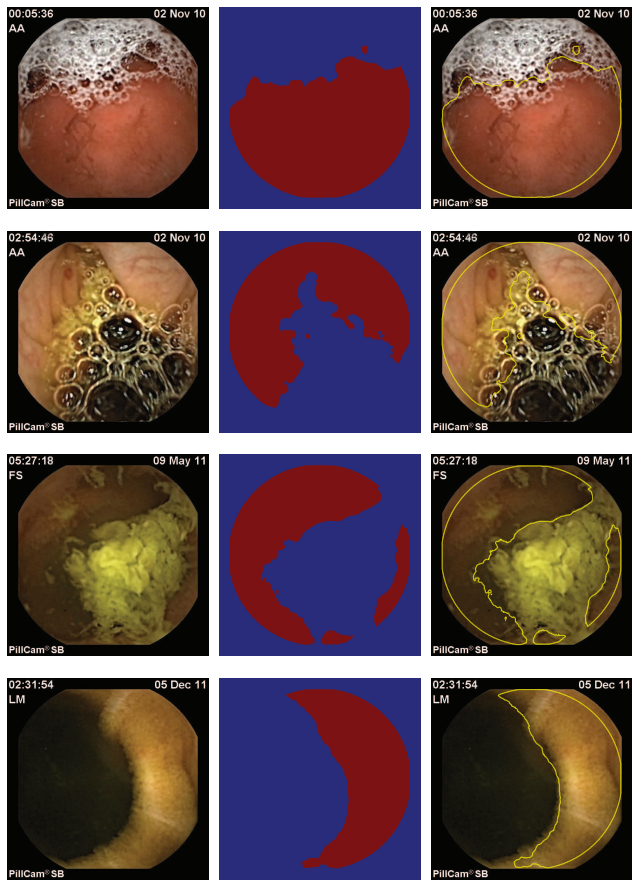


FIGURE 3. (a) Original image. (b) Segmentation mask. (c) Segmentation curve over the original image.

bright background) and similar magnitudes, whereas if the point belongs to an elongated structure (like a ridge) one of the eigenvalues is close to zero and the other's absolute value is large (see [11]).

Detector B_1. The bleeding detector B_1 is just the normalized diffused Frobenious norm defined before

$$B_1 := \frac{\|\widehat{F^a_{s_1}}\|_{L^\infty(\Omega_F)}}{\|\widehat{F^a_{s_1}}\|_{L^1(\Omega_F)}}, \tag{10}$$

for a fixed scale s_1 (in all the numerical experiments $s_1 = 10$).
Detector B_2. We now define

$$\widehat{B_2} := \max_{s_{min}\le s\le s_{max}} \left((\lambda^1_s)^2 + (\lambda^2_s)^2\right) \tag{11}$$

with s varying in a discrete set of values, between s_{min} and s_{max}. Then the second bleeding detector B_2 is

$$B_2 := \frac{\|\widehat{B_2}\|_{L^\infty(\Omega_F)}}{\|\widehat{B_2}\|_{L^1(\Omega_F)}}. \tag{12}$$

Detector B_3. We now assume that the eigenvalues of the Hessian of the diffused Frobenius norm $\widehat{F^a_s}$ are ordered at each point, thus without loss of generality the magnitude of λ^1_s is considered smaller than the magnitude of λ^2_s, *i.e.*, $|\lambda^1_s| \le |\lambda^2_s|$. We define the functions

$$f^1_s = \left(1 - \exp\left(-\rho\Big((\lambda^1_s)^2 + (\lambda^2_s)^2\Big)\right)\right) \text{ and } f^2_s = \left(1 - \exp\left(-\gamma\left(\frac{\lambda^1_s}{\lambda^2_s}\right)^2\right)\right), \tag{13}$$

where ρ and γ are two fixed positive constants, which act as regulating parameters. The formula for f^1_s expresses that it is close to one in a region, where there is a very big variation in $\widehat{F^a_s}$ and close to zero otherwise; similarly f^2_s is close to one in a blob-like region and close to zero elsewhere. The regulating parameters control the sensitivity of the functions f^1_s and f^2_s, and their values are fixed to $\rho = 10^{-2}$ and $\gamma = 1$, in all the numerical experiments. Then, for enforcing the detection of blob-like regions, we combine f^1_s and f^2_s, by defining at each point (x, y) in the image domain Ω, the function

$$B_{3s}(x,y) = \begin{cases} 0, & \text{if } \lambda^1_s(x,y)\lambda^2_s(x,y) < 0 \text{ or } |\lambda^2_s(x,y) - \lambda^1_s(x,y)| > \delta \\ f^1_s(x,y)\, f^2_s(x,y), & \text{otherwise} \end{cases} \tag{14}$$

where δ is a very small fixed parameter. Afterwards, we take the product of B_{3s} with $\widehat{F^a_{s_1}}$ (the scale s_1 was fixed before) and compute the maximum for each scale (as in (11)), that is

$$\widehat{B_3} := \max_{s_{min}\le s\le s_{max}} \left(B_{3s}\, \widehat{F^a_{s_1}}\right). \tag{15}$$

Finally

$$B_3 := \frac{\|\widehat{B_3}\|_{L^\infty(\Omega_F)}}{\|\widehat{B_3}\|_{L^1(\Omega_F)}}. \tag{16}$$

Detector B_4. The last bleeding detector is related to the Laplacian of $\widehat{F^a_{s_1}}$. We set

$$L_s = -s^2 \Delta G_s * \widehat{F^a_{s_1}}, \tag{17}$$

where G_s is a Gaussian kernel of variance s, Δ is the Laplacian operator, $*$ represents the convolution, and s_1 is an already fixed scale. Taking again the maximum with respect to the scales

$$L = \max_{s_{min}\le s\le s_{max}} L_s, \tag{18}$$

we define

$$B_4 := \frac{\|L\|_{L^\infty(\Omega_F)}}{\|L\|_{L^1(\Omega_F)}}. \tag{19}$$

We observe that the values of the bleeding detectors $\widehat{B_2}$, $\widehat{B_3}$ and L, in (11), (15) and (18), respectively, will be maxima at a scale s, that is close to the size of the blob to be detected. In the numerical experiments, the scales s take the values $8, 10, 12, 16$.

In Figures 4 and 5 we can visualize the results of the functions $\widehat{B_2}$, $\widehat{B_3}$ and L for bleeding and non-bleeding images, respectively. We remark that, by definition, they are contrast enhancements of the diffused Frobenius norm $\widehat{F^a_{s_1}}$ (compare column (d) in Figure 1 with columns (b), (c) and (d) in Figure 4); their role is particularly important in datasets where the bleeding regions appear with very low contrast with respect to the surrounding normal mucosa.

3. Algorithm

We outline now the main steps of the proposed algorithm, for automatic detection of bleeding regions in WCE images. For each frame we perform the six steps enumerated below.

(1) The original WCE image, I, possesses some details, such as the date and time of the image, the name of the patient and the name of the manufacturing company. These details can interfere and may produce false results. To remove these details we clip around the circular view of the original image.