



Polyp Localization and Segmentation in Colonoscopy Images by Means of a Model of Appearance for Polyps

A dissertation submitted by **Jorge Bernal del Nozal** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor of Philosophy**.

Bellaterra, November 4, 2012

Director	F. Javier Sánchez Pujadas Universitat Autònoma de Barcelona Dept. Ciències de la Computació & Computer Vision Center
Co-director	Fernando L. Vilariño Freire Universitat Autònoma de Barcelona Dept. Ciències de la Computació & Computer Vision Center
Thesis Committee	Dr. Panagiota Spyridonos Dept. Medical Physics, University of Ioannina Medical School University of Ioannina, Greece Dr. Debora Gil Dept. Ciències de la Computació and Centre de Visió per Computador Universitat Autònoma de Barcelona Dr. Stephen Patchett Consultant Gastroenterologist, Clinical Associate Professor Beaumont Hospital, Dublin, Ireland Dr. Santiago Seguí Centre de Visió per Computador Dr. Aura Hernández Dept. Ciències de la Computació and Centre de Visió per Computador Universitat Autònoma de Barcelona



This document was typeset by the author using L^AT_EX 2_ε.

The research described in this book was carried out at the Computer Vision Center, Universitat Autònoma de Barcelona.

Copyright © MMXII by Jorge Bernal del Nozal. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the author.

ISBN 978-84940231-9-4

Printed by Ediciones Gráficas Rey, S.L.

Dedicado a todo aquel que persigue un sueño y no para hasta alcanzarlo.
To anyone that follows a dream and never gives it up.
A tot aquell que persegueix un somni fins que ho aconsegueix.

Acknowledgements

Aún recuerdo esa mañana del 6 de Octubre de 2008 cuando yo, de frente a la puerta del CVC, me preguntaba si todo esto de venirse a Barcelona a hacer el doctorado no era una locura. Dudé bastante en entrar, pensando que a lo mejor todas mis expectativas respecto al doctorado eran exageradas y que estaba mejor en mi casita con mi familia. Hoy, 3 de Noviembre de 2012, no sólo no me arrepiento de haber entrado aquel día, si no que me gustaría entrar todos los días del resto de mi vida laboral. Y todo esto es gracias a mucha gente que me ha estado apoyando y ayudando desde que me metí en esta locura que es la realización de una tesis doctoral y espero que estas líneas sirvan para reconocer su indispensable labor.

Por supuesto nada de esto habría sido posible sin el apoyo incondicional de mi familia, que han servido de ejemplo y en ciertos momentos de refugio cuando parecía que había tormenta. Papi, Mami, muchas gracias por haber estado ahí cada día, escuchando, aconsejando, ayudando. Sois los mejores padres que se pueden tener y todo un ejemplo para mí. Os quiero mucho.

Una de las peores cosas que ha tenido el venirme a Barcelona ha sido el distanciarme físicamente de mis hermanos, a los que he echado mucho de menos. Muchas gracias Pepe por hacer que cuando vuelva a casa todo siga igual que antes de irme. Jaime, muchas gracias por esas pequeñas conversaciones mañanaderas, esas confianzas compartidas, esa comprensión.

También quisiera agradecer el apoyo de mi madrina Mariví, por haber estado siempre al otro lado dispuesta a escuchar y aconsejar. Eres la mejor. También agradecer esos ratos de diversión que paso con mi trolete, mi primo Alejandro. Muchas gracias Abuela Mari y tío Alejandro por hacer de cada visita una vuelta a casa. Agradezco también el apoyo y cariño del resto de mis tíos, tías, primos y primas.

Parte de esta tesis va dedicada a mi abuela Jose. Allá donde estés, me acuerdo de tí todos los días.

La verdad es que llevo siendo estudiante más de 20 años (y espero no dejar de serlo nunca) y durante este tiempo he conocido a personas que se han convertido en amigos para toda la vida. Mis amigos del colegio: Alfonso, Javier (¡quién nos iba a decir que acabaríamos los dos trabajando en colonoscopias!), Gabriel, Darío y Loreto (este año tampoco toca :D). Da gusto saber que hay gente con la que puedes estar cierto tiempo sin verte y, cuando nos encontramos, sentir como si el tiempo no hubiese pasado.

No puedo olvidar a mis amigos de la Universidad: Alberto, muchas gracias por haber estado siempre ahí, dispuesto a escuchar, a hacer locuras, a todo. ¡Vuelve ya

para España! Alicia, gracias por estar ahí cuando he vuelto a Valladolid, por quedar en el kiosko a menos cuarto, por buscarme cuando me pierdo. Ángela, gracias por haber estado siempre a la altura, por esas tardes de Talismán, por esas conversaciones por teléfono. Maite, por enseñarme en su día lo que es la verdadera amistad y por saber cuidarla. Gracias a todos por estar ahí siempre esperándome aunque me haya ido un poco lejos.

Otros momentos memorables en este doctorado han sido las reuniones con mis antiguos compañeros del C.E.E.T. Muchas gracias Marina, Félix, ese eje playero con Félix y Luis y mis compañeros de delegación de Valladolid, Pablo, Leticia y Yaiza. Gracias Natalia por el apoyo y el interés.

Yo jamás podría haber pensado que iba a encontrarme gente como la que trabaja en el CVC. Para mí muchos de ellos se han convertido en mi familia catalana y han sido sin duda un apoyo imprescindible durante estos cuatro años.

Ana María, aún recuerdo la primera vez que nos conocimos. Me hiciste dar la vuelta entera a la UAB para acabar yendo al edificio de al lado. Por estas cosas, y por estar siempre a mi lado desde el primer día, no concibo ir a trabajar al CVC y que tú no estés. Has sido mami, hermana y amiga a la vez, muchas gracias. Carlos, no he conocido en mi vida un amigo tan sincero como tú, que dice las cosas como tienen que decirse en el momento justo. Eres grande tío. Claire, pupi, pequeño pony ... gracias por dar a mi vida un toque de emoción e improvisación. Por estar siempre dispuesta a tomar un café. ¡Porque si no existieras, habría que inventarte!

Raulete, porque los telecos molamos mucho, por esas tardes en la Maquinista, por preocuparte. Rubén, porque yo me consideraba friki hasta que te conocí. Por esos planes de gimnasio que empezábamos y abandonábamos juntos. Raquel, por estar dispuesta siempre a quedar y a charlar un rato, por saber escuchar. Miguel, por esas discusiones futbolísticas. ¡Es hora de que demuestres al mundo el madridista que eres! Helena, muchas gracias por ofrecerme siempre trabajillos y ferias, por escuchar siempre mis ideas. Porque se que en el fondo te encanta Visio. Ferran, muchas gracias por el apoyo prestado tanto en el primer año como ahora.

A todas las chicas de administración: Raquel, Mireia, Eva (esos desayunos de Miércoles), Mari (por ayudarme a encontrar el camino), Pili, Gigi, Montse. ¡Sin vosotras el CVC no sería lo mismo! Al departamento de Marketing, Ainhoa y Marc, por ayudarme en los múltiples Jorge Tours que hemos hecho. A Joan Masoliver por ayudarme con esos problemillas informáticos de cada día, y por los ratos en el Turó. A Fownie (y a Marc) por llevarme a mundos a los que no se si quiero ir. Algún día volveré al WoW.

A mis compañeros del labo, que fueron mi familia desde el primer día. Anna, muchas gracias por las palabras de apoyo, por el cariño. Coen, por acercarte a mí desde el primer día, ¡porque eres un tío muy grande! Joanra, por esas conversaciones en el Turó, por darme mi primer pase de gol en el CVC. ¡Gracias! Enric, per fer-me parlar en català, per entener-me bé! ¡Joan Mas, por enseñarme que la Coca Cola Zero es el mal! Gracias a mi torracollons particular, Jordi, por elevar el nivel friki de algunas conversaciones. Gracias también a David Rotger, Marc Castelló, Albert Andaluz, Henry y Joan Soto (¡compi de piso!).

Por supuesto, a toda la generación del 2008 que empezamos el doctorado juntos. A Javi por ese abrazo cuando nos encontramos al pie del Snowdonia. ¡Me has hecho

sudar estos años! A Pep por ser tan buen amigo, aunque culé. Por ese final de El Internado. A David por esos ratos de risas viendo ciertas fotos, por esas conversaciones en las cenas del CVC. Thanks to 'guest star' Naila, for always giving me a smile! Thanks Noha and Wenjuan for all the plans we made together back in the days. Gracias a Fernando por ser mi compañero todos estos años. Gracias a Jaume, Albert (¿hacemos un Aribau?), Diego, David, Hilda, Shida, José Carlos, Mohammad, Farshad y el resto de compañeros del máster, por la ayuda y comprensión prestada. Gracias Sergio y Raúl por esas conversaciones, comidas, cenas, donde me lo he pasado tan bien.

Al resto de doctorandos que se han incorporado estos años (y otros que estaban ya). Gracias a Bhaskar, Marco, Nataliya, Zhanwu, Toni, Jon, David (muchas gracias por muchas cosas, ¡vecino!), Lluís Pere, Carles. Muchas gracias Mónica por esas conversaciones, esos cotilleos, esas terapias. Moltes gràcies Joan (Manel) per tota la comprensió, l'ajuda i per fer-me més madridista cada dia. Gracias Germán y Camp por ser de los que saludáis cada día y os acordáis del doctorando escondido en el rincón.

Gracias Gemma, por ayudarme en el mundo de los compiladores, por esas conversaciones en las comidas. Gracias a Ernest, Xavier Otazu, Jaume y Alicia por esos debates a la hora de comer. Muchas gracias a Aura por siempre devolver una sonrisa. Muchas gracias Debora por hacerme despertar cuando me has visto dormido. Moltes gràcies Poal per el suport que me has donat sempre, per les nostres converses. Gracias a Antonio López por ese detector de valles que ha dado tanto juego y por no dejarme parar un segundo cuando jugamos al fútbol. Gracias a Josep, Antoni y demás responsables del centro por el apoyo prestado en las actividades que he realizado.

Thanks to all the people i worked with during my research stage in Ireland. Thank you Gerry for the support and guidance. Thank you Mirko for your friendship, for all these days in the hospitals, for our conversations, for winning me playing chess. Thank you Ola, Stefan, Ziggy, Atul and the rest of the GV2 people. Gracias Claudia por cuidar de mí los últimos días de mi estancia.

Gracias a la música, a ese grupo de cinefrikis, a los compañeros de català, al sitio molón. Gracias Jeroni por ayudarme a salir de la oscuridad. Gracias Marta por cuidar de mis maltrechos músculos. Gracias Miguel Bote por aparecer siempre en momentos puntuales de mi vida.

No creo que existan palabras para definir el agradecimiento que tengo hacia tí Fernando. Muchas gracias por impulsar mi carrera investigadora, por darme siempre tan buenos consejos, por hacer el papel de revisor duro, por esas noches en Estambul y las Palmas de Gran Canaria. Por saber frenarme y, sobre todo, por comprenderme. Esto no hubiera sido lo mismo sin tí.

Si hoy estoy aquí, es por tí Xavi. Por elegirme en su día, por esa cara de pillo cuando vienes a verme para proponerme ideas, por esa conversación en la cafetería de Sant Cugat, porque eres muy grande como investigador y como persona. Por ayudarme en la docencia, en la tesis y en la vida. T'estimo molt, tio.

Dedico esta tesis a mis dos directores, Javier y Fernando, porque me habéis hecho sentirme muy cómodo durante estos años de doctorado. Esto no ha hecho más que empezar. ¡Muchas gracias!

This work was supported in part by a research grant from Universitat Autònoma

de Barcelona 471-01-3/08, by the Spanish Government through the founded project "COLON-QA" (TIN2009-10435) and by research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018).

Abstract

Colorectal cancer is the fourth most common cause of cancer death worldwide, with about 143,460 new cases expected in 2012 by recent estimates of the American Cancer Society. Colon cancer's survival rate depends on the stage in which it is detected, decreasing from rates higher than 95% in the first stages to rates lower than 35% in stages IV and V, hence the necessity for an early colon screening. In this process physicians search for adenomatous growths known as polyps, in order to assess their degree of development. There are several screening techniques but colonoscopy is still nowadays the gold standard, although it has some drawbacks such as the miss rate. Our contribution, in the field of intelligent system for colonoscopy [8], aims at providing a polyp localization and a polyp segmentation system based on a model of appearance for polyps. In this sense we define polyp localization as a method which given an input image identifies which areas of the image are more likely to contain a polyp. Polyp segmentation aims at selecting the region of the image that contains a polyp. In order to develop both methods we have started by defining a model of appearance for polyps, which defines a polyp as enclosed by intensity valleys [13]. The novelty of our contribution resides on the fact that we include in our model other elements from the endoluminal scene such as specular highlights and blood vessels, which have an impact on the performance of our methods and also other elements that appear as a result of image formation, such as interlacing. Considering this we define our novel Depth of Valleys image which integrates valley information with the output of the morphological gradient and also takes into account the presence of the before mentioned elements of the endoluminal scene. In order to develop our polyp localization method we accumulate the information that the Depth of Valleys image provides in order to generate accumulation energy maps. In order to obtain polyp segmentation we also use information from the energy maps to guide the process. Our methods achieve promising results in polyp localization and segmentation. In order to validate our methods we also present an experiment which compares the output of our method with physician's observations captured via an eye-tracking device. The results show to be close to physician's observations which point out a potentially inclusion of our methods as part of a future intelligent system for colonoscopy.

Resumen

El cáncer de colon es la cuarta causa más común de muerte por cáncer en el mundo, presentando alrededor de 143.460 nuevos casos en 2012, según datos estimados por la American Cancer Society. La tasa de mortalidad del cancer de colon depende de la etapa en que éste es detectado, decreciendo desde tasas mayores del 95% en las primeras etapas a tasas inferiores al 35% en las etapas cuarta y quinta, lo cual es muestra de la necesidad de una exploración temprana del colon. Durante la exploración el médico busca crecimientos adenomatosos que se conocen bajo el nombre de pólipos, con el fin de averiguar su nivel de desarrollo. Existen varias técnicas para la exploración del colon pero la colonoscopia está considerada aún hoy en día como estandar de facto, aunque presenta algunos inconvenientes como la tasa de fallos. Nuestra contribución, encuadrada dentro del campo de sistemas inteligentes para la colonoscopia [8], tiene como objetivo el desarrollo de métodos de localización y segmentación de pólipos basándose en un model de apariencia para los pólipos. Definimos localización de pólipos como el método por el cual dada una imagen de entrada se proporciona como salida donde se señalan las áreas de la imagen donde es más probable encontrar un pólipo. La segmentación de pólipos tiene como objetivo definir la región concreta de la imagen donde está el pólipo. Con el fin de desarrollar ambos métodos se ha comenzado desarrollando un modelo de apariencia para pólipos, el cual define un pólipo como una estructura limitada por valles en la imagen de escala de gris [13]. Lo novedoso de nuestra contribución radica en el hecho de que incluimos en nuestro modelo otros elementos de la escena endoluminal tales como los reflejos especulares o los vasos sanguíneos que tienen un impacto en el rendimiento global de nuestros métodos así como elementos derivados del proceso de formación de la imagen, como el interlacing. Teniendo esto en cuenta definimos nuestra imagen de profundidad de valles que integra la salida del detector de valles con el gradiente morfológico, añadiendo asimismo la presencia de los ya mencionados otros elementos de la escena endoluminal.

Para desarrollar nuestro método de localización de pólipos acumulamos la información que la imagen de profundidad de valles proporciona con el fin de crear mapas de energía. Para obtener la segmentación del pólipo también usamos información de los mapas de energía para guiar el proceso. Nuestros métodos alcanzan resultados prometedores tanto en localización como en segmentación de pólipos. Con el fin de validar nuestros métodos presentamos también un experimento que compara nuestro método con las observaciones de los médicos, obtenidas mediante un dispos-

itivo eye-tracker. Los resultados muestran que nuestros métodos son cercanos a las observaciones de los médicos, lo cual indica el potencial de los mismos de cara a ser incluidos en futuros sistemas inteligentes para la colonoscopia.

Resum

El càncer de còlon és la quarta causa més comuna de mort per càncer en el món, havent al voltant de 143.460 nous casos el 2012, segons dades de l'American Cancer Society. La taxa de mortalitat d'aquest càncer depèn de l'etapa en què sigui detectat, decreixent des de taxes majors del 95% en les primeres etapes a taxes inferiors al 35% en les etapes quarta i cinquena, la qual cosa demostra la necessitat d'una exploració precoç del còlon. Durant l'exploració, el metge busca creixements adenomatosos, que es coneixen sota el nom de pòlips, per tal d'esbrinar el seu nivell de desenvolupament. Hi ha diverses tècniques per a l'exploració del còlon, però la colonoscòpia és considerada encara avui com estàndard de facto, encara que presenta alguns inconvenients com una elevada taxa de errors.

La nostra contribució està enquadrada dins del camp de sistemes intel·ligents per a la colonoscòpia [13] i té com a objectiu el desenvolupament de mètodes de localització i segmentació de pòlips basant-se en models d'aparença. Definim la localització de pòlips com el mètode pel qual donada una imatge d'entrada es proporciona com a sortida una altra imatge on es assenyalen les àrees de la imatge on és més probable trobar un pòlip. Per altra banda, la segmentació té com a objectiu definir la regió concreta de la imatge on es troba el pòlip.

Per tal de desenvolupar ambdós mètodes s'ha començat desenvolupant un model d'aparença per a pòlips, el qual defineix un pòlip com una estructura limitada per valls en la imatge de escala de gris [9]. La novetat de la nostra contribució rau en el fet que incloem en el nostre model altres elements de l'escena endoluminal, com els reflexos especulars o els gots sanguinis, que demostren tenir un impacte en el rendiment global dels nostres mètodes i en elements derivats del procés de formació de la imatge com l'interlacing. Tenint això en compte, definim una imatge de profunditat de valls que integra la sortida del detector de valls amb un gradient morfològic, afegint-hi els altres elements de l'escena endoluminal. Per desenvolupar el nostre mètode de localització de pòlips capturem la informació de la imatge de profunditat de valls amb la finalitat de crear mapes d'energia. Per obtenir la segmentació del pòlip també fem servir aquests mapes de energia per guiar el procés.

Els nostres mètodes arriben resultats prometedors tant en localització com a segmentació de pòlips. Per a validar la metodologia emprada, presentem un experiment que compara el nostre mètode amb observacions fetes per metges obtingudes mitjançant un dispositiu de eye-tracking. Els resultats mostren que el rendiment de la

contribució d'aquesta Tesi és comparable a l'obtingut a partir de les observacions dels metges, la qual cosa indica el potencial de la nostra proposta en futurs sistemes intel·ligents de colonoscòpia.

Contents

Acknowledgements	i
Abstract	v
Resumen	vii
Resum	ix
1 Introduction	3
1.1 Colon cancer facts and figures	3
1.2 Colonoscopy as the gold standard for colon screening. Another alternatives for colon screening.	5
1.3 Current challenges in colon cancer detection by means of colonoscopy	8
1.4 Scope of the research: intelligent systems for colonoscopy	9
1.5 Objective of the thesis	11
1.6 Introduction of polyp localization and segmentation methods	12
1.7 Contributions of this thesis	13
1.8 Structure of the thesis	14
2 Endoluminal scene description: challenges and methods	17
2.1 Introduction to the Endoluminal Scene. The elements of the endoluminal scene	17
2.2 Introduction to the review of methods devoted to endoluminal scene object identification	19
2.3 Image Formation and Enhancement	20
2.3.1 Image Preprocessing and Enhancement	20
2.3.2 Specular highlights detection and inpainting	21
2.3.3 Non-informative frame definition	23
2.4 Endoluminal scene object description	25
2.4.1 Lumen Detection	25
2.4.2 Polyp Detection	27
Shape-based approaches	28
Texture-based approaches	32
2.4.3 Blood vessels characterization	34
2.5 Domains of application of endoluminal scene description methods . . .	35

2.5.1	Computer-aided Diagnosis as a potential domain of application of our methods	36
2.5.2	Other possible domains of application	37
3	A model of appearance for polyps	41
3.1	Introduction	41
3.2	Definition of polyp	41
3.3	Phong's illumination model	42
3.4	Cues on polyp characterization by means of an illumination-based model	46
3.4.1	Valleys surrounding the polyp	47
3.4.2	Contours and edges in lateral view	47
3.4.3	Internal shadings of the polyp	50
3.4.4	Specular highlights	51
3.4.5	Discussion	52
4	Valley detection for polyp localization and segmentation	55
4.1	Introduction	55
4.2	Valley detection	55
4.2.1	Introduction	55
4.2.2	Second Derivative of Gaussians	56
4.2.3	Multilocal Creaseness-based Valley Detector	58
4.2.4	Comparison between Second Derivative of Gaussians and Multilocal Creaseness	59
4.3	Issues on polyp boundary characterization by means of valley information	61
4.3.1	Introduction	61
4.3.2	Inpainting method	62
4.4	Image preprocessing to enhance polyp boundary-originated valleys. . .	64
4.4.1	Image formation-derived artifacts	65
	Color channel misalignment	65
	Deinterlacing	65
	Inpainting of black mask	66
4.4.2	Effect of elements of the endoluminal scene	67
	Specular highlights	67
	Blood vessels mitigation	70
4.5	Depth of Valleys image as a method to capture the boundaries of polyps	73
4.6	Discussion and summary of the chapter	75
5	Polyp localization	79
5.1	Introduction	79
5.2	Depth of Valleys accumulation based on the shape of polyp boundaries	81
5.2.1	Ellipse Fitting Depth of Valleys Accumulation (EF-DOVA) . .	81
5.3	Depth of Valleys accumulation based on the continuity of polyp boundaries	85
5.3.1	Sector Accumulation Depth of Valleys Accumulation (SA-DOVA)	85
5.3.2	Windows Median Depth of Valleys Accumulation (WM-DOVA)	89
5.3.3	Valley Orientation Depth of Valleys Accumulation (VO-DOVA)	94

5.4	Discussion and Qualitative comparison	99
6	Polyp segmentation	105
6.1	Introduction	105
6.2	Polyp segmentation by smart region merging	106
6.2.1	Image preprocessing	106
6.2.2	Watershed segmentation	107
6.2.3	Region and boundary labeling	107
6.2.4	Region Merging	109
	Region information-based region merging	110
	Depth of valleys-based Region Merging:	111
6.3	Polyp segmentation from the output of an energy map	113
6.4	Visual comparison of segmentation methods. Discussion	117
7	Experimental Results	121
7.1	Introduction	121
7.2	Experimental setup.	122
7.2.1	Ground truth and database building	122
	Building up of a database	123
7.2.2	CVC COLON DB database	123
7.3	Image preprocessing results	127
7.3.1	Experimental results on specular highlights detection and in-painting	127
7.3.2	Preliminary study on blood vessels mitigation	132
7.4	Polyp localization results	135
7.4.1	Polyp localization results on the original input image	135
7.4.2	Impact of image preprocessing techniques on polyp localization results	143
7.5	Polyp Segmentation results	152
8	Use of a polyp localization method to model physicians' observations	163
8.1	Motivation	163
8.2	Objective of the analysis	164
8.3	From eye tracking to saliency theory. Integrated analysis of fixations via task-driven saliency maps.	165
8.3.1	Analysis of the fixations	166
8.3.2	Introduction to saliency theory	166
8.4	Linking eye tracking with saliency. Task-driven saliency maps	167
8.5	Experimental Setup	169
8.5.1	COLON-PHYS database	169
8.5.2	Metrics used in the analysis	170
8.6	Eye tracking data analysis	172
8.6.1	Reaction time	172
8.6.2	Dwelling time	176
8.6.3	Energy concentration rate	177
8.7	From polyp localization to saliency maps: DOVA saliency maps	180

8.8	Comparison of DOVA saliency maps with physicians' task driven saliency maps	182
8.8.1	Reaction time	182
	Case by case analysis of videos	184
8.8.2	Dwelling time	190
8.8.3	Concentration rate	190
8.9	Preliminary results of the use of DOVA saliency maps as polyp detectors	193
8.10	Discussion	194
9	Conclusions and Future Work	197
9.1	Conclusions	197
9.2	Future Work	198
10	List of Publications	201
10.1	Journals, Book chapters	201
10.2	International Conferences	201
10.3	National Conferences	202
10.4	Internal Conferences and Technical Reports	202
A	Technical data	203
A.1	Technical equipment	203
A.2	Image preprocessing and Valley Detection	203
A.3	Polyp localization	203
A.4	Polyp segmentation	204
	Bibliography	207

List of Figures

1.1	Colon cancer survival rate	4
1.2	Adenoma-carcinoma sequence overview	5
1.3	Example of colonoscopy intervention	6
1.4	Real colonoscope devices	7
1.5	Wide angle field of view in colonoscopy	8
1.6	Research lines and potential applications in the scope of intelligent systems for colonoscopy.	10
1.7	Polyp localization processing scheme.	12
1.8	Example of polyp localization output	13
1.9	Polyp segmentation processing scheme.	13
1.10	Example of polyp segmentation output	14
2.1	Endoluminal scene	18
2.2	Color channel misalignment examples	20
2.3	Specular highlights examples	21
2.4	Non-informative frame definition	24
2.5	Operation shot detection system	25
2.6	Examples of lumen in colonoscopy frames	26
2.7	Taxonomy of Feature Descriptors	29
3.1	Type of polyps	42
3.2	Decomposition of the light reflected by the surface into the three components: ambient; diffuse and specular.	43
3.3	Model of appearance for polyps	44
3.4	Theoretical model applied on real images	45
3.5	Challenges in polyp detection	46
3.6	Valley detection as a cue to guide our methods	47
3.7	Model of appearance for lateral polyps	48
3.8	Theoretical lateral polyps model applied on real images	49
3.9	Variation of grey level inside the polyp	50
3.10	Specular highlights as a cue to guide our methods	51
3.11	Challenges that present the use of valley detectors	52
4.1	Use of second derivative of Gaussians as valley detector	57
4.2	Application of <i>multilocal creaseness</i> detector	59

4.3	Graphical comparison between <i>multilocal creaseness</i> and <i>second derivative of gaussians</i>	60
4.4	Valley detection on an endoluminal scene	62
4.5	Graphical example of the extension of the M mask.	64
4.6	Examples of color channel misalignment	65
4.7	Losing part of polyp contour by eliminating all black mask content . .	66
4.8	Black mask inpainting	67
4.9	Specular highlights detection and inpainting	68
4.10	Extension of specular highlights mask	70
4.11	Specular highlights detection and inpainting examples	71
4.12	Difference in contrast of blood vessels according to the channel observed	72
4.13	Valley information in separate channels of the RGB image	72
4.14	Difference in performance between gradient and morphological gradient	73
4.15	Morphological gradient in colonoscopy frames	74
4.16	Depth of Valleys image	75
4.17	Image preprocessing and valley detection processing scheme	76
4.18	Improvement on the valley image achieved by addressing the challenges presented	77
5.1	Polyp localization processing scheme	79
5.2	Definition of interior and exterior points to objects	80
5.3	Graphical explanation of the depth of valleys accumulation algorithm	83
5.4	EF-DOVA algorithm on synthetic images	83
5.5	Examples of EF-DOVA energy maps superimposed to the original image	84
5.6	Examples of bad performance of EF-DOVA energy maps	84
5.7	Graphical explanation of accumulation by SA-DOVA	86
5.8	Explanation of SA-DOVA algorithm	87
5.9	Examples of SA-DOVA accumulation superimposed to the original image	88
5.10	Example of SA-DOVA energy maps	89
5.11	Example of SA-DOVA energy maps	90
5.12	Graphical interpretation to SA-DOVA and WM-DOVA results	91
5.13	Graphical scheme of WM-DOVA accumulation algorithm	93
5.14	Examples of WM-DOVA energy maps superimposed to the original image	94
5.15	Example of cases where WM-DOVA output fails	95
5.16	Example of the definition of VO-DOVA energy map	95
5.17	Example of VO-DOVA output	97
5.18	Examples of VO-DOVA energy maps	98
5.19	Discussion of EF-DOVA energy map results	100
5.20	Discussion of SA-DOVA energy map results	101
5.21	Discussion of WM-DOVA energy map results	101
5.22	Discussion of VO-DOVA energy map results	102
5.23	Visual comparison of DOVA	104
6.1	General processing scheme of polyp segmentation	106
6.2	Image preprocessing stage	106
6.3	Watershed segmentation starting with gradient image	107

6.4	Example of region labeling and region merging	109
6.5	Region merging stage in region segmentation	111
6.6	Polyp segmentation after region merging method results	112
6.7	Problems in segmentation	113
6.8	Polyp segmentation after region merging method results	113
6.9	Examples of polyp segmentation from the output of an energy map . .	114
6.10	Graphical representation of contour improving method	115
6.11	Softening of boundaries by median filtering in polar space	116
6.12	Examples of our contour refinement method	117
6.13	Polyp segmentation from the output of an energy map results	118
6.14	Comparison between the segmentation performed by the two methods presented	119
6.15	Segmentation errors achieved by the two methods presented	120
7.1	Colonoscopy data acquisition	124
7.2	CVC Colon DB database examples	125
7.3	Content of CVC COLON DB database	126
7.4	Specular highlights mask	128
7.5	Comparison of specular highlights detection methods	131
7.6	Specular highlights inpainting comparison	132
7.7	Blood vessels database examples	139
7.8	Blood vessel impact on several channels	140
7.9	EF-DOVA performance results	141
7.10	Improvement of WM-DOVA against SA-DOVA and VO-DOVA	142
7.11	Improvement of polyp localization results by inpainting of the black mask	144
7.12	Improvement of polyp localization results by correcting specular high- lights	145
7.13	Improvement of polyp localization results by mitigating blood vessels effect	146
7.14	Improvement of polyp localization results by suppressing contribution of non-informative regions	147
7.15	Improvement of polyp localization results by combining all the methods explained	150
7.16	Errors of polyp localization results	151
7.17	Examples of AAC and DICE results	154
7.18	Qualitative comparison of segmentation results	156
7.19	Comparison of segmentation results after applying our two segmenta- tion methods (I)	157
7.20	Comparison of segmentation results after applying our two segmenta- tion methods (II)	158
7.21	Segmentation results obtained by polyp segmentation from the output of an energy map.	159
7.22	Comparison of final classification results after applying our two seg- mentation methods	161

8.1	Graphical representation of the scope of the analysis.	165
8.2	Creation of task-driven saliency maps	169
8.3	Ground-truth for Eye Tracking analysis	171
8.4	Difference in reaction time	175
8.5	Difference in concentration rate between experts and novices during all the frames with polyp.	180
8.6	DOVA saliency maps from the output of polyp localization	181
8.7	Analysis of source of DOVA retard for video 3	186
8.8	Analysis of source of DOVA retard for video 3	187
8.9	Analysis of source of DOVA retard for video 3	188
8.10	Analysis of source of DOVA retard for video 3	189
8.11	Some reasons for errors on locating polyps	191
A.1	Processing time graphs	206

List of Tables

2.1	Summary of shape-based descriptors for polyp detection.	31
2.2	Summary of texture descriptor-based methods for polyp detection. . .	34
6.1	Region labeling criteria.	108
6.2	Criteria used to merge regions.	110
7.1	Database description.	126
7.2	Methods use in the specular highlights detection and inpainting experiments.	128
7.3	Comparison of specular highlight detection methods (I).	129
7.4	Comparison of specular highlight detection methods (II).	129
7.5	Comparison of specular highlights inpainting methods.	129
7.6	Possible input images.	133
7.7	Combination of parameters used for the ridges and valleys detector. .	134
7.8	Summary of results.	134
7.9	Summary of results.	134
7.10	Summary of results.	135
7.11	Parameters of EF-DOVA.	136
7.12	Parameters of SA-DOVA.	136
7.13	Parameters of VO-DOVA.	137
7.14	Parameters of WM-DOVA.	137
7.15	Comparison between the different DOVA maps for polyp localization.	137
7.16	Processing time comparison results.	138
7.17	Impact of black mask inpainting on polyp localization results.	143
7.18	Impact of specular highlights on polyp localization results.	145
7.19	Impact of blood vessels mitigation on polyp localization results (sB channel as input).	145
7.20	Impact of blood vessels mitigation on polyp localization results (sG-sB as input).	146
7.21	Impact of non informative region identification on polyp localization results.	147
7.22	Combined impact of image preprocessing and non informative region identification on polyp localization results.	148
7.23	Comparison of segmentation results with some state-of-the-art methods	155

7.24	Comparison between the results obtained by our two methods.	156
7.25	Comparison between the results obtained by our two methods only for images with correct placing of maxima of DOVA.	158
7.26	Comparison between the results obtained by our two methods.	160
8.1	Number of interventions performed by each physician.	170
8.2	Experts' fixation results.	173
8.3	Novices' fixation results.	173
8.4	Difference Reaction time.	174
8.5	Dwelling time results.	176
8.6	Number of frames with polyp for each video.	177
8.7	Dwelling time percentage results.	177
8.8	Difference in concentration rate in the experts' first fixation frame. . .	178
8.9	Difference in concentration rate in the novices' first fixation frame. . .	179
8.10	Difference in concentration rate in the experts' first fixation frame. . .	179
8.11	DOVA saliency maps fixation results.	183
8.12	Comparison of DOVA saliency maps' reaction time with novices. . . .	183
8.13	Comparison of DOVA saliency maps' reaction time with experts. . . .	184
8.14	Dwelling time results.	190
8.15	Difference in concentration rate in the DOVA saliency maps' first fix- ation frame.	192
8.16	Difference in concentration rate in the DOVA saliency maps' first fix- ation frame.	192
8.17	Comparison of mean of WM-DOVA energy map for all the image, frames with and without polyp.	193
8.18	Comparison of maxima of WM-DOVA energy map for frames with polyp and without polyp.	194
8.19	Comparison of maxima of WM-DOVA energy map for frames with polyp, inside and outside the polyp mask.	195
A.1	Computation time of preprocessing methods	204
A.2	DOVA energy map processing times	204
A.3	Computation time of polyp localization	204
A.4	Computation time of segmentation	205
A.5	Polyp segmentation computation time	205

Chapter 1

Introduction

Before starting to develop the core of this thesis, it is necessary to introduce the scope of our research, which is the development of polyp localization and segmentation methods by means of a model of appearance for polyps in the context of colonoscopy. Without entering into concrete details of our method, we will pave the way in this chapter by making a first introduction to the problem we aim to solve. Therefore we will present the basic aspects of colon cancer and colonoscopy by then introducing the basics of our work, including a first description of our processing schemes, along with sketching the contributions that will be explained in this thesis.

1.1 Colon cancer facts and figures

Colorectal cancer is the third most common in incidence and the fourth most common cause of cancer death worldwide, with about 143.460 new cases expected in 2012 by the most recent estimates of the American Cancer Society for the number of colorectal cancer cases in the United States only [2]. Based on demographic trends, the annual incidence is expected to increase by nearly 80% to 2.2 million cases over the next two decades and most of this increase will occur in the less developed regions of the world. Fortunately, experience in Europe [95] has shown that systematic early detection and treatment of colorectal lesions before they become symptomatic has the potential to improve the control of the disease, particularly if they are effectively integrated into an overall programme of comprehensive cancer control.

Colon cancer's survival rate depends on the stage in which it is detected, decreasing from rates higher than 95% in the first stages to rates lower than 35% in stages IV and lower than 5% in stage V [106] (see Figure 1.1); hence the importance of detecting it on its early stages by using screening techniques, such as colonoscopy [44].

Colorectal cancer (CRC) is suitable for screening. The disease is believed to develop in a vast majority of cases from non-malignant precursor lesions called adeno-

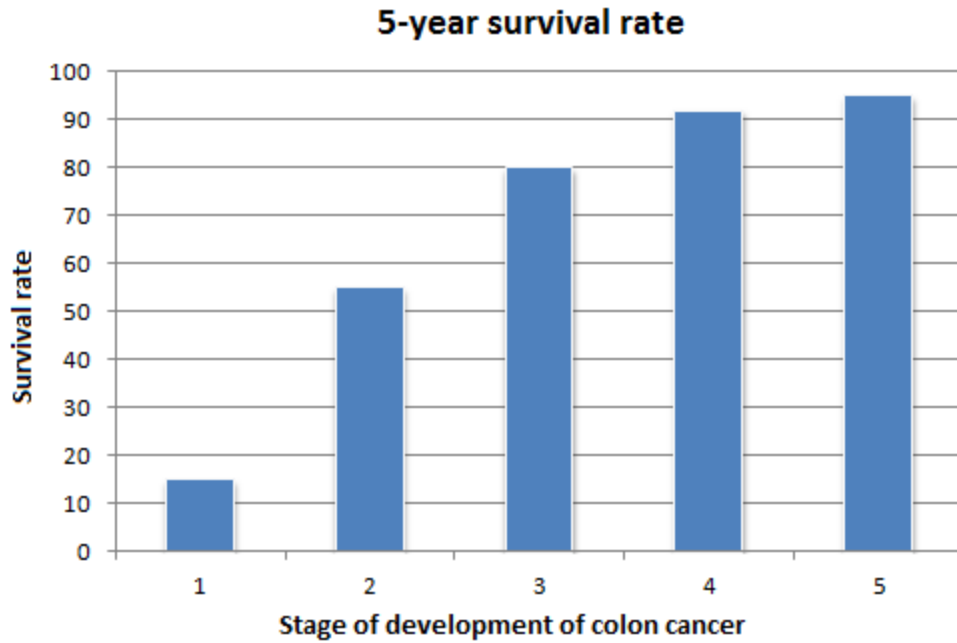


Figure 1.1: Survival rate of colon cancer according to the stage it is detected on.

mas, according to the adenoma-carcinoma sequence, as can be seen in Figure 1.2. Adenomas can occur anywhere in the colon rectum after a series of mutations that cause neoplasia of the epithelium. Adenomas are most often polypoid, but can also be sessile or flat. An adenoma may grow in size and develop into high-grade neoplasia.

At a certain point in time, the adenoma can invade the submucosa and become malignant. Using the same stage-based scale than the one mentioned before, at first this malignant cancer is not diagnosed initially and does not give symptoms yet (preclinical stage). Then it can progress from localised (stage I) to metastasised (stage IV) cancer, until it causes symptoms and is diagnosed. Although the average duration of the development of an adenoma to cancer is unobserved, it is estimated to take at least 10 years, which provides an extensive window of opportunities for early detection of the disease. Screening techniques are needed to observe the colon and, consequently, prevent any lesions.

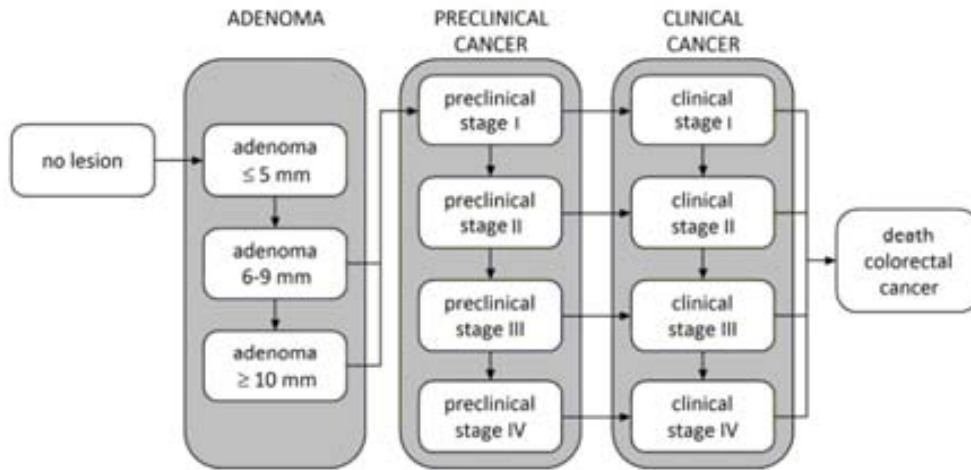


Figure 1.2: Schematic overview of the adenoma-carcinoma sequence [95].

1.2 Colonoscopy as the gold standard for colon screening. Another alternatives for colon screening.

The aim of screening is to lower the burden of cancer in the population by discovering diseases in its early latent stages. It is necessary to mention the potential harm that may be caused by CRC screening, which may include the creation of unnecessary anxiety and morbidity, inappropriate economic cost, and exposure to the risk of invasive procedures for detection and diagnosis as well as for removal of lesions detected in screening [61].

One of the most techniques for colon screening is optical colonoscopy (see Figure 1.3). Colonoscopy is useful in diagnosing and treating patients with neoplasms, strictures or colonic mucosal disease previously diagnosed on radiological imaging. Other uses include the evaluation of patients with gastrointestinal hemorrhage (hematochezia and occult bleeding) ([32] or [56]), screening and surveillance for colonic neoplasms ([116, 60]) or diagnosis and surveillance of inflammatory bowel disease ([6], among others.

Colonoscopy [44] consists of a procedure to see inside the colon and rectum and it is able to detect inflamed tissue, ulcers, and abnormal growths [44]. By means of colonoscopy, polyps as small as one millimetre or less can be removed in order to be studied later to determine if they are precancerous or not. During colonoscopy, patients lie on their left side on an examination table [36]. The physician inserts a long and flexible tube called colonoscope into the anus and guides it slowly through the rectum and into the colon. A small camera is mounted on the scope and transmits a video image from inside the large intestine to a computer screen, allowing the doctor to examine carefully the intestinal lining. The colonoscopy procedure itself consists

of several stages, such as bowel preparation or sedation of the patient, apart from the navigation through the colon. The former has a great impact in the overall performance of the colonoscopy because if patient's preparation is done poorly, the physician will not be able to detect easily adenomas. During this process the physician can remove polyps and later test them in a laboratory to look for signs of cancer.



Figure 1.3: Graphical example of a colonoscopy intervention [35].

It is important to understand how colonoscopy images are acquired as the resulting frame will be the input of the several algorithms that we will describe in this thesis. The colonoscope (see Figure 1.4) is a device that consists of a proximal housing, a flexible insertion tube, and an 'umbilical cord' connecting the light source and the proximal housing. The proximal housing, which is designed to be held in one hand, typically includes the eyepiece (fiberoptic models only), controls for distal tip (bending section) angulation and suction, and the working channel port. Colonoscopes have several hollow channels for suction, water and air delivery, and insertion of accessory instruments and cannulae. The distal tip of video colonoscopes includes a charge-coupled device (CCD) that serves as a small camera and electronically transmits the image from the CCD to an external video-processing unit. Video colonoscope insertion tubes contains a fiberoptic light bundle, which transmits light from the light source to the tip of the endoscope. Each fiberoptic bundle consists of thousands of

individual glass fibers coated with glass causing internal reflections that allow light transmission through the fiber even when it is flexed. The light is used to illuminate the field of view in the patient's colon. Video images are detected by the CCD and are then transmitted to the video processor and then display monitors or recording devices.



Figure 1.4: Example of real colonoscopes: (a) Olympus; (b) Pentax.

There are several companies that produce colonoscopes, being the most relevant Olympus [82] and Pentax [83]. As it can be seen in Figure 1.5 and in the manufacturer's websites cited before, there is a trend that consists of increasing the size of the field of view (up to 170° in some models) to allow a more comprehensive observation of the colon which may facilitate scope maneuvering for shorter examination times and enhanced efficiency. The use of wide angle technology allows the physician to cover a greater part on the endoluminal scene but it has a drawback: there are some parts of the image where the content is not clear enough to be shown on a screen. The most common solution consists of adding a black border around the image to cover these unclear areas.

Colonoscopy is not the only alternative for colon screening. For instance, *sigmoidoscopy* [94] is also used, being the difference between this and colonoscopy related to the parts of the colon that each of them can examine: by using colonoscopy we can examine the entire colon whereas by using sigmoidoscopy we can only examine the distal part. Nevertheless sigmoidoscopy can potentially be used for preventing colon cancer, as the benefits to cancer survival of colonoscopy have been limited to detecting lesions in the distal part of the colon [7]. Another alternative is the use of *virtual colonoscopy* [30], which uses 2D and 3D images reconstructed by means of either Computed Tomography (CT) or Nuclear Magnetic Resonance (NMR) scans. This procedure is meant to be totally non-invasive and needs of a smaller patient exploring time (around 10 minutes) but it has a big drawback: if a lesion is detected it can not be removed without having to practice again an intervention to the patient (a biopsy), leading frequently to performing a conventional colonoscopy after a virtual one. As of today, conventional colonoscopy remains as the gold standard for colon screening.

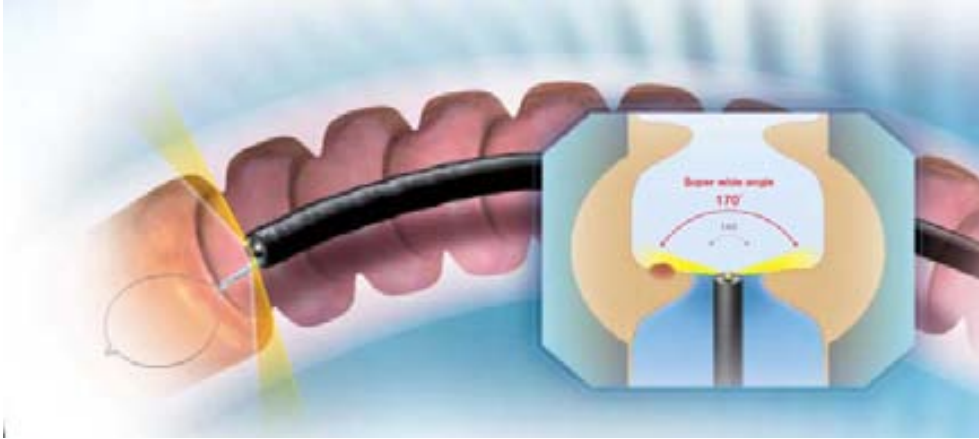


Figure 1.5: Example of the use of wide angle field of view to acquire colonoscopy images [82].

Other approaches include *fecal occult blood test* or fecal immunochemical test (FIT) [87], because one of the symptoms of large polyps or colon cancer is bleeding. Blood loss into the colon may be slow and chronic and not visible to the naked eye and a stool test can be performed to occult blood loss. If blood is detected in fecal samples, a colonoscopy is recommended to determine the cause of the bleeding. Another option is *double contrast barium enema* [66], which consists of an X-ray test similar to virtual colonoscopy but, in this case, barium is put into the colon and rectum using an enema. The liquid barium enables a healthcare provider to see a detailed outline of the colon and rectum in an X-ray. From the outlines of the procedure a health-care provider may be able to detect polyp presence. A more newer method is *fecal DNA testing* [64] in which a stool sample is collected and sent to a specialized lab to search for genetic mutations that can arise in large colon polyps and colon cancer. Finally another alternative is the use of wireless capsule endoscopy [49, 111] which provides a way to record images of the digestive tract by means of a capsule which is introduced in the patient. This capsule contains a camera and after it is swallowed by the patient, it takes pictures of the inside of the gastrointestinal tract.

Although colonoscopy has been widely accepted as the gold standard for colon screening, it has some drawbacks, such as the risk of perforation, the intervention cost, or visualization difficulties among others. We will detail some of these drawbacks in the next section.

1.3 Current challenges in colon cancer detection by means of colonoscopy

Miss-rate is one of the most relevant problems associated to colonoscopy meaning this that some polyps are still missed. The colorectal cancer miss rate of colonoscopy has

been reported to be as high as 6% [20] and the miss rate for adenomas larger than 1 cm ranges from 12% to 17% rates [85]. There are several causes to the failures in the detection, being the most relevant the complications in bowel preparation which will not be discussed in this thesis. An interested reader may find useful information about this topic in [70].

Since the scope of our thesis will be focused on the development of polyp localization and segmentation methods, we should take into account the clinical importance of missing a polyp. We can see how the size of the polyp has a direct relation with the severity of the lesion, which results on having separate miss rates for the different types of polyps. It is also necessary to be mentioned that physicians find it difficult to deal with small (6 to 9 mm) or diminutive (less or equal than 5 mm) adenomas [29]. As reported in the cited editorial and also mentioned in the previous section of this chapter, there are two techniques in the field of colon screening that coexist: *optical colonoscopy* and *virtual colonoscopy*. Referred to this last group, patients with any polyp with diameter higher than 10 mm or more than three 6 to 9 mm adenomas on virtual colonoscopy are later driven to optical colonoscopy. Still nowadays the removal of polyps with diameter smaller than 5 mm leads to inconsistencies on whether to remove them or not, although it seems to be a trend on removing them, as mentioned in the previously-cited editorial. The clinical significance of diminutive adenomas is small, which rarely harbor cancer [27] although they should be detected because they can later grow up and develop into cancer.

Finally there are some other drawbacks that need to be mentioned although as they are out of our scope of research such as the risk of bleeding or perforation [88] or the high intervention cost, which also suffers from bowel preparation problems [90]. We will present in the next section the scope of this thesis, which aims at providing tools that can aid on mitigate one of the drawbacks of colonoscopy, the miss rate on polyp detection.

1.4 Scope of the research: intelligent systems for colonoscopy

Before presenting the objective of this thesis, it is necessary to put in context our research, which could be enclosed into the field of intelligent systems for colonoscopy. During the last decades there is a trend that consists of developing intelligent systems for medical applications. Intelligent systems are currently being used to assist in other medical interventions. For instance, there are systems that can interpret medical data automatically, such as KARDIO [18], which was developed to interpret electrocardiograms. It is possible to find many examples of intelligent systems built to assist in cancer detection. The interested reader can consult some works in the field of breast cancer detection [115] or prostate cancer detection [113], only to mention a few examples.

The question that arises now is: **how can intelligent systems help in colonoscopy?** What kind of applications these systems can be built for? In Figure 1.6 we depict

some of the potential areas related to colonoscopy where an intelligent system can play a key role.

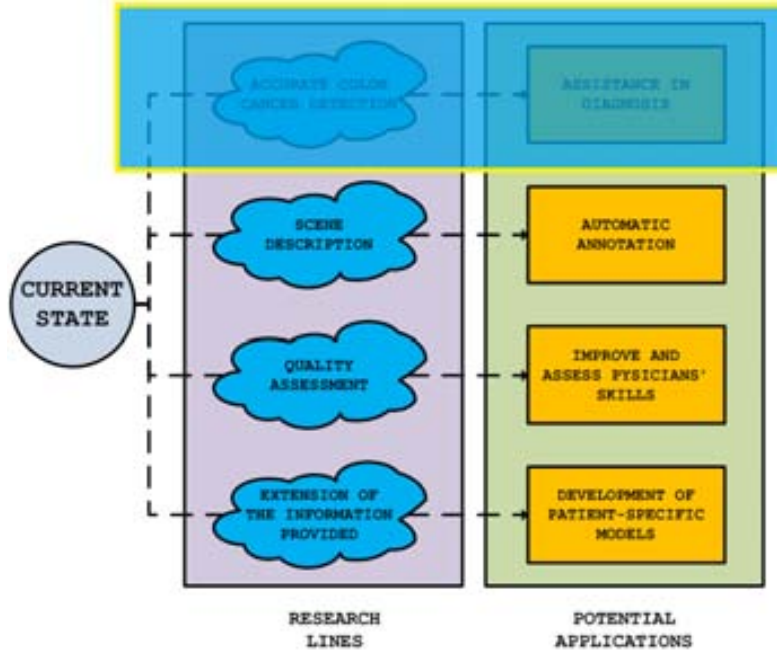


Figure 1.6: Research lines and potential applications in the scope of intelligent systems for colonoscopy.

As shown in Figure 1.6, we foresee four different areas where an intelligent system can be introduced and add significant value to the colonoscopy procedure [8]:

1. The most manifest application of this kind of systems could be the **assistance in the diagnosis procedure** during the intervention or in post-intervention time. This could be very useful in order to reduce the miss rate associated to polyp identification.
2. We can make use of the scene description provided by an automatic system -including the presence of salient traits, such as informative frames, anatomical structures, insertion or withdrawal phases, etc.- in order to **automatically annotate colonoscopy videos**. This would potentially provide a very efficient way of case annotation, with multiple uses in different applications.
3. In addition, an intelligent system may offer a quality assessment of the colonoscopy intervention, which could provide a non-subjective way of assessment. This could also be used as a way to train physicians in a way such they can **assess and improve their skills** without the cost associated to a real interventions, and it would allow to compare different performance metrics objectively.

4. We can also think about intelligent systems that extend and provide additional information from colonoscopy data. Belonging to this area we can think of applications such as the **development of patient-specific models**, that can be re-used later, when a new study arrives, to check for coincidences that can help in the diagnosis and enrich in this way a final case report.

1.5 Objective of the thesis

Out of all the domains of application that we have presented in the previous subsection, the methods presented in this thesis could be enclosed in the first and second topics. We present a polyp localization method in which given an input image we offer as output an energy image where the brighter areas corresponds to those parts of the image where the likelihood of having a polyp is higher. We also provide a polyp segmentation method in which given an input image we offer as output an image where we highlight the pixels that constitute the polyp.

To achieve good performance of both methods we have defined a robust model of polyp appearance that does not only take into account how polyps appear on colonoscopy images but also the presence of some elements of the endoluminal scene, such as blood vessels or specular highlights, which could affect the performance of our system. More specifically, we present a model of polyp appearance which is based on how polyps do appear in colonoscopy frames. This implies a different path than some other approaches in the literature as will be seen in Chapter 2 which are devoted to detect polyps based on fitting to concrete shapes. Our model, along with a study on how colonoscopy images are created, leads to infer an illumination model which models polyps as structures delimited by intensity valleys. Taking this into account we develop our algorithms by paying special attention to the valleys that surround three dimensional objects such as polyps or folds, differentiating them from other structures such as blood vessels or specular highlights. We also undertake the task of eliminating specular highlights information in order to provide an input image (which is the novel depth of valleys image) to our system where the majority of valley information would ideally be constituted by polyp boundaries.

Once our input depth of valleys image is presented, we develop a polyp localization method by integrating depth of valleys information. We offer two alternatives to polyp localization, depending on whether we are interested in the shape of the boundaries or their completeness. After polyp localization is performed, we offer again two possible solutions to segment the polyp in the image, which will be explained in depth in later chapters of this thesis.

We have built our methods on a model of polyp appearance which needs of a deep study of several colonoscopy studies in order to define a general model. This model needs to be tested on a large database and we offer in this thesis an introduction to the first public available database which we have generated from a series of colonoscopy studies. Finally, as our aim is to provide tools that can be useful for several applications (being in this case Computer Aided Detection (CAD)), we compare

the performance of our polyp localization method with the response that physicians provide, which is captured by means of an eye tracking device.

We will present in the next section our polyp localization and polyp segmentation processing schemes.

1.6 Introduction of polyp localization and segmentation methods

Our polyp localization and segmentation methods will vertebrate the structure of this thesis.

Polyp localization aims at providing, given an input image, an energy map which indicates which part of the image is more likely to contain the polyp inside. In order to achieve this, a first stage of image preprocessing is needed. In this first stage the effect of several elements of the endoluminal scene is addressed, such as specular highlights or blood vessels, in order to mitigate their contribution in terms of valley information. As a result of this preprocessing and the application of the valley detector we obtain the novel the Depth of Valleys (DoV) image. We will introduce this novel DoV in chapter 5, along with introducing all the preprocessing algorithms before mentioned.



Figure 1.7: Polyp localization processing scheme.

In order to obtain the final energy map that indicates the 'likelihood' of having a polyp in a given are of the image, we integrate the values of the DoV image to elaborate the DOVA energy map which will be superimposed to the original input image to highlight interesting areas. We can see the whole polyp localization processing scheme in Figure 1.7 and we can observe an example of the output of our polyp localization method in Figure 1.8.

Polyp segmentation, on the other hand, aims at providing an output image where the pixels inside the polyp are highlighted. The difference between polyp localization and polyp segmentation is that the first method does not label which pixels on the image are inside the polyp region and which are not, whereas the second clearly divides the image in two regions: a polyp region and the rest. Polyp segmentation processing scheme, as can be seen in Figure 1.9, needs of a first iteration of the polyp localization scheme in order to identify seed points for the posterior polyp segmentation algorithm, which is built on how the energy map is calculated.

In order to understand better what is the objective of our polyp segmentation system we show a graphical example on Figure 1.10. As it can be seen, given an input image our polyp segmentation method gives as output an image where the

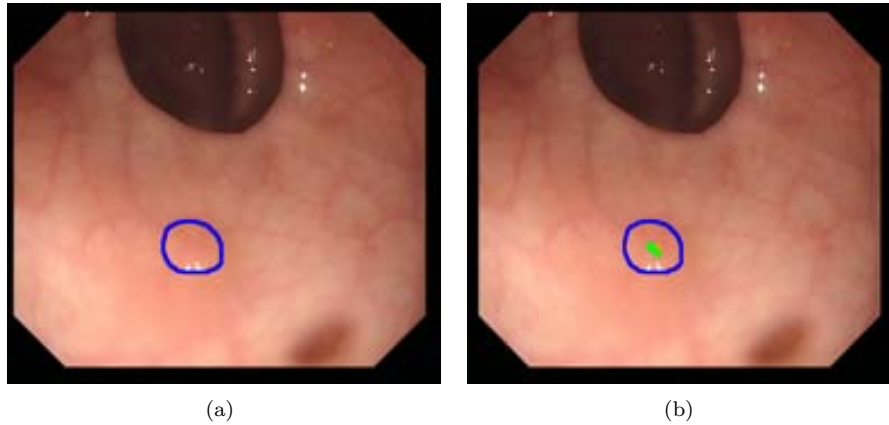


Figure 1.8: Examples of: (a) Input image; (b) Output image where the polyp is located by means of a green area.



Figure 1.9: Polyp segmentation processing scheme.

polyp region is highlighted.

1.7 Contributions of this thesis

The main contributions of this thesis are:

1. Definition of a model of polyp appearance based on how polyps appear in colonoscopy frames.
2. Definition of the novel depth of valleys image.
3. Improvement of the previous state-of-the-art on specular highlights detection.
4. Definition of an inpainting method which is used for the black mask surrounding the endoluminal scene and for specular highlights.
5. Preliminary study of the effect of blood vessels on the creation of the depth of valleys image and their effect on polyp localization and polyp segmentation results.
6. Development of a polyp localization method by means of DOVA energy maps.
7. Development of a polyp segmentation method.

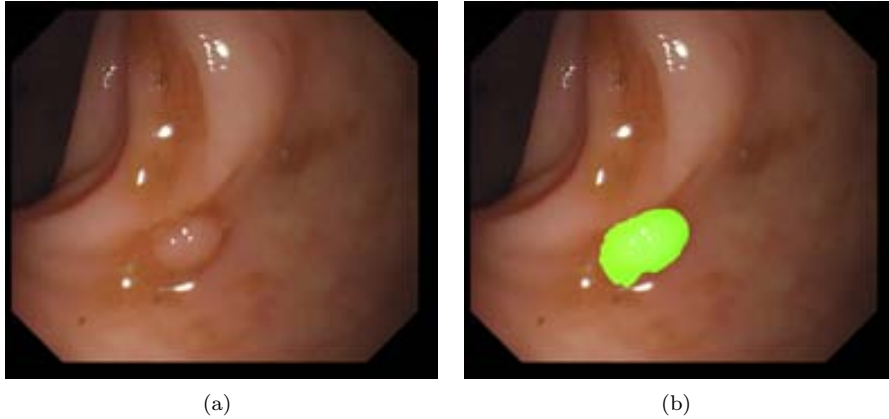


Figure 1.10: Examples of: (a) Input image; (b) Output image with polyp region highlighted in green.

8. Creation of COLON-DB: A database for assessment of polyp localization and polyp segmentation.
9. Study of the effect of specular highlights and image formation and acquisition on polyp localization and polyp segmentation results.
10. Development of a framework and study to compare polyp localization results with physician's observations.

1.8 Structure of the thesis

This thesis is structured as follows:

1. Chapter 1 includes the introduction to colon cancer and colonoscopy by following with the presentation of the objectives of the thesis.
2. Chapter 2 contains a review of the existing literature on endoluminal scene objects identification along with presenting the domains of application in which the methods that we propose in this thesis could be included.
3. Chapter 3 is dedicated to the introduction and explanation of the model of appearance for polyps.
4. Chapter 4 is centred on presenting how valley detection can be used to guide the development of future polyp localization and polyp segmentation methods.
5. Chapter 5 comprises the explanation of the polyp localization methods.
6. Chapter 6 introduces the two proposed solutions to polyp segmentation.

7. Chapter 7 presents the experimental setup and the result of the experiments that have been carried out in the scope of the research.
8. Chapter 8 sketches how to make the transition from the research lab to real life applications by introducing a framework and a study to compare the output of our polyp localization methods with physicians' observations.
9. Chapter 9 closes this thesis by exposing the main conclusions that can be extracted along with sketching the future lines of research that could be followed.

Chapter 2

Endoluminal scene description: challenges and methods

The objective of this thesis is to develop polyp localization and polyp segmentation methods, which will be based on a model of polyp appearance. Polyps are not the only part of the endoluminal scene and, as will be shown later, there are other elements that can play a key role in the performance of our methods. Before starting with the explanation of our method, we will review the existing works on describing some of the elements of the endoluminal scene such as polyps, lumen or specular highlights. We will also extend in this chapter the concept of intelligent systems for colonoscopy as our method could potentially be incorporated for some applications such as Computer Aided Detection (CAD).

2.1 Introduction to the Endoluminal Scene. The elements of the endoluminal scene

As depicted in Figure 2.1 the endoluminal scene consists of several elements that we will define next:

1. Lumen: In biological contexts lumen is defined as the interior space to a tubular structure, such as the intestine. The lumen is the cavity where digested food passes through and from nutrients are absorbed.
2. Folds: The haustral folds represent folds of mucosa within the colon [52]. They are formed by circumferential contraction of the inner muscular layer of the colon.
3. Blood vessels: They are the part of the circulatory system that transports blood through the body. They can be identified by their tree-like shape with ramifications.

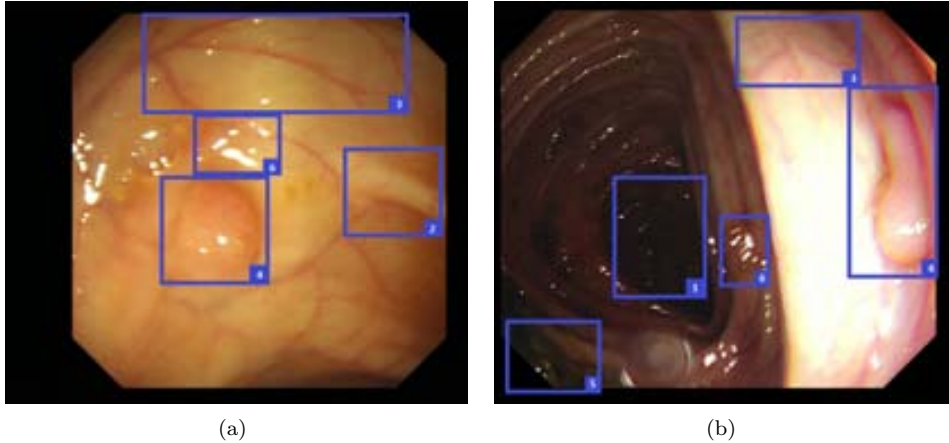


Figure 2.1: Elements of the endoluminal scene: (1) Lumen; (2) Wrinkles and folds; (3) Blood vessels; (4) A polyp; (5) Fecal content; (6) Specular highlights

4. Polyp: A polyp is an abnormal growth of tissue projecting from a mucous membrane. If it is attached to the surface by a stalk, they are referred as pedunculated polyps. If there is not stalk, they are known as sessile polyps. They are found in the colon, stomach, nose among other locations. In the case of colon, untreated colorectal polyps can develop into colorectal cancer. We will extend the definition of polyp in the next chapter.
5. Fecal content: They are a waste product from the digestive tract which is expelled through the anus during the process of defecation.
6. Specular highlights: They are the bright spots of light that appear as a result of the illumination on shiny objects.

Other elements of the endoluminal scene can be diverticulosis are shown as cavities or holes in the intestinal wall. The lesions related with bleeding are generally identified by its characteristic color. Polyps present a large variety in shapes, and seldom show a discriminative change in texture and/or color in comparison to the surrounding area. Due to the flexible and extendible nature of the colon, and in part owed to the impact of the probe insertion or withdrawal in its deformation, it is difficult to find a perfect tubular appearance in the colon lumen because intestinal walls can be bent and folded. In addition, the wrinkles and folds associated to the colon physiological structure appear in the scene as radial protuberances which modify the flat surface of the intestinal walls. On the intestinal walls, blood vessels are observed with their characteristic tree ramifications, presenting a certain variability associated to their width.

Despite a preparation is required for most of the colonoscopy interventions - with the aim of eliminating all fecal matter so that the physician conducting the colonoscopy can have a clear view- in many cases intestinal content is still present

after the preparation procedure, and this intestinal content will hinder the right visualization of the intestinal walls. The procedure of elimination of the remaining fecal matter, consisting of the direct injection of water through the colonoscope in order to dilute the intestinal contents, turns out into the blurring of the video sequence and the appearance of bubbles. Finally, during the time of intervention, some tools used by the physician for different tasks -i.e., biopsy, cauterization, etc-. can be part of the visual scene too.

Although in this thesis we will focus on the analysis of colonoscopy video frames, we should not forget that there exist diverse endoscopy techniques, such as capsule endoscopy (both for small bowel and colon), bronchoscopy, gastroendoscopy, etc. that also show different endoluminal scenes, each of them with particular features. Besides that, there is a wide variety of imaging methods used to enhance particular physiological targets, which is the case for *narrow band imaging* or *chromoendoscopy*, just to mention a few. This situation sets up a heterogeneous scenario from the perspective of automatic analysis using computer vision, and makes it not feasible to tackle the endoscopic image problem as a whole. However, it is possible to take some of the methods used in a given technique and adapt them to the specific particularities of colonoscopy video. For example, the automatic detection of intestinal content is a topic addressed in the bibliography of capsule endoscopy [108] by means of the analysis of color distribution and texture, and its equivalent to the detection of intestinal content in colonoscopy would require relatively minor modifications.

In addition to the difficulties associated to the characterization of the colonoscopy scene due to its high variability and complexity, there are many visual artifacts the impact of which should be taken into account in order to tackle a robust system for the automatic analysis of colonoscopy video, such as color channel misalignment or specular highlights, which will be covered in the next section.

2.2 Introduction to the review of methods devoted to endoluminal scene object identification

We can group the existing literature related to computer vision applied to colonoscopy in two main groups, namely: 1) Image preprocessing and non-informative frame definition and 2) Endoluminal scene objects description. In this section we will cover both groups, paying more attention to the second, considering that the aim of the work presented in this thesis is to provide a polyp localization and segmentation method. Considering this, the majority of the existing literature devoted to endoluminal scene description in colonoscopy video could be grouped according into two different topics, lumen segmentation and polyp detection. We will also include in this review the literature that studies the impact of specular highlights and blood vessels, since both topics will be tackled later in Chapter 4. All the methods exposed in this review correspond to optical colonoscopy unless otherwise stated. We have included some methods used in CT or *virtual colonoscopy*, based on the potential application of these techniques for the optical modality.

2.3 Image Formation and Enhancement

Two different topics are studied here: 1) Image preprocessing and enhancement methods, which aim at improving the input to a posterior processing scheme, and 2) Non-informative frame definition, which aim at defining which frames on the whole video do not contain useful information.

2.3.1 Image Preprocessing and Enhancement

Image preprocessing methods are needed in order to eliminate or minimize the impact of image artifacts associated to colonoscopy video, which fundamentally consist of the color phantoms, interlacing and the presence of specular highlights.

The problem of color phantoms associated to the temporal misalignment of the color channels has been addressed in literature [31]. Color phantoms occur because most colonoscopy devices use monochrome CCD cameras in which the RGB components are taken at different times. This causes a worsening in the quality of the images, as can be seen in Figure 4.6, which may difficult posterior image analysis tasks. The method presented in [31] involves both color channels equalization and the estimation and compensation of the camera motion. The experimental results show a global improvement in the quality of the images, failing only in cases when the quality of the original image is very low, although the evaluation is done qualitatively.

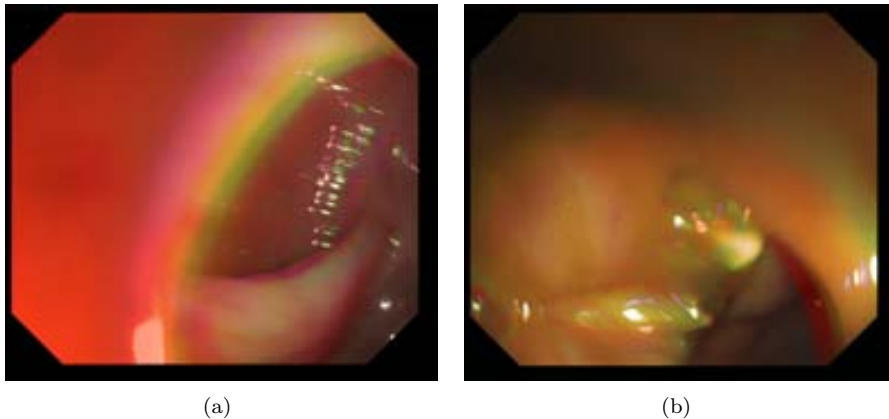


Figure 2.2: Examples of color channel misalignment.

One particular problem that also has to be addressed is interlacing. This technique doubles the perceived frame rate without consuming extra bandwidth. By the use of progressive scan, an image is captured, transmitted and displayed following a similar path that we do when reading a page of a book: line by line and from top to bottom. The interlaced scan pattern does the same process but only for every second line, that is, one line out of two, from the top left to the bottom right of a display. Once this is finished, the process is repeated but in this new iteration the scanned lines are those

which were not scanned in the first one. The process of scanning every second line is known as interlacing, which takes advantage of the fact that human visual system is less sensitive to flickering details than to large-area flicker [33]. Deinterlacing doubles the vertical-temporal sampling density and aims at removing the first repeat spectrum caused by the interlaced sampling of the video. Although it may seem to be an easy-to-solve problem in fact it is not, as usual TV signals do not fulfill the demands of the sampling theorem.

In our case, we are concerned of some effects that interlacing may have in our videos, such as the apparition of false contours/edges caused by the minimal time difference between two consecutive frames, and for this reason we will address this problem in Chapter 4.

2.3.2 Specular highlights detection and inpainting

The specular highlights on the polyp can affect texture features obtained from the polyp surface and may therefore interfere in the output of several methods such as polyp detection. A negative effect of specular highlights was also reported by [78], in the context of the detection of indistinct frames in colonoscopic videos. There are several approaches to segment specular highlights in images, which are usually based either on detecting grey scale intensity jumps or sudden color changes. For instance, the work presented in [19] present a detection of specularities based on a characterisation of Lambertian surfaces. In computer graphics, Lambertian reflection is often used as a model for diffused reflection. Lambertian reflection from polished surfaces are typically accompanied by specular highlights (gloss) where the surface luminance is highest when the observer is situated at the reflection direction. In our case, this happens when the light of the probe impact into a prominent surface, as can be seen in Figure 2.3.

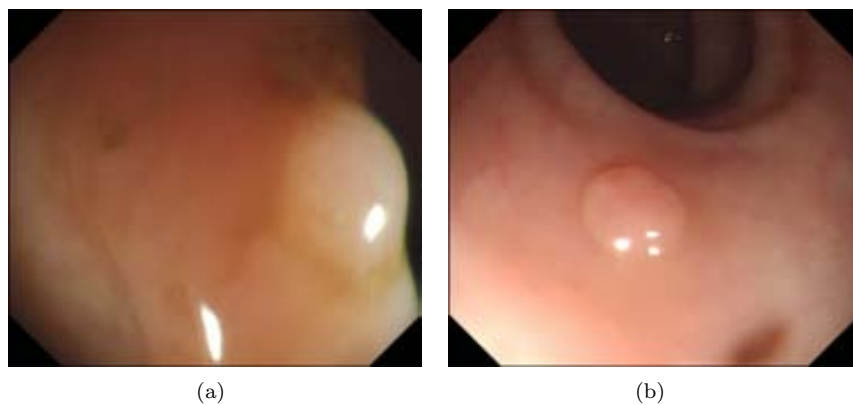


Figure 2.3: Example of apparition of specular highlights in colonoscopy video frames: (a) Specular highlights inside and outside the polyp; (b) Specular highlights inside the polyp.

There are several works devoted to specular highlights detection and inpainting. For instance the work presented in [40] uses sudden color changes as a cue to detect specular highlights. A similar work can be found in [57], which faces the problem that specular highlights cause in image segmentation. The work of [74] addresses the problem of re-rendering images to high dynamic range (HDR) displays, where specular highlights present a problem because of the operations made when capturing and rendering to standard displays. A more recent approach is the one presented by [101], where specular highlights are defined as a combination of diffuse and specular reflection, which need of an explicit color segmentation.

Following a similar line of research based on obtaining the diffuse color component, the work of [119] estimates the maximum diffuse chromaticity values of the specular pixels by directly applying low-pass filter to the maximum fraction of the color components of the original image. The method presented in [120] is built on the fact that specular-free two-band image are introduced to deal with specular reflections. For given input images, specular-free two-band images are generated by using simple pixel-wise computations in real-time. More recently, the work presented in [92] proposes a method that is based on a wavelet-based histogram denoising followed by adaptative thresholding to isolate the specular spike. Once this spike is found, an intensity descent is performed to select the specular lobe. This approach also exploits chromatic information but uses histograms to speed up the process. The work of [50] proposes a method for estimating the scene illuminant spectral power distributions of multiple light sources under a complex illumination environment. The authors note that specular highlights on inhomogeneous dielectric object surfaces include much information about scene illumination according to the dichromatic reflection model. Although our domain of application is very different, we can observe the work that the authors refer on specular highlights detection. In the cited paper, the authors describe three different methods, namely: 1) use of variable thresholding of luminance; 2) use of luminance and hue components and 3) use of a polarization filter.

But, as it also happens in our case, we are not only interested in detecting the specular highlights in the images but also in providing as input to our algorithm an specular-free image. This can be achieved by means of an inpainting method. Inpainting consists of filling the missing areas or modifying the damaged ones in a non-detectable way for an observer not familiar with the original images. For instance, the authors of [92] base their inpainting method in a previous work [81], which used a very simple isotropic diffusion model extended with the notion of user-provided diffusion barriers. This method provides results comparable to other methods existing but is two to three orders of magnitude faster. In order to work properly, the method needs of a mask that should include exactly the region to be retouched, hence the importance of a correct specular highlights detection method. If the mask is smaller its boundary will contain spurious information which will be incorporated into the restored area and if it is bigger some important information can be discarded.

An interesting approach is the algorithm presented in [3], which addresses both specular highlights detection and inpainting:

- Detection: The proposed method consists of two different modules. The first

module uses color balance adaptive thresholds to determine the parts of specular highlights that show too high intensity to be part of nonspecular image content (saturated parts of the image), taking also into account that specular highlights can appear only in one of the three RGB channels. The second module compares every given pixel to a smoothed nonspecular surface color at the pixel position, which is estimated from local image statistics. This module aims at detecting the less intense parts of the specular highlights in the image in a way such by looking at a given pixel, the underlying nonspecular surface color could be estimated as the representative color of an area that surrounds the pixels, only in cases when it was known that this particular area does not contain specular highlights or at least which pixels in the area lie on specular highlights. The authors also consider that, in some cases, the mucosal surface in the close neighborhood of the camera can appear saturated without showing really specular reflection but it is picked up by the detection algorithm. In order to avoid this problem, the authors make use of the property that the image area surrounding the contour of specular highlights generally shows strong image gradients.

- **Inpainting:** Inpainting is done in two levels. First the specular region are filled by replacing all detected pixels by the mean color of the pixels within a certain distance range of the outline. This image is then filtered by using a Gaussian kernel, which results in a strongly smoothed image free of specular highlights. For the second level the binary mask marking the specular regions is converted to a smooth weighting mask, in a way such the smoothing is performed by adding a nonlinear decay to the contours of the specular regions. The weights depend on the distance from each pixel to the contour of the specular highlight region. That is, the weight will be higher if the point is far from the contour.

The main conclusion that can be extracted from the review of the bibliographic sources is that specular highlights detection and inpainting is a problem that appears for a wide range of images and has plenty of different solutions, depending on the domain or the scope of the application. Unfortunately some of the most well-known approaches (such as [101] or [119]) can not be used in their entirety as the inpainting of the saturated pixels (which are part of the specular highlights) are based on having information in the channels that are not saturated, which is not our case. We will present our specular highlights detection and inpainting methods in Chapter 4.

2.3.3 Non-informative frame definition

As can be suspected from the description above, the analysis of the endoluminal scene is a complicated task. There are several elements that can appear and sometimes bad visualization or quality of the image can affect the performance of the algorithms. In order to separate the analysis of frames where the content can be clearly seen and those where the analysis is difficult, non-informative frames are defined. In this domain of application, non-informative frames can be defined as those that, either their quality is so damaged (by the artifacts, hindering intestinal content, etc.) that

it is difficult to extract information from them, or they are clinically uninteresting for a given task. For instance, frames where the instrumental take up great part of the image may not be relevant for polyp detection tasks. An accurate detection of the non-informative frames could also lead to a great reduction in the processing time of a stored colonoscopy intervention. Fundamentally, this information may be used for automatic video annotation and efficient video indexing and retrieval. Examples of non-informative and informative frames are shown in Figure 2.4.

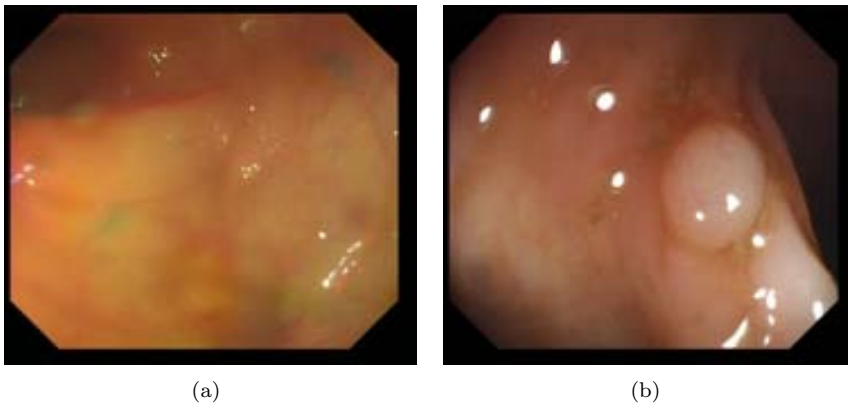


Figure 2.4: Non-informative frame definition: (a) Non-informative frame; (b) Informative frame.

There are a few works that are centered on the identification of non-informative frames. The work of [4] addresses the identification of clinically uninteresting frames by analyzing the energy of the detail coefficients of the wavelet decomposition of a given image, which is used as the input to the classification system. In this case non-informative frames are those which do not carry any useful clinical information, such as those that occur when the camera is covered with liquids or when it is very close (even touching) the mucosa. These cases do occur frequently in colonoscopy procedures leading to extremely blurry images. This method is based on the 2D discrete wavelet transform which results in a set of approximation and detail coefficients. The approximation coefficients represent the low frequency content of the image while the detail coefficients hold the complementary high frequency information. The authors use detail coefficients to distinguish between informative and non-informative frames holding on the fact that the norm of the detail coefficients will be lower for low contrast images, making them more likely to be classified as non-informative.

The work of [22] presents a method that extract those frames which correspond to a diagnostic or therapeutic operation, following work done in other domains (i.e., detecting important semantic units such as scenes and shots). This work takes profit of several characteristics that colonoscopy videos present, such as the presence of many blurred frames due to the frequent shifts of the camera position while it is moving along the colon. The identification of the operation shots is based on the detection of diagnostic or therapeutic instruments. In this case the authors map the problem of

detecting instruments to the problem of detecting the cables of these instruments as they are present in the operation, regardless of their type. The architecture scheme shown in Figure 2.5 consists of five different steps which involve: 1) image preprocessing, to remove the effects of the specular highlights; 2) identification of the insertion direction of an instrument; 3) region filtering, where regions that are not part of the cable are removed; 4) region merging, which combines regions where parts of the instrument appears and 5) region matching, which matches the candidate regions in the image with the cable and without the cable.

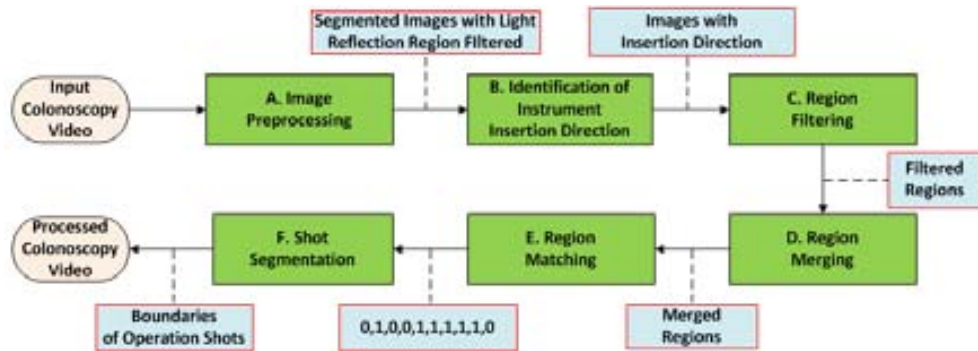


Figure 2.5: A system architecture for operation shot detection as described in [22].

Apart from the two methods presented related to the identification of non-informative frames, other approaches have been carried out such as the work of [80] where a measure called the *isolated pixel ratio* (IPR) is used to classify the frames into informative, ambiguous and non-informative. The IPR measure is calculated from the edges of the image: an edge pixel that is not connected to any other edge pixel is defined as an isolated pixel. Those isolated pixels are counted for each frame and are put in relation with the total number of edge pixels to obtain the IPR ratio.

Finally, an example of an endoscopic full multimedia information system for video annotation implementing many of these approaches is described in the work of [65].

2.4 Endoluminal scene object description

The existing literature concerned on describing the elements that constitute the endoluminal scene will be covered here. We will pay special attention here to methods devoted to the description of three of the elements, namely lumen, polyps and blood vessels.

2.4.1 Lumen Detection

The detection of the lumen and its position can be crucial, for example, in post-intervention video processing. Frames where the proportion of lumen out of all the

image is large can be related to the progression of the colonoscope through the gut (Figure 2.6 a-b)). On the other hand, frames where the amount of lumen presence is low (Figure 2.6 c)) may potentially indicate areas of the image where the physician has paid more attention. In addition to that, an efficient lumen segmentation may lead to remove great part of the image for a further computational analysis.

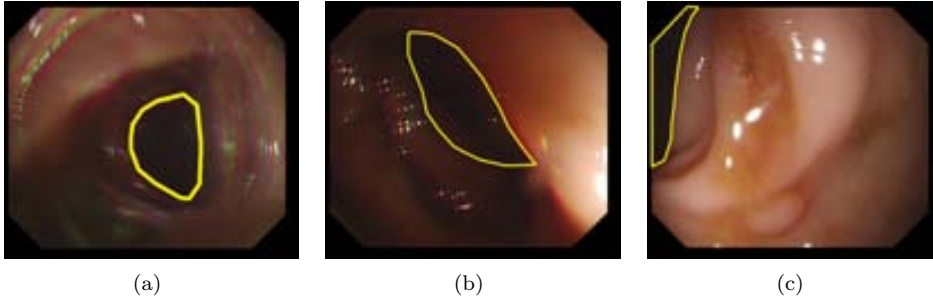


Figure 2.6: Examples of lumen (surrounded by a yellow boundary): a) and b) full view and c) partial view.

Several works are centered on lumen detection, such as the work of [42], which aims at decomposing the tissue image (where lumen may be present) into a set of primitive objects and segment glands making use of the organizational properties of these objects. In this approach, an image is first decomposed into its tissue components that, as they are difficult to locate, are approximately represented transforming the image into a set of circular objects (nucleus and lumen objects). The results presented in the paper show that the use of object-based information, instead of using pixel-based information alone, leads to more robust segmentations with respect to the presence of imaging artifacts. This is attributed to pixel intensities being more sensitive to the noise that arises from the staining, fixation, and sectioning related problems.

Following a similar line of research, the work of [103] presents an automatic segmentation algorithm for lumen region and boundary extraction from endoscopy images. The proposed algorithm consist of two different steps: first a preliminar region of interest that represents the lumen is segmented via adaptative progressive thresholding. Later on an adaptative Iris filter is applied to the previously segmented region of interest in order to determine the actual region. Experimental results show that by applying this method unclear boundaries can be enhanced and detected, resulting in a more accurate lumen region. The authors also present a method to improve boundary extraction, based on an heuristic search on the neighborhood pixels. Following this trend, the work of [15] addresses lumen segmentation by first estimating the center-line, which can be achieved by first removing the background and then extracting air regions with a threshold filter.

Some other works are devoted to automatic segmentation of the colon, which includes the extraction of the lumen region although they belong to CT or *virtual colonoscopy* fields. This task is not easy, as the high-contrast gas/tissue interface

that exists in the colon lumen makes segmentation relatively easy but two factors prevent from making this process automatic. First of all, colon is not the only gas-filled organ in the data volume (lungs and stomach, among others, also meet this criterion). User-defined seeds placed in the colon lumen have previously been required to isolate spatially the colon. Secondly, portions of the lumen may be obstructed by peristalsis, large masses and fecal content. These factors require user interaction in order to isolate colonic segments. The work presented in [117] uses an automatic method that locates automatically the seed points and segment the gas-filled lumen sections without user supervision, aiming at an automatic colon segmentation. The method presented also includes an improvement to lumen segmentation by digitally removing contrast-enhanced fluid. Still in the field of CT colonography and also exploring the use of seeds to guide lumen segmentation, we can see in the work of [89] how lumen segmentation is done automatically by placing seeds automatically inside the colon lumen. The seeds expand until they reach the bowel wall. A Computer Aided Diagnosis (CAD) system is built having this lumen segmentation method as a feature.

2.4.2 Polyp Detection

The main objective of the colonoscopy procedures is to check the status of the colon and to find possible lesions and cancer polyps on it. Therefore a polyp model is needed in order to guide polyp detection. As it will be shown in the next chapter there are several types of polyps, which differ on their general shape and relative position with respect to the colon wall. We will develop more about this later but in this section we will make a review of several polyp detection methods by means of computer vision. The direct application of the methods presented in this section is the potential assistance in the diagnosis, both during and in post-intervention time. In this case we will focus on methods based on applying a certain feature descriptor because the majority of available works can be enclosed into this kind of classification.

In the context of image processing, features can be defined as singular visual traits, associated to the visual primitives that constitute an object, such as edges, corners or lines, among others. The usual procedure is to use *feature detection* methods to locate the potential *regions of interest* (ROIs) of the image and then describe them using one or many *feature descriptors*. After doing an extensive research on the different types of feature descriptors [14], we have divided them into four groups: shape, texture, color and motion. As can be seen from Figure 2.7, there are many *feature description* algorithms and, of course, mostly all of them are being used currently. Again, the problem that we want to solve will tell us which *feature descriptors* we want to test. Maybe we only need to find out the shape of an object to get to a decision but in more complicated tasks (i.e. classification of images) some other characteristics, like texture or colour can be discriminative as well.

We have divided the existing methods in four general classes: *Shape Descriptors* (subdivided in Contour-based and Region-based, depending on which part of the object we use to get to a description), *Colour Descriptors*, *Texture Descriptors* and *Motion Descriptors*. For the concrete case of polyp detection in colonoscopy videos we

will only explore approaches belonging to the Shape and Texture Descriptors groups, although there are some works that combine texture and color information as will be mentioned later.

Shape-based approaches

This group of approaches observe the structure of polyps as they appear in images and find the shapes which polyps commonly have. More precisely, some of the methods belonging to this group aim at approximating the boundaries detected to general polyp shapes. As it can be thought, this leads to a big challenge, considering that in many times we do not have a perfect shot of the polyp but an image where its pose, size and appearance can vary largely. Thus, many of the approaches presented try to detect polyps not by detecting its whole shape but by detecting parts of the image that may indicate polyp presence.

For instance, flat polyps are meant to have elliptical shapes [114] so one way to detect polyps is trying to find which structures in the image are surrounded by boundaries that constitute ellipses. The difficulty in this case is that in many occasions we do not have complete boundaries or the concrete view of the polyp that we have makes it difficult to fit elliptical shapes (such is the case of lateral views). The works presented in this subsection could also be classified into two categories, namely: a) detection by curvature analysis ; b) detection by ellipse fitting and c) combination of curvature analysis and ellipse fitting.

a) Detection by curvature analysis

In this case curvature is defined as the amount by which a given object differs from being flat, for the case of 3D objects, or straight, in the case of a line. In the context of polyp detection curvature can be used in several ways, as it will be shown next. For instance, we can check the curvature profile of the boundaries that appear in the image, which may have been detected by means of an edge detector. An example of the former can be consulted in the work of [59]. The proposed method is built in the following way: image contours correspond to haustra creases are extracted and curvature of each contour is then performed after a non-parametric smoothing. Zero-crossings of curvature along the previously extracted contour are detected. In this case the method defines an abnormality when there is a contour segment between two zero-crossing that have opposite curvature sign (that is, one concave and another convex) to their neighboring segments. Experimental results show that the proposed method can be useful when detecting lesions present along the creases of the haustra. The method also provides a solution for specular highlights, which are discarded through curvature analysis.

The work of [122] elaborates on the use of curvature-based shape measures (such as the shape index, curvedness or mean curvature) to analyze the local shapes in the colon wall. The proposed method estimates curvature by means of kernel methods. The problem in this case may appear in spurious calculations indicating high curvature, which is observed when the kernel contains two surfaces. This problem is

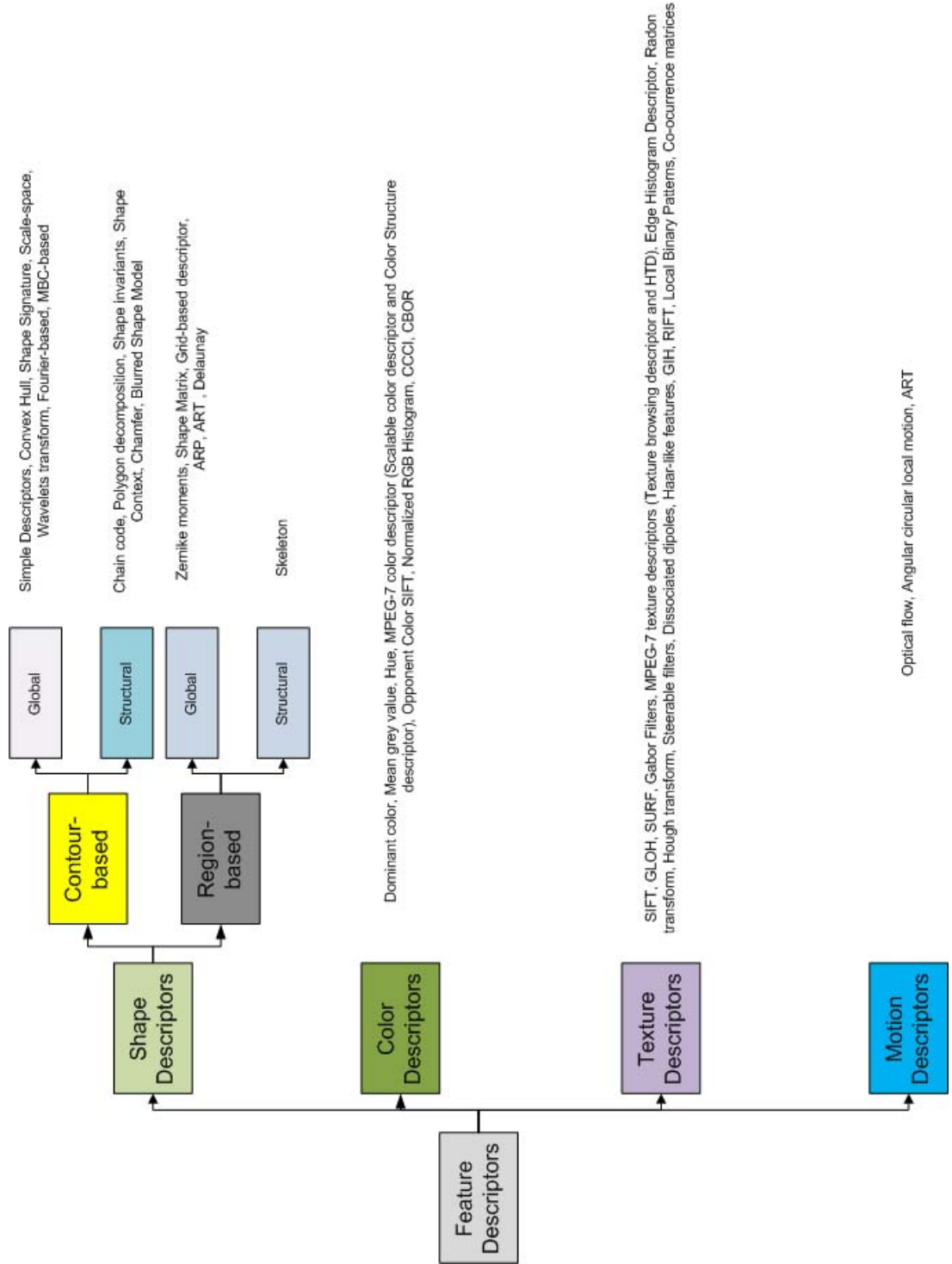


Figure 2.7: Taxonomy of Feature Descriptors, highlighting those used in polyp detection

solved via the Knutsson mapping method, which consists of solving discontinuities by observing the change in the surface normal. Another problem that can appear is the discontinuities in curvature, which appears when the gradient magnitude necessary to calculate the curvature vanishes.

One of the problems that optical colonoscopy image analysis present is that it is very difficult to estimate how protruding an object is without having no information about the distance between the light source and the colon walls. By using Virtual or CT colonoscopy we can have a 3D map of the whole colon, which can be useful for detecting protruding structures such as polyps. One example of this can be found in the work of [107]. The authors present a method that enables automated detection and segmentation of colorectal polyps, proposing a method that measures the amount of protrudedness of a candidate object by means of scale adaptive methods.

Finally, during all the research that has lead to this thesis, we explored the use of contour information in a way that seems similar to some of the alternatives presented here. More precisely we designed our Valley Orientation DOVA which uses the orientation of the intensity valleys that constitute the boundaries of objects. This method will be fully explained in Chapter 5.

b) Detection by ellipse-fitting

The methods presented above explore the changes in curvature to detect them but they are not concerned about the concrete shapes of the objects they are detecting as polyps. We will present here some methods that, starting by contour information, aim at detecting polyps by approximating their boundaries to ellipses.

Belonging to this group we can observe in the work of [53]. The presented method introduces a processing scheme which aims at providing a real-time image processing system for endoscopic applications. Although the system has three different modules, we will focus on the image processing one. This module consist of several tasks: 1) Contrast enhancement (to improve the visibility of the different structures); 2) Image segmentation (by means of finding the edges on the image, which is performed in the R, G and B channels separately); 3) Feature extraction (using features such as the area, color or object's shape) and 4) Object classification. It is important to mention that the objective is not to univocally identify polyps but to highlight the regions of the image which are more likely to contain polyp-like objects. Experimental results show good performance qualitative performance although concrete polyp detection results are not given.

One of the contributions that we will present in this paper belongs to this type of techniques [11]. Our Ellipse Fitting DOVA method defines as polyp containing those areas on the image enclosed by valleys where we can fit an ellipse.

c) Combined methods

For instance an approach that combines both curvature and ellipse fitting can be found in the work of [47]. The method presented consists of fitting ellipses into the frontiers obtained after a first segmentation, and then classifying candidate regions by considering curvature, distance to edges and intensity value. Without entering into

many details, in order to detect the ellipses an edge image is needed in which desirable edges should be grouped. Taking into account the challenges that colonoscopy images present, only some parts of the polyp boundary will have strong edge information so, based on this, the method uses the marker-controlled watershed algorithm [112] for polyp segmentation because it can handle the gap between broken edges properly. Then, using the edges in each segmented region, the method generates an ellipse by means of an ellipse fitting method. Finally the number of final ellipses is reduced by removing those which do not represent actual polyps filtering by curve direction and curvature, by edge distance and by intensity value.

The work of [34] also starts with a watershed segmentation but it performs its detection scheme by using color information. MPEG-7 descriptors are used in polyp detection tasks as it is introduced in the work of [28]. *Region-based shape descriptor* (RBS) belongs to the broad class of shape-analysis techniques based on moments. A set of separable angular radial transformation (ART) basis functions is defined that classifies shape along various angular and radial directions. The RBS descriptor obtains 35 coefficients from the ART transform.

Finally, the work of [58] is devoted to describe polyp appearance. Several parameters are evaluated, such as the response in the red channel of the image (which may indicate the presence of malignant tumors), the perimeter, the enclosed boundary area or the *form factor*, which can give indication of possible presence of abnormalities in the colon (the more irregular the shape of the lumen, the smaller the value of the *form factor*).

We present a summary of the shape-based polyp detection methods in Table 2.1, including the datasets used and the classification methodology applied.

Author	Method	Classification	Datasets
Krishnan et al. [59]	Edge detection to extract contours. Curvature	-	2 normal and 4 abnormal images
Hwang et al. [47]	Ellipse fitting. Curvature	-	27 polyp shots
van Wijk et al. [107]	Amount of protrudness. Curvature	5 measures (MaxIntChange, LongAxis, ShortAxis, MinHU, MaxHU), distances	84 studies, 168 scans, 108 polyp shots
Dhandra et al. [34]	Segmentation of color images followed by watersheds	-	50 normal and 50 abnormal images
Zhu et al. [122]	Curvature-based shape measures	By segmentation results	1 phantom image and 1 colon patient
Bernal et al. [12]	Valley orientation accumulation	By segmentation results	COLON-DB database
Bernal et al. [11]	Ellipse fitting accumulation	By segmentation results	COLON-DB database

Table 2.1: Summary of shape-based descriptors for polyp detection.

Texture-based approaches

The use of texture descriptors on polyp detection has been gaining interest during last years. There is a number of works that are based on the use of **wavelet descriptors**. In this case the wavelet transform is calculated for each frame and the attention is put on the detail and approximation coefficients. Wavelets are defined as a representation of square-integrable function by a certain orthonormal series generated by a wavelet. A wavelet is a wave-like oscillation where the amplitude goes up from zero to come back decreasing to the original value [21].

In this context we can observe the works of [54]. In these works the first operation that is done to the image is wavelet transformation, which is combined with other texture descriptors, such as co-occurrence matrices or local binary patterns [9]. The presented method is based on a color feature extraction scheme used to represent the different regions that appear on the frame. This method is built on wavelet decomposition by means of color wavelet covariance and it is supported by a linear discriminant analysis proposed to characterize image regions along video frames. The objective of this work is to detect abnormal regions in order to indicate physicians area where they may pay more attention. The method is tested on video sequence that contain relatively small polyps and experimental results are compared with histological results. The same group of researchers developed a tool to detect colorectal lesions in endoscopic frame, which was named CoLD (colorectal lesions detector, [72]). This tool provides a graphical user interface so both novice and experts user can take advantage of its use. In this case wavelets information is used to discriminate amongst regions of normal and abnormal tissue.

The work of [63] takes into account when detecting abnormalities in colonoscopic images, the location, shape and size of the abnormal regions in the image. The concrete locations are unknown and vary across images therefore it is difficult to determine the appropriate patch-size to use for searching. In this case the solution is to use multi-size patches and ensemble them in order to achieve good performance. Experimental results show that combining classifiers that use multiple-size patches leads to a better abnormal region recognition. The classification is done by means of a binary Support Vector Machine (SVM). The features extracted from these patches are taken from both approximating and detail coefficients from wavelet decomposition of the image patches in the three channels of the CIE-Lab color space.

There are some other texture descriptors that have been used to develop polyp detection method, such as the already mentioned local binary patterns or co-occurrence matrices. The work of [1] combine both of them, with the novel use of local binary patterns in opponent color space. As the authors state, texture can be seen as a local property and therefore, each image is divided into small image patches and four different methods were implemented, which combine co-occurrence matrices (using different statistical measures such as energy, homogeneity or entropy) and local binary patterns.

As in the case of shape-based approaches, MPEG-7 also offers texture descriptors that can be used to build polyp detection methods. In the work of [28], although

applied to a different type of endoscopic process, several texture and color descriptors are presented. In the sub-field of color descriptors, methods such as *dominant color*, *scalable color* or *color structure* are presented (see [14] for a further explanation of them). Related to texture descriptors, *homogeneous texture* and *local edge histogram* are introduced. These methods are evaluated in a big database and, in order to quantify the performance of each descriptor, several measures were used such as descriptor's redundancy or the variation of the descriptors' value. The experimental results show the superiority of *scalable color* over other color descriptors due to its higher resolution. On the other hand we have the apparently strong *local edge histogram* that performs worse than other simpler approaches, such as *homogeneous texture*, since it pays too much attention to the small texture variations in the image.

All the methods presented above are based on the use of a certain texture descriptor (wavelets, local binary patterns) to build their polyp detection methods but there is a separate group of techniques that do not follow this path. For instance, the work of Tjoa et al. ([105, 104]) introduces the concept of texture unit (TU) and texture unit number (NTU). Texture units characterize the local texture information for a given pixel and its neighborhood, and the statistics of all the texture units over the whole image reveal the global texture aspects. Without entering into details, each pixel value is compared with the value of the pixels in its neighborhood and then the value for this pixel in the TU matrix is assigned according to the comparison. The texture information is presented in the texture spectrum histogram, which is obtained as the frequency distribution of all the texture units. Six statistical measures are used to extract new features from each texture spectrum, which include energy, mean, standard deviation, skew, kurtosis and entropy.

In Table 2.2 we can see a summary of the main characteristics of recent approaches for texture-based polyp detection,. As it happens with shape-based methods, it is very difficult to make a comparison between them, due to the fact that there is no common local database where to test the different methods. Therefore, we could only compare different methods in terms of certain measures, such as precision or recall, always taking into account that these measures will have been calculated for a specific dataset.

Author	Method	Classification	Datasets
Coimbra et al. [28]	MPEG-7 descriptors: (Shape) Region-based shape descriptor (texture) homogeneous texture, local edge histogram (color) dominant color, scalable color, color structure.	Mean of descriptor values for each event group	Blood (832 images from 17 events), ulcers (400 images from 21 events), polyps (899 images from 31 events), and normal (55000 images from an entire uneventful exam) images
Karkanis et al. [54]	Wavelets applied in different color spaces (CWC features)	LDA	5 different videos
Li et al. [63]	Division of image in patches. Mean and standard deviation of DWT coefficients	Support Vector Classifiers	46 colonoscopic images with multiple categories of abnormal regions and 12 normal
Tjoa et al. [104]	Measures extracted from texture spectra in chromatic and achromatic domain. Texture units	Helped by PCA, Nearest Neighbor	12 normal and 54 abnormal images
Ameling et al. [1]	Local binary patterns (also in opponent color space) and grey-level co-occurrence matrices.	SVM	4 videos

Table 2.2: Summary of texture descriptor-based methods for polyp detection.

2.4.3 Blood vessels characterization

As will be shown in Chapter 4, blood vessels do have an impact in the performance of our method and in this thesis we will provide a preliminary study on how to mitigate their effect. Many different methods have been used to provide a segmentation of blood vessels in two-dimensional images. Most of them have been tested in retinal or angiography images. Despite the wide variability of enhancement steps and segmentation methods they are usually separated in two big groups: pixel-based methods and tracking-based methods [73]. In the former category different approaches are included such as kernel-based methods, model-based techniques, classifier-based methods or morphology-based strategies. Kernel-based methods are based on the convolution operator with some kind of kernel designed according to a model. Different kernels and filters have been proposed. The aim of the convolution is usually to extract vessel borders or centerlines. A matched filter approach based on Gaussian kernels is used in some methods to model the cross-section of a blood vessel [23] [45]. These methods use Gaussian-shaped templates in different orientations and scales to identify vessel profiles. An example of model-based technique was presented in [51], where the authors proposed a knowledge-guided adaptive thresholding framework where binarization is used to generate object hypotheses. Those hypotheses are only accepted if they pass a verification procedure.

Classifier-based methods intend to assign each pixel in the image to the vessel or non-vessel class. In this group we find the primitive-based method in [100]. In this method a ridge detection is performed as a first step to achieve a segmentation of the image. Afterwards, that information is considered to classify regions and pixels. In [99] a bayesian classifier is used after computing feature vectors obtained by Wavelet Gabor responses. A neural network is used in [71] after computing a feature

vector based on moment invariants-based features. Morphology-based techniques use morphological operators to take advantage of shape characteristics of blood vessels. Morphological operators are usually combined with other techniques. The extraction of vessel centerlines combined with local information such as the vessel length is followed by an iterative vessel filling phase based on morphological filters in [73].

In [121] mathematical morphology is combined with curvature evaluation to differentiate vessels from other structures. Tracking-based methods aim to obtain the vasculature structure using local information to follow vessel centerlines. Tracking techniques trace vessels from selected starting points. At each point a neighborhood is evaluated to decide whether they are vessel candidate pixels regarding some kind of local information. The process finishes when the pixels evaluated are considered to be end points. Other approaches that can be included in this category are based on deformable or snake models. This techniques place an active contour model near the aimed contour and evolve it iteratively to fit the desired object [38]. Many methods using techniques in different categories can be found. For instance, some approaches combine a classification based on support vector machine followed by a tracking stage based on the Hessian matrix [118].

2.5 Domains of application of endoluminal scene description methods

The application of the methods that have been described in the previous sections could be enclosed in the field of intelligent systems for colonoscopy. Intelligent systems for colonoscopy are defined as systems which add key information to the colonoscopy procedure, whether it is by assisting the physician during the intervention or helping to provide metrics to assess the performance of the procedure.

It is clear that the most manifest use of this kind of systems could be the **assistance in the diagnosis procedure** during the intervention or in post-intervention time. This kind of systems could lead to a reduction of the miss rate associated to polyp identification. Another possible domain of application could be the **automatic annotation of colonoscopy videos** by providing a complete scene description. This description may include the presence of salient traits such as anatomical structures (blood vessels, folds, polyps) and could also indicate the limits between insertion or withdrawal phases.

It is important to mention that an intelligent system could also be used to offer automatically metrics that may be used to offer a **quality assessment of the colonoscopy intervention**, which could provide a non-subjective way of assessment. We can also think of using elements of the scene description to generate testing sequences that could be used as a way to train physicians in a way such they can **assess and improve their skills** without the cost associated to a real interventions, and it would allow to compare different performance metrics objectively.

Finally, we could also use intelligent systems to extend and provide additional in-

formation from colonoscopy data. Belonging to this area we can think of applications such as the **development of patient-specific models**, that can be re-used later, when a new study arrives, to check for coincidences that can help in the diagnosis and enrich in this way a final case report.

In this section we will introduce the role that our contributions may play in the previously mentioned domains of application along with presenting works by other authors.

2.5.1 Computer-aided Diagnosis as a potential domain of application of our methods

The justification of the methods like the one we propose in this thesis is to aid in the diagnosis. We can think of a system that highlights, in real-time, which part of the image is more likely to contain a polyp inside. In this case we could think of two different approaches, depending on the type of solution that we want to provide. We can opt for only highlighting a given part of the image as having a high likelihood to contain a polyp inside or to provide a segmentation of the polyp. In this thesis we will introduce our polyp localization and polyp segmentation method, which could potentially be incorporated into a Computer Aided Diagnosis system.

Intelligent systems have been already used to assist in other medical interventions. For instance, there are systems that can be used to develop the knowledge bases used by expert systems, such as KARDIO [18], which was developed to interpret electrocardiograms. Another example can consist of developing a system that, in the context of anesthesia, provides a robust/reliable control system that could determine the optimal infusion rate of the drugs (muscle relaxant, anesthetic, and analgesic) simultaneously, and titrate each drug in accordance to its effects and interactions. Such a system would be a valuable assistant to the anesthetist during the procedure. An example of such a system can be found in the work of [77]. More close to our topic of interest, colonoscopy, we can find many examples of intelligent systems build to assist in cancer detection. Such is the case of breast cancer detection [115] or prostate cancer detection [113].

In particular, we think that the requirements that an intelligent system for colonoscopy must fulfill to be incorporated into a computer-aided diagnosis system are:

- **Real time performing:** If we want to provide useful information to the physician while he/she is performing the colonoscopy, we need to develop our systems to be fast enough. This means that the processing time of each frame has to be as fast as the frame rate in order not to slower the procedure. In this case maybe we should not aim a giving a perfect segmentation of the polyp even if it is hardly visible, but a good objective may be to indicate potential areas with polyp presence.
- **High specificity:** In medical imaging, we have to be specially careful about the kind of results that we are providing. In our case, it is clear that we prefer to

highlight bigger areas of the polyp even if the polyp only constitutes a small part of them than not highlighting the polyp at all. Considering this, we should aim at a system that has a low number, the closest possible to zero, of False Negatives, although we should also try to keep the number of False Positives low.

- **Easy user interface:** We have to think that our system should not be intrusive to the physician's work. We cannot simply fill the colonoscopy frame with lots of indicators that may distract the physician. Instead of that, we should aim at provide just an easy-to-see but not invasive indicator. We can think of highlighting the part of the image where we predict there is a polyp or just some kind of arrow indicator to guide the physician.

The methods that we present in this thesis could be directly incorporated into a computer aided diagnosis system. For instance we foresee as a use of our polyp localization method to provide the physician with indications of which areas of the image are more likely to contain a polyp inside by means of highlighting them. It is clear that the methods presented work faster if they have to deal with smaller regions of the image, which correspond to the informative areas. Regions such as the interior of the lumen should be automatically discarded. We should also take into account the clarity of the image [4], which can give us hints about if we can find useful information on a particular frame or, because of its quality, we should indicate only the physician to look for a better view.

As it has been presented in the previous sections, we count with a lot of theoretical tools to build CAD systems for colonoscopy but we also have some key constraints. As of now the majority of approaches only cover polyp detection and, when giving performance results, they are only referred to their own polyp databases which makes it difficult to compare between different approaches. We will present in Chapter 6 our contribution in terms of database building, hoping that it will be useful for future researchers in order to develop good polyp localization, segmentation and detection methods. More importantly, we provide in this thesis with the first transition between the research lab and real life by means of comparing the results of our methods with physicians observations, as will be shown in Chapter 8.

2.5.2 Other possible domains of application

As mentioned before, we foresee CAD as the main domain of application of our methods but there are other domains where they may play a role, which will be presented in this subsection.

- **Automatic scene description:** although polyp localization is by itself a method that could be used for CAD, we also provide in this thesis our contributions to identify some other elements of the endoluminal scene such as blood vessels or specular highlights. These methods, along other existing in the literature for elements that have not been covered in this thesis such as lumen

or folds, could be used to provide a map of the elements that appear on the endoluminal scene.

- **Quality assessment:** Currently, there are several metrics for the assessment of the quality of the colonoscopy intervention, such as the insertion time and withdrawal time. For instance, current ASGE (American Society for Gastrointestinal Endoscopy) and ACG (American College of Gastroenterology) guidelines suggest that, on average, withdrawal time should last a minimum of 6 minutes. Other works propose the use of additional metrics that include the quality of preparation, among others [75]. In the case of Europe, a very good work on quality assessment in colonoscopic interventions can be found in the work of [95], which defines from how to prepare conveniently the patient to an intervention to a classification of the polyps that can be found. These metrics can be potentially used into training programs for physicians, in order to assess their skills. Unfortunately, there is not a lot of information about what metrics could be extracted from a colonoscopy video in terms of computer vision analysis. One interesting approach can be found in the work of [46], which was later extended in [79]. These works presents a method to measure automatically the quality metrics for colonoscopy videos, based on analysis of a digitized video file created during colonoscopy and produces information regarding insertion time or withdrawal time.
- **Training:** one possible domain of applications of the methods that we propose could be to provide information to build up training systems for the physicians to improve and test their skills. The work of [109] proposes the evaluation of the skills of the trainees, and their evolution during learning processes, by using eye-tracking methodologies as a tool for the assessment of abilities such as active visual search and reaction time to the presence of polyps, among others. This study presents a novel method which compares visual search patterns between the skilled specialists and the trainees. This is done by tracking the eye position of two groups of physicians (experts and novices) while they are shown a set of colonoscopy videos. Several measures were computed by analyzing the eye-tracker results, such as eye movement speed or number of fixations. The obtained results show that colonoscopy experts and novices show a different behavior in their visual search patterns, and therefore the proposed eye-tracking based procedure can provide automatic and objective measures for their evaluation. A method similar to the one presented in [109] can be potentially used both for assessment of the skills of the trainees during their learning process or to assess the quality of the whole procedure in intervention time. In addition, the inclusion of the models of appearance and the item categorization from the tools for scene description can provide an objective ground-truth against which to check the abilities of the trainee. This can potentially be implemented by analyzing the extent to which the trainee identifies the regions of interest, which could be provided by means of our polyp localization method.
- **Development of patient-specific models:** Since some of the methods presented allow the detection, segmentation and characterization of anatomical

structures, lesions and physiological behavior, there is a manifest potential to use these strategies in order to endow current techniques with architectures ready to work with patient-specific models. The patient-specific approach has been one of the main trends in clinical research lately and it has been one of the pillars of the research funding schemes for Information and Communication Technologies related to health care in Europe during the last Framework Programs [48]. The patient-specific orientation focuses on the adaptation of existing methodologies so that they can take profit of the particular information, traits, clinical details or characteristics associated to each patient. Thus, the patient-specific viewpoint aims at the focalization of the (general) outcomes provided by each technique onto the (particular) specificities of each case. The extent to which this perspective can be exploited by using intelligent systems in colonoscopy is an open field of work. Here, we expose only as a matter of example a tentative list of a few prospective ideas. On the one hand, the use of feature detection in colonoscopy video could provide a way to the characterization of the inner walls of the colon, based on the identification of unique traits, which could be used for the tagging or annotation of physiological features as markers, and apply this information in a further step for the identification of the exact place of the situation of region close to a polyp. These visual traits of the colon could be used in order to find those very specific locations when a new colonoscopy intervention is performed on that patient. This could provide a method for a precise spatial localization of regions of interest. The straightforward application of this potential implementation would be oriented to the registration and study of evolution of lesions in time (or whatever other item of interest) in the sequential routine interventions carried out on a particular patient, by automatically providing the specialist with a measure of certainty about the location of those lesions. The generalization of this methodology could be addressed towards the definition of a patient-specific atlas of the colon, in a way in which the specialist could have track of landmark positions in intervention time. This perspective presents a scenario in which the specialist is endowed with a road map for the navigation in intervention time, allowing the specialist to address specific targets with high reliance, reduced time and a potential shrinking of the miss rates.

Chapter 3

A model of appearance for polyps

3.1 Introduction

In order to develop our polyp localization and polyp segmentation methods we need a robust definition on what a polyp is and how it appears in colonoscopy frames in order to develop accurate polyp localization and segmentation methods. These methods will be depicted in other chapters and we will dedicate this current chapter to introduce our model of appearance for polyps, which is based on how polyps appear in colonoscopy frames. Before entering into details of our model, we will introduce and define what a polyp is and how they appear in actual frames, along with presenting some of the challenges that polyps present. After this, we will explore the connection between how images are acquired and Phong's illumination model, which leads to one of the contributions of this thesis, which is a model of appearance for polyps. We will also introduce some tools that could be used to aid in polyp localization and polyp segmentation. We will finish this chapter by introducing our polyp localization and segmentation schemes, which will vertebrate the following chapters.

3.2 Definition of polyp

Polyp is a general term used to describe a benign (non-cancerous) growth on the lining, or inside, of a mucous membrane which includes those that are found in the colon. In general polyps are considered pre-cancerous but if left untreated, they may develop into cancer. There are several different classifications for polyps, depending on the scope of the analysis. If our aim is to classify polyps by their shape, the most general classification divides them into pedunculated and sessile [16]. Pedunculated polyps are mushroom-like tissue growths which are attached to the mucous membrane by a peduncle or stalk. On the other hand, sessile polyps lay directly on the surface of the mucous membrane. Depending on their protrudness they can be subdivided into

salient, flat and depressed polyps. We can observe graphical examples of the several polyp types in Figure 3.1. In our case we aim at localize and segment polyps like the ones shown in Figure 3.1 (a), (b) and (c).



Figure 3.1: Type of polyps: (a) Pedunculated polyp; (b) Salient sessile polyp; (c) Depressed sessile polyp; (d) Flat sessile polyp .

As we have mentioned, polyp classification by means of shape is not the only possible classification [16]. Another criteria is the type of polyp. The most common types of colon are inflammatory, adenomatous, hyperplastic, villous, lymphoid and juvenile. In our case we are interested in the first classification criteria because it provides an initial cue to localize polyps in colonoscopy frames, which is shape. It is clear that by means of this first polyp classification, we obtain some general shapes that may have but, as we are dealing with frames from colonoscopy studies, it is rare that we will obtain clear perfect shots of polyps which are easy to localize. In our case we will build our model of appearance for polyps by taking a look at how polyps do appear in colonoscopy frames, which also considers how the colonoscopy frames are acquired.

3.3 Phong's illumination model

The colonoscope has a light source and a camera attached to it in a way such both the camera and the source of light are in the same direction. We will use this in order to develop our model of appearance for polyp. For this aim, we need both an a priori model about the polyp and a model of the illumination. For the sake of simplicity let us consider a polyp as a semi-spherical shape protruding from the colon wall plane. It is important to mention that polyps and folds are the only elements on the endoluminal scene that could be considered as having a three-dimensional shape whereas other elements such as blood vessels and specular highlights could not. We will also consider that the polyp surface is regular and that its reflectance can be approximated by the Phong's illumination model [84].

Phong reflection is an empirical model of local illumination, which describes the way a surface reflects light. This model defines the illumination of each surface point as a combination of the diffuse reflection of rough surfaces with the specular reflection of shiny surfaces. It is based on Bui Tuong Phong's informal observation that shiny surfaces have small intense specular highlights, while dull surfaces have large highlights that fall off more gradually. Phong's illumination model also includes an ambient term to account for the small amount of light that is scattered about the

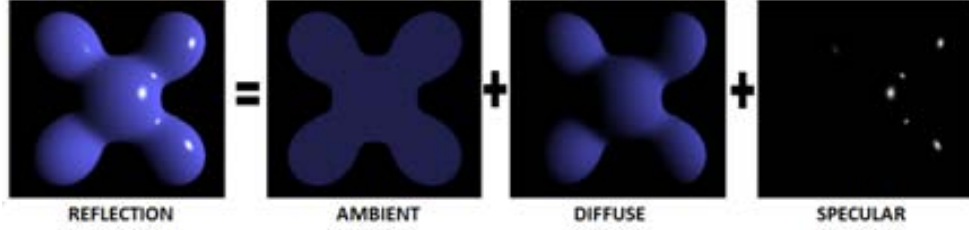


Figure 3.2: Decomposition of the light reflected by the surface into the three components: ambient; diffuse and specular.

entire scene. We can see in Figure 3.2 a graphical example that decomposes the light reflected into three components, namely: ambient, diffuse and specular.

Combining how the image is acquired ¹ and Phong's illumination model (see Figure 1.5 and Figure 3.2) it is somehow easy to infer a model of polyp appearance, as polyps can be approximated as protruding surfaces where the light is reflected. We can model the colonoscope itself by a pinhole camera and a punctual illumination source placed in the same position. Figure 3.3 (a) shows a schematic representation of this scenario. Under such assumptions, the image provided following these approaches is calculated as:

$$I = I_a * K_a + f_{att} * I_p * [K_d * \cos \theta + W(\theta) * \cos^n \alpha] \quad (3.1)$$

where I is the light reflected by the surface towards the camera, I_a is the ambient intensity, K_a is the ambient reflection constant, f_{att} is the attenuation factor, I_p is the punctual intensity, K_d is the diffuse reflection coefficient, θ the angle between the surface normal and the illumination source, $W(\theta)$ is the fraction of light reflected in a specular way, α is the angle between the surface normal and the camera, and n modulates the decay in the specular reflection. This model is implemented with a set of arbitrary values for the sake of visualization in Figure 3.3 (b) and (c), which show a rendering of a synthetic polyp for tilt angles of $\alpha = 0^\circ$ and $\alpha = 60^\circ$. In this scenario, sharp edges, gradual shading and specular reflections are created (these same elements can be visualized in the real example of Figure 3.3 (d)).

To close this section and to assess the validity of our model in real images, we present in Figure 3.4 a comparison between the grey level intensity under a given region of the image for both a model image and a real image. As it can be seen in Figure 3.4 (c) and (d), the boundaries between the polyps and the rest of the image appear as jumps in the grey level, which confirms our initial assumption of shadings surrounding polyps and therefore constituting polyp boundaries.

¹We have explained how colonoscopy frames are acquired in Chapter 1.2 of this thesis

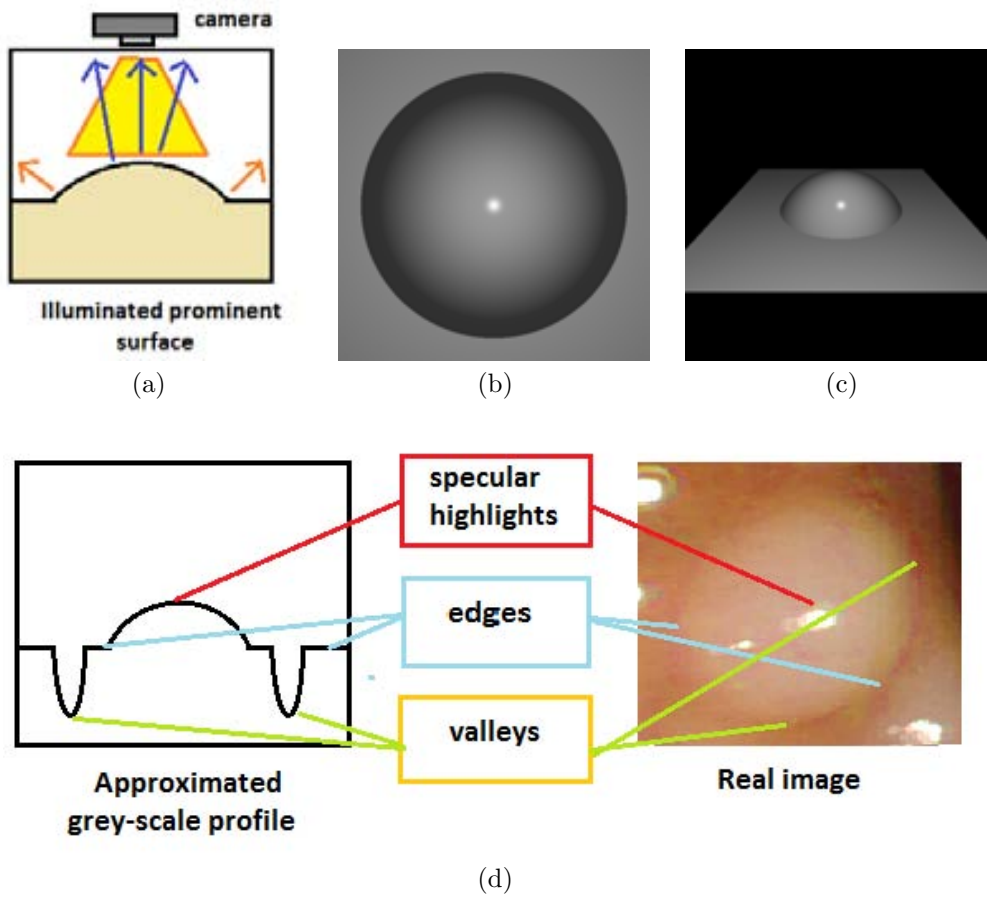


Figure 3.3: Model of appearance and illumination of polyps: (a) Scheme of an illuminated prominent surface; (b) and (c) synthetic model rendering for 0 and 60 degrees. (d) a real example.

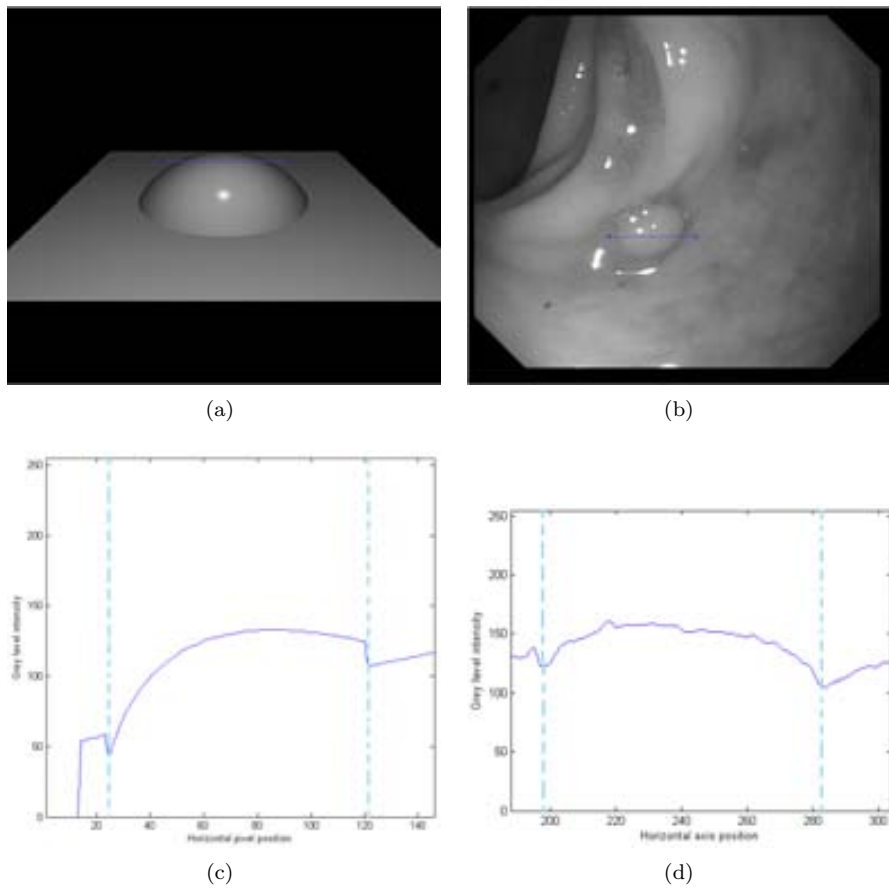


Figure 3.4: Validation of the theoretical model in real images: (a) Synthetic 3D model of a polyp; (b) Real polyp example; (c) Grey level of the model image; (d) Grey level of the real image. In this Figure the blue line represents the portion of the image which gray level variation we will check and the red line marks the boundary of the polyp.

3.4 Cues on polyp characterization by means of an illumination-based model

The characterization of the polyp is obtained through the shadings, which are related to valleys in the intensity image, but these are not the only cues that can be used to guide a model of appearance for polyps. Moreover, colonoscopy frames present some challenges which can hinder polyp localization and segmentation that have to be taken into account, as they result on a great variability in polyp appearance in colonoscopy videos. These challenges are:

1. Non-uniform appearance of polyps (see Figure 3.5 a-b).
2. Variability in shape: sessile or peduncular (Figure 3.5 a-c).
3. Effects of image acquisition, such as changes in pose, blurring, occlusions, specular highlights (Figure 3.5 d-e-g).
4. High similarity between the tissues inside and outside the polyp, which disables the possibility of relying only on texture or color cues (Figure 3.5 h).

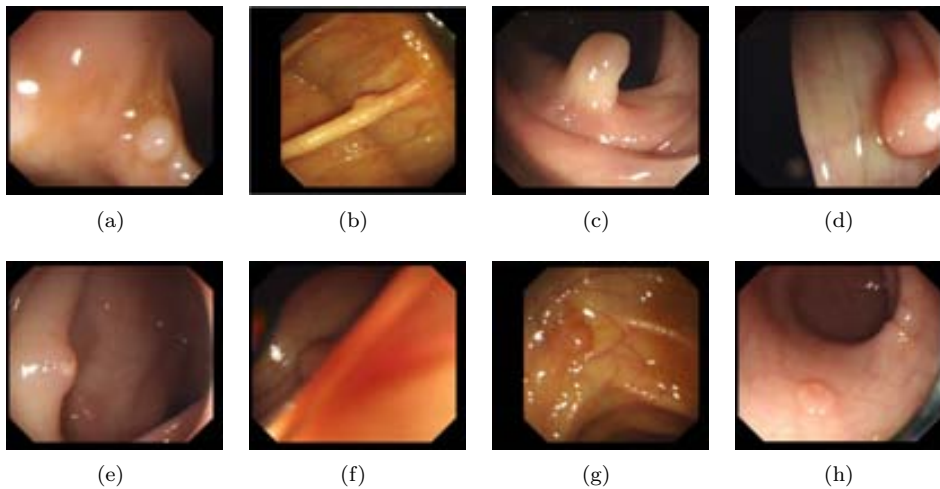


Figure 3.5: Challenges in polyp detection: (a-d) non uniform appearance, (e) partial (lateral) views, (f) blurred images, (g) specular highlights, and (h) uniform texture and color inside and outside the polyp.

By observing several polyp examples from the available studies and acknowledging the challenges that we have already presented, we have compiled a list of cues that can be used to develop our model of appearance for polyps, which are enumerated next.

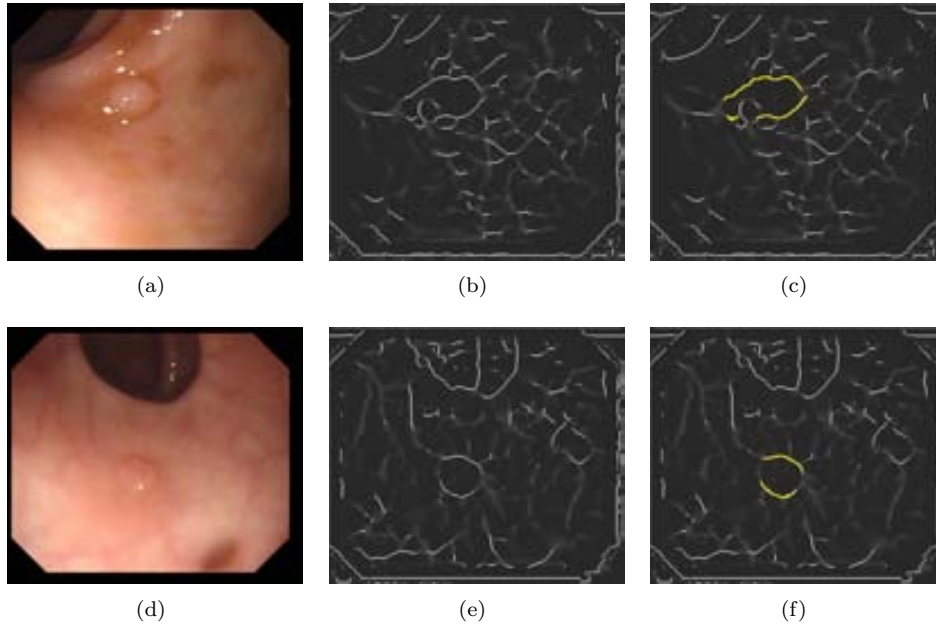


Figure 3.6: Valley detection to guide a model of appearance for polyps: (a-d) Original images; (b-e) Valley images and (c-f) Valley image with polyp boundary valleys highlighted in yellow.

3.4.1 Valleys surrounding the polyp

We can see in Figure 3.6 some examples that illustrate how the output of a valley detector, conveniently treated, can be useful to guide a model of appearance for polyps.

As it can be seen from the examples, the output of a valley detector can be linked with the boundaries of polyps (painted in yellow in Figure 3.6 (c) and (f)) although, as we can see, in some cases we do not have a high valley response for every pixel of the boundary.

3.4.2 Contours and edges in lateral view

The model presented previously defines polyps as surrounded by shadings, which constitute valleys on the intensity image. But depending on the view of the polyp that we are observing, this definition may fall a little short, as shown in Figure 3.7. In this case we observe that for the case of lateral polyps we do not have complete valley information. Even more, for the part of the polyp that is closer to the lumen, we do not have any valley information at all, as it can be seen in Figure 3.7 (d).

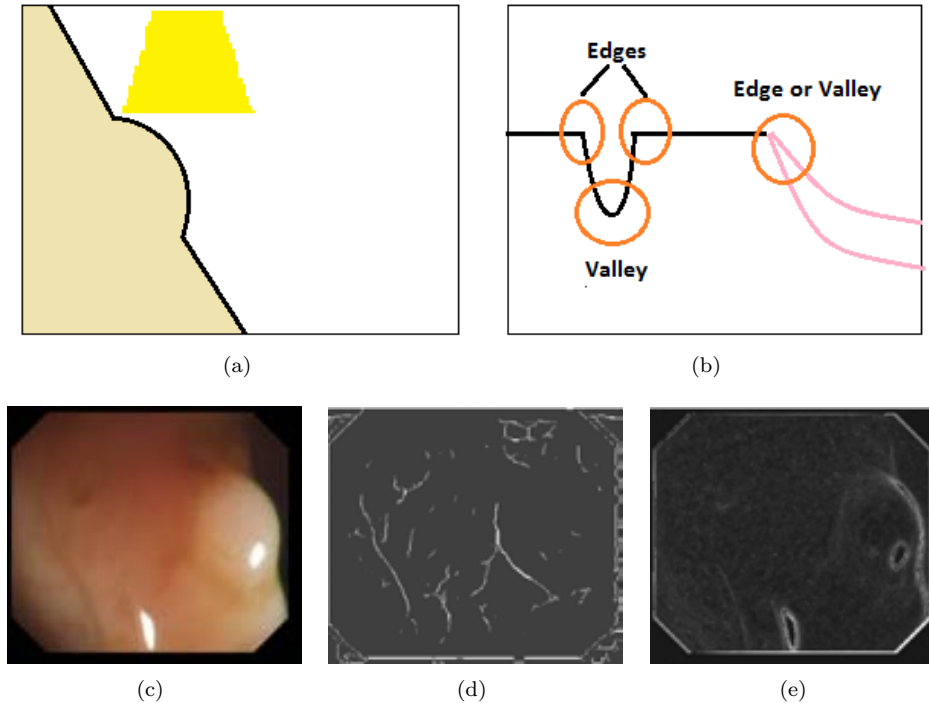


Figure 3.7: Model of appearance applied to lateral polyps: (a) Scheme of an illuminated prominent surface; (b) Intensity profile and (c) a real example (d) Valley image and (e) Gradient image.

Figure 3.7 shows that, for the case of lateral polyps some of the boundaries do not convey any valley information but they provide results by means of a contour or edge detector. We can how this is represented on the grey level intensity image in Figure 3.8. As it can be seen from Figure 3.8 (c) and (d) we obtain valley response for one of the extremes of scope, but for the other we only see a great decay in the intensity level which indicates edge or gradient information.

The solution that we propose is to use gradient or edge information. As can be seen in Figure 3.7 (e) by doing this, we can recover one part of the boundary of the polyp that was missed by using only valley detector information. But, as we can also see, by using only gradient information we lose the boundary that we have already obtained by means of a valley detector. Therefore we are in a situation where by using separate valley or gradient information we only obtain half of the needed information. As it can be suspected, a good solution to this will be to combine both types of information and we will develop more about this in the Chapter 4.

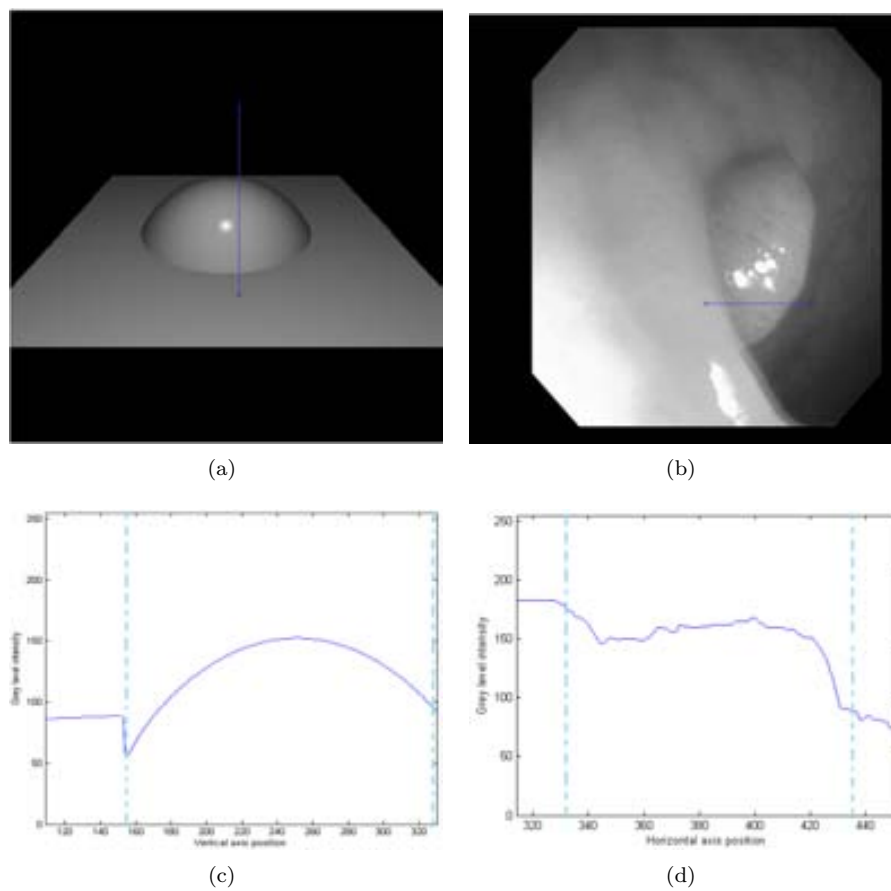


Figure 3.8: Validation of the theoretical lateral polyps model in real images: (a) Synthetic 3D model of a polyp; (b) Real polyp example; (c) Grey level under the selected region for the model image; (d) Grey level under the selected region for the real image. In this Figure the blue line represents the portion of the image which gray level variation we will check and the red line marks the boundary of the polyp.

3.4.3 Internal shadings of the polyp

Another cue that could be used to guide our methods will be the content of the polyp regions by means of its internal shadings. But, as it can be seen in Figure 3.9, it is very difficult to see internal shadings as a key cue, as the grey level inside the polyp does not vary enough to make it relevant.

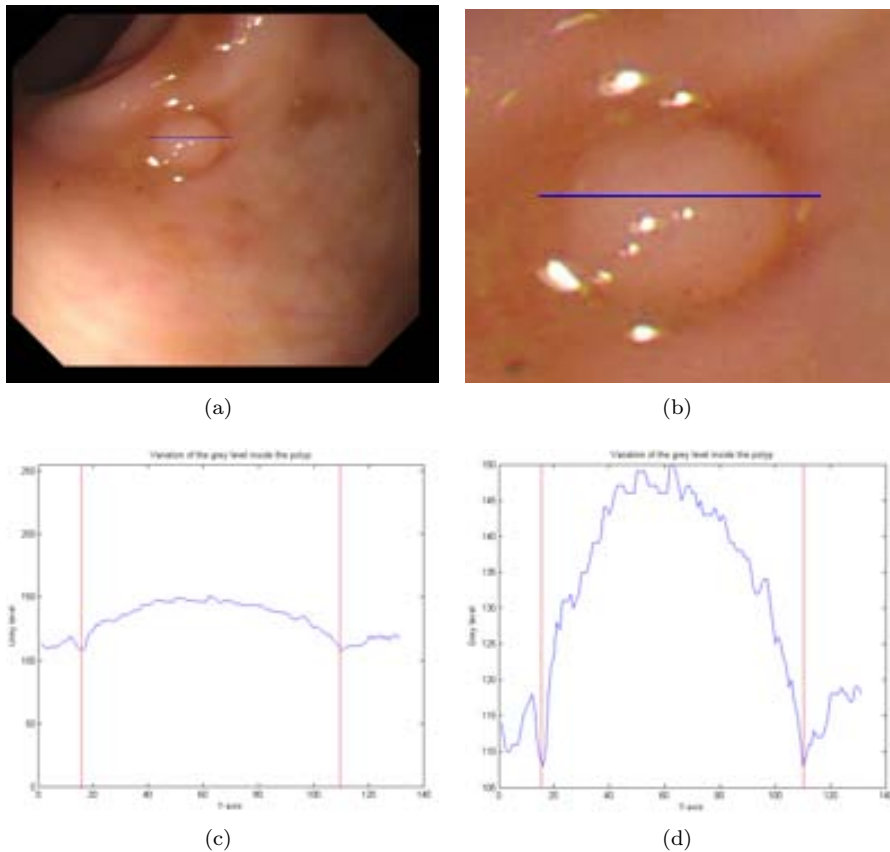


Figure 3.9: Variation of grey level inside the polyp. : (a) Original image; (b) Zoom of original image; (c) Overall gray level below the blue line; (d) Zoom of gray level below the blue line. In this Figure the blue line represents the portion of the image which gray level variation we will check and the red line marks the boundary of the polyp.

As we can see from Figure 3.9 there is not a great variation of gray level inside the polyp and we even have to make a zoom (see Figure 3.9) to find some differences, which are not high anyway. We do see a great jump in terms of gray level that coincides with the boundary of the polyps (marked with a red line in both Figures 3.9 (c) and (d)), which also leads to the use of measures based on difference of intensity

levels such as valleys or gradient to determine polyps' contours.

3.4.4 Specular highlights

Specular highlights appear as the light falls perpendicularly into protruding objects, such as polyps which return part of the light that impinges them, as it can be seen from Phong's illumination model. Because of this, one possible cue to guide a model of appearance for polyps could be the presence of specular highlights.

The property that we want to take advantage from here is that protruding 3D objects such as polyps reflect the light as it falls into them therefore if we find specular highlights in some area of the image we can assume that a protruding object is present. Unfortunately, after observing frames with polyp from our available studies, we have found that specular highlights are not always present. We show in Figure 3.10 some examples of frames with polyp and with or without presence of specular highlights.

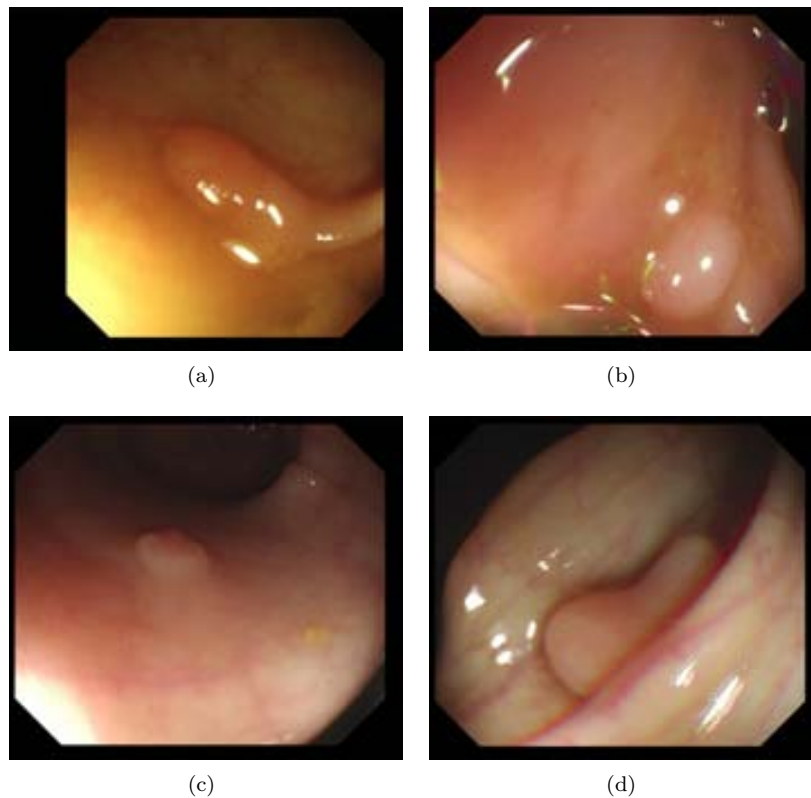


Figure 3.10: Use of specular highlight as a cue to define a model of appearance for polyps. (a-b) Examples of frames with polyps and specular highlights. (c-d) Examples of frames with polyps and without specular highlights.

Taking this into account and considering that there are some examples where specular highlights are not present, we can not affirm that the presence of specular highlights could be included as a necessary condition that can indicate a polyp presence, even considering that this is true for the majority of the cases. And we can not forget that there are other protruding objects on the endoluminal scene apart from polyps, such as wrinkles and folds, which also contribute to the presence of specular highlights.

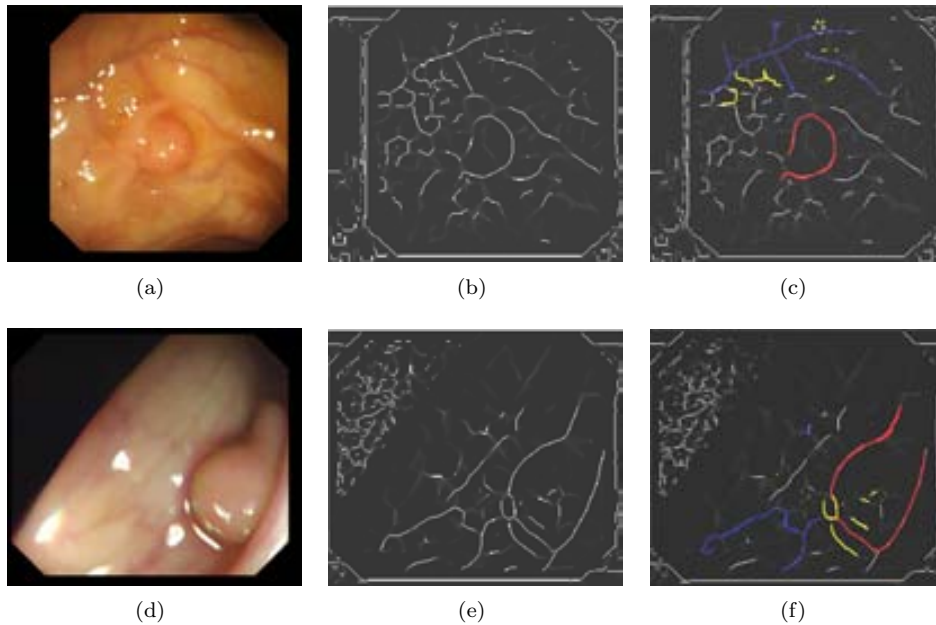


Figure 3.11: Challenges of using valley detection to guide a model of appearance for polyps. (a-d) Original images; (b-e) Valley detection images; (c-f) Valley detection marked images. Blood vessels valleys are marked in blue and specular highlights-caused valleys in yellow, whereas polyp contour valleys are marked in red.

3.4.5 Discussion

We have detailed in this section some cues that could be used to guide our polyp localization and segmentation method and we will close this section by exposing which of them we will consider for our algorithms. Our model defines a polyp as *a prominent shape enclosed in a region with presence of valleys around its frontiers*. Therefore the use of the output of a valley detector as a cue seems logical, as the examples provided show that valleys do indeed constitute part of the polyp's boundaries. But we have also shown that for some views of the polyp this valley information may not be enough to define polyp boundaries and, because of this, we will also use gradient information to complement the output of a valley detector.

As was shown in Figure 3.9 we can not obtain enough useful information from internal shadings of the polyp that can be used to differentiate polyp regions from the rest therefore we will not use this kind of information. The case of specular highlights is different in some way: they appear for many of the images where polyps are present but they also appear in parts where there is no polyp presence. Moreover, not even for all the polyps we have a specular highlight, as it has been shown in Figure 3.10. Taking this into account we will not use specular highlights as a cue to indicate polyp presence but, as shown in Chapter 4, we will address the effects that they produce in our methods.

Finally, it is important to mention that although we will base our methods on valley information that polyp boundaries provide, we have also taken into account the contribution in terms of valley information of image acquisition and other elements of the endoluminal scene in order to develop our polyp localization and segmentation methods. Consequently, we will have to expand our model or treat these elements separately in order to make our algorithms work as intended. As an example, we show in Figure 3.11 the output of a valley detector on a complete endoluminal scene.

Chapter 4

Valley detection for polyp localization and segmentation

4.1 Introduction

We will base our polyp localization and polyp segmentation methods on a model of appearance for polyps. This model defines a polyp as a prominent shape enclosed in a region with presence of edges and valleys around its frontiers. We have also explained how the use of valley detectors seems suitable for our concrete case although this model has to be extended in order to overcome some of the challenges that the sole use of valley detection present, such as the presence of other elements from the endoluminal scene such as blood vessels and specular highlights.

We will present in this section how we will use the information that valley detection methods provide to guide polyp localization and segmentation, starting with the definition of what a valley detector is and presenting some alternatives that can be used. Once our final valley detector is introduced, we will detail how we do solve the challenges that were sketched at the end of the previous chapter, in order to both generate our enhanced input and depth of valleys images.

4.2 Valley detection

4.2.1 Introduction

As we are going to base our polyp localization and segmentation methods on the shadings that surround polyps and their relationship with the output of a valley detector, it is clear that we need to define what a valley detector is and present some different approaches that we can use to obtain a valley image. The characterization of the polyp is obtained through the shadings, which are related to valleys in the

intensity image. But, as was shown in Chapter 3, the model of intensity valleys as boundaries of the polyp is not valid for a certain type of polyp appearance such as lateral polyps, as it can be seen in Figure 3.7.

Therefore, we need to extend our initial approximation of intensity valleys as the shades that define polyp boundaries in order to cope with cases such as the one showed in Figure 3.7. We will use in this thesis the term *valleys*, although in the literature the general concept is known as ridges and valleys detection. In mathematics the ridges of a smooth function of two variables are the set of curves which points are local maxima of the function in at least one dimension. For our case, we are interested in the valleys that surround polyps, as the grey level in these valleys will be lower than the intensity of the pixels in a neighborhood.

There are several approaches to detect intensity valleys in images. In our case, we have decided to use *multilocal creaseness* [67], although other alternatives such as *second derivative of gaussians* [41] could have been considered. We will present both methods in the next subsections.

4.2.2 Second Derivative of Gaussians

The difference in contrast between structures such as the boundaries of the polyp point out the potential that second derivatives of anisotropic Gaussian kernels as valley detector. The kernel values are defined by the oriented Gaussian function described by:

$$G_{(\sigma_x, \sigma_y), \theta} = \frac{1}{(2\pi)\sigma_x\sigma_y} e^{-\left(\frac{\tilde{x}^2}{2\sigma_x^2} + \frac{\tilde{y}^2}{2\sigma_y^2}\right)} \quad (4.1)$$

where (σ_x, σ_y) are the scales in the corresponding axis and θ is the rotation angle of the filter. \tilde{x} and \tilde{y} are the coordinates given by the rotation angle. Hence they are defined as:

$$\begin{aligned} \tilde{x} &= x \cos \theta + y \sin \theta \\ \tilde{y} &= x \sin \theta - y \cos \theta \end{aligned} \quad (4.2)$$

As we use anisotropic Gaussians with $\sigma = \sigma_x = 2\sigma_y$ the Gaussian function results in:

$$G_{\sigma, \theta} = \frac{1}{(2\pi)2\sigma^2} e^{-\left(\frac{\tilde{x}^2}{2(2\sigma)^2} + \frac{\tilde{y}^2}{2\sigma^2}\right)} \quad (4.3)$$

Therefore, as we are modelling polyp boundaries with second derivatives of anisotropic Gaussian kernels, the kernel will be defined as:

$$\partial_{\tilde{y}}^2 G_{\sigma, \theta} = \frac{\tilde{y}^2 - 1}{\sigma^4} G_{\sigma, \theta} \quad (4.4)$$

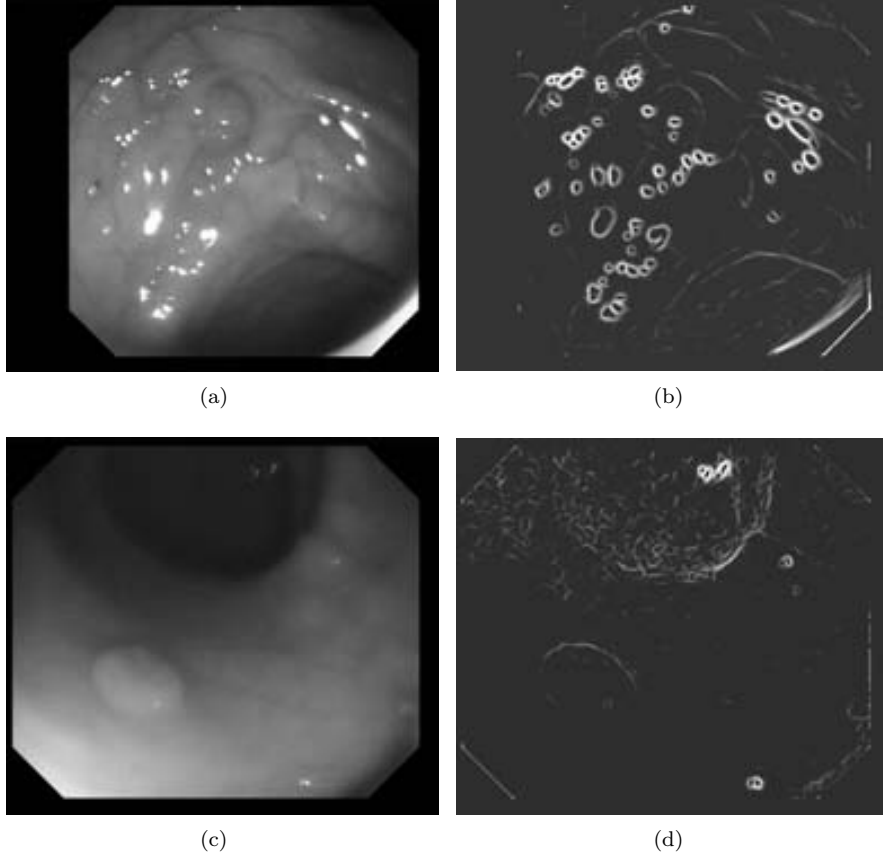


Figure 4.1: Second derivative of Gaussians applied on colonoscopy frames: (a) and (c) Original image; (b) and (d) Corresponding valleys image.

We apply a normalization so that the geometry of the polyp boundaries is prioritized:

$$G_{\sigma,\theta}^N := \frac{\|\partial_y^2 G_{\sigma,\theta} * I\|}{\|\partial_y^2 G_{\sigma,\theta}\| \|I\|} \quad (4.5)$$

where $\|\cdot\|$ stands for the L^2 integral norm and $*$ denoting the convolution operator.

The kernels are applied for 8 equally distributed orientations and scales $\sigma = [2, 4, 8]$, which cover the majority of polyp contours from our test dataset. It all means we have 24 output images, each of them corresponding to a determined orientation and scale. Hence, the output $I_{valleys}$ must be a combination of all of them, defined as follows:

$$I_{valleys} = \max_{i,j} (G_{\sigma^i,\theta^j}^N) \quad (4.6)$$

We show some examples on valley detection by means of *second derivative of gaussians* in Figure 4.1.

4.2.3 Multilocal Creaseness-based Valley Detector

As mentioned in the previous chapter, by combining Phong's illumination model with a priory shape model of polyps in colonoscopy images, we can approximate the shadings that protruding objects generate as intensity valleys. Ridges and valleys in n -dimensional images are commonly identified as loci of minimum gradient magnitude along the relief's level curves [67]. If $|\lambda_1| \geq \dots \geq |\lambda_d|$ are the eigenvalues of $\nabla \nabla L$ and v_1, \dots, v_d their corresponding eigenvectors, then a n -dimensional crease ($1 \leq n \leq d$) is characterized as:

$$\forall i \in \mathbf{I}_{d-n} \quad \nabla L \cdot v_i = 0 \quad (4.7)$$

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

In 2D, ridges/valleys can be also identified as positive maxima/negative minima of the curvature of the relief's level curves. Maxima are connected from one level to the next therefore constituting a subset of the vertex curves. In d dimensions we generalize the level curves of L to level sets. A level set of L consists of the set of points $S_l = \{x \in \Omega : L(x) = l\}$ for a given constant l . Then, if $|k_1| \geq \dots \geq |k_d|$ are the principal curvatures of the level hypersurface S_l and t_1, \dots, t_d their corresponding principal directions, a nD crease ($1 \leq n \leq d$) is characterized as ([67]):

$$\forall i \in \mathbf{I}_{d-n} \quad \nabla k_i \cdot t_i = 0 \quad (4.8)$$

and

$$t_i^t \cdot \nabla \nabla k_i \cdot t_i < 0 \quad \text{and } k_i > 0 \quad \text{if ridge} \quad (4.9)$$

$$t_i^t \cdot \nabla \nabla k_i \cdot t_i > 0 \quad \text{and } k_i < 0 \quad \text{if valley} \quad (4.10)$$

The condition of positive maxima or negative minima of k can be translated to high values of $|k|$, where values greater than 0 measures ridgeness and values smaller than 0 measures valleyiness. In 2D when the height condition holds, we can equal v_1 to v and λ_1 to L_v , where $v = (L_y - L_x)^t$ is the tangent vector of the level curves of L . Consequently L_{vv} can be seen as a creaseness measure: if its value is high there are higher chances that the highest value in magnitude of the second order directional derivative is reached along v . The measures L_{vv} and k are related by

$$k = -L_{vv}/L_w = \frac{(2L_x L_y L_{xy} - L_y^2 L_{xx} - L_x^2 L_{yy})}{(L_x^2 + L_y^2)^{3/2}} \quad (4.11)$$

where $w = (L_x, L_y)^t$ is the 2D gradient vector of L . By this we can consider L_{vv} as the measure k weighted by the gradient magnitude in order to eliminate its response in isotropic regions. In our case we are interested in those pixels that have creaseness value smaller than 0 in the creaseness image, as it can be seen in Figure 4.2.

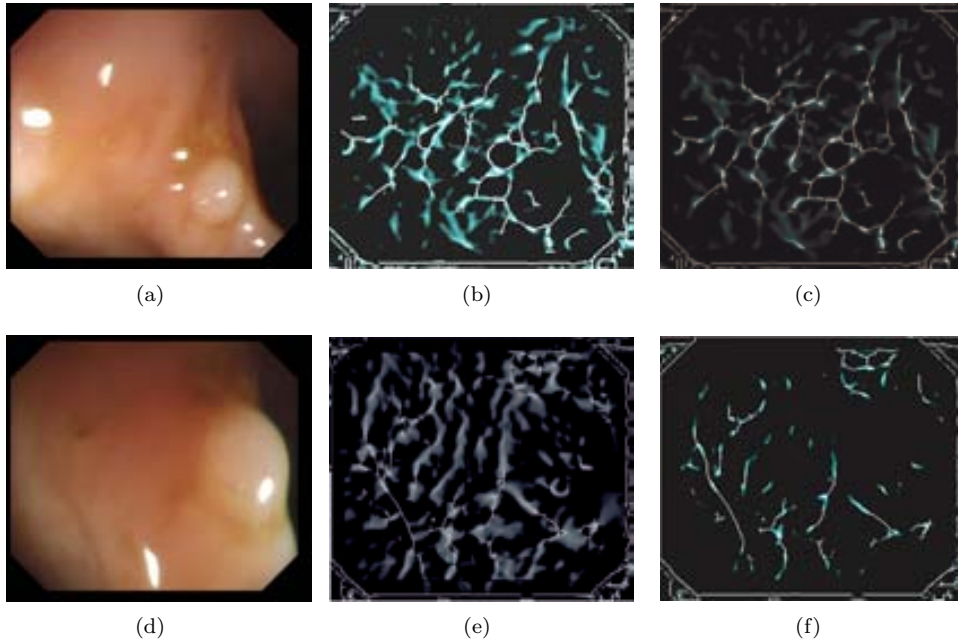


Figure 4.2: Intensity valleys in colonoscopy frames: (a) and (d) Original image; (b) and (e) Creaseness image; (c) and (f) Valleys image.

4.2.4 Comparison between Second Derivative of Gaussians and Multilocal Creaseness

Once we have presented two of the available methods of valley detection we will make a brief comparison of the output that they provide for several colonoscopy frames. We have to mention that for the sake of the comparison both methods have been applied with the default parameter values. We show some representative examples in Figure 4.3.

The scope of this thesis is not to select which of the available valley detector performs better because an analysis of this type will require to define a complete experimental setup and would involve a fine tuning of the several parameters that both methods need. Preliminary qualitative studies show that the use of multilocal creaseness provide with less noisy valley images while both give similar output in terms of polyp boundary detection. For instance we can see from Figures 4.3 (b) and (c) where we can see that *second derivative of gaussians* seems to give much more

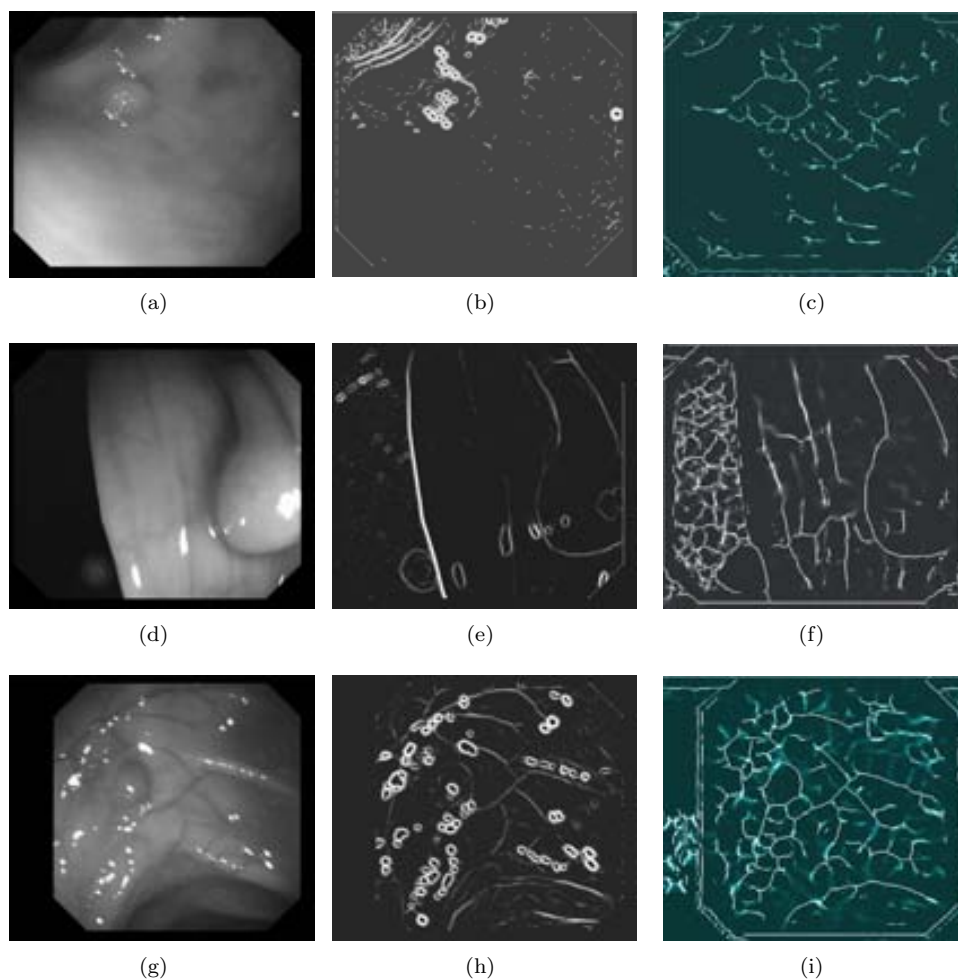


Figure 4.3: Intensity valleys in colonoscopy frames: (a),(d) and (g) Original image; (b),(e) and (h) *second derivative of gaussians* output; (c),(f) and (i) *multilocal creaseness* output.

importance to valleys constituted by non-interesting elements of the endoluminal scene such as the black mask or specular highlights whereas *multilocal creaseness* gives higher importance to polyp boundaries. We can see from Figures 4.3 (e) and (f) that blood vessels seems to impact more the output of *multilocal creaseness*, but the three examples that we present show that in order to get complete polyp boundaries *multilocal creaseness* is the option to take. We have to take into account that *multilocal creaseness* also gives high valley value to some other elements of the endoluminal scene which we will address in the next section.

4.3 Issues on polyp boundary characterization by means of valley information

4.3.1 Introduction

As we mentioned in last chapter, although the use of valley detection methods seem to suit our model of appearance for polyps, there are some challenges that valley output images present that should be addressed in order to aid in our polyp localization and polyp segmentation methods. As it can be seen in Figure 4.4 there are several elements on the image, whether they are elements from the endoluminal scene or just artifacts caused by the image acquisition process, that have an impact in terms of valley detection and present challenges that have to be addressed. These elements are:

1. Interlacing: Because of the interlacing process, the time difference between the difference in acquisition time of the odd and even lines of a frame may result in small differences in the image, which may give response in terms of valleys.
2. Black mask: Because of the image acquisition process and the lens shape [102], there is some part of the image where we do not have scene information. As it can be seen in Figure 4.4, the black mask that surrounds the endoluminal scene do generate a high response in terms of valley information.
3. Specular highlights: Specular highlights appear as a result on the light of the camera falling into protruding objects which reflect the light back to the camera. As specular highlights result on highly saturated regions of the image, they do generate valleys surrounding them which may be taking into account when developing our methods.
4. Blood vessels: Blood vessels appear as darker line segments due to its lower reflectance with respect to colon walls. Because of this, they constitute valleys as the grey level below them is low compared to the grey level of the pixels on their neighborhood.
5. Lateral views: As mentioned in Chapter 3, lateral views of polyp result on not having valley response in the contour of the polyp that touches the lumen.

Because of this we may lose some polyps as the decrease in terms of polyp contour information may be high.

In the next sections we will provide solutions for all the presented challenges in order to improve our valley image but we will start by making a brief explanation on how colonoscopy images are created and which elements of the endoluminal scene are generated directly from the image acquisition process. In this case we will divide the solutions in two groups: 1) Image preprocessing to enhance valleys generated by polyps and 2) Extension of the valley image to provide with better valley quantification and completeness.

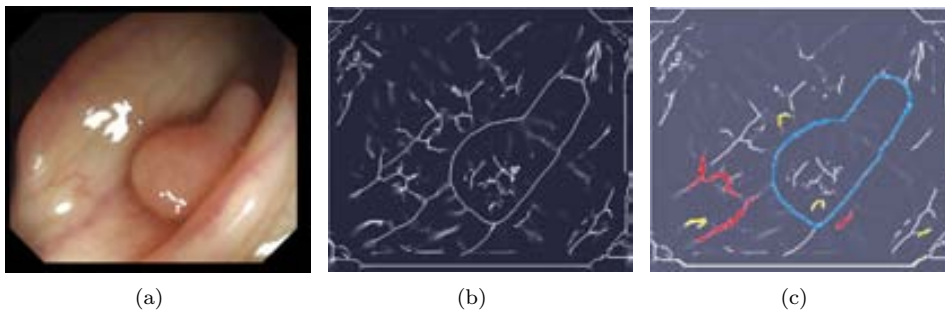


Figure 4.4: Valley detection on an endoluminal scene: (a) Original image (b) Valleys image. Black mask-generated valleys are marked in green, specular highlights-generated in yellow and blood vessels-generated in red. Polyp contour valleys are marked in blue.

As we will apply an inpainting operation to both the black mask and the specular highlights we will explain our inpainting method before starting with explaining the solutions to the challenges.

4.3.2 Inpainting method

To do the inpainting of a certain area we first need to define which pixels of the image are going to be inpainted. Once this mask is obtained, we can start with the inpainting algorithms which consists of two stages, namely diffusion and obtention of the final inpainted image.

- Diffusion: The objective of this stage is, once detected the pixels that constitute the specular highlight, to diffuse values from the original image into the specular highlights mask in the following way:

Algorithm 1: Inpainting diffusion algorithm

Data: Diffusion(I_o, M, s_{th})
Arg: (I_o : input image, M : input mask, s_{th} : stop threshold)
Result: Diffused image(I_d)

```

1  $Vm = \neg M$ ;
2  $I_d = I_o$ ;
3 repeat while the image is modified over  $s_{th}$ 
4    $mod = false$ ;
5   forall the  $\vec{x} \in I, M(\vec{x})$  do
6      $Neigh = \{\vec{p} | \vec{p} \in Neighborhood(\vec{x}), Vm(\vec{p})\}$ ;
7     if  $\#Neigh > 0$  then
8        $nv = \frac{\sum_{\vec{p} \in Neigh} I_{\vec{p}}}{\#Neigh}$ ;
9       if  $Vm(\vec{x})$  then
10        | if  $|nv - I_d(\vec{x})| > s_{th}$  then  $mod = true$ ;
11        | else
12        |    $mod = true$ ;
13        | end
14        |  $I_d(\vec{x}) = nv$ ;
15        | end
16     end
17   end
18 until  $\neg mod$ ;
```

The algorithm starts by creating an inner matrix which keeps track of the positions of the pixels under the input M mask and then initializes the diffused image with the input one (lines 1 and 2 of the algorithm). For each pixel under the original input mask we do the following: we obtain a neighborhood around the pixel (line 6) and change its original value by the mean value of the neighbor pixel's values (line 8). It has to be mentioned that we only take account in the mean operation those pixels which were not part of the original input mask, that is, pixels which have zero value in the M mask. This process is repeated until every pixel with 1 value in the M mask has obtained a new value. Once this happens, we repeat the process until the stop criteria is met. The stop criteria halts the algorithm when the difference between the new value and the previous value under the pixels with 1 value input mask M is smaller than a s_{th} threshold value (line 10).

- Obtention of the final inpainted image: In order to obtain the final image we take into account that if we want to assure that the pixels under M do have their new value on the final image but we also have to consider that if we do a direct substitution there will still remain a clear frontier between pixels inside and outside the final image. In order to solve this we create an extended mask which ponders the way we combine I_o and I_d in the I_{np} image. This M_1 mask is created by dilating the original M mask with a circle structural element and later convolving the result with a gaussian kernel (see Eq. 4.3.2).

$$M_1 = M \oplus C_\sigma * G_\sigma \quad (4.12)$$

Once this mask is obtained the final inpainted image I_{np} is calculated in the following way:

$$I_{np} = M_1 \cdot I_o + (1 - M_1) \cdot I_d \quad (4.13)$$

As can be seen in Figure 4.5, I_o value of pixels under the original M mask are completely replaced by their corresponding values in the I_d . On the other hand, as we depart from the original M mask, the contribution of the original I_o values increases.

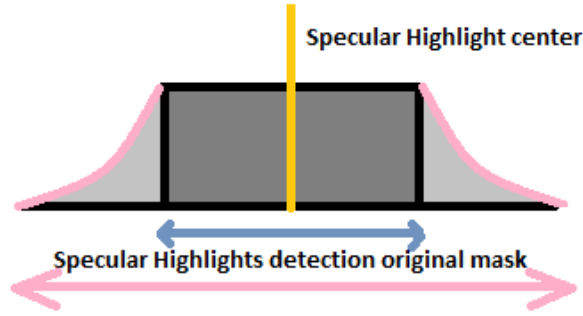


Figure 4.5: Graphical example of the extension of the M mask.

where $I_o(x, y)$ and $I_n(x, y)$ correspond, respectively, to the original image and the image where the values of the pixels belonging to pixels under the original M mask have been changed and α corresponds to the decay factor of the mask. The α factor is used to weight in the final image the contribution of the original version of the image and its smoothed version. By doing this, pixels close to the boundary of the mask will have more content of $I_n(x, y)$ image and pixels further from the mask will keep their original value.

4.4 Image preprocessing to enhance polyp boundary-originated valleys.

We cover in this section those challenges that are derivated from either artifacts generated in the image formation process or derivated from other elements of the endoluminal scene, such as specular highlights or blood vessels.

4.4.1 Image formation-derived artifacts

Color channel misalignment

As mentioned in Chapter 2, due to the way color images are created in colonoscopy videos, color channel misalignment could be present in colonoscopy video frames, specially in those where the movement of the colonoscope is abrupt (see Figure 4.6). There are methods devoted to correct color channel misalignment [31] but we have not considered the problem of misalignment as we will deal with still colonoscopy video frames which are chosen to test our model of polyp appearance. As it will be shown later in this thesis, we have selected frames that cover the widest range possible of polyp appearances and one of the constraints we have self-imposed was to provide a clear view of the polyp, which is difficult to achieve in frames from an abrupt transition. In this case color channel misalignment is out of the scope of this thesis. As color channel misalignment is directly related with strong motion which blurs the image and therefore damages its quality, we prefer to discard directly those non-informative frames (by using some of the methods detailed in Chapter 2) and continue with the processing of the informative ones.

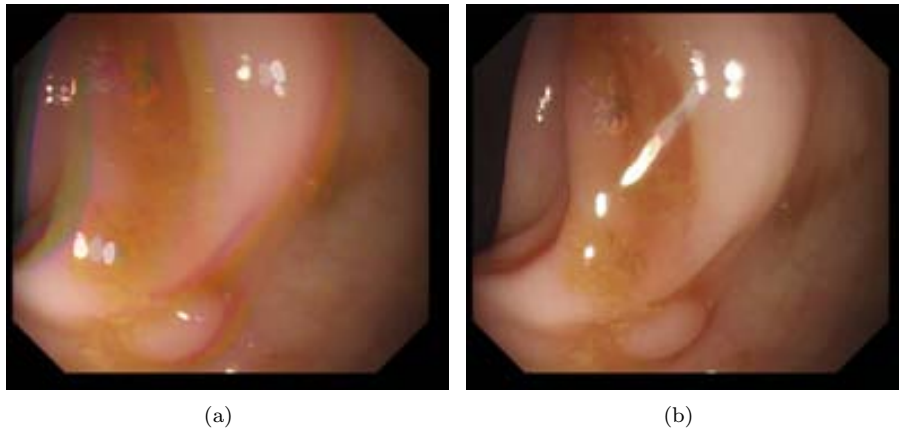


Figure 4.6: Examples of color channel misalignment: (a) Frame presence of color channel misalignment. (b) Frame without presence of color channel misalignment from the same scene.

Deinterlacing

Interlacing is also due to the image acquisition process. We have already introduced the problem of interlacing in Chapter 2 and, as it was mentioned there, because of interlacing we may find contours/edges on images caused by the minimal time difference between two consecutive frames. Although we suspect that for the majority of the frames the difference will be minimal, we have decided to address the problem of interlacing in a rather drastic way. In order to prevent the possible effects of time

misalignment between odd and even lines, we will only take one of each two lines and we will resize the image in order to keep the proportions. The process is as it follows: The original size of the image is 1148×1000 . If we take only one of each two lines, we obtain two images of size 1148×500 . We discard one of them and, in order to keep the proportion we have to pass from size 1148×500 to size 574×500 , which can be done by resizing along the horizontal direction by a 0.5 factor.

Inpainting of black mask

Due to the shape of the lens of the camera that acquires the image, a black mask surround the endoluminal scene. As our model of appearance for polyps is based on the output of a valley detector, we have to take into account that these black borders do provide a high valley response, as there is a great difference in grey level inside and outside the black border. This difference results in a valley placed in the separation between the black mask and the endoluminal scene.

There are several ways to mitigate the effect of this black mask. One alternative could be to cut the image in a way such we eliminate all the black mask (see Figure 4.7), resulting in an smaller image but keeping as much original image information as possible. The problem that this alternative present is that we may lose interesting parts of the image (even part of polyp contours) while cutting the image. Because of this, we have decided to extend the image in the area that the black mask occupies via an image inpainting method that has already been presented. We only show here some graphical examples of the inpainting method applied to the black mask in Figure 4.8.

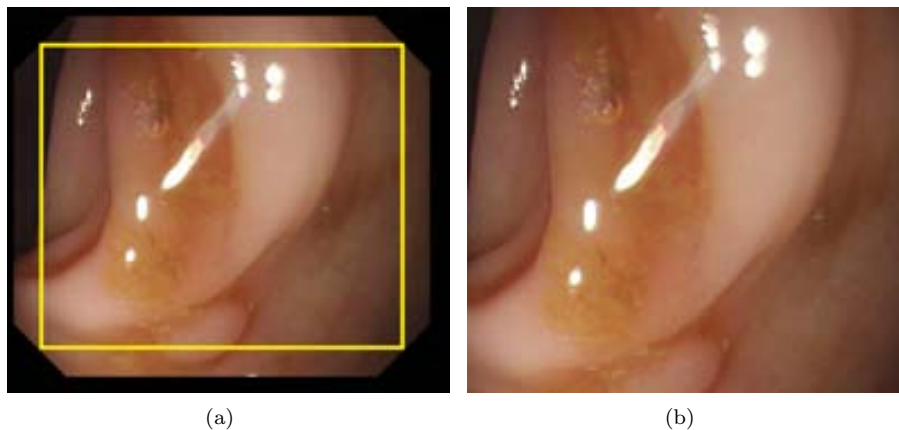


Figure 4.7: Losing part of polyp contour by eliminating all black mask content: (a) Original image (with superimposed cutting mask); (b) Cropped image.

Because of the problems that cutting the image presents,

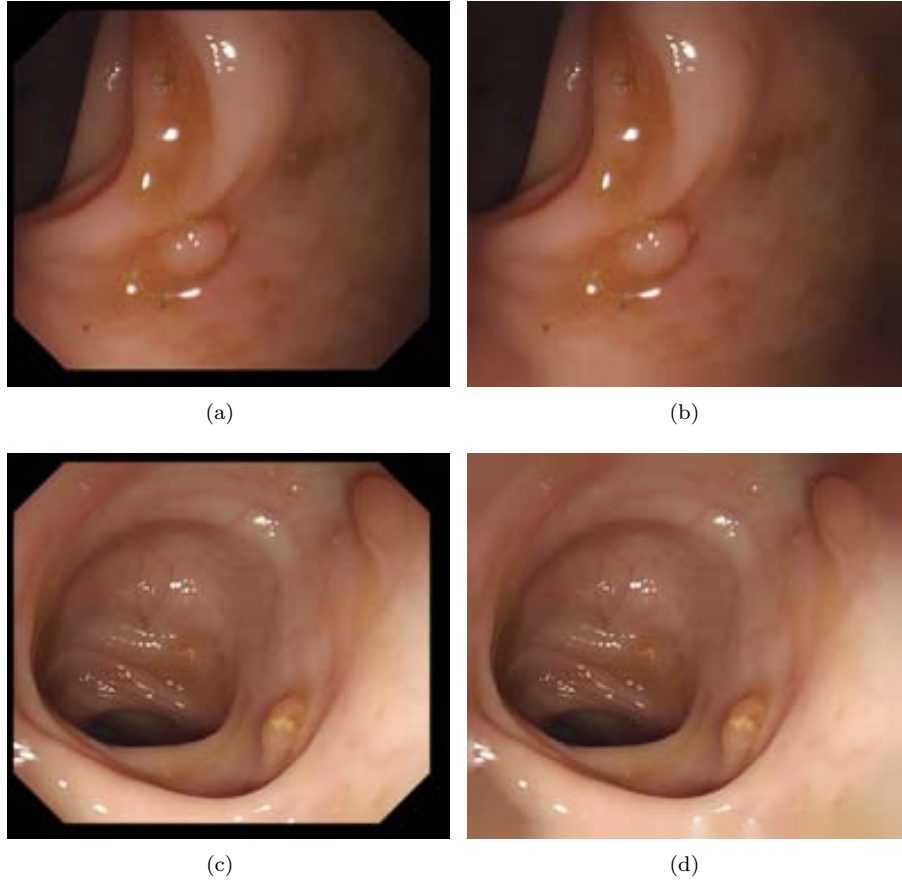


Figure 4.8: Example of black mask inpainting; (a-c) Original images; (b-d) Image with inpainting of black mask.

4.4.2 Effect of elements of the endoluminal scene

Specular highlights

As can be seen in Figure 4.9, specular highlights do have an incidence in our polyp detection scheme as their apparition produces intensity valleys around the boundaries of the specularities.

Taking into consideration that the depth of valleys image that we will use to in our polyp localization and polyp segmentation methods is directly related with the output of a valley detector and that specular highlights alter the valley image, we need to address their effect. In this case we will combine both specular highlights detection and inpainting, as our aim is to mitigate as much as possible their effect.

Specular highlights detection method

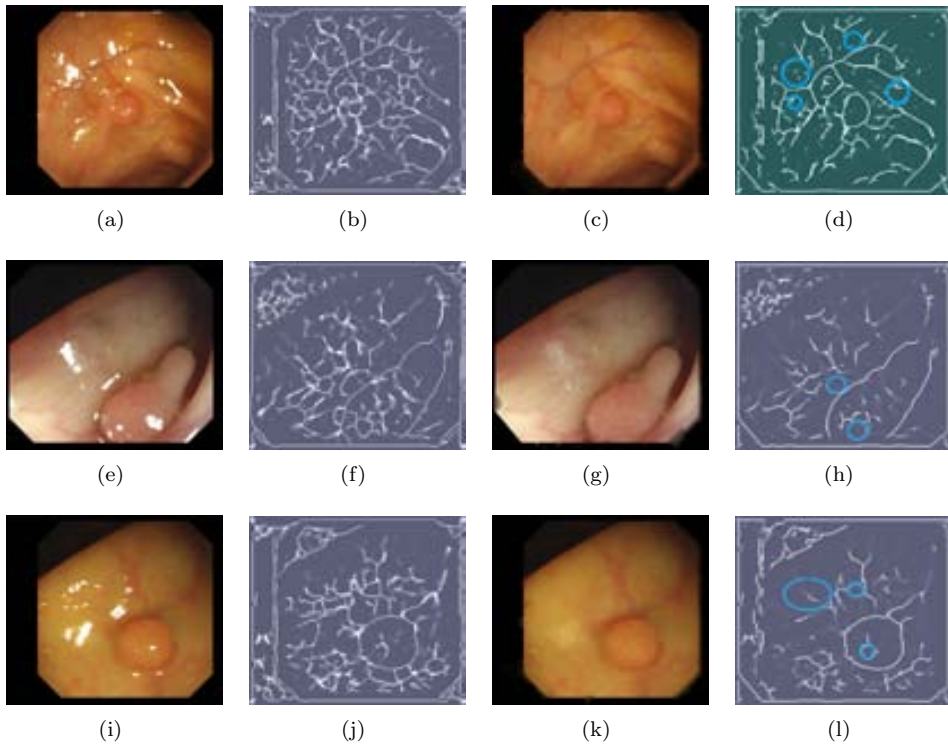


Figure 4.9: Specular highlights detection and inpainting (1st column) Original images; (2nd column) Valley image; (3rd column) Specular-free images; (4th column) Valley image (the most significant differences with the valley image calculated with specular highlights are marked with a blue circle).

We base our specular highlights detection method on the method explained by [3], which has two different modules: the first one uses color balance adaptative thresholds to determine the parts of specular highlights that present too high intensity to be part of nonspecular image content, that is, the saturated pixels on the image. The second module refines the previous specular highlight detection by including pixels nearby to saturated regions of the image that appear to be either shadows of the original artifacts or correspond to the less intense parts of the specular highlights in the image.

Our contribution to specular highlight detection aims at improving the already good results achieved after completing the second module. In our case, we are concerned on those pixels that are suspected to be part of a specular highlight but they can not be easily labelled as to be part of them due to a simple threshold value. We use the following assumption: the intensity value inside the specular highlight is higher than its surroundings. Even considering that, pixels nearby to specular highlights will continue having higher intensity values, although smaller than inside the specular highlight. We aim at finding these pixels by calculating the difference between the original image and its median but in this case we do not consider pixels we already know that they are part of the specular highlight.

By doing this, we will find which pixels or regions in the image have a intensity value marginally higher than its neighborhoods and, applying a threshold value, we can only keep those where the difference is higher. We also assume that the pixels that we will denote as specular highlights by using this method will be connected with regions that were originally labeled as specular highlights. We will compare the performance of several specular highlight detection methods in Chapter 7.

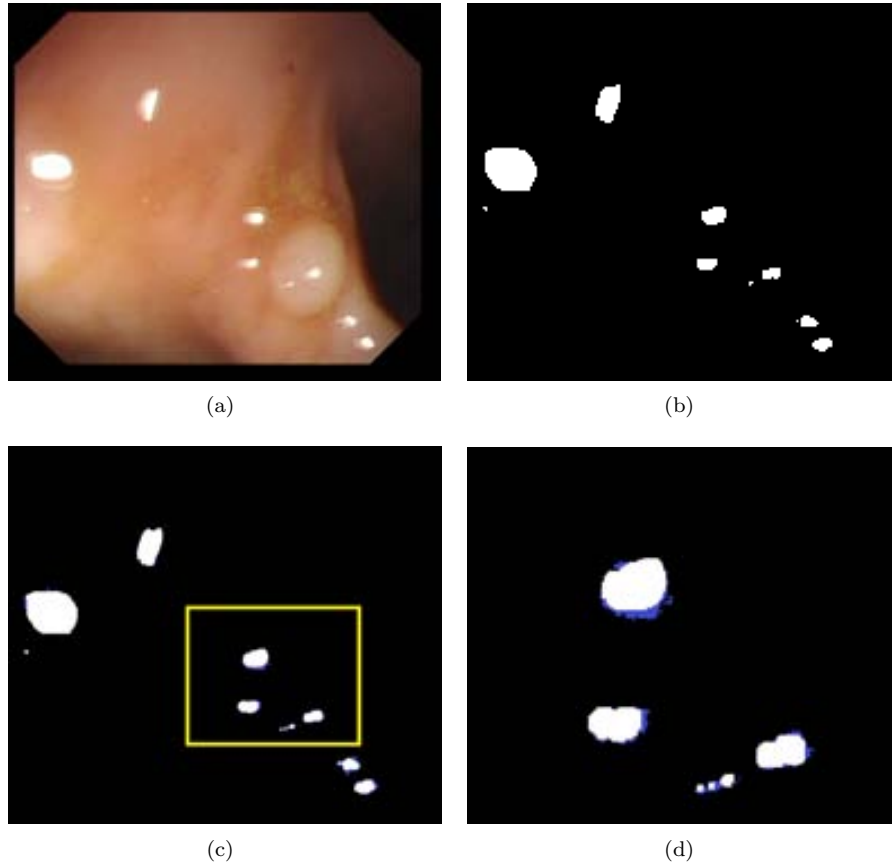


Figure 4.10: Extension of the specular highlights mask: (a) Original image; (b) Original specular highlights mask; (c) Extended specular highlights mask (differences with the original are painted in blue). (d) Zoom of the extended specular highlights.

Specular highlights inpainting method

To do the inpainting of specular highlights we have also used our inpainting method that has been explained before and used for the inpainting of the black mask. We show in Figure 4.11 some examples on specular highlight detection and inpainting.

Blood vessels mitigation

We have introduced in Chapter 2 the bibliography related to blood vessels detection but, as we have mentioned, as far as we know there is no single work with deals with blood vessels in colonoscopy images. In order to tackle the effect of blood vessels in terms of valley information we have carried out a preliminary study which will be detailed in Chapter 7.

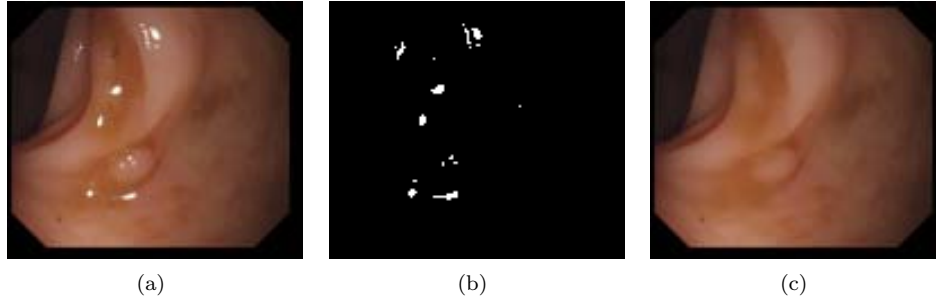


Figure 4.11: Specular highlights detection and inpainting qualitative results. (a) Original images; (b) Specular highlights detection masks; (c) Final inpainted images.

As blood vessels segmentation is a complicated task that is not in the scope of this thesis, we have opted to mitigate their effect by observing their impact on the valley image. Blood vessels do generate valleys that could be confused with polyp boundaries therefore their mitigation could be crucial for our localization and segmentation algorithms. The objective of this experiment is to check if blood vessels-generated valleys have different presence depending on the color channel that we use. In our case at this moment we do not aim as far as to detect accurately blood vessels but to mitigate their response in the valley image. The preliminary study indicates that the contrast between blood vessels and the rest of the image is higher in channel G of the RGB image (Figure 4.12 (c)) and smaller for channel B (Figure 4.12 (d)) so if we are interested in mitigating their contribution we should use channel B.

We will develop more on how blood vessel information can be mitigated according to several cues (color channel, type of information) in Chapter 7 but we show in Figure 4.13 how, using the same example than in the figure above, we observe some differences in the valley images generated by each of the channels of the RGB image. We can see that the valley image for the green channel seems to provide higher contrast for blood vessels valleys than other channels.

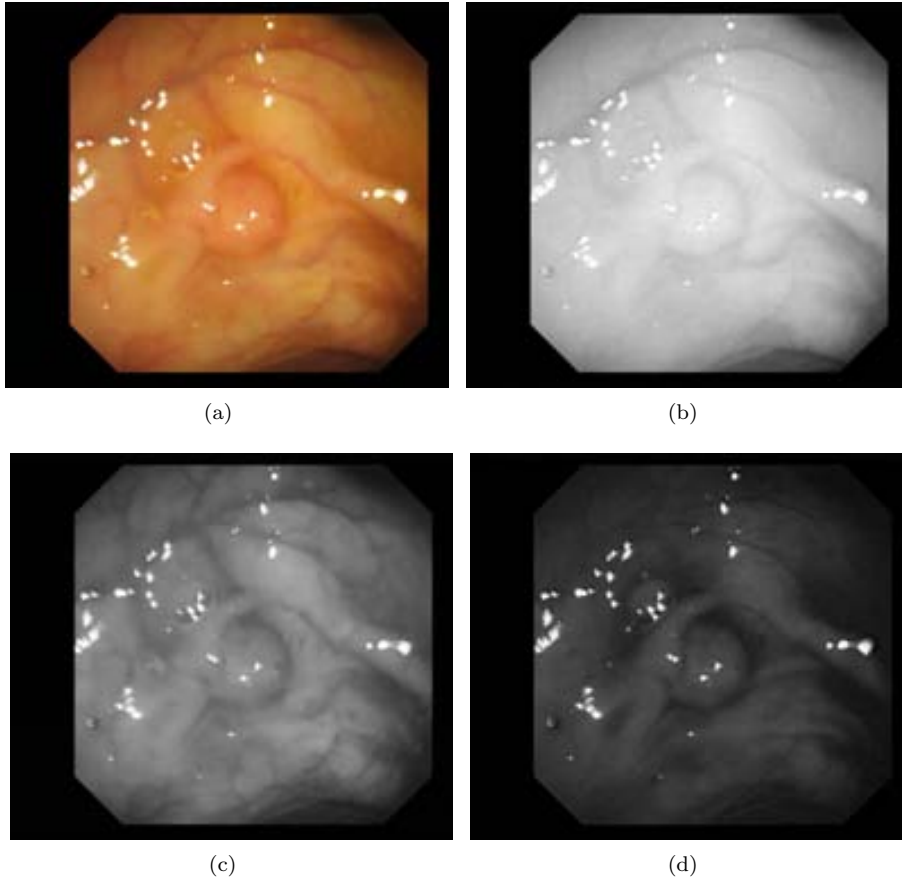


Figure 4.12: Difference in contrast of blood vessels according to the channel observed: (a) Original image; (b) R channel (c) G channel; (d) B channel.

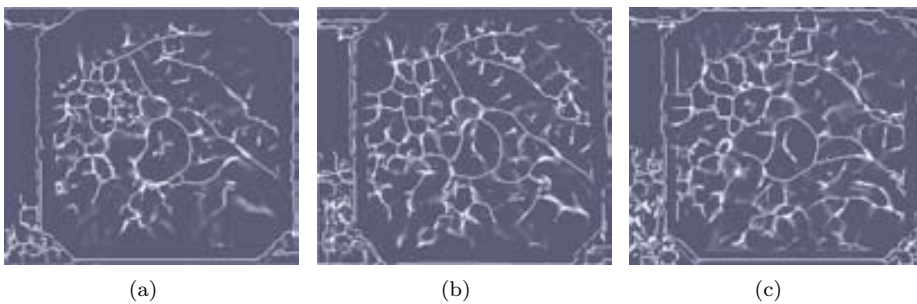


Figure 4.13: Valley image obtained from separate channels of the RGB image. (a) R channel valley image; (b) G channel valley image (c) B channel valley image.

4.5 Depth of Valleys image as a method to capture the boundaries of polyps

As mentioned before, by using only valley information we may lose some polyp boundaries. In order to solve the potential problem that lateral views of polyps generate and also to provide with more meaningful metrics, we have extended our initial approximation into the novel Depth of Valleys image (DoV image).

The objectives behind the creation of the DoV image are:

- Provide strong polyp boundaries by means of valley detection, following the model of appearance for polyps.
- As the output of *multilocal creaseness* detector is good in terms of valley localization but not in terms of valley intensity, incorporate another measure in order to define how deep a point is inside a valley.

The idea behind the DoV is to complement the information that a valley detector provides with the morphological gradient in order to achieve a method that enhances both valley and contour information (which appear in some problematic cases such as lateral polyps, as (Figure 3.7 (b)). The rationale of this approach is that in certain type of views -in general, in lateral views- we do not have a whole valley surrounding the polyp, but still non-connected edges are available in these cases. By combining valley and edge/gradient information we are able to keep all the contour pixels that we need in order to define the polyp boundary. We also have to stress that the output of the *multilocal creaseness* valley detector informs us about where valleys are in the image but not about their depth inside the valley. We use morphological gradient because as its performance is more useful for our case than normal gradient's. More precisely, as can be seen in Figure 4.14, by means of morphological gradient not only we obtain values for the case of abrupt contours in lateral views but also for those parts of the image where there is valley information.

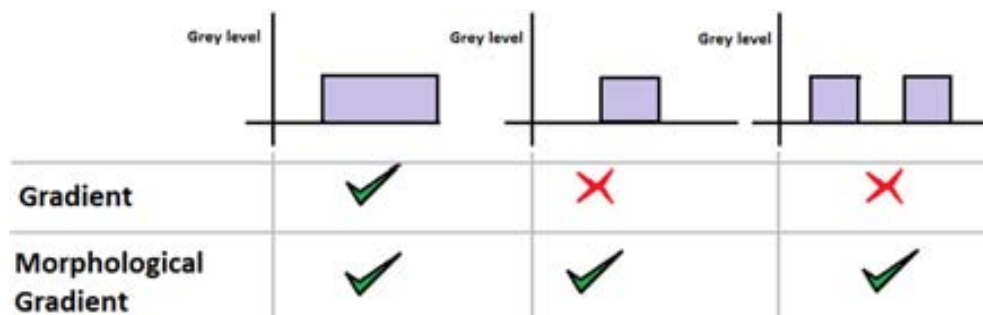


Figure 4.14: Difference in performance between gradient and morphological gradient. The image shows three different grey level profiles that should be taken into account.

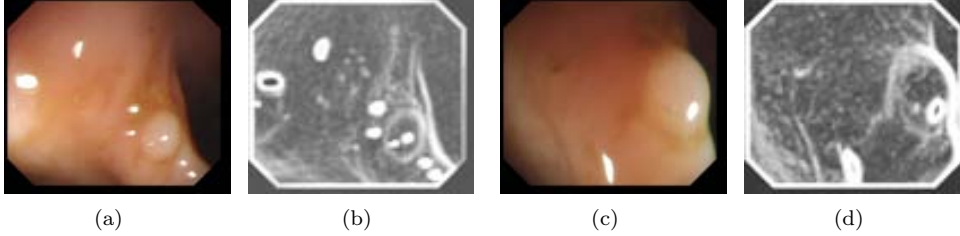


Figure 4.15: Morphological gradient in colonoscopy frames: (a) and (c) Original image; (b) and (d) Morphological gradient image

The morphological gradient is defined as the difference between the dilation and the erosion of a given image [96] and it gives as output an image where each pixel value indicates the contrast in intensity in the close neighborhood of that pixel. As depicted in Figure 4.15 by using the morphological gradient we can also solve some of the issues of our initial valley model such as the absence of valley information in some part of the contours of lateral polyps (see Figure 4.2 (f) and Figure 4.15 (d)).

Therefore if we combine valley information, which gives us information on which are the pixels that constitute the boundaries of the pixel, and morphological gradient information, which helps by weighting the intensity of the valley information along with completing the contours for the case of lateral polyps, we get as result the DoV image. As shown in Figure 4.16, the depth of valleys image is achieved through the pixel-wise multiplication of the ridges/valleys detector and morphological gradient. In the points where we have a valley (marked as a green line) and the morphological gradient value is high the result is high value of the depth of valleys image. Conversely, in the points where there exist a valley but the morphological gradient is low (or vice versa) the depth of valley will not result in a maximal value. The orange line surrounds an area with no local features, and the yellow circles surround the frontal reflections with a typical saturation pattern. The mathematical representation for this definition of the depth of valleys image is as follows:

$$DV = V(\sigma_d, \sigma_i) \cdot MG = V(\sigma_d, \sigma_i) \cdot (I \oplus C_{\sigma_i} - I \ominus C_{\sigma_i}) \quad (4.14)$$

where DV stands for the depth of valleys image, V for the output of the ridges and valleys detector, I for the original input image and C for structural element used in the dilation and erosion morphological operations that constitute the morphological gradient MG calculation, both normalized to unit. Formulated in this way, the depth of valleys has higher values in the points that constitute the relevant valleys of the image and lower to points inside the valley.

It must be noticed that in order to calculate the morphological gradient, the ridge and valley extractor needs two parameters to be set in correspondence to the size of the structural element (sd) [67]. These parameters are the differentiation scale σ_d and the integration scale σ_i . In our case, the structural element is a disk. More precisely, σ_i should span the of the same size as sd in order to work in the same scale. If this

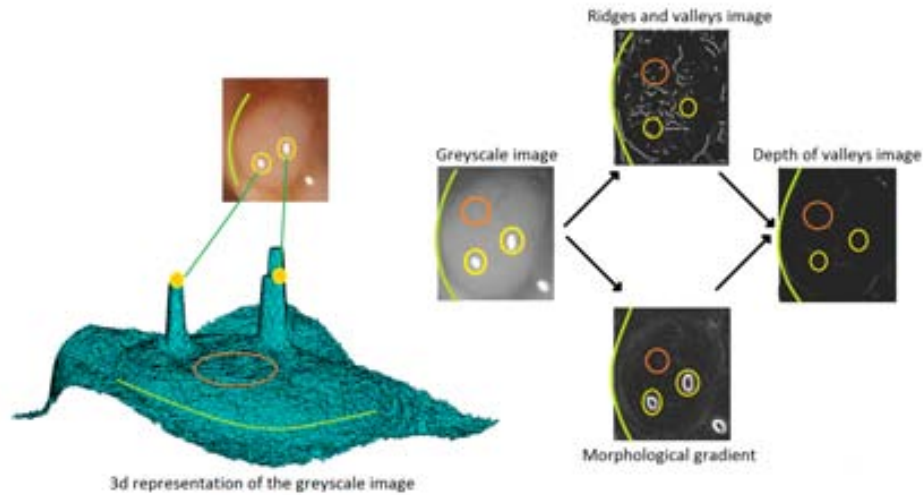


Figure 4.16: Example of the calculation of a depth of valleys image.

does not happen, maximal points of the ridge and valley extractor could be located in places where the morphological gradient is not maximal, and therefore the desirable properties of the resulting depth of valleys image would be lost.

We designed the DoV image in order to keep with high value those pixels which constitute the boundaries of the polyp (shades in the original image) while leaving with near zero value those pixels which are not part of a protruding object. This DoV image will be the input image that we will use to guide our polyp localization and polyp segmentation methods, which will be detailed in the next chapters. We also have to mention that as DoV image is composed by the output of a valley detector, all the improvements that we are detailing in this section (black mask, specular highlights or blood vessels) will also affect the final DoV image.

4.6 Discussion and summary of the chapter

The objective of this chapter was to show how the output of the valley detector information, considering some of the challenges that we introduced in the previous chapter, can be potentially useful to guide further polyp localization and segmentation methods. As it has been mentioned, polyps are not the only elements of the endoluminal scene that produce valley information and there are some of them which may have a high impact on the valleys image. Furthermore, there are some undesired results on the valley image that are caused by means of how the colonoscopy frames are acquired, such the superimposition of the black mask around the endoluminal scene or interlacing.

We show a complete scheme of all the preprocessing and valley detection operations

that we apply to the input image in Figure 4.17.

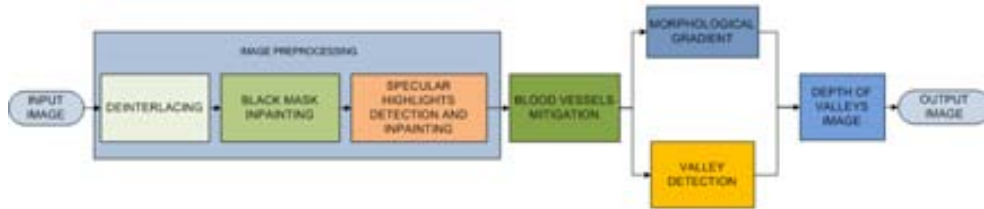


Figure 4.17: Image preprocessing and valley detection processing scheme.

First of all we have to solve the effects of interlacing which is achieved by selecting one of each two lines and resampling. Black masks appear in the image due to the lens of the shape. In this case we opt for inpainting them because a simple cut will result in a loss of information. Another direct effect of the image formation process are specular highlights which also generate response in the valleys image. In order to mitigate their effect we also perform an inpainting method after a specular highlights detection. We also address the problem of blood vessels mitigation by exploring their impact according to the color channel that we use.

Finally, in order to cope with incompleteness of boundaries due to lateral views of polyps we extend our valley image to the novel Depth of Valleys image, which combines the information of the valley detector with the one provided by the morphological gradient. This combination not only helps at recovering some parts of the boundaries in lateral views that are lost by means of valleys but also provides with a better intensity metric which reflects how deep a given pixel is inside a valley.

Although we will show in Chapter 7 a full experimental analysis on the influence of the endoluminal scene objects that we have sketched, we present in Figure 4.18 to end this chapter a qualitative comparison of the valley image before and after addressing the mentioned challenges.

We can see in Figure 4.18 how the most significant improvement on the Depth of Valleys image is the elimination of the valleys generated by the borders and specular highlights. We have marked some of the suppressed valleys by circling them by a green shape. As it can be seen, we eliminate some valleys in the image but in the next chapter we will develop more on how to use this valley information in our polyp localization and segmentation algorithms.

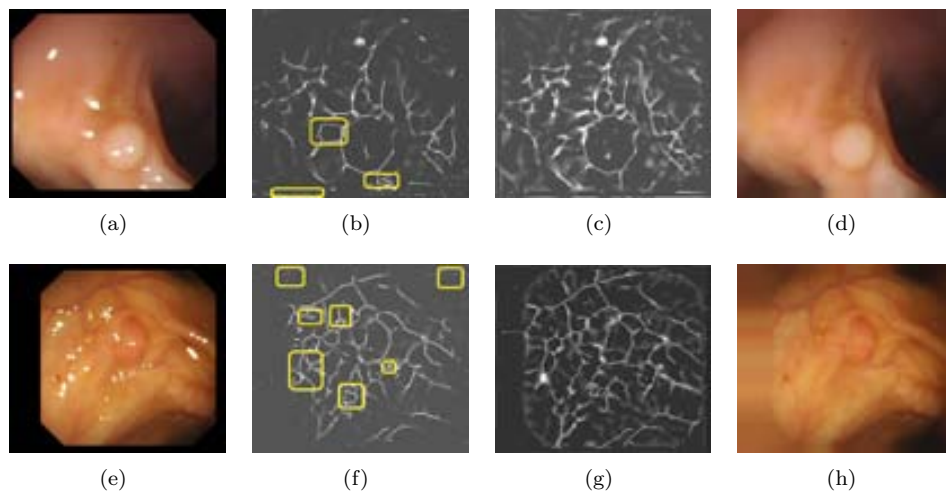


Figure 4.18: Improvement on the valley image. (First column) Original images; (Second column) Depth of Valleys image before addressing challenges: 1) Black mask; 2) Specular highlights detection and inpainting; 3) Blood vessels mitigation; (Third column) Depth of Valleys image after addressing challenges; (Fourth column) Enhanced input image;

78 VALLEY DETECTION FOR POLYP LOCALIZATION AND SEGMENTATION

Chapter 5

Polyp localization

5.1 Introduction

In this chapter we will present algorithms for polyp localization, which general processing scheme can be seen in Figure 5.1.

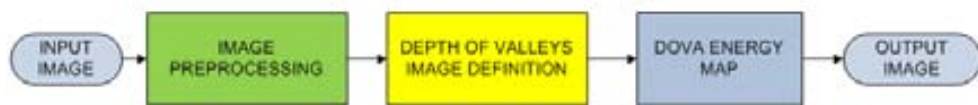


Figure 5.1: Polyp localization processing scheme.

Polyp localization algorithm consists of several stages, namely: 1) Image preprocessing, which consists of image deinterlacing, inpainting of the black mask and specular highlights detection and inpainting; 2) Valley detection; 3) Development of depth of valleys image (which includes blood vessels mitigation and the use of morphological gradient) and 4) DOVA energy map calculation. The first three stages have been explained in the previous chapter and we will dedicate this to the definition and calculation of DOVA energy maps, which accumulate the information that the DoV image provides.

The general idea that we will use for polyp localization is, considering that polyp boundaries appear as intensity valleys in the image due to the illumination process, to define whether a pixel is interior or exterior to an object based on either the shape that the boundary that surrounds it or the completeness of the cited boundary. As it may seem logical, we need to have a Depth of Valleys (DoV) image as clean as possible from undesired contributions from other elements of the endoluminal scene, which is performed via stages 1 to 3. Once this is achieved we can take the next step and start to develop our polyp localization method.

In order to define which pixels are inside and outside an object we still can use

the information that the DoV image provides by measuring in how many different directions a given pixel is surrounded by high DoV image values. We can see in Figure 5.2 a synthetic example where the pixel represented by the red square is surrounded by a high number of directions, creating a boundary constituted by pixels with high value of the depth of valley image. On the other hand we can see that the pixel represented by the blue square is surrounded only by a few pixel values with high value of the DoV image. It is clear that there are some other factors to consider, such as the distance to a given point to a boundary or the shape of the objects that we want to detect.



Figure 5.2: Definition of interior and exterior points to objects.

We may take two different approaches: a) Boundary-shape approaches, where the objective is to fit boundaries to previously-known shapes that represent a polyp and b) Boundary-completion approach, where we define points interior to objects according to how complete are the boundaries that surround them.

Both approaches present at a glance their own pros and cons. For instance, we have developed a method that fits the boundary to a generic shape of a polyp (such as ellipses) but then we process a video where the polyp that appears never fits a general shape so the method fails. Furthermore, we may define erroneously a point as interior to an object which is not a polyp due to the real polyp on the image has less complete boundaries than the false positive.

We have developed 4 different algorithms for polyp localization, namely: 1) EF-DOVA (Ellipse Fitting Depth of Valleys Accumulation), which aims at fitting ellipses in the DoV image; 2) SA-DOVA (Sector Accumulation Depth of Valleys Accumulation), which defines a metric based on accumulating the maxima of DoV in a series of sectors around a given point; 3) WM-DOVA (Windows Median Depth of Valleys Accumulation), which extends SA-DOVA in order to look for more complete boundaries and 4) VO-DOVA (Valley Orientation Depth of Valleys Accumulation), which pays attention to the orientation of the valleys that constitute the boundaries of the objects in the image. We will detail the four algorithms, ordered in increasing level of complexity, along the following sections of this chapter.

5.2 Depth of Valleys accumulation based on the shape of polyp boundaries

5.2.1 Ellipse Fitting Depth of Valleys Accumulation (EF-DOVA)

On Chapter 2 we studied several shape-based polyp detection methods. We subdivided the existing methods in two different categories, namely: detection by curvature analysis and detection by ellipse-fitting. Methods belonging to this last group aim at fitting previously detected contours by means of ellipses. Following this same line of research, we have explored fitting the points with high value on our depth of valley image to ellipses, which should be constituted by protruding object contours.

Therefore, our Ellipse-Fitting Depth of Valleys Accumulation Algorithm (EF-DOVA) algorithm belongs to the boundary-shape fitting approaches. EF-DOVA relies on a general assumption that polyps tend to have elliptical shapes in order to define as polyp boundaries those that fit the equation of the ellipse. The algorithm that calculates the likelihood of each point being the center of a polyp (measured by the number of boundaries points that coincide with the ellipse equation) consist of five different steps:

Algorithm 2: EF-DOVA algorithm

Data: EF-DOVA($DoV, sp_{flag}, DoV_{th}, d_b, tol, Acc_{th}$)

Arg: (DoV : Input depth of valleys image, DOV_{th} : Depth of valleys threshold, d_b : distance threshold, tol : tolerance threshold, Acc_{th} : accumulation threshold)

Result: Accumulation image(Acc)

- 1 Definition of starting points by means of the starting point flag;
 - 2 Eliminate the contribution of pixel with DoV value less or equal than DoV_{th} ;
 - for** *Each starting point* **do**
 - 3 Obtain the positions of the maxima of DoV image (distance from starting point should be higher than d_b) to calculate major axis' position;
 - 4 Calculate the minor axis of the ellipse;
 - 5 Calculation of the ellipse points following equation 5.1;
 - for** *Each ellipse point* **do**
 - 6 Definition of a search neighborhood of size 3×3 centred on the ellipse point;
 - 7 In order to define the ellipse point as boundary point check if the closer maxima of DoV image is at distances less or equal than tol from it;
 - 8 To eliminate contribution of noise check if at least $Acc_{th}\%$ of the neighborhood points are above the DoV_{th} threshold;
 - end**
 - end**
-

As can be seen EF-DOVA needs of 5 different parameters, namely: 1) Starting point flag (which defines if the starting point is any pixel on the image or the min-

ima of the creaseness image); 2) Depth of Valleys threshold (DoV_{th}) which aims at eliminating the contribution of pixel with low DoV value; 3) Minimum distance from the starting point to the approximated ellipse (d_b); 4) Tolerance in distance from the calculated ellipse points to pixels with higher-than-threshold DoV value (tol) and 5) Percentage of pixels in the neighborhood around the ellipse point that have a higher-than-threshold DoV value (Ac_{th}).

The algorithm starts by calculating the starting point and, for each of them, calculates the distance to the closest maxima of the DoV image, which will constitute the major axis. The minor axis is calculated by rotating the position of the major axis 90° . Once the axis are defined, the equation of the ellipse can be defined by means of Eq 7.3. From this equation we can calculate all the points that we want from the ellipse although for our experiments, like is shown in Figure 5.3, we have considered 8 directions that go from $\Theta = 0^\circ$ to $\Theta = 360^\circ$ separated by 45° each point from another. In order to calculate the position of the rest of the points we have to equal (see eq. 7.3) the equation of our approximated ellipse (eq. 5.1) at the equation of the line with slope the tangent of each angle (eq. 5.2). In this case we will consider that the ellipse may not be placed above the x-y axis but it can be rotated by a ϕ angle.

$$x^2 \left(\frac{\cos^2 \varphi}{a^2} + \frac{\sin^2 \varphi}{b^2} \right) + y^2 \left(\frac{\sin^2 \varphi}{a^2} + \frac{\cos^2 \varphi}{b^2} \right) + xy (\sin 2\varphi) \left(\frac{1}{a^2} - \frac{1}{b^2} \right) = 1 \quad (5.1)$$

$$y = \tan \Theta \cdot x \quad (5.2)$$

$$x = \frac{1}{\sqrt{\left(\frac{1 + \cos 2\varphi}{2} \right) \left(\frac{1}{a^2} + \frac{1}{b^2} (\tan \theta)^2 \right) + \left(\frac{1 + (\tan \theta)^2}{b^2} - \frac{1}{a^2} \right) + \sin 2\varphi \tan \theta \left(\frac{1}{a^2} - \frac{1}{b^2} \right)}}; \quad (5.3)$$

Once we have all the ellipse points calculated we start checking if in a neighborhood around each of them we find a maxima of the DoV image in a distance less or equal than tol from the ellipse point and if a given Ac_{th} percentage of the neighborhood pixels are above the DoV_{th} threshold value. If both conditions are fulfilled we increment the accumulation value for that pixel in 1, as it can be seen in Figure 5.3, where we show how EF-DOVA would work on an ideal binary depth of valleys image. Accumulation values go from 2 (the approximated ellipse only touches the depth of valleys image only in minor and major axis) to 8. In our experiments we will denote as high accumulation values those higher or equal than 5. As it was shown in Figure 4.16 abrupt changes in texture may result in artificial high values in the depth of valleys image so, to avoid this, we consider a distance factor that gives less weight to accumulation in points of the image very close to maxima of the depth of valleys image.

In Figure 5.4 we can see how our algorithm works on synthetic images that resemble to some of the regular and irregular shapes that we may find (with both closed and unclosed contours). As we can see EF-DOVA places the maxima of accumulation (represented by a circle) near of the center of the structures and, what is more important, does not place any maxima inside non-closed structures such as the one on the

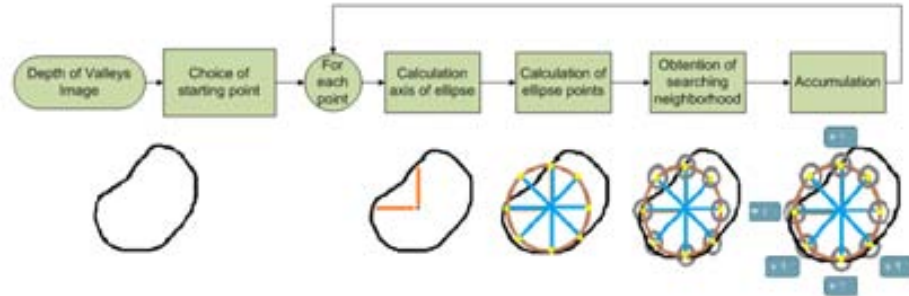


Figure 5.3: Graphical explanation of the depth of valleys accumulation algorithm.

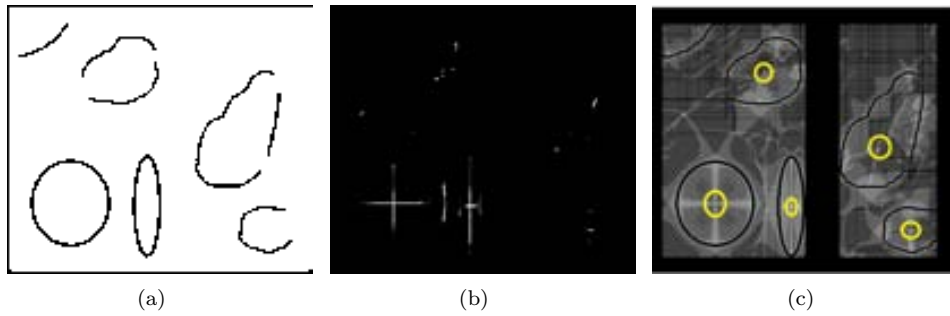


Figure 5.4: EF-DOVA algorithm on synthetic images: (a) Original image (b) Accumulation image (threshold 5). (c) Original image with accumulation values superimposed. Global maximas of accumulation are surrounded by a yellow circle.

upper left corner of the image. The energy DOVA image will be directly composed by the accumulation values for each pixel and its range of values will depend on how many directions we are considering. As we have already mentioned, we consider a distance factor to accumulate less those points close to high values of the DoV image but we have also included tolerance values that measure how far is a given ellipse point to its closer boundary. By using this tolerance value we can obtain those intermediate accumulation values therefore achieving better maximas points in those centers of ellipse that fits perfectly an ellipse.

We show some preliminary EF-DOVA results in Figure 5.5. In this case we have superimposed in the original image the polyp mask via a blue circle. In the output image, apart from the polyp mask, we have painted in green those pixels with high accumulation value (higher than 5) that are inside the polyp mask (True Positives (TP)) and in red those which are outside the polyp mask (False Positives (FP)). As it can be seen EF-DOVA performs better when the view of the polyp is zenithal (see Figure 5.5 (c) and (d)) and the approximated polyp contour has an elliptical shape although the number of pixels with high accumulation value outside the polyp mask

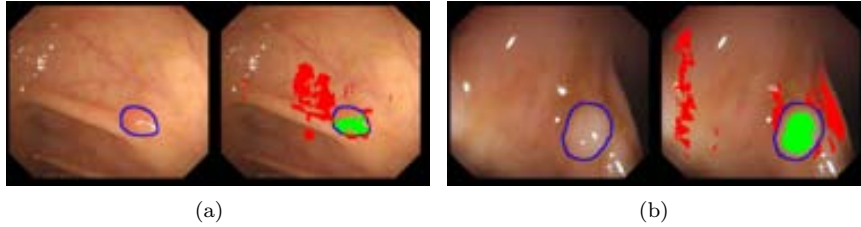


Figure 5.5: Examples of EF-DOVA energy maps superimposed to the original image. Green points represent TP whereas red points represent FP.

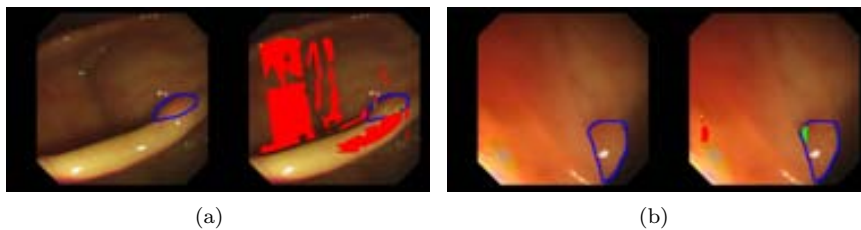


Figure 5.6: Examples of bad functioning of EF-DOVA energy maps. Green points represent TP whereas red points represent FP.

is relatively higher.

We can see from Figure 5.4 that EF-DOVA works good when the polyp that we have in the image has an elliptical shape but, as we can see in Figure 5.6, when the polyp's shape departs from elliptical, EF-DOVA performance is damaged. We can see from Figure 5.6 (a) that when we have a lateral view of a polyp, the lack of one half of the contour results on a bad accumulation by means of EF-DOVA. And, even for cases where we have a complete boundary (see Figure 5.6 (b)), as the polyp cannot be approximated by an ellipse the accumulation process fails.

As it has been shown EF-DOVA performs well in cases where the view that we have of the polyp let it be approximated by an ellipse but as we diverge from elliptical shapes its performance is damaged. Considering this we proceed to explain in the next chapter different accumulation methods which are not based or focused on concrete polyp shapes but aim at accumulating based on the completeness and continuity of contours.

5.3 Depth of Valleys accumulation based on the continuity of polyp boundaries

5.3.1 Sector Accumulation Depth of Valleys Accumulation (SA-DOVA)

The method presented in the previous section aims at fitting boundary pixels obtained by means to the depth of valleys image to ellipses. Although EF-DOVA could work perfectly for a certain group of polyp views, we can not guarantee that the polyp is fully surrounded by strong valleys in all directions due the different points of view of the colonoscope and the potential presence of occlusions. In this case we will change the way we calculate our accumulation image by taking into account the completeness of the boundaries in the following way. We will measure the completeness of a boundary by counting in how many directions (expressed as angles from 0° to 360°) we find a high value of the depth of valleys image. We also take into account that pixels tend to that constitute the boundary of a polyp should be in the same range of distances from the polyp's center, we define a series of radial sectors. These radial sectors, centred on a given point, are used to accumulate the maxima of the DoV image under the sector's area of influence. Therefore we calculate our accumulation operator as:

$$Acc(x) = \int_{\alpha=0}^{\alpha=2\pi} \max_{r \in [R_{min}, R_{max}]} DV(x + r * (\cos(\alpha), \sin(\alpha))) d\alpha \quad (5.4)$$

where DV is the Depth of Valleys image, x represents the coordinates of pixel in the image, R_{min} and R_{max} define the area in which the valleys are searched, and α allows the angular integration of the maximum values in all directions.

Acc will be large when the structure present strong valleys in all directions. However, the presence of multiple valleys in the same angle will not affect the value of Acc . Because of that, weak valleys coming from thin vessels or wrinkles will not be computed at a given angular orientation, and only the strong valleys associated to the polyp would tend to provide high outputs, as can be seen in the graphical example shown in Figure 5.7. We can see in Figure 5.7 (b) how by using our new accumulation method we enhance the accumulation output for structures such polyps, which should provide in their boundaries with high values in the DoV image whereas the output for also closed structures such as blood vessels is smaller, due to the difference in the DoV value that each shape's boundaries convey.

It is important to notice that the ring structure defined by R_{min} and R_{max} provides invariance to rotation. Defined in this way, this approach will perform well in circular or elliptical patterns, but the Acc operator will not be linked to any particular shape, as long as a substantial amount of (strong) valleys reside under the ring. This property makes Acc robust to occlusions and potential lack of information for a number of directions. Acc can be digitally implemented in an efficient way as the sum of the

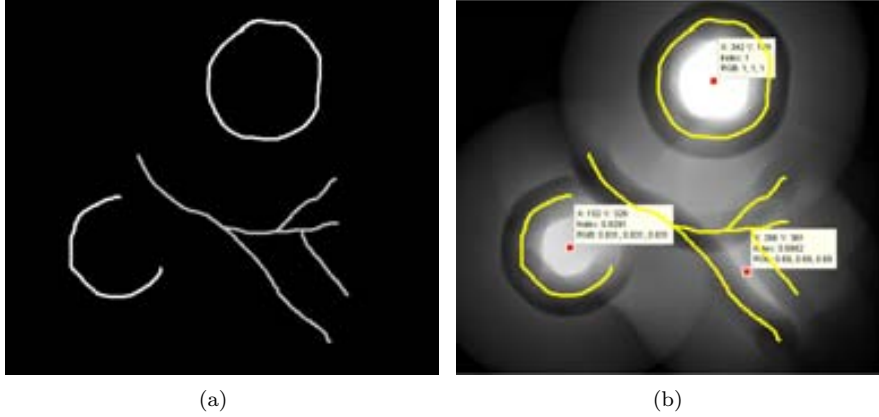


Figure 5.7: Graphical explanation of SA-DOVA accumulation: (a) Input depth of valleys image with two polyp boundaries and blood vessels; (b) SA-DOVA accumulation map. The position of the maxima of accumulation is painted as a red square. The original shapes are painted in yellow.

valley maxima that are found in the sectors by following the method that will be described next.

SA-DOVA needs of 3 different parameters which are related to the radial sectors that we use to accumulate the maxima of the DoV image. The parameters are: 1) Minimum radii of the sectors (r_{min}); 2) Maximum radii of the sectors (r_{max}) and 3) Number of sectors ($n_{sectors}$). Conversely to EF-DOVA, we do not have to define which points will be the seed of the algorithm as the current implementation let us do the whole accumulation process at once.

In order to accumulate values from the DoV image, an array of sectors centered in each pixel position is built taking into account the parameters mentioned before. The circular shape of the detector provides invariance to rotation, which is mandatory in the case of polyp detection. The number of sectors is linked to the definition of structures to be detected, and minimum and maximum radii are associated to an implicit scale factor in which the detector provides optimal results. For each sector of the array we will accumulate, in each pixel position, the value of the maxima of the DoV image that falls under it. This process can be performed in an efficient way for all the pixels in the image by approximating the acquisition of the maxima to a dilation operation, using each sector as the structural element, and then adding up all the contributions from each partial dilation.

The way this algorithm works can be better understood by following a graphical example, as shown in Figure 5.8. We start with the original image and we calculate the DoV image (see Figure 5.8 (b)) as it was mentioned in Chapter 4. Once this is done, we start with the accumulation method. The sectors that guide the accumulation process are directly applied to calculate the accumulation image. In Figure 5.8 (c) we can see in yellow the area that each different sector covers (in this case, in order to

enhance the understanding of the algorithm, we only plotted one third of the sectors). We can also see how, for each sector, there are some pixels that are painted in blue. This represents the value of the DoV image that will be accumulated for the given pixel. We can see in Figure 5.8 (d) our resulting accumulation image, where brighter areas correspond to higher values in the accumulation image. The final DOVA energy map will correspond to the accumulation image calculated in this way.

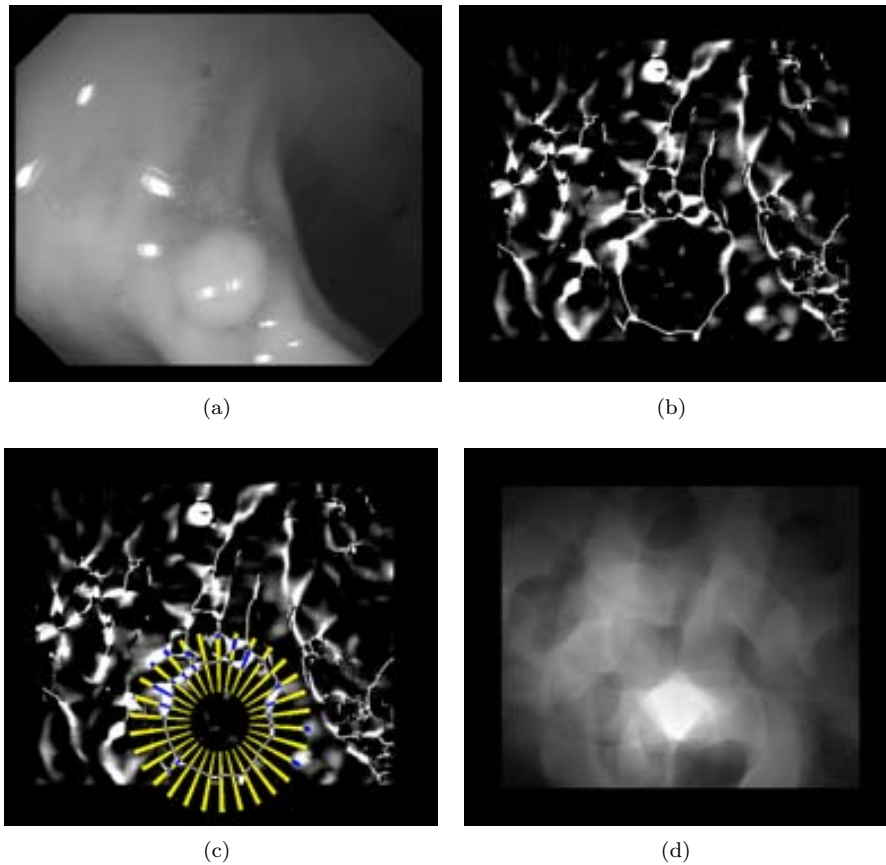


Figure 5.8: Explanation of SA-DOVA algorithm: (a) Original image; (b) depth of valleys image; (c) sector accumulation, and (d) final accumulated image.

We show some preliminary SA-DOVA results in Figure 5.9. Following the same criteria than with EF-DOVA, we have superimposed in the original image the polyp mask via a blue circle. In the output image, apart from the polyp mask, we have painted in green those pixels with high accumulation value that are inside the polyp mask and in red those which are outside the polyp mask, As it can be seen SA-DOVA performs well for several types of polyp shapes, giving a smaller number of maxima of accumulation outside the polyp.

Although we have shown how by means of SA-DOVA we can cover more general

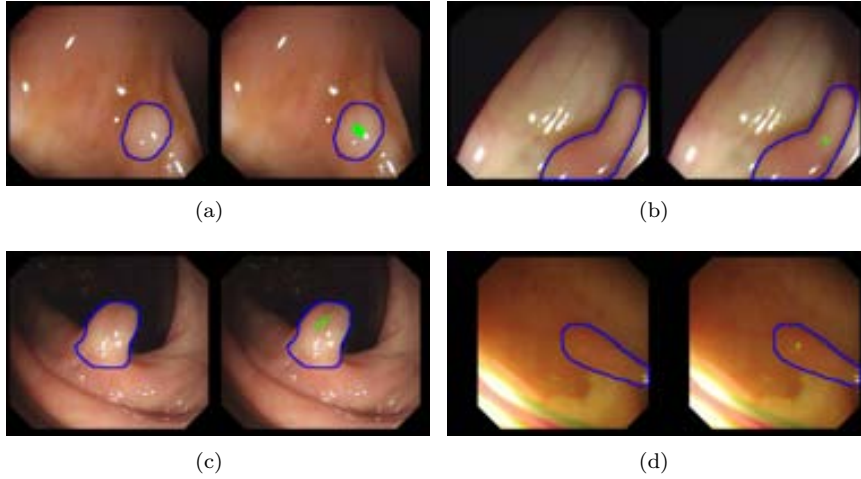


Figure 5.9: Examples of SA-DOVA accumulation superimposed to the original image. Maxima of accumulation inside the polyp are painted in green (TP) and maxima outside the polyp are painted in red (FP).

shapes, the method still has some drawbacks: 1) SA-DOVA does not differentiate (in terms of maxima of accumulation) between a boundary composed by medium values of DoV image and noise generated by an isolated pixel with very high value in the DoV image (see Figure 5.10 (a) and (b)); 2) SA-DOVA does not take into account the completeness and continuity of the boundaries, giving a similar output in terms of maxima of accumulation in examples like the one shown in Figure 5.10 (c) and (d). In order to overcome this drawbacks we propose in the next section our improve WM-DOVA algorithm.

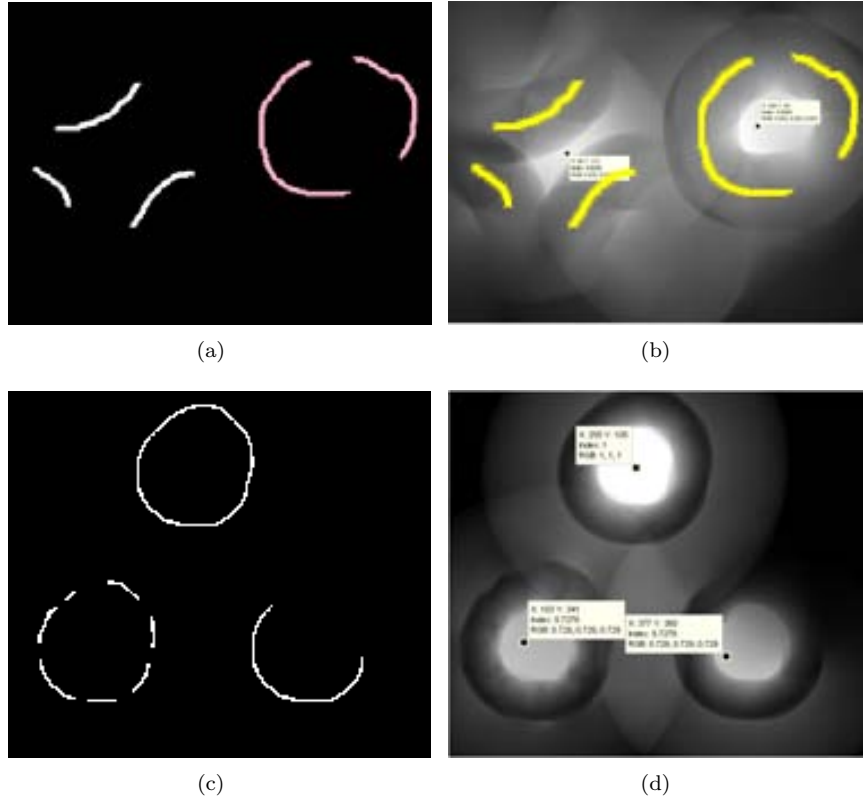


Figure 5.10: Examples of bad performance of SA-DOVA energy maps. (a) and (b): Effect of isolated pixels of high DoV value; (c) and (d): Completeness and continuity of boundaries.

5.3.2 Windows Median Depth of Valleys Accumulation (WM-DOVA)

Median Depth of Valleys Accumulation

In order to obtain the final accumulation image until now we have simply added the maxima of DoV image below each sector but this way of accumulating may lead to some errors like the ones we show in Figure 5.10. By using a sum-based accumulation method there is no difference in terms of maxima of accumulation between the two shapes that we show in the image. In this case, having few pixels with a very high value of the DoV image equals in terms of maxima of accumulation the output of another profile where we have a more complete boundary constituted by pixels with smaller value of the DoV image. One possible solution to this will be to use instead of the sum of the maxima of DoV image the median. In this case we punish those boundaries constituted by few pixels with high value.

In this case, we will calculate the accumulation as:

$$MaxList(x, \alpha) = \max_{r \in [R_{min}, R_{max}]} DV(x + r * (\cos(\alpha), \sin(\alpha))) \quad \alpha \in [0, 2\pi] \quad (5.5)$$

$$Acc(x) = Median(MaxList(x, :)) \quad (5.6)$$

We show in Figure 5.11 a comparison between the results obtained by SA-DOVA and median-based accumulation. We can see that by using median-based accumulation the accumulation output differs from the one obtained by using the sum and by means of this new accumulation the difference of maxima of accumulation inside the two shapes is higher than in the previous case. Considering this new accumulation method and also having as objective to strengthen the notion of continuity of boundaries, we have developed a new iteration of our DOVA energy maps which will be explained next.

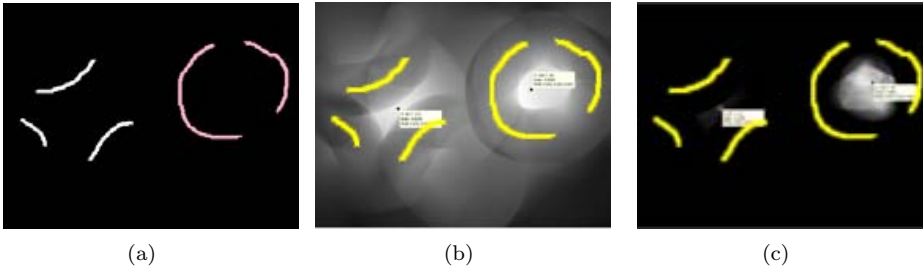


Figure 5.11: Improvement of median DOVA against SA-DOVA energy maps: (a) Synthetic example image; (b) SA-DOVA energy map; (c) Median DOVA energy map.

In order to solve the other drawback that SA-DOVA presents, the assessment of continuity and completeness of boundaries we take a new approach by considering not only the maxima of DoV under a sector but under a whole window of sectors, as it will be explained next.

WM-DOVA algorithm

The objective of WM-DOVA is to add a sense of continuity and completeness to the boundaries we are looking for. In this case, instead of just accumulating the maxima of the DoV image in each direction, we will take into account if the maxima is located near maximas for consecutive angles. This can be better understood by taking a look at Figure 5.12. We see that if we use our SA-DOVA accumulation algorithm, there is not a big difference in terms of maxima of accumulation between the output of the two different contours. If we apply the DOVA iteration that we present here, Window Median Depth of Valley Accumulation (WM-DOVA), we get a more useful accumulation profile, where the accumulation value depends on the number of consecutive directions we find close maxima points of the DoV image.

This results on a more concentrated and accurate DOVA profile, as it will be seen later, and we obtain a DOVA profile more accurately place on the polyp's center.

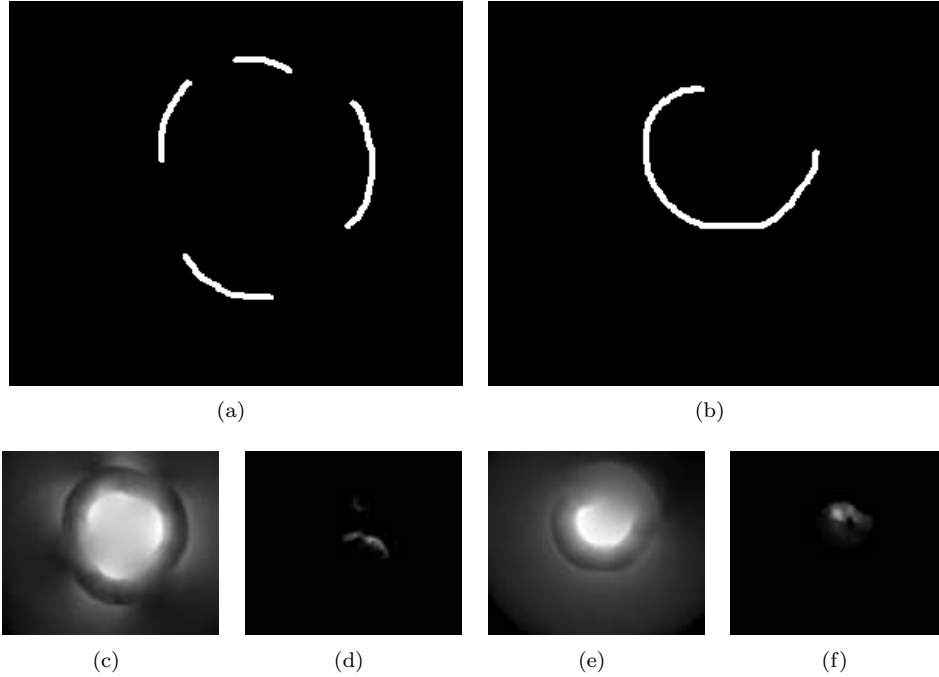


Figure 5.12: Graphical interpretation to SA-DOVA and WM-DOVA results: (a) Synthetic DoV image with an incomplete non-consecutive contour; (b) Synthetic DoV image with an incomplete consecutive contour; (c) and (e) SA-DOVA accumulation image; (d) and (f) WM-DOVA accumulation image.

Therefore, we build this final iteration of DOVA by using the following assumption: boundaries that enclose a polyp are constituted by pixels with high value in the Depth of Valleys image. The pixels that constitute the boundaries are located in a position such the difference between pixels with consecutive direction is small. This means that the distance from the center point of the polyp, which should be the maxima of accumulation for the whole image, and each boundary pixel position should be similar or, at least, have no significative jumps from one direction to another. As it has been seen in the previous section if the pixels that constitute the boundary follow a similar distance pattern we get an smooth polyp boundary as result but if they do not convey to a distance pattern, the final boundary will have abrupt jumps. WM-DOVA algorithm is as follows:

As can be seen from the algorithm, WM-DOVA starts by defining a search window which comprises for each direction sectors up to distance $(+/-)w_{size}/2$, which is shown as a blue cone in Figure 5.13. The general idea is to define complete contours by means of nearby pixel positions with high DoV value. Once the search window is defined, we obtain the maxima of DoV image under the current direction's sector.

Algorithm 3: WM-DOVA algorithm

Data: $DoV, r_{min}, r_{max}, n_{sec}, w_{size}$
Arg: (DoV : input DoV image, r_{min} : minimum radii of the sectors, r_{max} : maximum radii of the sectors, n_{sec} : number of sectors, w_{size} : size of the search window) **Result:** Accumulation image

- 1 Define the $n_{sectors}$ sectors with minimum radii r_{min} and maximum radii r_{max} ;
- for** *Each pixel position* **do**
 - for** *Each sector* **do**
 - 2 Define a search window from sector's corresponding angle $-w_{size}/2$ to sector's corresponding angle $+w_{size}/2$;
 - 3 Obtain the maxima of DoV image below current sector max_i ;
 - 4 Obtain the maximas of accumulation of each sector below the search window (Eq. 5.3.2);
 - 5 Calculate the median of the maximas of accumulation of each sector under the search window med_i (Eq. 5.3.2);
 - 6 Accumulate the median pondered as depicted in Eq 5.3.2;
 - 7 Ponder the value to accumulate according to the distance to the position of med_{i-1} ;
 - end**
- end**

The next step is to obtain the median of the maxima of DoV image under the different sectors contained in the search window med_i . The final accumulation value is obtained by pondering this maxima with respect to two different criteria: 1) Difference between the difference between med_i and max_i and 2) Distance between the position of max_i and the median of distances from the sector's center to each direction's maxima of DoV image. We can see in Figure 5.13 that in this case the accumulation value pixel labeled as A will not be damaged by the distance criteria whereas the pixel labeled as B will be punished in this sense.

$$MaxList(x, \theta) = \max_{r \in [R_{min}, R_{max}]} DV(x + r * (\cos(\theta), \sin(\theta))),$$

$$\theta \in [\alpha - \frac{winsize}{2}, \alpha + \frac{winsize}{2}] \quad (5.7)$$

$$Acc(x) = Median(MaxList(x, :)) \quad (5.8)$$

$$Acc(x) = Acc(x) \cdot \frac{1}{|dist_i - dist_{med_{i-1}}|} \quad (5.9)$$

WM-DOVA needs of four different parameters. In this case r_{min} , r_{max} and $n_{sectors}$ play the same role than in SA-DOVA algorithm. The new parameter is the

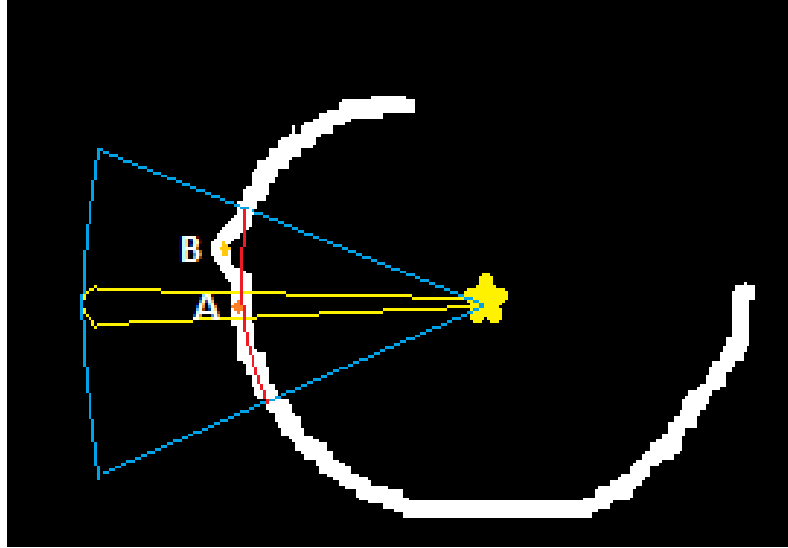


Figure 5.13: Graphical scheme of WM-DOVA accumulation algorithm.

size of the window search w_{size} which defines the width of the angular search window. We can understand better how this works by taking a look at the scheme presented in Figure 5.13: If we start at the given direction, which is determined by the yellow sector with center in the start point, we find that the maxima of accumulation below the current sector is found in point A. We assume that the median of position of the maximas is marked by the red line and therefore the point A falls on this line. Consequently the decay factor (calculated by Eq. 5.3.2) applied to the accumulation value will be close to zero. On the other hand, for the point B we see that the position of the maxima of the DoV image does not fall above the median distance's line, therefore the decay factor in this case will be different from zero.

To close with this section, we have developed WM-DOVA as a way to reward those boundaries that apart from complete, keep continuity. In this case we have defined continuity not only by checking if the boundaries are closed but also measuring how far are the boundary pixels from a possible object's center, as it has been described in the algorithm. We show several examples of WM-DOVA performance in Figure 5.14.

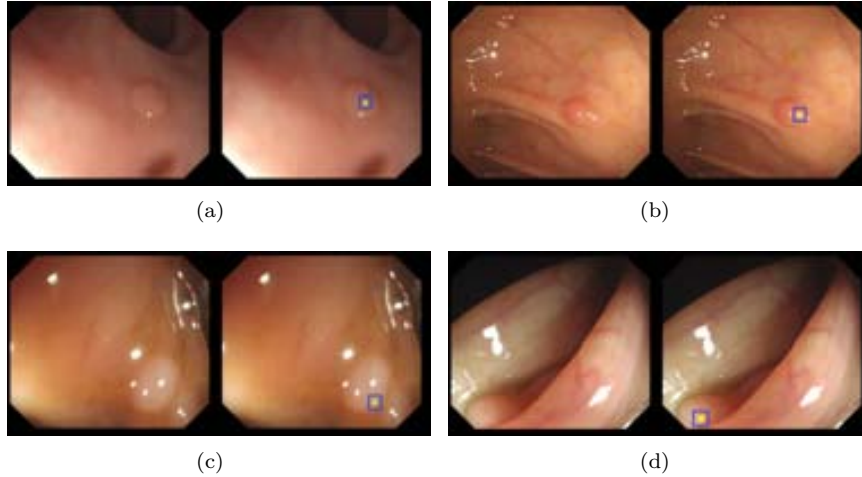


Figure 5.14: Examples of WM-DOVA maxima of accumulation superimposed to the original image. Maxima of accumulation are represented as yellow squares surrounded by a blue square to ease the visualization.

We have shown in this subsection how by means of WM-DOVA we have incorporated a sense of continuity and completeness to the boundaries we are looking for. We can also observe that by using WM-DOVA we obtain DOVA energy maps where the energy is really concentrated inside the boundaries, with a great reduction of the energy outside. But there are some cases where even WM-DOVA still provides with erroneous outputs, as it can be seen in Figure 5.15. We can see in this image that in terms of accumulation the difference between a point interior to the four objects and a point in the middle of the four objects is not the one we would expect. Our DOVA energy map should give a zero accumulation in the point in the middle of the four objects. As a final improvement of our algorithm, we explore the use of the orientation of the valleys, as they should point to the interior of the object and, for the case of the point in the middle of the objects, the orientations will annulate themselves therefore correcting the accumulation value in this point.

5.3.3 Valley Orientation Depth of Valleys Accumulation (VO-DOVA)

Considering that polyp's boundary pixels are meant to appear as valleys in the intensity image, those valleys will be oriented towards areas of the image with higher intensity value. If we think of intensity profiles inside the polyp look like the ones shown in Figure 3.9, we could assume that valleys on the boundary pixels of the polyp have radial orientation. We have developed our Valley Orientation Depth of Valleys Accumulation (VO-DOVA) algorithm based on this assumption taking into account, when accumulating, not only the value of the boundary pixel on the Depth of Valleys image but if its orientation is towards the polyp or that it follows the same orientation

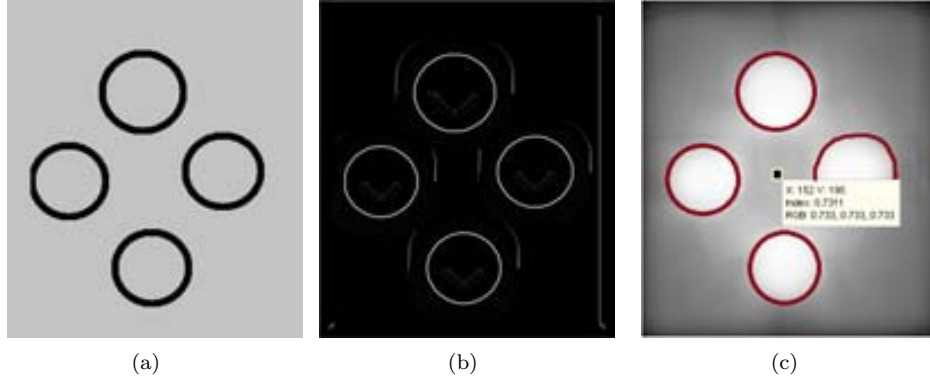


Figure 5.15: Examples of cases where WM-DOVA fails: (a) Original synthetic image; (b) Corresponding Depth of Valleys image; (c) WM-DOVA accumulation.

profile than neighbour boundary pixels.

We can see how VO-DOVA would ideally work by observing Figure 5.16. For each point we define a series of sectors (which are the same sectors that we defined for SA-DOVA), shown in red, and we only accumulate those maxima points with similar orientation to the particular sector (see the algorithm for more detail on this). In this case we will only accumulate those maxima whose orientation coincides with the range of angles covered by the sector (depicted as green arrows surrounded by yellow boxes) and not those whose orientation is very different (depicted as blue arrows).

VO-DOVA algorithm consists of the following steps:

VO-DOVA needs of 4 different parameters, the already mentioned r_{min} , r_{max} and ang_t and the novel ang_{th} , which indicates the maxima deviation that we accept

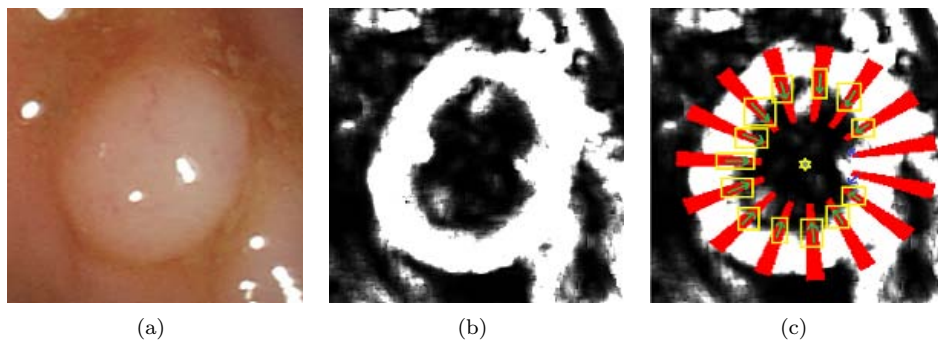


Figure 5.16: Example of calculation of VO-DOVA energy map: (a) Original image; (b) Depth of valleys image. (c) VO-DOVA , where points with suitable orientation are marked with green arrows and surrounded by a yellow box, points with wrong orientation are marked with blue arrows and sectors are shown in red.

Algorithm 4: VO-DOVA algorithm

Data: Input DoV image, Valley Orientation, r_{min} , r_{max} , $n_{sectors}$, ang_{th}
Result: Accumulation image
Define the $n_{sectors}$ sectors with minimum radii r_{min} and maximum radii r_{max} ;
for *Each pixel position* **do**
 for *Each sector* **do**
 Place the sector's center in the pixel position;
 Obtain the maxima of accumulation below the sector;
 if *valley orientation in the maxima of accumulation belongs to range*
 [$sector_angle - ang_{th}$, $sector_angle + ang_{th}$] **then**
 Increment the accumulation value of the given pixel by the maxima
 of accumulation previously obtained;
 end
 end
end

from the current sector's direction. We show in Figure 5.17 how VO-DOVA algorithm would perform in the whole frame. We can see in Figure 5.17 (c) VO-DOVA accumulation results. Brighter areas in the image correspond to pixels whose accumulation value is high, conversely to dark areas which correspond to pixels with low accumulation value. In order to make the results more understandable, we shown in Figure 5.17 (d) how VO-DOVA places points with high accumulation value inside the polyp region, placing also the maxima of the accumulation inside the polyp region.

It is clear that VO-DOVA is based on the strong assumption that all the pixels from the boundary of a polyp share a same objective in terms of valley orientation. This assumption needs consequently a good definition of valley orientation, which will depend on the output of a valley detector. This method also needs again a good input image (in this case, a clean DoV image) in order to obtain the correct pixels which orientation we want to check. We will close this section by showing in Figure 5.18 some preliminary visual results on how VO-DOVA works for a series of images. In this case, as the number of maxima points for each image is very reduced (for most of the cases is 1) we show a green square centered in the maxima point.

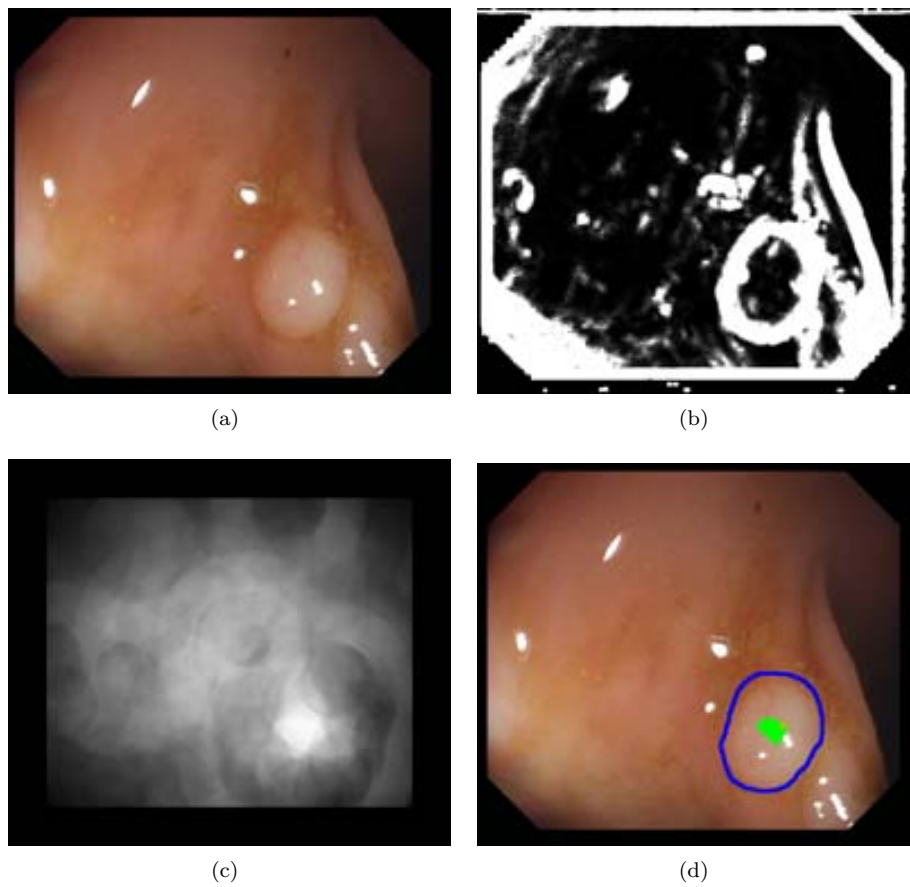


Figure 5.17: Example of VO-DOVA (a) Original image; (b) Depth of Valleys image, and (c) Accumulation image. (d) Result image, where the polyp is surrounded by a blue line, the points with high accumulation value are shown in green and the maxima of the VO-DOVA descriptor is marked as an orange square.

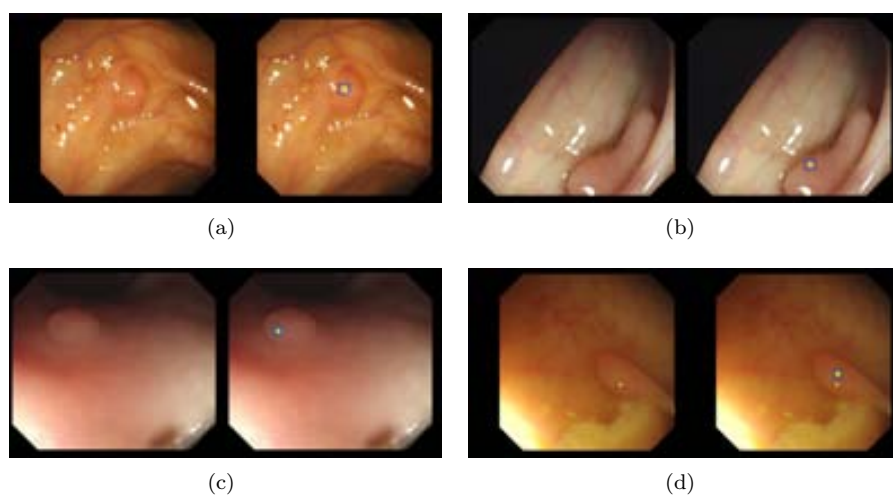


Figure 5.18: Examples of VO-DOVA energy maps. Maxima of accumulation are represented as yellow squares surrounded by a blue square to ease the visualization.

5.4 Discussion and Qualitative comparison

We have exposed during the last sections all the different approaches that we have developed as our polyp localization method:

- **Ellipse Fitting-DOVA:** This method assumes that points with high values of the depth of valley image will constitute polyp's boundaries. Taking this into account, EF-DOVA aims at fitting those polyp boundaries to some general shapes that polyps present, such as ellipses.
- **Sector Accumulation-DOVA:** In this case the method does not assume polyps as having any concrete form and only checks if a given point is surrounded in an elevated number of directions by pixels with high value of the DoV image by means of placing radial sectors centred in each point and accumulating the maxima of DoV image that falls under the sector.
- **Windows Median-DOVA:** This different iteration of DOVA is not only concerned about the completeness of the boundaries like SA-DOVA, but also takes into account the continuity of contours. That is, it aims at finding continuous closed contours in order to accumulate the maxima of DoV image but also taking into account how the position of this maxima deviates from the mean distance of maxima to the centre point.
- **Valley Orientation-DOVA:** The final version of DOVA integrates information of the valleys that appear to surround protruding objects such as polyps. In this case, apart from accumulating the maxima of DoV image that falls under the radial sector, it also takes into account if it follows the same pattern of orientation that its neighbours maxima.

While an extensive analytical comparison will be shown in Chapter 7, we will close this chapter by doing an analysis of the pros and cons of every method along with presenting some comparative visual results.

We started this chapter by doing a separation between the methods according to how they use boundary information. The approach belonging to the first group, EF-DOVA, aims at fitting boundary pixels (which should consist of pixels with high value in the DoV image), performs well in zenithal views where polyp contour has a regular shape (Figure 5.19 (a) y (b)) but it fails to perform well in more irregular shapes, as it can be seen in Figure 5.19 (c). Even for the cases where the maxima of accumulation is placed inside the polyp, the number of false positives is high, as shown in Figure 5.19 (d).

As it can be seen, EF-DOVA works well where it is supposed to, but as the variability of polyp appearance is high, it seems to fall a bit short if we want to use it for a general polyp localization algorithm.

The other group of methods aim at accumulating values of the DoV image by checking boundaries continuity. We have exposed 3 different methods, which vary

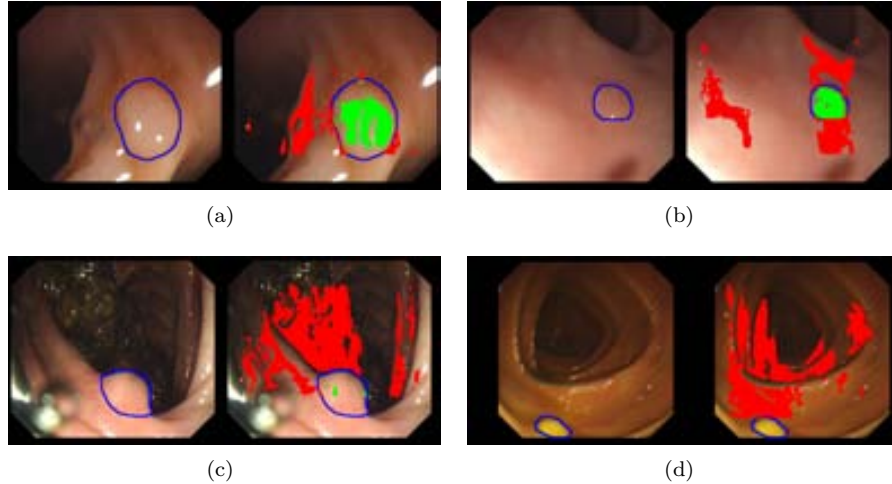


Figure 5.19: Discussion of EF-DOVA energy map results (a) and (b) show a high number of high accumulation pixels inside the polyp (painted in green), with some false positives (painted in red). (c) Shows a result with a non balanced number of true and false positive whereas (d) shows an image where no maxima is placed inside the polyp.

depending on how the accumulation is done. Starting from a basic sector accumulation (SA-DOVA), alternatives include the use of median window pondering (WM-DOVA) or the use of valley orientation. We show in Figure 5.20 additional examples on SA-DOVA performance, in order to analyze those images where it fails to put the maxima of the accumulation inside the polyp mask.

The results show that by using SA-DOVA we obtain good results for a high number of images, as it can be seen in Figure 5.20 (a) and (b), providing with a lower number of false positives than EF-DOVA (we will show examples on this at the end of this chapter). Nevertheless, there are some images where SA-DOVA fails. Among the causes of errors are: influence of lumen contours (Figure 5.20 (c)) or lateral views that in some cases, even by using morphological gradient, as it can be seen in Figure 5.20 (d).

As mentioned in previous sections, SA-DOVA may lead to inconsistencies due to accumulating only according to the number of directions where a maxima is found but not to the continuity. Because of this we developed WM-DOVA, results of which we show in Figure 5.21 to discuss its performance.

WM-DOVA results also show a good performance for many images and, in this case, we can see that the number of false positives is very low. In fact, for a great number of images there are no false positives at all (see Figure 5.21 (a) and (b)). There are some images where WM-DOVA fails at providing an energy map where the maxima is placed inside the polyp, caused again by lumen contours (Figure 5.21 (c)) or a high number of folds, blood vessels or the particular texture of the mucosa

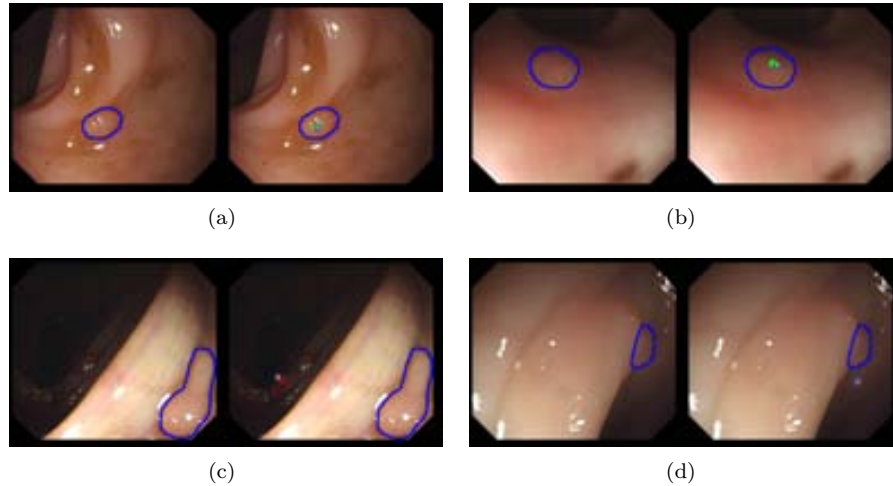


Figure 5.20: Discussion of SA-DOVA energy map results (a) and (b) show a high number of high accumulation pixels inside the polyp, with some false positives. (c) Shows a result with a non balanced number of true (in green) and false positive (in red) whereas (d) shows an image where no maxima is placed inside the polyp. Maxima of accumulation are represented as yellow squares surrounded by a blue square to ease the visualization.

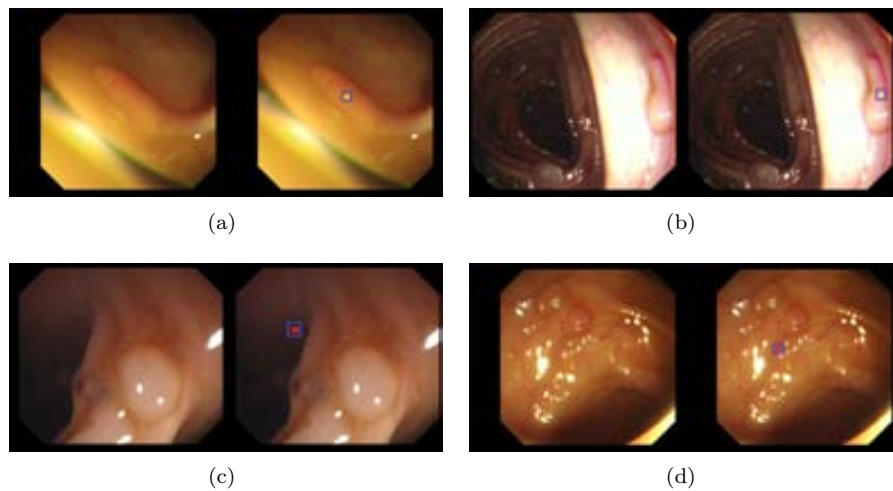


Figure 5.21: Discussion of WM-DOVA energy map results (a) and (b) show a high number of high accumulation pixels inside the polyp, with no false positives. (c) Shows a bad result due to lumen contour influence. (d) Shows an image where no maxima is placed inside the polyp because of the incidence of elements of the endoluminal scene such as folds or blood vessels not fully mitigated.

(Figure 5.21 (d)).

Finally we show a visual analysis of some VO-DOVA results. We can see that we achieve a good performance for several frames (Figure 5.22 (a) and (b)). There are several causes that make VO-DOVA fail, being the inherent reason bad placing of polyp boundary pixels with affect the orientation calculation. Some other causes of errors are the incidence of wrinkles and folds (Figure 5.22 (c) or blood vessels ((Figure 5.22 (d))

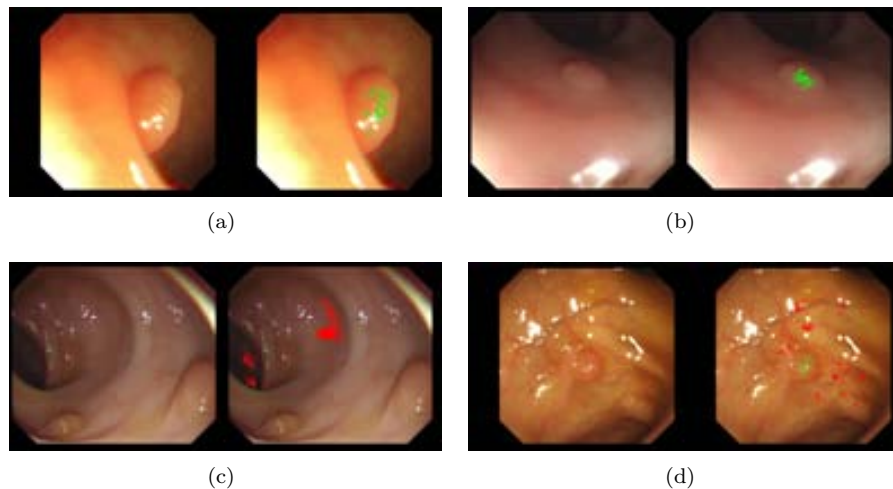


Figure 5.22: Discussion of VO-DOVA energy map results (a) and (b) show a high number of high accumulation pixels inside the polyp, with some false positives. (c) Shows a result with a non balanced number of true and false positive whereas (d) shows an image where no maxima is placed inside the polyp. True positive pixels are painted in green whereas false positives are painted in red.

We will close this chapter with a visual comparison of DOVA energy maps. We will make an analytical comparison on Chapter 7, but we can start to see some differences here. There are some cases where EF-DOVA performs better than SA-DOVA or VO-DOVA, as they are more affected by wrinkles and folds as their accumulation methods admit any particular shape (Figure 5.23 (a)). There are also some studies where all the methods perform well on placing the maxima of accumulation inside the polyp, although EF-DOVA and SA-DOVA provide a higher number of false positives (Figure 5.23 (b)). One special case that has brought our attention is the case of lateral views of polyps, which EF-DOVA, SA-DOVA and VO-DOVA solve but WM-DOVA does not, placing the maxima far from the polyp, as it can be seen in Figure 5.23 (c)). Figure 5.23 (d) summarizes well what happens in the majority of cases: all the methods place maxima of accumulation inside the polyp mask but EF-DOVA and SA-DOVA also places a high number of maxima of accumulation outside the polyp therefore they are less discriminative.

Without entering into details of the results, which will be presented in Chapter 7, by means of EF-DOVA we are able to place the maxima of accumulation inside

the polyp mask in about 32% of the images, by means of SA-DOVA in 42% of the images, by means of WM-DOVA in 54% of the images and finally by means of VO-DOVA we place the maxima inside the polyp mask in 45% of the images. These preliminary qualitative result, although each method performs qualitatively differently although WM-DOVA provide with the best results in terms of balance between TP and FP. We also have to mention that VO-DOVA performance is damaged not by the algorithm itself but by the valley orientation that is passed as input parameter therefore with a better definition of valley orientation (which is out of the scope of this thesis) may provide with different results. Finally it has to be mentioned that we have shown here results achieved with the best combination of parameters for all 4 methods where the best combination is the one that provides the highest number of true positives and true negatives and the lowest number of false positives and false negatives.

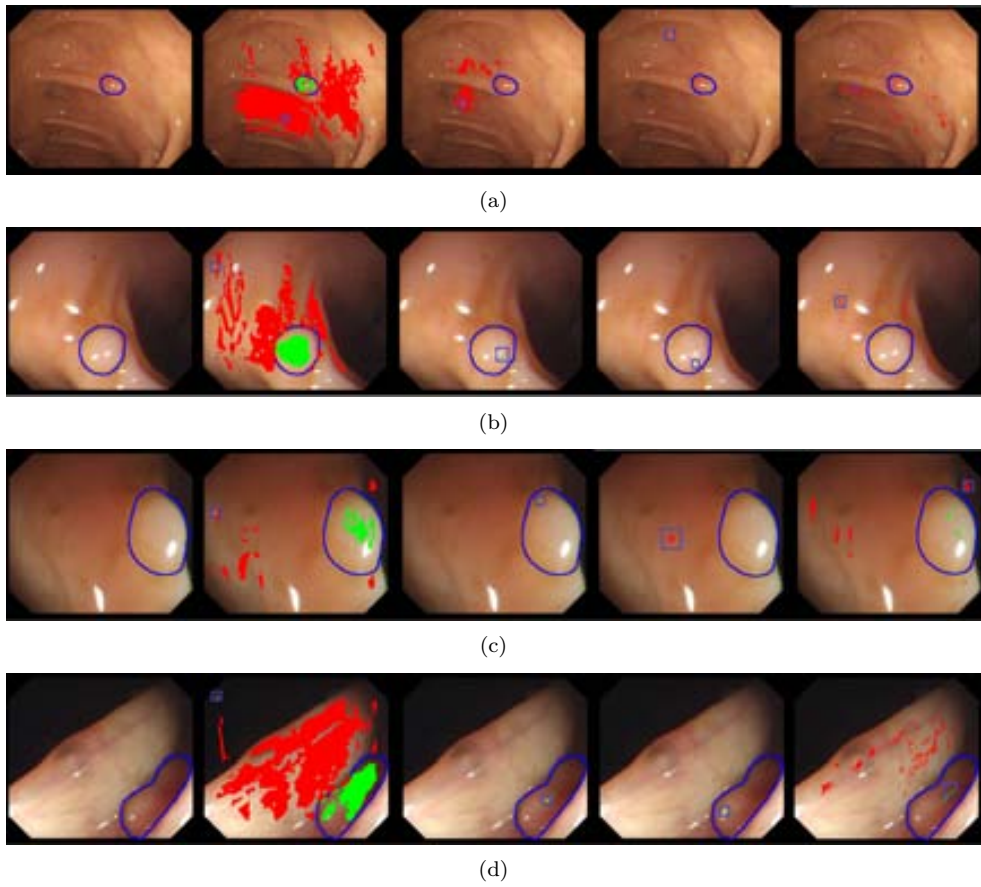


Figure 5.23: Visual comparison of DOVA. (First column) Original image; (Second column) EF-DOVA results; (Third column) SA-DOVA results; (Fourth column) WM-DOVA results; (Fifth column) VO-DOVA results. True positive results are marked in green, false positives in red and the maxima of accumulation is marked by an orange square surrounded by a blue square to ease the visualization.

Chapter 6

Polyp segmentation

6.1 Introduction

In the last chapter we have explained our method for polyp localization on which, given an input image, our algorithm gives as output an energy map where high values correspond to areas of the image where protruding objects, such as polyps, are present. We have to mention that the whole processing scheme is built on our model of appearance for polyps, which defines polyps as surrounded by valleys. Polyp localization aims at indicating which area of the image is more likely to contain a polyp but it does not provide information about which concrete pixels belong to the polyp region.

We present in this chapter our polyp segmentation algorithms which, given an input image, indicate the region of the image that corresponds to the polyp. We have tackled the segmentation problem in two different ways that will be explained here.

The first method, polyp segmentation after region merging, starts by means of an initial segmentation which is refined via region merging until obtaining a reduced number of regions. In this case we only incorporate specific polyp constraints in the last stage of the region merging process. Finally the polyp region will be the one where the maxima of WM-DOVA energy map falls on. The second method incorporates polyp-specific information from the beginning, describing as the initial region the one limited by the pixels that contributed to the maxima of WM-DOVA energy map. This initial segmentation is then refined by means of median filtering in the polar coordinates.

Both methods share the same polyp localization info but use it in different way, as it will be seen throughout the next sections.

6.2 Polyp segmentation by smart region merging

As mentioned in the introduction of this chapter, in this case we faced the problem of polyp region segmentation as a classic Pattern Recognition problem in which, given an input image, we segment it into a number of initial regions by then arriving to the final output region by discarding non-polyp containing regions according to the values of a certain region descriptor, in this case DOVA. Therefore the processing scheme consist of four separated stages, as it can be seen in Figure 6.1. As DOVA energy map creation has been covered in Chapter 5, we will focus here on the rest of stages.

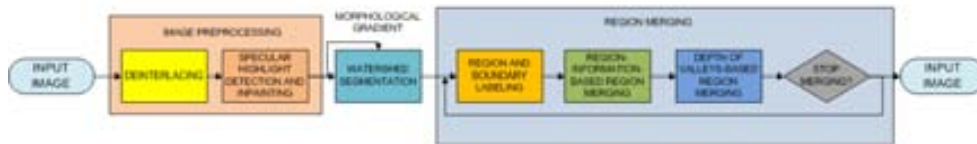


Figure 6.1: General processing scheme of polyp segmentation.

6.2.1 Image preprocessing

In the first stage, image preprocessing, we apply the following operations to the the input image in order to avoid possible errors in the segmentation: 1) Image deinterlacing (Figure 6.2 (b)) and 2) Specular highlights detection and inpainting (Figure 6.2 (c)). Both operations have been explained in Chapter 4. In this case we do not apply black mask inpainting because we will tackle its effect by means of our region labeling method.

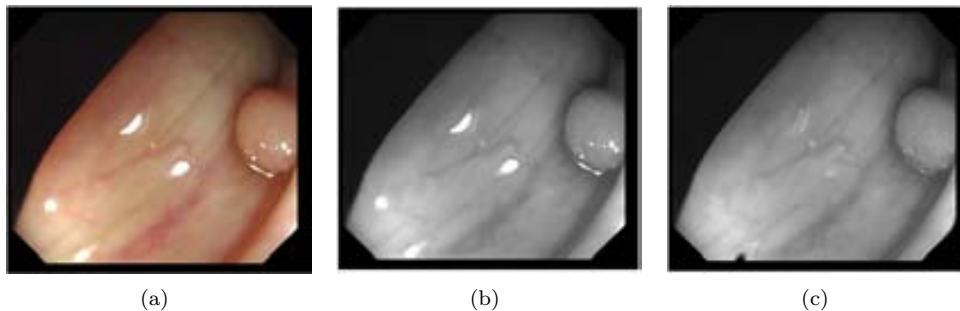


Figure 6.2: Image preprocessing stage: (a) Original image; (b) Greyscale image; (c) Specular highlights-corrected image.

6.2.2 Watershed segmentation

To perform initial segmentation we have chosen to apply watersheds [112], following the same line of thinking than protruding objects are surrounded by intensity valleys. One novel point of this initial segmentation is that we have explored the use of different versions of the input image in order to provide a better initial set of regions. More precisely, we have explored both the use of gradient and morphological gradient image. The first one is obtained by calculating the modulus of the output of the gradient function on each pixel position, and it is able to follow better the edges structure of the image. The second method corresponds to the morphological gradient (i.e. calculated by subtracting two versions of the image, one dilated and the other eroded), which can give us a first representation on the valley structure of the image. The results of watershed segmentation by using morphological gradient approaches suit better the structure of the image, as it can be seen in Figure 6.3 although they lead to deal with a higher number of regions [10].

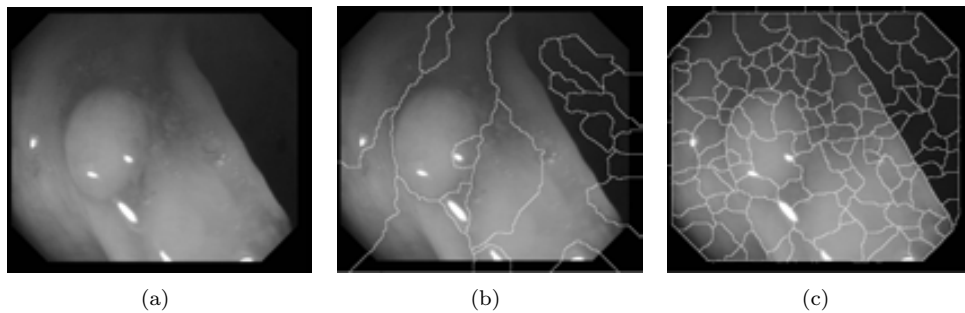


Figure 6.3: Comparison of starting segmentation by using the original image and morphological gradient image: (a) Original image; (b) Watershed segmentation of the complement of the original image; (c) Watershed segmentation on morphological gradient of the complemented original image.

6.2.3 Region and boundary labeling

Once we have a first segmentation of the image the following step is to reduce the number of regions until it is stabilized, which will be done by means of the region merging stage. We will merge regions according to two types of information: a) The degree of information of the region and b) The boundaries that separate them.

In the context of our region segmentation method, we define the degree of information of a region by observing the mean and the standard deviation of the intensity values of the pixels that belong to the particular region, following the criteria shown in Table 6.1.

[!h]

As it can be seen, this criteria let us discard directly some regions with a low

Region label	Mean grey value	Standard deviation
Non-informative dark (NID)	[0, 25]	[0, 10]
Informative dark (ID)	[25, 60]	[0, 40]
Informative (I)	[60, 195]	[0, 40]
Informative bright (IB)	[195, 230]	[0, 40]
Non-informative bright (NIB)	[230, 255]	[0, 10]

Table 6.1: Region labeling criteria.

visibility (NID regions) such as the parts of the image that are on the darker part of the lumen or the black mask that surrounds the endoluminal scene. Our preliminary study confirmed that our non-informative regions do not contain polyp information so discarding them for future processing steps causes no damage to the system's performance. We show an example of region labeling in Figure 6.4 (b).

The criteria to label the boundaries that separate the segmented regions is:

1. Non-informative dark boundary (NIDB): boundary pixels present both a low mean and standard deviation of grey level. They will typically constitute dark contours difficult to distinguish, which will be located in dark areas of the image such as the lumen.
2. Non-informative dark boundary (NIBB): boundary pixels present both a high mean and standard deviation of grey level.
3. Informative boundary (IB).

We can see an example of region labeling in Figure 6.4 (b). We can see a complete example of the application of region merging by means region and boundary labeling in Figure 6.4 (d).

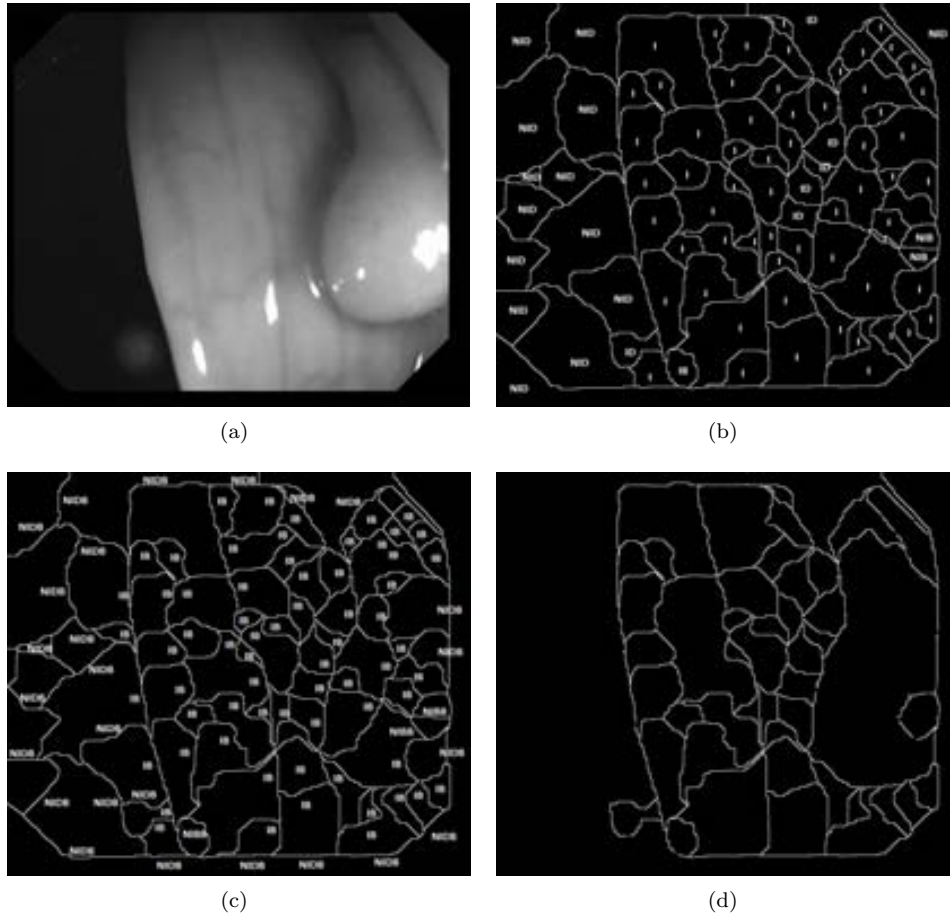


Figure 6.4: Example of region labeling to prepare the image for region merging: (a) Original image; (b) Watershed segmentation with region labeling; (c) Contour labeling. (d) Final segmentation after region merging

6.2.4 Region Merging

The objective of this region merging stage is twofold: 1) Reduce the number of regions so the final segmentation consist of a few regions one of them containing the polyp and 2) Label the resulting images so we can discard some of them and not process them in later stages of the processing scheme. In order to merge regions, we have followed the criteria presented in Table 6.2:

Our region merging stage consists of two different phases:

	NID	ID	I	IB	NIB
NID	If separated by NIDB	Merge if separated by NIDB and similar grey level profile	Not merge	Not merge	Not merge
ID	Merge if separated by NIDB and similar grey level profile	Merge if similar grey level profile	Not merge	Not merge	Not merge
I	Not merge	Not merge	Not merge	Not merge	Not merge
IB	Not merge	Not merge	Not merge	Merge if similar grey level profile	Merge if separated by NIDB and similar grey level profile
NIB	Not merge	Not merge	Not merge	Merge if separated by NIDB and similar grey level profile	Merge if similar grey level profile

Table 6.2: Criteria used to merge regions.

Region information-based region merging

In this first phase we start by calculating the neighborhood map of the image, which is obtained directly from the region labeling. Once we know this neighborhood map we can start with this first stage, which aims at merging close small regions into bigger ones. In this case in order to merge two regions we calculate a frontier weakness measure (Eq. 6.1) which is applied only to the boundary pixels that separate these two regions.

$$FrontierWeakness = \alpha * gradient + \beta * median \quad (6.1)$$

The frontier weakness measure is composed by two different terms. The first one, pondered by means of the α variable, takes into account the strength of the gradient in the boundary pixels. The second term is built on the assumption that if a frontier between two regions is weak, if we apply two consecutive increasing window size median filter to the whole image this frontier would have disappeared in the filtered image. If this proves to happen, the contribution of the second term will stop the merging of the two involved regions. α and β optimal values were set respectively as 1 and 2 in a previous experiment [10] and both the gradient and median magnitudes have been normalized to 1 so the frontier weakness measure goes from value 0 (weak frontier) to 3 (strong frontier).

We merge regions until one of the following stop criteria is met:

- The number of final regions after two consecutive iterations remains the same.

- There are no weak frontiers left. In this case our frontier weakness threshold has been set to 1.

Depth of valleys-based Region Merging:

We introduce our model of appearance for polyps in this second stage of the region merging by means of using the DoV image (explained in Chapter 4). In this case we want to assess if the boundaries that separate the remaining regions are constituted by pixels with high DoV value or not. Our model of appearance for polyps state that protruding objects are surrounded by valleys so the boundaries that separate the protruding object from the rest of the image should have a high DoV value whereas points close to the center of the polyp should have a very low DoV value.

Taking this into account, we continue by merging compatible regions (according to Table 6.2) that are separated by boundary pixels whose mean DoV value is higher than a given threshold value DoV_{th} . We merge regions until there are no weak frontiers according to the depth of valleys threshold value or when the number of regions is stabilized.

We can see a graphical summary of the results of the region merging stage in Figure 6.5.

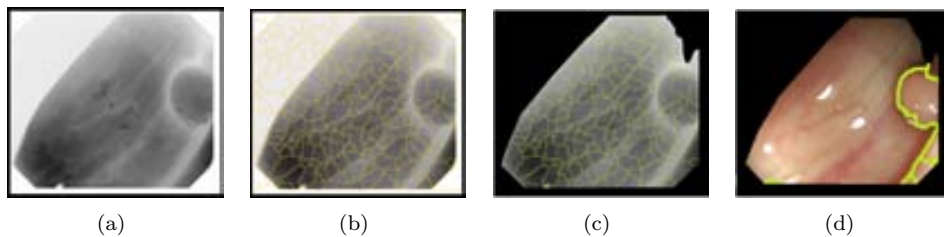


Figure 6.5: Region merging stage in region segmentation: (a) Complemented specular-corrected image; (b) Initial watershed segmentation; (c) Segmented image after region information-based region merging; (d) Final segmented image

The final step in this polyp segmentation by smart region merging is to decide which of the final regions that we will have will be the polyp region. At this step in the processing scheme we incorporate the information from DOVA energy maps. Taking this into account, the final polyp region will be the one where the maxima of the DOVA map falls on. Preliminary segmentation results are good for a high number of images but there are others where, although the final region contains the polyp in full, the region boundaries do not overlap the polyp contour, as it can be seen in Figure 6.7. We present some preliminary segmentation results in Figure 6.6.

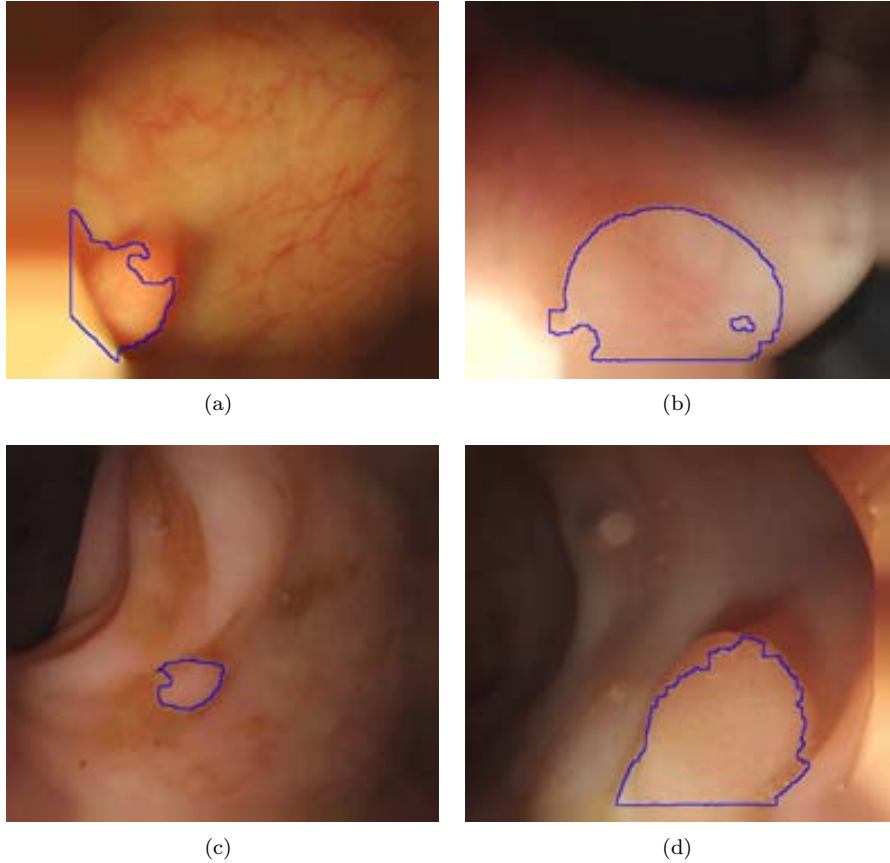


Figure 6.6: Polyp segmentation after region merging method results. For all the examples the polyp region mask is superimposed to the original image.

The experimental results showed that segmentation results depend on the DoV_{th} threshold that we apply to the DV image in a way such the higher the value, the lower the number of final regions. But, in this case, as the threshold is increased, the performance is damaged, as we can lose, for some images, some of the boundaries of the polyp region if we apply a high threshold value, as it can be seen in Figure 6.7 (b) and (c). This makes the method not robust and for this reason we propose an alternative method which does not start by a previous segmentation.

Another problem that this segmentation method presents is that the final polyp regions in some cases are much bigger than the actual polyp in the image, as can be seen in Figure 6.8. This happens because of two different reasons: first of all, the method does not incorporate polyp-specific constraints until the very last step of the region merging and secondly, and more important, by means of excluding this polyp specific information we have to rely on the DoV_{th} threshold value to keep polyp boundaries. The problem here is that there are some images where the polyp

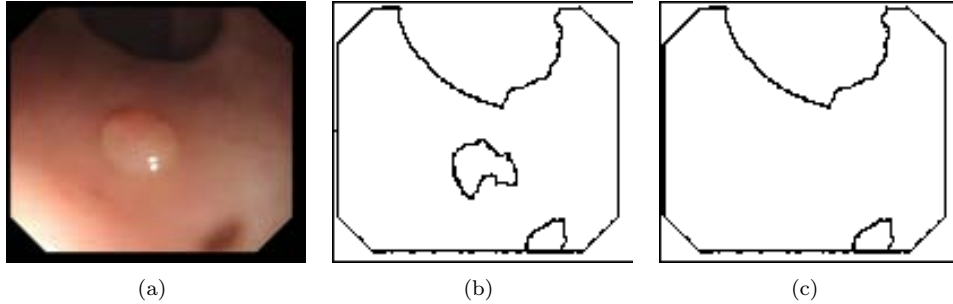


Figure 6.7: Problems in segmentation: (a) Original image; (b) segmentation result with depth of valleys threshold 0.7, and (c) Segmentation result with depth of valleys threshold 0.8.

boundary is weak and if we increase the DoV_{th} value, we will lose the whole boundary and our final polyp region will not approximate the actual polyp, as can be seen in Figure 6.8. Because of this we present in the next section another polyp segmentation algorithm that does take into account the model of appearance for polyps from its first step.

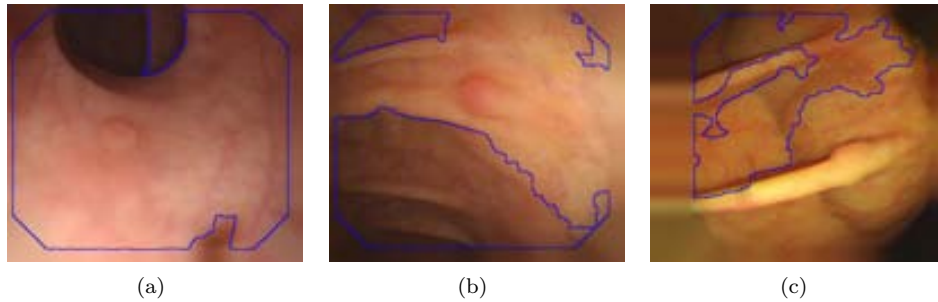


Figure 6.8: Polyp segmentation after region merging method results. For all the examples the polyp region mask is superimposed to the original image.

6.3 Polyp segmentation from the output of an energy map

As it has been mentioned before, we will take a different approach on how to identify which region in the image contains the polyp. In this case, we will start our polyp region identification method by calculating the DOVA descriptor. In the example we will use as DOVA energy map SA-DOVA descriptor which accumulates, for each pixel, the contribution in terms of DoV value that a group of pixels under a series of sector made. The idea that we present here is to obtain the pixels that constitute the boundary of the final polyp region by extracting the pixels that contributed to the

maxima of accumulation, as can be seen in Figure 6.9.

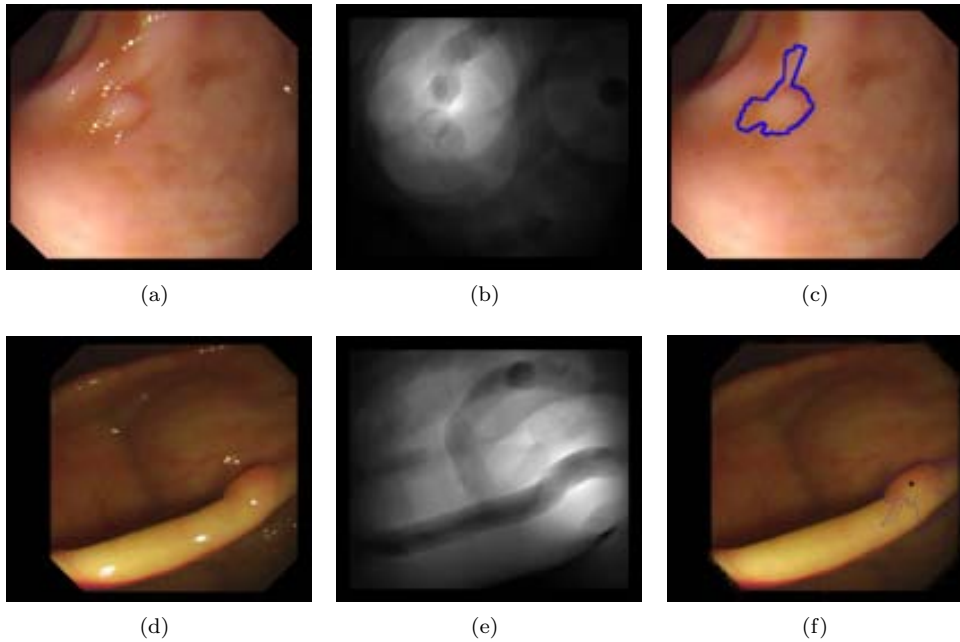


Figure 6.9: Examples of polyp segmentation from the output of an energy map: (first column) Original images; (second column) : Accumulation images; (Third column): Examples of obtaining the points that contribute to the maxima of SA-DOVA accumulation image.

As can be seen from the results shown in the Figure above, if we just join the points that contribute to the maxima of accumulation, we get for the majority of the cases irregular boundaries that do not fit the real contour of the polyp. Our proposal for improving these contours is based on the direct observation of the current results. As it is shown in Figure 6.10, when we obtain the maxima contributing pixels, we can observe some abrupt jumps. These jumps may have several sources such as high DoV values caused by folds, blood vessels not fully mitigated or simply that the radius parameter values of SA-DOVA let include further valleys.

Our polyp segmentation from the output of an energy map algorithm consists of the following steps:

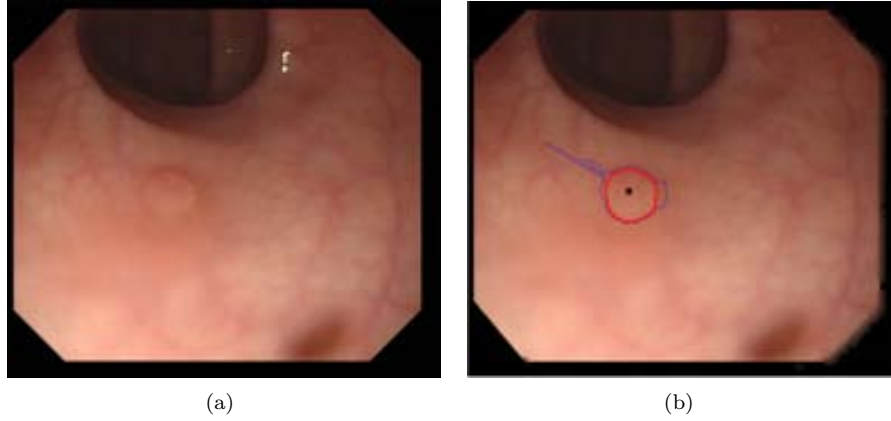


Figure 6.10: Example of contour enhancement: (a) Original image; (b) Original image with initial segmentation boundary pixels painted in blue.

Algorithm 5: Polyp segmentation from the output of an energy map algorithm

Data: I_o : Input image, I_{dov} : Input DoV image, ws : window size

Result: Polyp segmentation

- 1 $Acum = WM - DOVA(I_{dov})$ // Calculate WM-DOVA energy map;
 - 2 $c_x, c_y = maxima(Acum)$ // Calculate the position of the maxima of the WM-DOVA energy map;
 - 3 $\vec{p}_o = contrib(Acum, c_x, c_y)$ // Obtain the pixel position of the points that contributed to the maxima of WM-DOVA energy map;
 - 4 $[\rho, \phi] = polar_coord(\vec{p}_o)$ // Convert to polar coordinates;
 - for** Each different ϕ value **do**
 - 5 $Neig(\rho_i) = \{\rho_{i-ws/2}, \rho_{i-ws/2+1}, \dots, \rho_{i+ws/2-1}, \rho_{i+ws/2}\};$
 - 6 $\rho_{f_i} = median(\rho(Neig(\rho_i)))$ // Compute the new polar coordinate by means of the median of the ρ values of its neighbors;
 - end**
 - 7 $\vec{p}_f = cartesian_coord(\rho_f, \phi)$ // Revert the conversion to cartesian coordinates;
-

The algorithm first starts by calculating the WM-DOVA energy map of the input image, with all the preprocessing and definition of DoV image explained in Chapter 4. Once we have this energy map we obtain the first rough polyp segmentation by joining the points that contributed to the maxima of accumulation of the calculated WM-DOVA map (lines 2 and 3 of the algorithm). We showed before in Chapter 5 that we accumulate for each radial sector the maxima of DoV image that falls under the sector's area of influence. What we do here is to store the pixel position of the maximas of DoV image for each sector and by joining them in angle-increasing order we achieve the first segmentation. As we can see from Figure 6.11 (a) this first segmentation may present some spikes that make its shape far from regular. In order to solve this we pass the coordinates of the region boundary to polar domain and we apply a median filter of size ws to eliminate the abrupt jumps (see Figure 6.11 (c) and (d) (lines 5 and 6 of the algorithm). The final segmentation will be calculated by simply inverting the polar domain conversion in order to have final cartesian coordinates (line 7 of the algorithm).

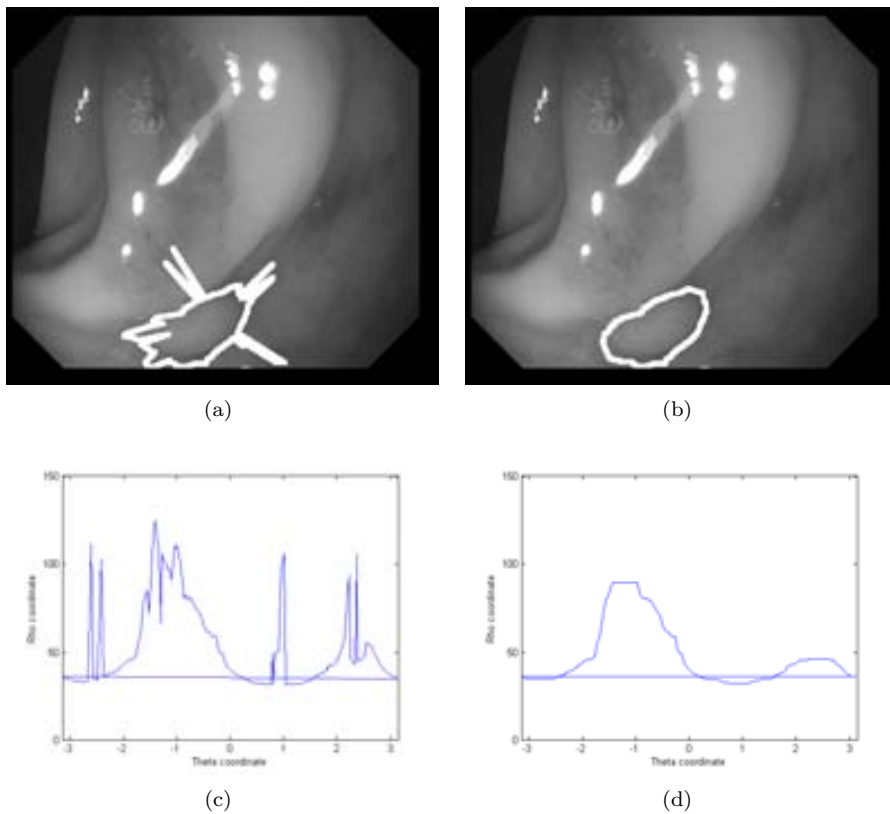


Figure 6.11: Softening of boundaries by median filtering in polar space: (a) Original image with initial segmentation marked in white; (b) Original image with final segmentation marked in white; (c) Polar representation of the initial segmentation; (d) Polar representation of the final segmentation.

As can be seen, we base our method on the fact polyp contours tend to follow a closed uniform path where the distances between two consecutive contour points are around a certain value and if we find the distance between two consecutive points is higher than usual we will have an abrupt jump. We show in Figure 6.12 a complete example of the contour refinement algorithm.

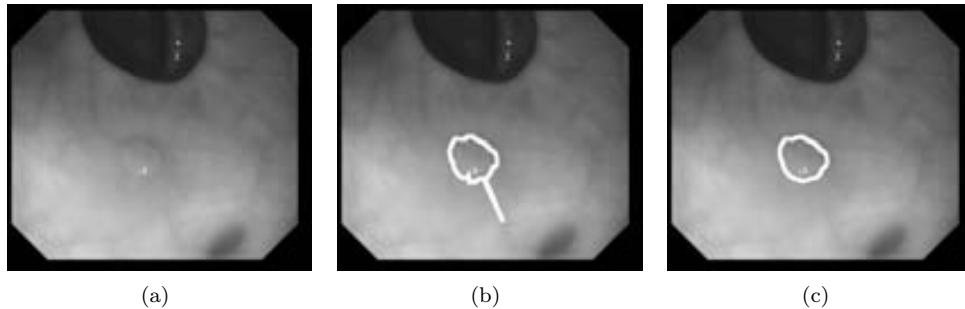


Figure 6.12: Examples of our contour refinement method: (a) Original image; (b) Original image with initial segmentation marked in white; (c) Original image with final segmentation marked in white.

Finally we show in Figure 6.13 some examples of this method's performance. As can be seen, the method leads to good segmentation results whenever the maxima of WM-DOVA energy map is placed inside the polyp (Figure 6.13 (a), (b) and (c)) but when this does not happen, the results obtained do not represent the polyp region at all (Figure 6.13 (a), (b) and (d)).

6.4 Visual comparison of segmentation methods. Discussion

We have detailed in the previous two sections the two different approaches that we have undertaken to provide a polyp segmentation method. Although both approaches share the same information, which is the DOVA energy map, they use it in a different way. The first method uses DOVA information to decide which of the available regions contains the polyp whereas the second method uses it to directly guide the segmentation process.

Although a more extensive comparison of the performance of both methods will be carried out in the next chapter, we will close this chapter by offering a preliminary visual comparison of the outputs of both methods. We can see in Figure 6.14 that we obtain similar results for some images, specially in those where we can see clearly the polyp boundary and DOVA energy map performs well.

As can be seen from Figure 6.14, there are some images where the results for both methods are very similar (Figure 6.14 (b) and (c)) and some where even considering that the maxima of the DOVA is common, the second method covers more polyp area

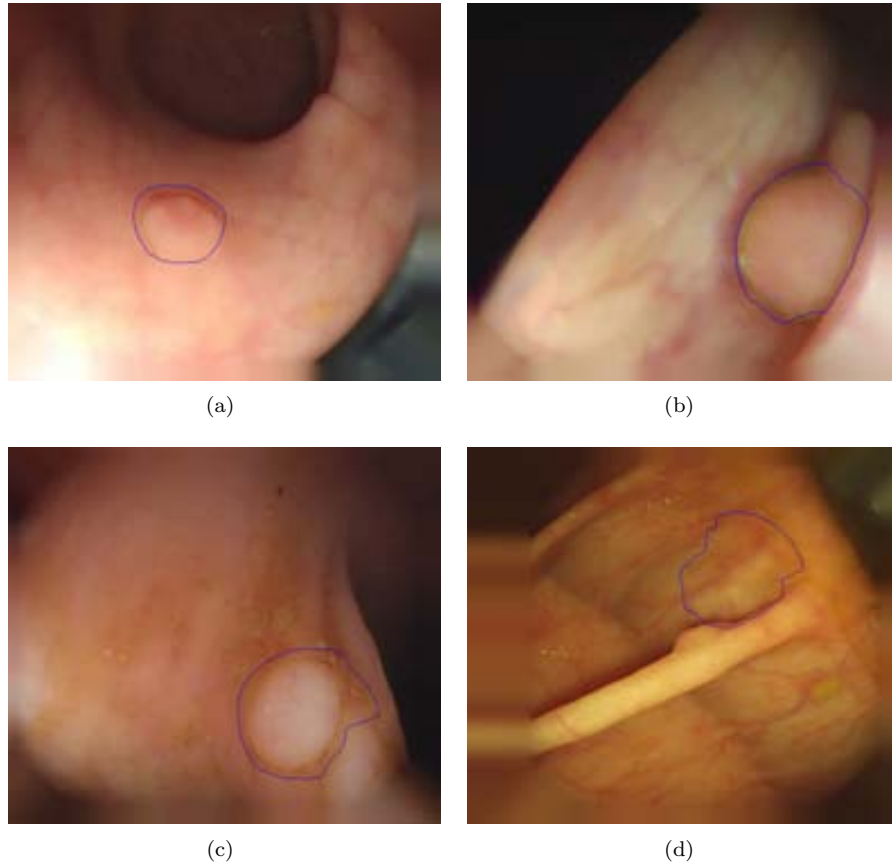


Figure 6.13: Polyp segmentation from the output of an energy map results.

(Figure 6.14 (h) and (i)). In general, when the maxima of DOVA is placed inside the polyp, calculating the final polyp segmentation from the contributions of the maxima of DOVA leads to better segmentation results. As they share the same DOVA energy map, although used differently, they inherit the same behaviour, that is, if DOVA fails to place the maxima inside the polyp, both segmentations will provide bad results, as it can be seen in Figure 6.15.

As can be seen, where DOVA fails the segmentation follows therefore one of the main sources of segmentation errors are those that make DOVA fail, which include contours generated by lumen or the presence of wrinkles and folds. In the next chapter we will evaluate both polyp segmentation algorithms, along with presenting our complete Experimental Setup.

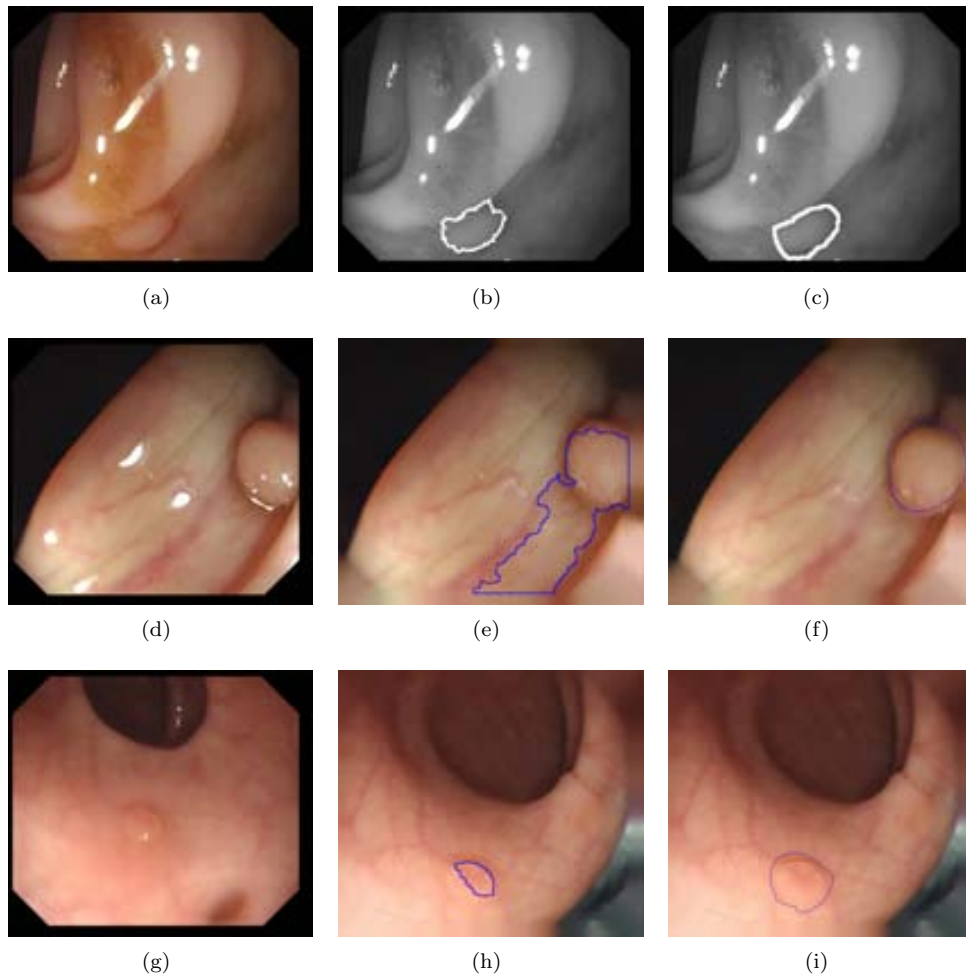


Figure 6.14: Comparison between the segmentation performed by the two methods presented: (First column) Original images; (Second column) Polyp segmentation by smart region merging results; (Third column) Polyp segmentation from the output of an energy map results.

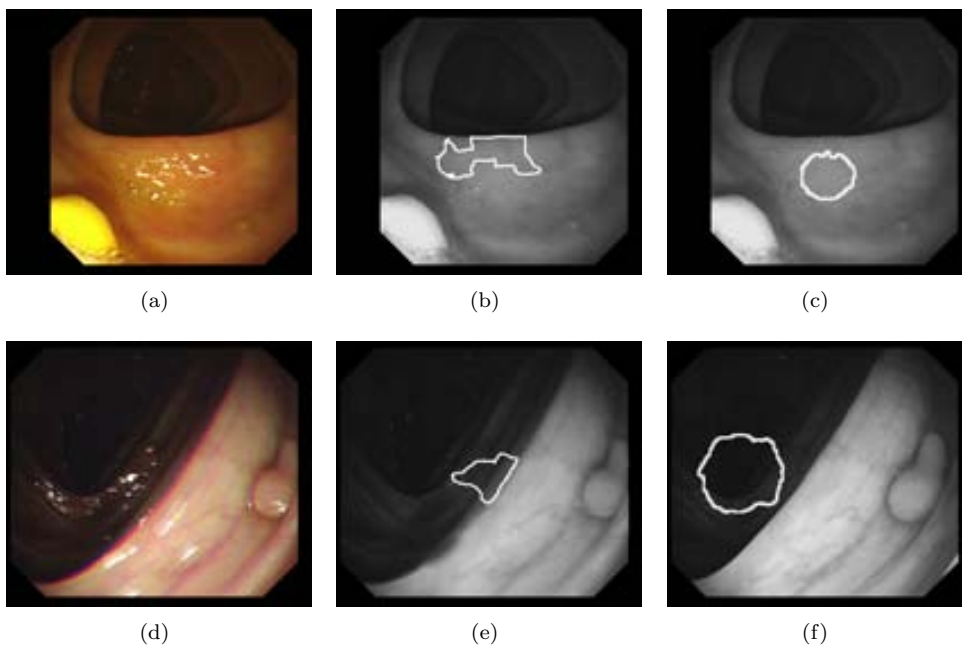


Figure 6.15: Segmentation errors achieved by the two methods presented: (First column) Original images; (Second column) Polyp segmentation by smart region merging results; (Third column) Polyp segmentation from the output of an energy map results.

Chapter 7

Experimental Results

7.1 Introduction

The objective of this thesis is to provide with polyp localization and segmentation methods based on a model of appearance for polyps, also taking into account some other elements of the endoluminal scene. We have presented our complete methodology in Chapters 4 to 6 and we expose in this chapter experimental results to assess the performance of the methods explained.

The aim of this chapter is to assess the performance of each of the methods that we have proposed in this thesis, from image preprocessing to polyp segmentation. In some of them we have presented several approaches to the solution and by means of the results presented here we provide the evidences that led us to decide among them. More precisely, the experiments presented in this chapter cover:

1. Image preprocessing results:
 - Specular Highlights detection and inpainting.
 - Blood vessels mitigation.
2. Polyp localization results.
3. Effect of image preprocessing on polyp localization results:
 - Impact of black mask inpainting.
 - Impact of specular highlights correction via detection and inpainting.
 - Impact of blood vessels mitigation on polyp localization results.
 - Combined effect of preprocessing on polyp localization results.
4. Polyp segmentation results.

As it can be seen, in order to ease the understanding of the experimental results combined with the theoretical explanation that has been exposed in the previous chapters, we have used the same structure going from preprocessing to polyp localization and polyp segmentation but also taking into account the separate effect of some of the preprocessing methods that have been explained. We will present the complete experimental results in the following sections, giving for each of them the following information: 1) Objective of the experiment; 2) Experimental Setup; 3) Metrics; 4) Experimental results and 5) Analysis and discussion of the results.

In order to truly assess the performance of these methods we need first to define a complete experimental setup, which will be exposed in the first sections of this chapter. Moreover, we introduce here the first public database available in our domain of research.

7.2 Experimental setup.

Before presenting the results of our experiments, it is necessary to define on which database we are working. Sadly up to our knowledge there is no public database that can be used to assess the performance of methods like the ones we have introduced and, in order to fix this, we developed the first public database on colonoscopy videos along with the corresponding ground truth, which will be introduced later in this section after doing a brief review on how to create databases such as the one that we present.

7.2.1 Ground truth and database building

In order to carry out an objective assessment of a given method or system, a ground-truth must exist. The ground truth consists of set of samples from a given number of case studies, with the corresponding annotations provided by an expert or group of experts. In our context, a video annotation can be of different natures, among which we can highlight, only to mention a few: 1) a whole frame, indicating that it is that frame which contains a particular event -e.g., the first image in a sequence showing a polyp-; 2) a given region of interest (ROI) -e.g., indicating the bounding box surrounding the polyp itself-; 3) any textual information -e.g., a qualitative assessment of the clinical relevance of a polyp-, etc. These annotations are used to check the performance of a new expert or a new method against the results provided by the annotator, who is considered the reference. In the ideal case, the annotation procedure should be repeated by each expert, in order to get an intra-observer variability measure, and by different experts, in order to get an inter-observer variability measure. A good database with a solid ground-truth is an invaluable resource and a key point for the objective assessment of different methods under a common context of evaluation.

Unfortunately, databases of annotated colonoscopy videos are scarce, and even the access to small databases is very restricted (few examples can be found at [76]). The reason of this (without taking into account the natural motivations related to ethical

and administrative issues) has to do with the generalized fact that colonoscopy video interventions are not routinely saved, since no a-posteriori analysis is needed after the intervention. In many cases, the only image saved consist of a single picture of the ileo-cecal valve, which serves as a prove of the its achievement during the phase of introduction and indicates the start of the withdrawal phase [69]. In the computer vision bibliography, some authors proposed pilot approaches that were validated in a few frames, with no significant inference for the case of a large video. In other cases, when the number of cases was higher, the database used for the results was not available. As mentioned before, we address this problem by publishing publicly our database.

Building up of a database

The building-up of a colonoscopy database consists of two different parts, namely: 1) The video acquisition system, and 2) the video annotation procedure.

1. The video acquisition system must be able to grab HD frames from the colonoscopy source and store them to hard disk, with no lose of frame rate or frame quality. Although the posterior analysis of the frames must not need HD size, by storing the HD version of the video we assure the maximum image quality provided by the device. In order to capture the HD frames, a HD frame grabber must be installed into a PC which will play the role of data repository. Finally, in order to keep the frame rate and video quality, the frames must be compressed with a fast compression codec to be efficiently stored later.
2. The video annotation procedure can be performed in different ways. In the case of frame annotation, a keyboard interaction can be potentially enough to select the desired frames. A navigation system must be implemented if the the expert is allowed to go forward and backwards in the video sequence. If the annotation task consist of the definition of ROIs, a mouse, a digital pen, or a tactile device can be used. More sophisticated techniques, such as the use of eye-tracking [110], can be implemented in case that the video is to be annotated by using attention/perception models -see Figure 7.1 b) for a general scheme.

7.2.2 CVC COLON DB database

Our CVC Colon DB database, has been generated from 15 different video studies [68]. For each study a sequence containing a polyp was extracted. In Table 7.1 we present the key data the videos that compose our database: length, number of frames, polyp shape (flat or peduncular) and correspondence with the number of frame in the published database.

As the work presented in this thesis is focused on testing the validity of our model of polyp appearance, our database consists only of frames which contains a polyp. Consequently, a set of frames were selected from each sequence, paying particular

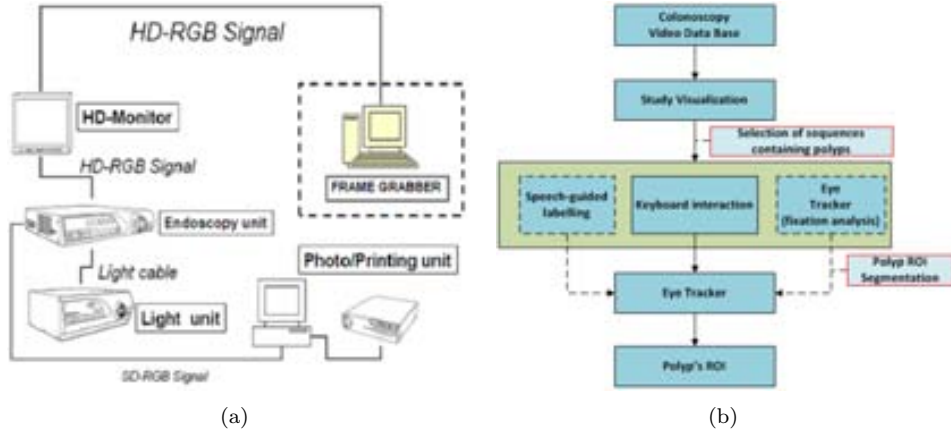


Figure 7.1: a) Schematics of the HD colonoscopy data acquisition system. b) Data annotation scheme. [110]

attention in showing several points of view of the polyp. More precisely, a random sample of around 20 frames per sequence was obtained, with frame size of 1000×1148 pixels. The final size of CVC Colon DB database is of 380 images. The experts guaranteed that all the selected frames showed a significantly different point of view within the scene by rejecting similar frames. As mentioned in Chapter 4, these frames come from an interlaced video so the final frames that will compose our database are already deinterlaced therefore the final size of the frames of our database is 500×574 . In Figure 7.2 the reader can observe the great variability between the different types of polyp appearance along the different videos. This allows us to maximize the variability of the images used, while not jeopardizing any bias.

Along with the selection of the frames that constitute our database, the expert also provided with ground truth for the polyp mask and non-informative regions (following the criteria expressed in Chapter 6). As we also present results on specular highlights detection, the expert also labeled the specular highlights for every frame of the database. We can see some examples of the frames of our database and their corresponding ground truths in Figure 7.3:

However, not all the sequences annotated in CVC Colon DB are suitable for our experiments and we rejected frames belonging to some videos because of the following reasons:

- Sequence 4 presents fecal content which makes it difficult to provide a proper polyp mask.
- Sequence 8 also presents a lot of fecal content.
- Sequence 12 has a bad quality of visualization of the polyp.
- Sequence 13 does not present frames with enough quality of point of view.

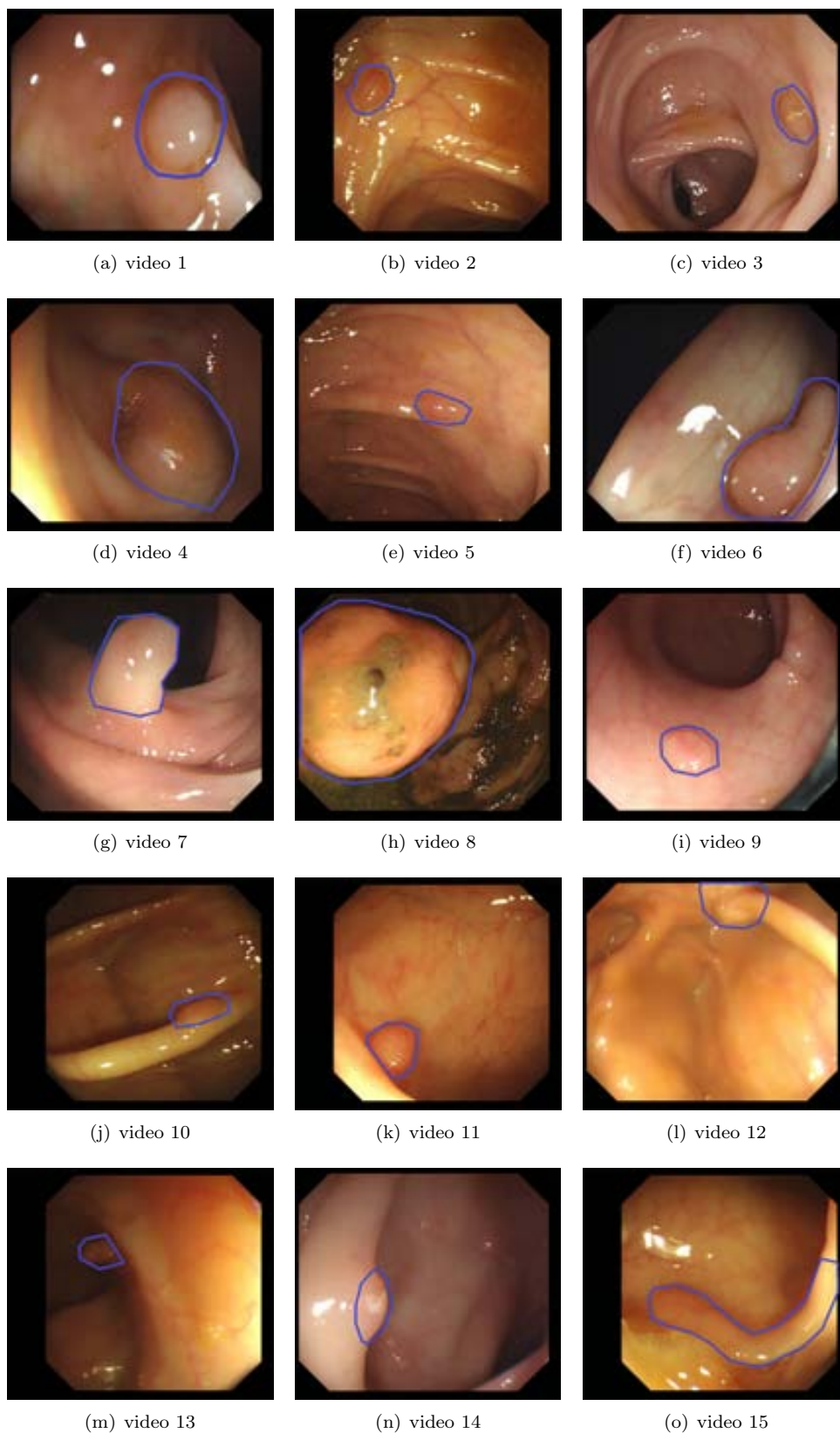


Figure 7.2: Examples of the type of polyp present in each colonoscopy video. Polyps are surrounded by a blue contour.

Video	Length (mins)	Frames	Shape	Correspondence
1	1:02	1570	Flat	1-38
2	1:30	2252	Flat	39-60
3	1:07	1679	Flat	61-76
4	1:05	1648	Peduncular	301-320
5	1:15	1885	Flat	77-97
6	1:09	1773	Peduncular	98-148
7	0:54	1361	Peduncular	149-155
8	2:44	4121	Flat	321-340
9	1:05	1639	Flat	156-203
10	2:36	3912	Flat	204-208
11	0:37	947	Flat	209-263
12	1:21	2038	Flat	341-360
13	0:36	923	Flat	361-380
14	0:43	1097	Flat	264-273
15	0:35	879	Peduncular	274-300

Table 7.1: Database description.

Taking this into account, the database that we use in our experiments is composed by 300 frames and consequently we will be mention it as CVC 300 Colon DB. CVC 300 Colon DB database is composed by the frames selected by the experts for the rest of the videos, but the expert did not select the same number of frames from each of

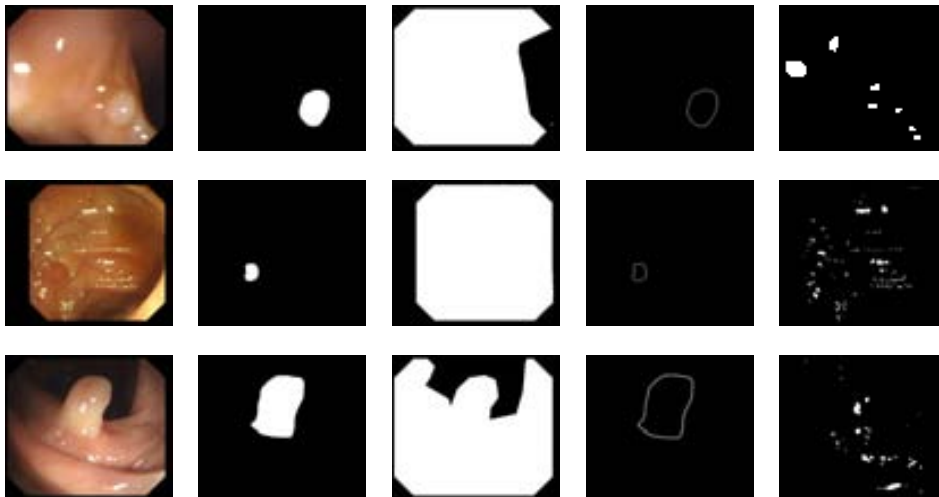


Figure 7.3: Some examples of the content of COLON-DB database: (first column) Original images; (second column) Polyp masks; (third column) Non-informative regions; (fourth column) Polyp contour masks. (fifth column) Specular highlights masks.

them. Sequences 1,9 and 15 are especially rich in terms of number of different views of the polyp. Sequences 3 and 10 are short, they have few frames with polyp presence, and for this reason the number of different views is lower.

Finally, as another of our objectives is to mitigate the impact of blood vessels, the expert labeled 29 frames of the CVC 300 Colon DB database which were rich in vessel content.

7.3 Image preprocessing results

7.3.1 Experimental results on specular highlights detection and inpainting

Objective of the experiment

As it has been mentioned in chapters 3 and 4, specular highlights are meant to have an impact on our polyp localization and polyp segmentation results, as their presence generate response in the valleys image (see Figure 4.9 for a graphical example. As our polyp localization and segmentation methods have as input the depth of valleys image, created by combining the output of a valley detector with the morphological gradient, the valley response generated by specular highlights will be propagated through the following stages of our algorithms.

In terms of specular highlights detection we will compare the several approaches by measuring how many specular highlights pixels are detected with respect to the ground truth. For the case of specular highlights inpainting, as our objective is to mitigate as much as possible the response of specular highlights in terms of valley detection, we will measure the creaseness of a region centered in each specular highlight to check which of the available methods reduces most the response of specular highlights.

Experimental setup

The expert segmented manually the specular highlights (pure and saturated) from the 300 images. An example of some images and their corresponding specular highlight masks can be seen in Figure 7.3. For the case of specular highlights inpainting, we will use as specular highlights influence area a dilated version of the ground truth for specular highlights, as it is shown in Figure 7.4.

As it can be observed from the figure, by dilating the original specular highlights mask we achieve a bigger ground truth. The objective of this dilation is to have a ground truth that covers the area of influence of the specular highlight, as valleys will surround them. Therefore, we will measure the creaseness under this extended specular highlights mask (see Figure 7.4 (d)) in order to compare different inpainting methods.

We present in Table 7.2 the approaches that we will compare in this experiment.

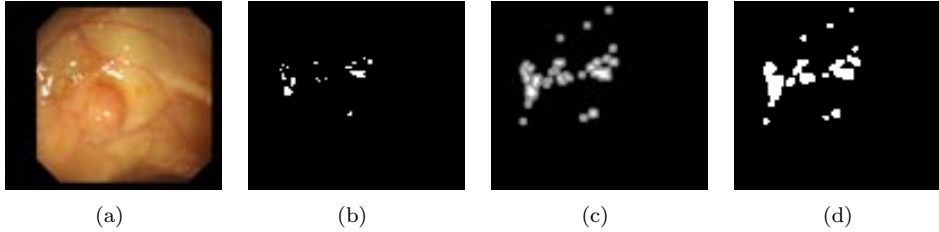


Figure 7.4: Inpainting of specular highlight masks. (a) Original image; (b) Specular highlights mask; (c) Extension of the specular highlights mask; (d) Dilated specular highlights mask.

Method	Operation
Specular Highlight Removal Using Bilateral Filtering [119]	Detection
Correspondence Search in the Presence of Specular Highlights Using Specular-Free Two-Band Images [120]	Detection
Extension of Specular Highlights Detection masks by observing neighborhoods around specular items (Our contribution, Chapter 4)	Detection
Automatic segmentation and inpainting of specular highlights for endoscopic imaging [120]	Detection and Inpainting
Automatic specular highlights inpainting by means of weighted contributions of the original and specularity-less image (Our contribution, Chapter 4)	Inpainting

Table 7.2: Methods use in the specular highlights detection and inpainting experiments.

We also tried the method described on [101] but after having several conversations with the author, it has been discarded because it does not suit our type of images.

Metrics of the experiment

The metrics that we will use to make the comparison between the different methods are:

- Specular highlights detection: We count the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) and we also provide two additional measures, the Detection Rate and the Error. We define Detection Rate as the percentage of specular highlights pixels that have been detected by each method and we define as Error the percentage of specular highlights pixels that have not been detected by each method.
- Specular highlights inpainting: As we want to obtain as a result an inpainting image where the intensity valleys caused by the specular highlights have less

impact, we will compute the sum of valley energy below the dilated mask before and after applying the inpainting operation.

Experimental Results

In our first experiment we compared the output of four specular highlight detection methods. Experimental results on specular highlights detection can be consulted at Table 7.3 and Table 7.4.

Method	TP	FP	TN	FN
[119]	561.890	208.956	83.669.107	612.371
[120]	697.358	698.172	83.179.891	476.903
[3]	826.203	1.461.264	82.416.799	348.058
Our contribution (Chapter 4)	868.210	1.457.537	82.420.526	306.051

Table 7.3: Comparison of specular highlight detection methods (I).

Method	DR %	Error %
[119]	53.04%	46.95%
[120]	42.12%	57.87%
[3]	81.44%	18.55%
Our contribution (Chapter 4)	84.2%	15.55%

Table 7.4: Comparison of specular highlight detection methods (II).

Results from Table 7.3 show that our contribution improves state-of-the-art specular highlight detection [3] in the two most important categories for detection, which are TP and FN, although we provide a slightly higher number of FP. There is another approach that wins in two categories ([119]) but in this case the good performance that it provides in terms of FP and TN is not compensated by the decrease in performance in terms of TP and FP. We can see more clearly in Table 7.4 the difference in performance by observing the Detection Rate and Error measures, where our contribution outperforms current state-of-the-art results, detecting above 84% of the specular highlights in the images.

Method	E_0	E_{inp}	$\%(E_{inp}/E_0)$
[3]	1083.99	574.38	52.98%
Our contribution	1083.99	445.84	41.13%

Table 7.5: Comparison of specular highlights inpainting methods.

We present in Table 7.5 experimental results on specular highlights inpainting. In this case we have only compared our method with the state-of-the-art on specular highlights inpainting in colonoscopy. There are other inpainting approaches, most of them linked with some of the specular highlights detection methods that we have compared, but after contacting several authors we have decided not to use these methods for comparison, as the type of images we are working on seem to generate problems with inpainting, as big part of the specular highlights are actually saturated and, because of this, reliable information under these pixels cannot be recovered by applying methods such as the one proposed in [101].

We have to mention that for this second experiment we have used as input to both algorithms the same specular highlights mask, so E_0 the valley energy under the dilated mask will be the same. This is done aiming at a fair comparison where the differences in performance are only caused by the inpainting algorithms and possible errors from specular highlights detection could be dragged. As it can be seen from Table 7.5, by using our method we improve the mitigation of specular highlights-originated valleys. More precisely, we can see that, by means of our method, the energy after inpainting under the dilated mask E_{inp} is lower than the one achieved with [3]. More precisely we can see that by our inpainting method we are able to reduce the energy under the dilated mask to a 41.13% of the original value whereas the method presented in [3] only reduces it to a 52.98%.

Analysis and discussion of the results

For the case of specular highlights detection, it is true that the difference in performance is minimal nevertheless we improve current state-of-the-art in two key categories which are TP and FN, although we obtain a higher number of FP. We show some qualitative examples on specular highlights detection in Figure 7.5. In this case we only offer visual comparison of the two methods that have been developed with specular highlights in colonoscopy in mind. As it can be seen from the Figure, there are some images where the difference is significant, although we have to mention that both approaches seem to be on par and we do not have to forget that our contribution extends the other method's detection, that is why the number of TP is higher for our contribution in the majority of images (243 out of the 300 images have a positive difference).

Finally, in order to end this section, we offer comparative results on specular highlights inpainting in Figure 7.6. As it can be seen, by using our method we obtain smoother results than by using the one proposed in [3]. More precisely we can see clearly in Figure 7.6 (b) and (e) the contours of the specular highlights in the original image whereas the contours are more difficult to see in Figure 7.6 (c) and (f).

We can conclude that by means of our specular highlights detection and inpainting we have improved the performance of current state of the art methods in the scope of colonoscopy video analysis. Therefore, we will use our complete specular highlights correction scheme as a preprocessing stage for our polyp localization and polyp segmentation methods.

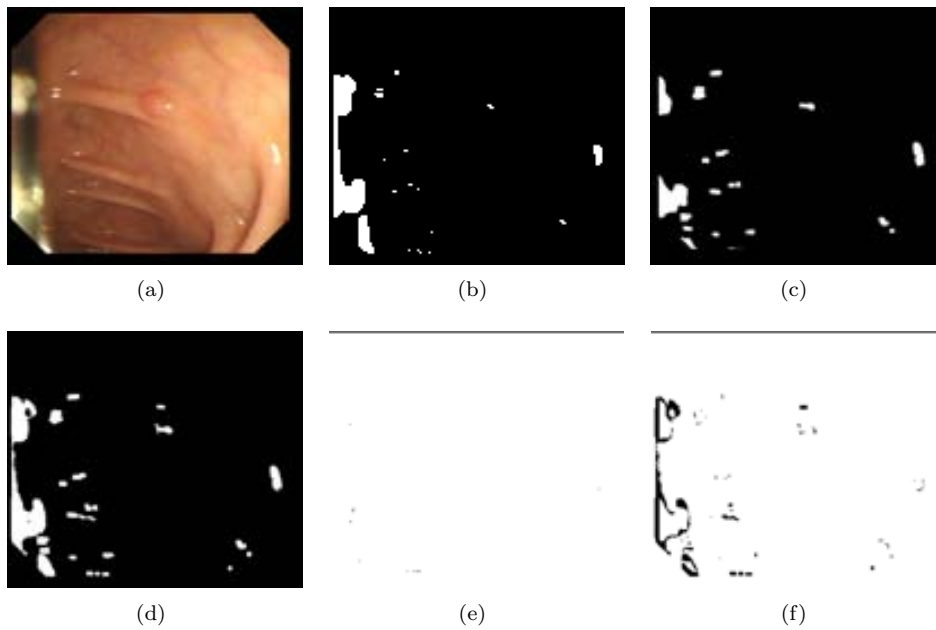


Figure 7.5: Specular highlights detection comparison. (a) Original image; (b) Ground truth; (c) [3] output (d) Our contribution's output. (e) Pixels detected by [3] but not by our contribution (complemented image to ease visualization of the difference pixels, shown here in black); (f) Pixels detected by our contribution but not by [3] (complemented image to ease visualization of the difference pixels, shown here in black).

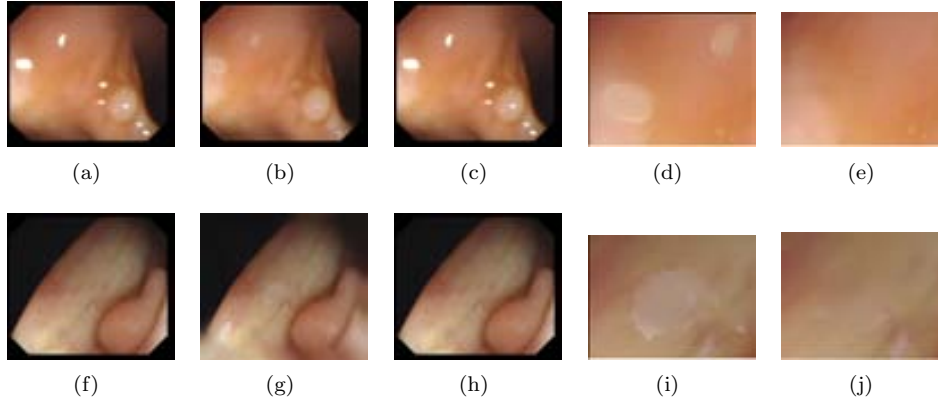


Figure 7.6: Specular highlights inpainting comparison. (First column) Original image; (Second column) Inpainting by means of [3] method; (Third column) Our contribution's inpainting (Fourth column) Zoom of results of [3] method (Fifth column) Zoom of results of our contribution's inpainting.

7.3.2 Preliminary study on blood vessels mitigation

Objective of the experiment

As mentioned in Chapter 3, one of the challenges that we must overcome when using valley detection for our polyp localization and polyp localization methods is that there are some other elements of the endoluminal scene which also generate valley information. We present in this Chapter a preliminary study on blood vessels mitigation, as an accurate detection and segmentation of blood vessels is out of the scope of this thesis.

Experimental setup

In order to study the effects that blood vessels have in the performance of our algorithms, we have created a blood vessels database, consisting of examples extracted from our CVC 3000 COLON DB database that are rich in blood vessels information (that is, presence of lots of blood vessels in the image). To do so, an expert has manually marked each pixel of the image that is part of a blood vessel. Some examples can be seen in Figure 7.7.

We will also use the polyp contour masks that were provided by experts. Our initial hypothesis consist of the assumption that blood vessels may potentially have a higher presence on certain color channels therefore we will test the effect of blood vessels in several well-known color spaces such as: sRGB, linear RGB, HSV, CieLAB or XYZ. Taking this into account, we show in Table 7.6 the 31 possible input images that we will use for the experiments detailed in this section:

Color space	Input images
sRGB	Grey scale (baseline), sR, sG, sB, sR-sG, sR-sB, sG-sB
IRGB	IR, IG, IB, IR-IG, IR-IB, IG-IB
XYZ	X, Y, Z, X-Y, X-Z, Y-Z
CieLab	L, a, b, L-a, L-b, a-b
HSV	H, S, V, H-S, H-V, S-V

Table 7.6: Possible input images.

where: sR, sG and sB are the three components of the sRGB image; IR, IG and IB are the three components of the linear RGB image; X, Y and Z are the three components of the XYZ image; C, a and b are three components of the CieLAB image and H, S and V are the three components of the HSV image.

The objective of this experiment is to test, by the two low-level image processing algorithms that we use to generate the DoV image (valley detection and morphological gradient), how we can mitigate the effects of blood vessels in the image without losing polyp information. To do so we will measure for each input image the difference in energy under both vessels and polyp contour masks from the original value obtained from the grey scale image. The final aim of this experiment is to find out if there is any input image where the blood vessels information is mitigated whereas polyp information is kept the same or enhanced.

Metrics of the experiment

We will use two measures in order to assess the performance of blood vessels in different color spaces. As we are using our Depth of Valleys images as input of our polyp localization and polyp segmentation algorithms, we will use the two tools that generate it, namely valley detection [67] and morphological gradient. In both cases we will check the energy under blood vessels masks in each different combination from all the possible input images that have been defined in 7.6.

The experiments will check for every possible input image the energy below blood vessels and polyp contour masks in order to compare the behaviour of both elements under different color spaces.

Experimental Results

Analysis of Valley information

For the valley detector we have used the following parameter values (Table 7.7). In this case we have considered that σ_d must be lower or equal than σ_i and that the difference between both parameter's values should not be high.

We have run a complete experiment to check if there is any difference in behaviour between pixels under polyp and vessels mask. In general, vessels valley information

Combination of σ	σ_d	σ_i
1	2	4
2	4	4
3	4	8
4	8	8

Table 7.7: Combination of parameters used for the ridges and valleys detector.

does not seem to depend highly on the parameters of the valley information whereas the relative increment/decrease for the polyps shows more difference, being the increase/decrease in the valley measure more strong when the sigma values are set to the higher values in combination 4. We present in Table 7.8 a summary of the most relevant results. In this case we aim at either an increase of blood vessel energy and decrease of polyp contour's energy or the contrary, a mitigation of blood vessels while enhancing polyp contour's information. We can observe from Table 7.8 that by using the B channel of the sRGB image we mitigate blood vessels' energy while enhancing polyp contours. We can also see that by using the difference between channels G and B of the sRGB image we achieve the contrary effect, mitigation of polyp contours and enhancement of blood vessels.

Measure	Input Image	Combination of parameters	Polyp	Vessel
Highest difference polyps-vessels	sB	4	20.33 %	-6.61 %
Highest difference vessels-polyps	sG - sB	4	-24.45 %	43.46 %

Table 7.8: Summary of results.

Analysis of Morphological Gradient information

In this case, and considering that we use the morphological gradient as one of the two necessary ingredients to cook the depth of valley images, we have taken the same 3 different values: [2, 4, 8] for the size of the structural disk than we take for the integration sigma, as both integration sigma value and size of the disk should be the same in order to work in the proper scale. After running a complete experiment, we have observed that in this case the morphological gradient information takes the following pattern: when we use the blue channel or the difference between sRGB's red and blue channels. In this case we can see that the size of the structural element has no impact on the results. As with valley information, we present in Table 7.9 a summary of the most important results of the global difference of performance for both polyp and vessels.

Measure	Input Image	σ	Polyp	Vessel
Highest diifference polyps-vessels	sB	4	9.21 %	-0.44 %
Highest difference vessels-polyps	sR - sB	4	-14.5 %	0.7 %

Table 7.9: Summary of results.

Analysis and discussion of the results

In Table 7.10 we show the best results (in terms of contrary difference in performance by using a certain image input and considering all the five types of contour information).

Type of contour information	Maximum difference polyps vs vessels	Input image
Valleys	26.95 %	sB
Morphological Gradient	9.65 %	sB
Type of contour information	Maximum difference vessels vs polyps	Input image
Valleys	67.91 %	sR, sG-sB %
Morphological Gradient	15.21 %	sR,sR-sB %

Table 7.10: Summary of results.

By observing this table we observe that the use as the input image the B channel leads to an enhancement for polyp contours information and a mitigation of blood vessels information for both valley detection and morphological gradient therefore we use this B channel as the input image for our algorithms. Finally we show in Figure 7.8 some examples of both the B and $G - B$ images for some of the database images.

7.4 Polyp localization results

7.4.1 Polyp localization results on the original input image

Objective of the experiment

Following the polyp localization processing scheme that was presented in Chapter 5, we offer in this section results for polyp localization over the CVC 300 COLON DB database. First of all we will present polyp localization results considering that no preprocessing operation has been done to the input image, in order to obtain results of the isolated performance of each method.

Experimental setup

For each of the polyp localization methods, we have set up a series of parameter values, which will be detailed next. As the tuning of parameters have been covered in our publications (EF-DOVA [11], SA-DOVA [13] and VO-DOVA [12]), we will only offer the best results achieved for all the combinations of a certain polyp localization method.

We carried out a research on finding the optimal parameter values for EF-DOVA [11]. The initial values of the parameters and the optimal are shown in Table 7.11.

SA-DOVA itself has only three parameters, but in order to find the combination of values that gives as a better results in terms of polyp localization, we added to

Parameter	Values	Optimal Value
DV threshold	[0.6, 0.8]	0.8
Distance	[20, 40]	40
Tol x	[8, 16]	8
Tol y	[8, 16]	16
Accum. threshold	[0.4, 0.6]	0.4

Table 7.11: Parameters of EF-DOVA.

the experimental setup the two parameters of the DoV image, differentiation and integration sigma [13]. Therefore, the parameters that SA-DOVA needs of are:

1. *Radius 1*: Minor radius of the sector.
2. *Radius 2*: Major radius of the sector.
3. *Number of sectors*: Number of sectors used to guide the accumulation method.

We have fixed 3 possible values for each parameter, which are shown in Table 7.12 along with the optimal value. In order to fix the optimal value for each parameter we performed a preliminary study on 30 different images from our database, which showed different views and types of polyps. For these images we run complete experiments varying the parameters values as shown in Table 7.12. The optimal parameters were those that lead to obtain the highest number of TP and lowest of FN, keeping FP number as low as possible.

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

Table 7.12: Parameters of SA-DOVA.

VO-DOVA uses the same parameters than SA-DOVA but it also adds one more which is the threshold difference between the angle of the sector and the orientation of the valley which maxima is found in this sector. As with SA-DOVA, we have fixed 3 possible values for each parameter, which are shown in Table 7.13 along with the optimal value.

Finally we show in Table 7.14 the possible parameter values for each parameter of WM-DOVA:

Metrics of the experiment

As it has been mentioned before, the objective of this experiment is to provide with several comparison criteria in order to decide which energy map is more discriminative

Parameter	Values	Optimal value
Differentiation sigma	[2, 4, 8]	8
Integration sigma	[4, 8, 10]	10
Rad 1	[30, 40, 50]	40
Rad 2	[80, 100, 120]	120
Number of sectors	[60, 120, 180]	180
Angle threshold	[5, 10, 15]	10

Table 7.13: Parameters of VO-DOVA.

Parameter	Values	Optimal value
Differentiation sigma	[2, 4, 8]	8
Integration sigma	[4, 8, 10]	10
Rad 1	[30, 40, 50]	40
Rad 2	[80, 100, 120]	120
Number of sectors	[60, 120, 180]	180
Window size	[4, 8, 12, 18, 20, 24, 36]	18

Table 7.14: Parameters of WM-DOVA.

in terms of polyp localization. In this case we will denote as good results those where the maxima of the DOVA energy map falls on the polyp mask. Consequently we will have a bad result whenever the maxima of the DOVA energy map falls outside the polyp mask.

Experimental Results

We offer a comparative of polyp localization results in Table 7.15:

Method	OK	OK [%]	No OK	No OK [%]
EF-DOVA	96	32%	204	68%
SA-DOVA	128	42.66%	172	57.33%
VO-DOVA	134	44.66%	166	55.33%
WM-DOVA	162	54	138	46

Table 7.15: Comparison between the different DOVA maps for polyp localization.

We can see from Table 7.15 how WM-DOVA offers the best performance, surpassing in more than a 20% the results of EF-DOVA, which coincides with our original hypothesis that constraining our method to a certain shape could lead to good particular result for those polyps that fit it, but a bad overall result. We can also see how WM-DOVA performs better than its radial counterparts SA-DOVA and VO-DOVA although in this case the difference is lower. Finally we can observe how VO-DOVA fails at performing best, damaged by a poor definition of valley orientation.

In order to provide more information to the comparison, we have calculated the

mean processing time per image. Results of this are shown in Table 7.16:

Measure	EF-DOVA	SA-DOVA	WM-DOVA	VO-DOVA
Number of experiments	9600	72900	218700	218700
Mean processing time per experiment (mins)	1521	15	80	30
Mean processing time per image (secs)	304.2	3	16	6

Table 7.16: Processing time comparison results.

Analysis and discussion of the results

Several conclusions can be extracted by observing Table 7.15:

- There is a big difference in results between the two groups, where EF-DOVA provides worse results than any other method for all the categories. As we can see in Figure 7.9, when the polyp boundary can not be approximated by an ellipse, EF-DOVA fails at locating its maxima of accumulation inside polyp mask for the majority of images.
- Once discarded the use of EF-DOVA for our polyp localization method, we continue by comparing the three remaining accumulation methods. As it can be seen from Table 7.15 the difference in performance between the three methods is not as big as between group of methods. We show in Figure 7.10 some examples where WM-DOVA has obtained a better localization result than SA-DOVA and VO-DOVA.
- VO-DOVA offers slightly better results than SA-DOVA but the difference in processing time is significant enough to propose its use among other faster alternatives.
- Finally, if we want our system to be incorporated into an intelligent system for colonoscopy (for instance, a CAD) we must be fast enough to provide a response to the input image. In this case SA-DOVA, WM-DOVA and VO-DOVA present comparable results, much faster than EF-DOVA. Between all the radial approaches, SA-DOVA is the fastest, followed by VO-DOVA and WM-DOVA.

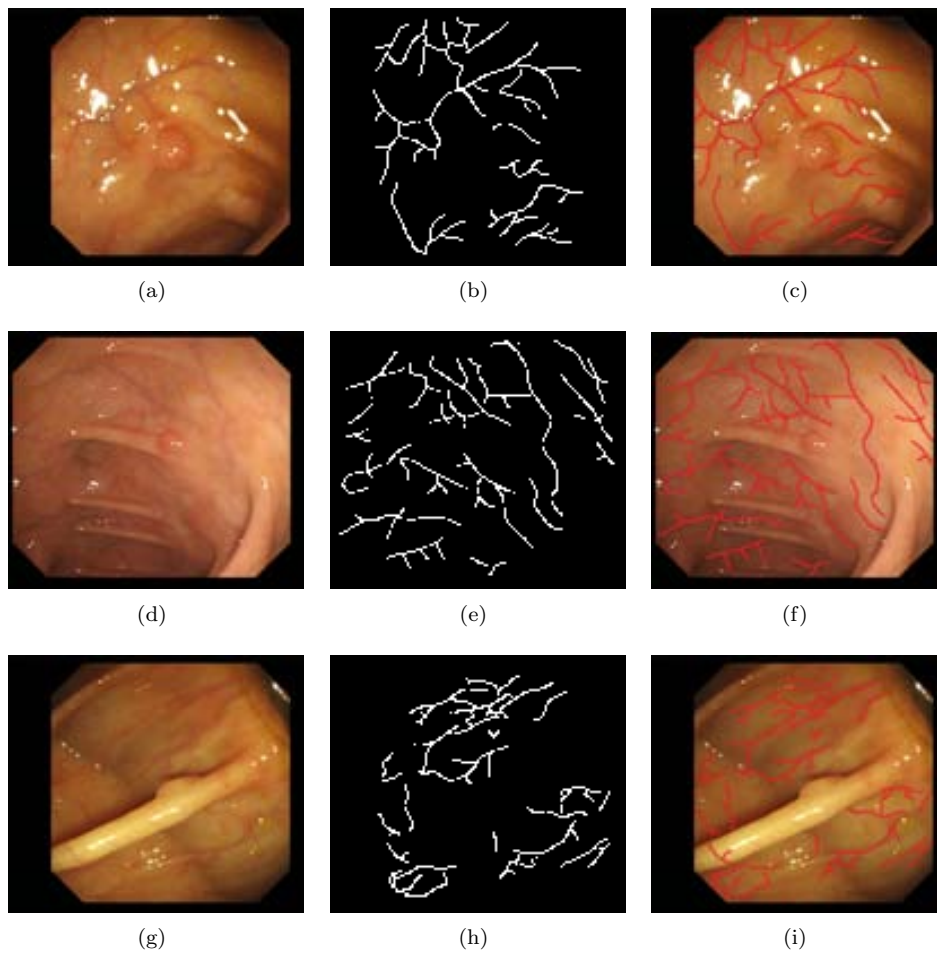


Figure 7.7: Examples of the blood vessels database: (1st column) Original images; (2nd column) Blood vessels masks examples; (3rd column) Blood vessels masks superimposed on the original image;.

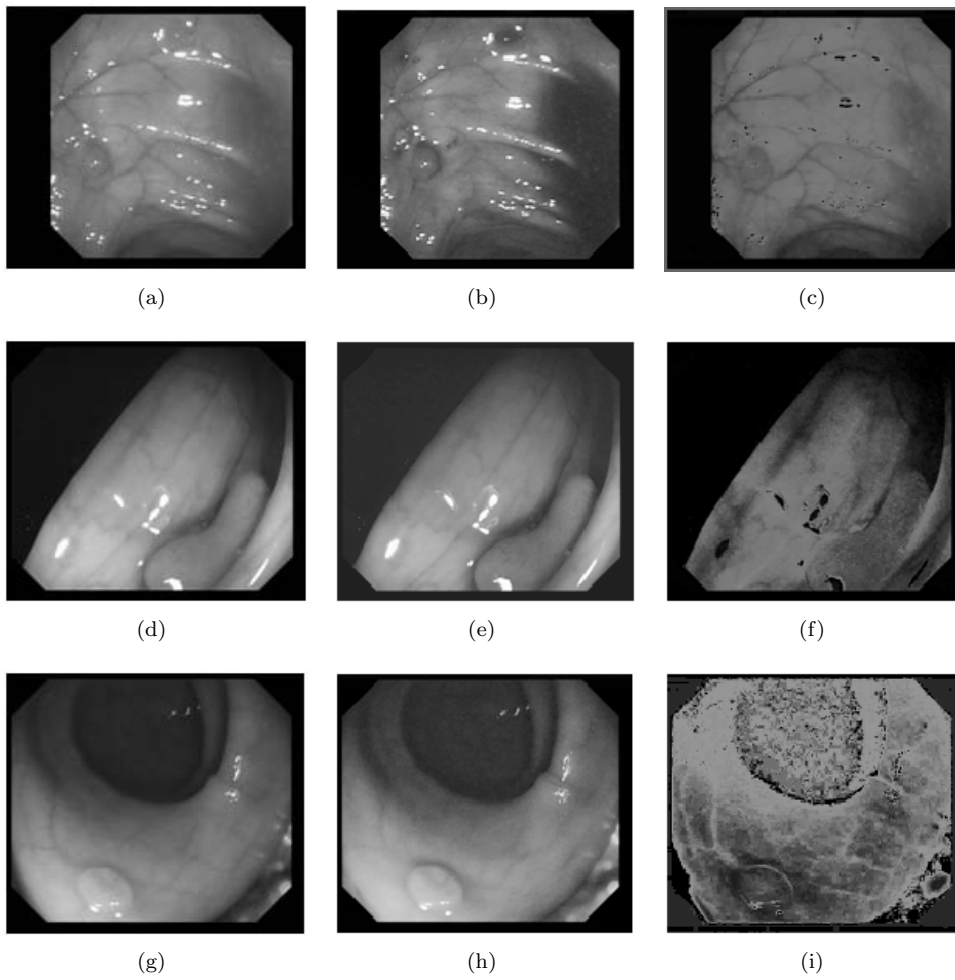


Figure 7.8: Blood vessel impact on several channels (left column) Original images; (center) Blue channel; (right column) Green - Blue channel.

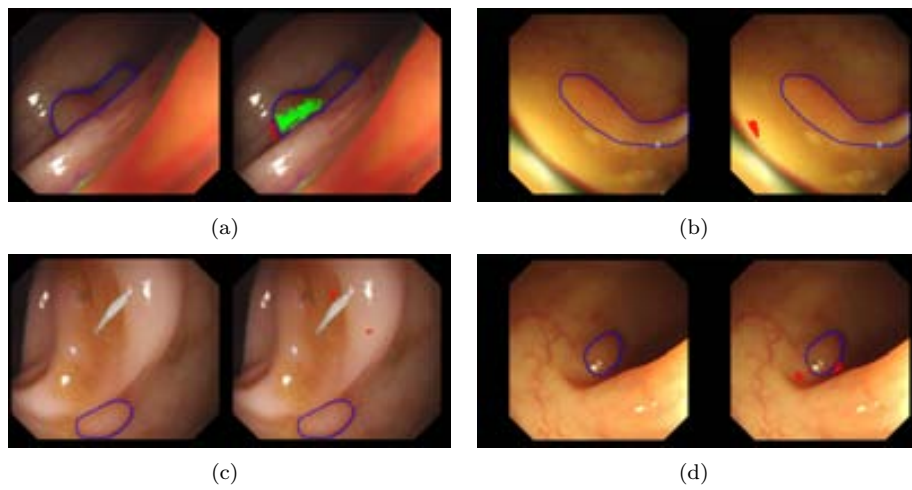


Figure 7.9: EF-DOVA performance results. We show for every example two images. The one on the left corresponds to the original image whereas we have superimposed in the image on the right the position of the maxima of accumulation, which is painted in green if it falls in the polyp mask and painted in red in the contrary case.

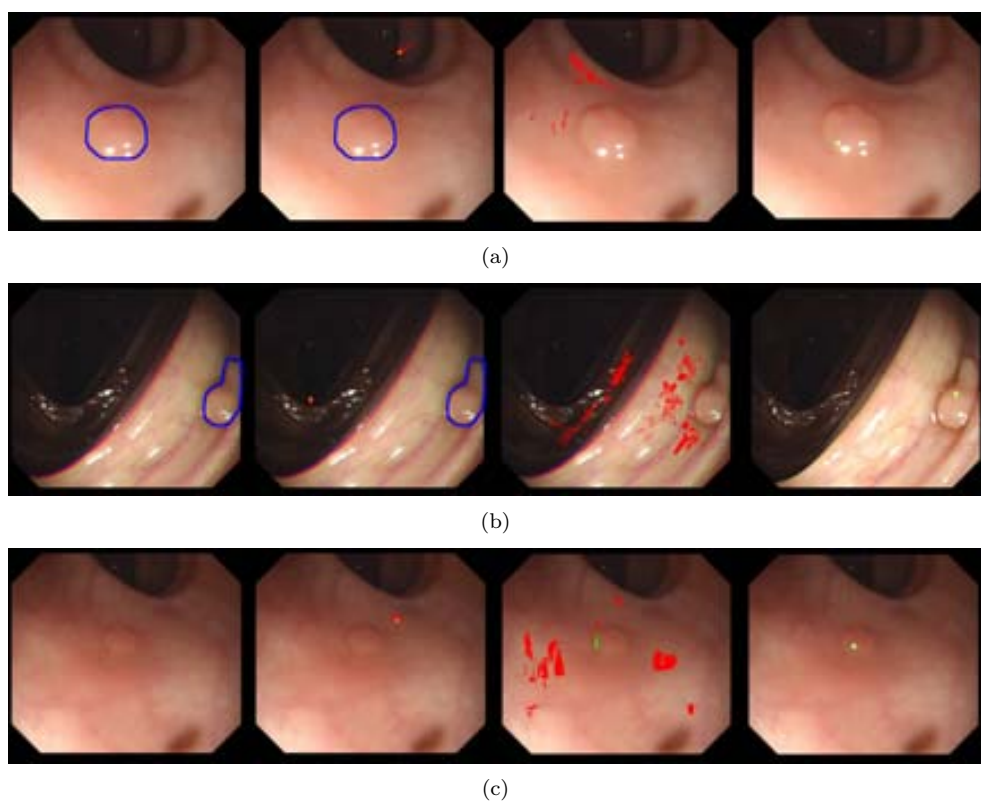


Figure 7.10: Improvement of WM-DOVA against SA-DOVA and VO-DOVA. (First column) Original image; (Second column) EF-DOVA results; (Third column) SA-DOVA results; (Fourth column) WM-DOVA results; (Fifth column) VO-DOVA results.

7.4.2 Impact of image preprocessing techniques on polyp localization results

Objective of the experiment

In this subsection we will measure the impact of each preprocessing technique (specular highlights detection and inpainting, black mask inpainting and blood vessels mitigation) in polyp localization results. We also measure the impact of a correct identification of non-informative regions may have in the performance of our polyp localization method.

Experimental Setup

In this case we will also use CVC 300 COLON DB, along with their corresponding polyp masks.

Metrics of the experiment

As the aim of this experiment is to assess the effect that some elements of the endoluminal scene have on overall polyp localization results, we will use the same metrics than in the previous section: number and percentage of good results and number and percentage of errors. We have to mention that we will only offer results for WM-DOVA energy map, as its superior performance has been proven in the previous subsection.

Impact of black mask inpainting

We see a comparison of results with and without black mask inpainting in Table 7.17:

Metric	Baseline	Baseline + Black mask inpainting	Difference
OK	162[54%]	163 [54.33%]	+1 [0.33%]
Error	138[46%]	137 [45.66%]	-1 [-0.33%]

Table 7.17: Impact of black mask inpainting on polyp localization results.

As we can see the inpainting of the black mask does not have a big impact on global polyp localization results. In fact there is only one image which benefits from the inpainting of the black mask (Figure 7.11).

Impact of specular highlights detection and inpainting

We present in Table 7.18 results on the impact of specular highlights detection and inpainting in polyp localization results.

In this case we can see that the correction of specular highlights do has an impact in polyp localization results. We have improved polyp localization in 11 images.

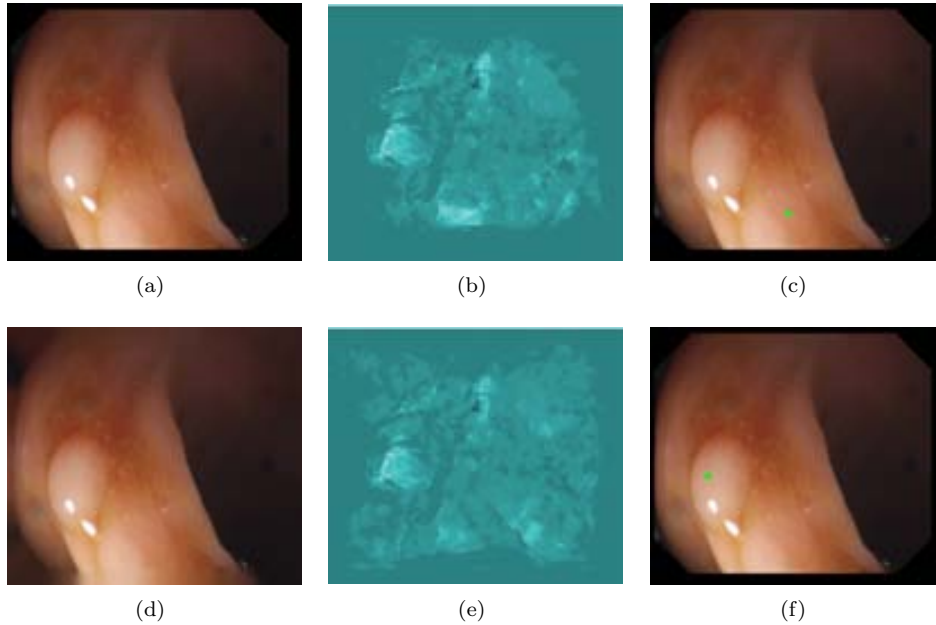


Figure 7.11: Improvement of polyp localization results by inpainting of the black mask. (a) Original image with black mask; (b) Original WM-DOVA energy map; (c) Polyp localization by means of WM-DOVA (maxima of WM-DOVA is painted as a green square); (d) Image with black mask inpainted; (e) WM-DOVA map after inpainting of black mask; (f) New polyp localization by means of WM-DOVA.

We show an example of this improvement in Figure 7.12, where we can see how the inpainting of specular highlights has an impact on the energy image, leading to bigger concentrations of energy inside polyps.

Impact of blood vessels mitigation

After a preliminary study we have concluded that they seem to have less presence in channel B or a higher contrast from polyp contours in the subtraction of channel G and channel B of the standard RGB image. We can observe polyp localization results by using as input channel sB or channel $sG - sB$ in Table 7.19 and Table 7.20:

In this case we have confirmed our initial assumption about using separate color channels to mitigate the effect of blood vessels. More precisely, we have shown in Table 7.20 how, by using the $G - B$ channel, we improve our polyp localization results in 11 images whereas if we use the B channel we improve our results in 20 images. We show some examples on blood vessels-rich images in Figure 7.13.

As can be seen from Figure 7.13 in images with heavy presence of blood vessels by using information of the B channel we are able to mitigate their presence and change the maxima of accumulation which, in this case, gives a clear indication on where the polyp is in the image.

Metric	Baseline	Baseline + Specular Highlights corrected	Difference
OK	162[54%]	173 [57.66%]	+11 [3.66%]
Error	138[46%]	127 [42.33%]	-11 [-3.66%]

Table 7.18: Impact of specular highlights on polyp localization results.

Metric	Baseline	Baseline + channel sB	Difference
OK	162[54%]	182 [60.66%]	+20 [6.66%]
Error	138[46%]	112 [39.66%]	-11 [-6.66%]

Table 7.19: Impact of blood vessels mitigation on polyp localization results (sB channel as input).

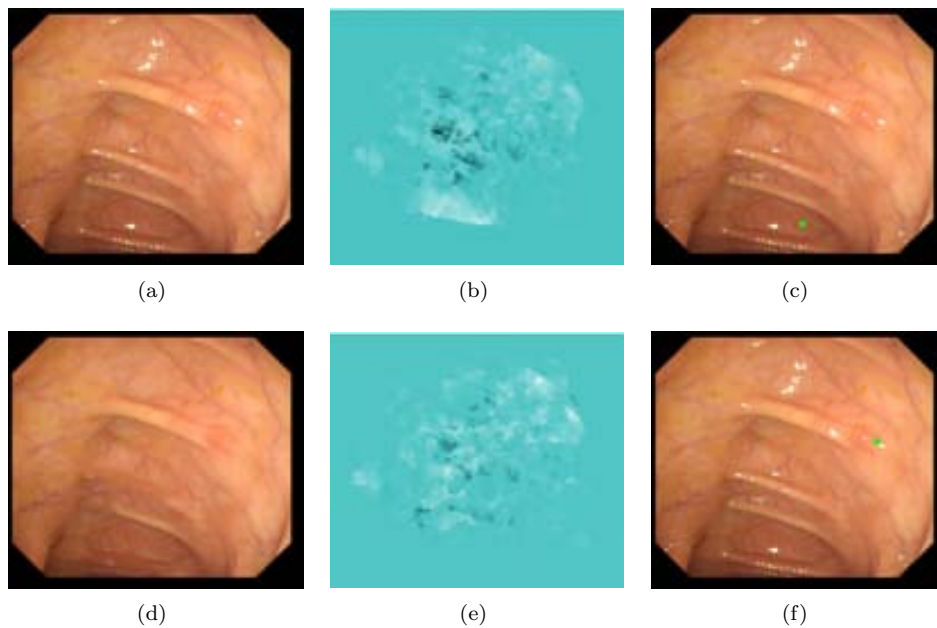


Figure 7.12: Improvement of polyp localization results by correcting specular highlights. (a) Original image with specular highlights; (b) Original WM-DOVA energy map; (c) Polyp localization by means of WM-DOVA (maxima of WM-DOVA is painted as a green square); (d) Image with specular highlights inpainted; (e) WM-DOVA map after inpainting of specular highlights; (f) New polyp localization by means of WM-DOVA.

Metric	Baseline	Baseline + sB-sG	Difference
OK	162[54%]	173 [57.66%]	+11 [3.66%]
Error	138[46%]	127 [42.33%]	-11 [-3.66%]

Table 7.20: Impact of blood vessels mitigation on polyp localization results (sG-sB as input).

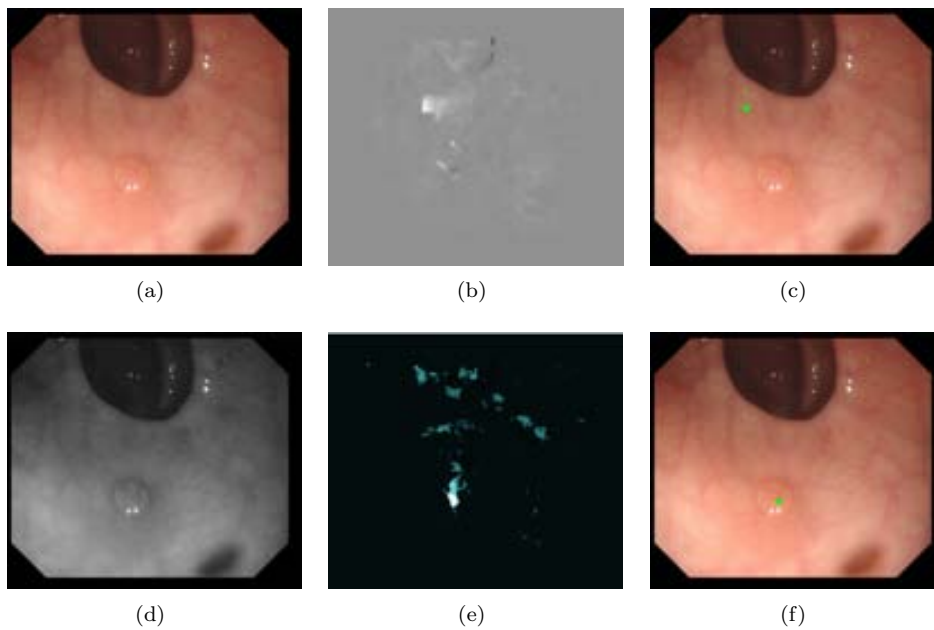


Figure 7.13: Improvement of polyp localization results by mitigating blood vessels effect. (a) Original image; (b) Original WM-DOVA energy map; (c) Polyp localization by means of WM-DOVA (maxima of WM-DOVA is painted as a green square); (d) Channel B of the original image; (e) WM-DOVA map after mitigating blood vessels effect; (f) New polyp localization by means of WM-DOVA.

Impact of non informative region identification

As mentioned in Chapter 6, one of the steps taken in order to reduce the number of final region is non informative region identification. In this experiment we will take into account this small part of the whole polyp segmentation method in order to check if polyp localization results could improve by just eliminating all the information belonging to these areas on the Depth of Valleys image. We present experimental results in Table 7.21:

Metric	Baseline	Baseline + Non-informative	Difference
OK	162[54%]	172 [57.33%]	+10 [3.33%]
Error	138[46%]	128 [42.66%]	-10 [-3.33%]

Table 7.21: Impact of non informative region identification on polyp localization results.

In this case we can see that the elimination of the contribution from non-informative regions do has incidence in polyp localization results. We have improved polyp localization in 10 images. We show an example of this improvement in Figure 7.14.

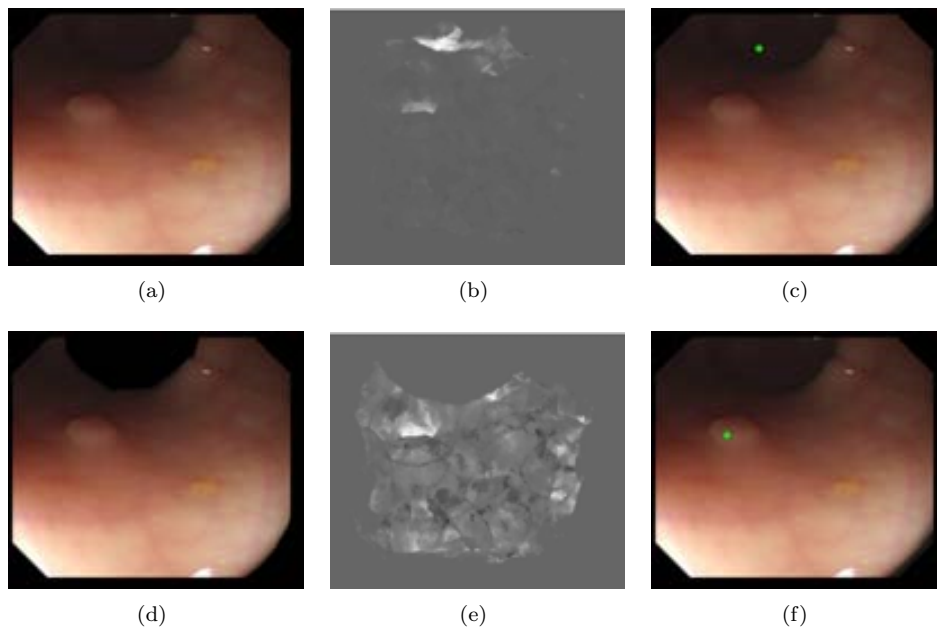


Figure 7.14: Improvement of polyp localization results by suppressing contribution of non-informative regions. (a) Original image; (b) Original WM-DOVA energy map; (c) Polyp localization by means of WM-DOVA (maxima of WM-DOVA is painted as a green square); (d) Image with non informative regions suppressed; (e) WM-DOVA map after eliminating contributions from non-informative regions; (f) New polyp localization by means of WM-DOVA.

Combined effect

We have shown the individual impact of image preprocessing and non-informative region identification in polyp localization and to finish this section, we present results obtained by combining all the techniques presented. Experimental results can be consulted at Table 7.22.

Metric	Baseline	Baseline + Preproc + Non-Inf	Difference
OK	162 [54%]	208 [69.33%]	+10 [15.33%]
Error	138 [46%]	92 [31.66%]	-10 [-15.33%]

Table 7.22: Combined impact of image preprocessing and non informative region identification on polyp localization results.

As it can be seen from the results, by combining all the methods that we have presented earlier, we improve our polyp localization results in 46 images, which correspond to an increment of 15% from our original results, leading to an almost 70% of success on polyp localization by only using the maxima of the WM-DOVA energy map. Of course, these results can be improved and we will show later negative examples in order to find out the causes of our errors. First of all we show in Figure 7.15 some examples where we have improved the results by combining all 3 enhancement methods.

As it can be seen from Figure 7.15 we have improved our polyp localization results even on images where many of the challenges that we mentioned back in Chapter 3 are present. For instance, we can see that Figure 7.15 (a) presents a high number of specular highlights and blood vessels and we are able to place the maxima of accumulation in the very center of the polyp. Our method also works for images with bad quality, such as Figure 7.15 (c) where the blurring leads to have specular highlight shadows that even touch the polyp boundary. Another of the challenges that we mentioned in Chapter 3 was that sometimes we do not have a zenithal view of the polyp but a lateral one, which could potentially cause a problem for our method which is based on the presence of valleys. We prove the efficiency of use of the Depth of Valleys image by taking a look at Figure 7.15 (d). We can see for this example how the initial high influence of specular highlights lead to a misplacing of the maxima of DOVA energy map, which is solved by applying the corrections before mentioned.

It is true that our method still have mistakes, images where the maxima of polyp localization is not placed inside polyp mask. We show some examples in Figure 7.16.

By observing some of the mistakes of our polyp localization method, we can sketch which future lines of research could be taken in order to improve our current results. For instance, we can observe in Figure 7.16 how the presence of folds and wrinkles do have an impact in our method. This can also be observed in Figure 7.16 (b) and somehow in Figure 7.16 (d), where we can see that even by inpainting both specular highlights and black mask and using the channel B of the image, we still fail at localizing the polyp. There are some other elements that can have an impact on our method and have not been covered on this thesis. The example shown in

Figure 7.16 (c) is a clear one: we have a lateral polyp which touches lumen which has high presence of fecal content. We have not covered at all lumen or fecal content detection and, although our results improve by applying the corrections we still do not locate the maxima of WM-DOVA inside the polyp. Finally, the case of Figure 7.16 (d), apart from highlighting the effect of folds, also points out the necessity of evolving our algorithms to cover this kind of sessile polyps in images where they do not protrude much from mucous membrane.

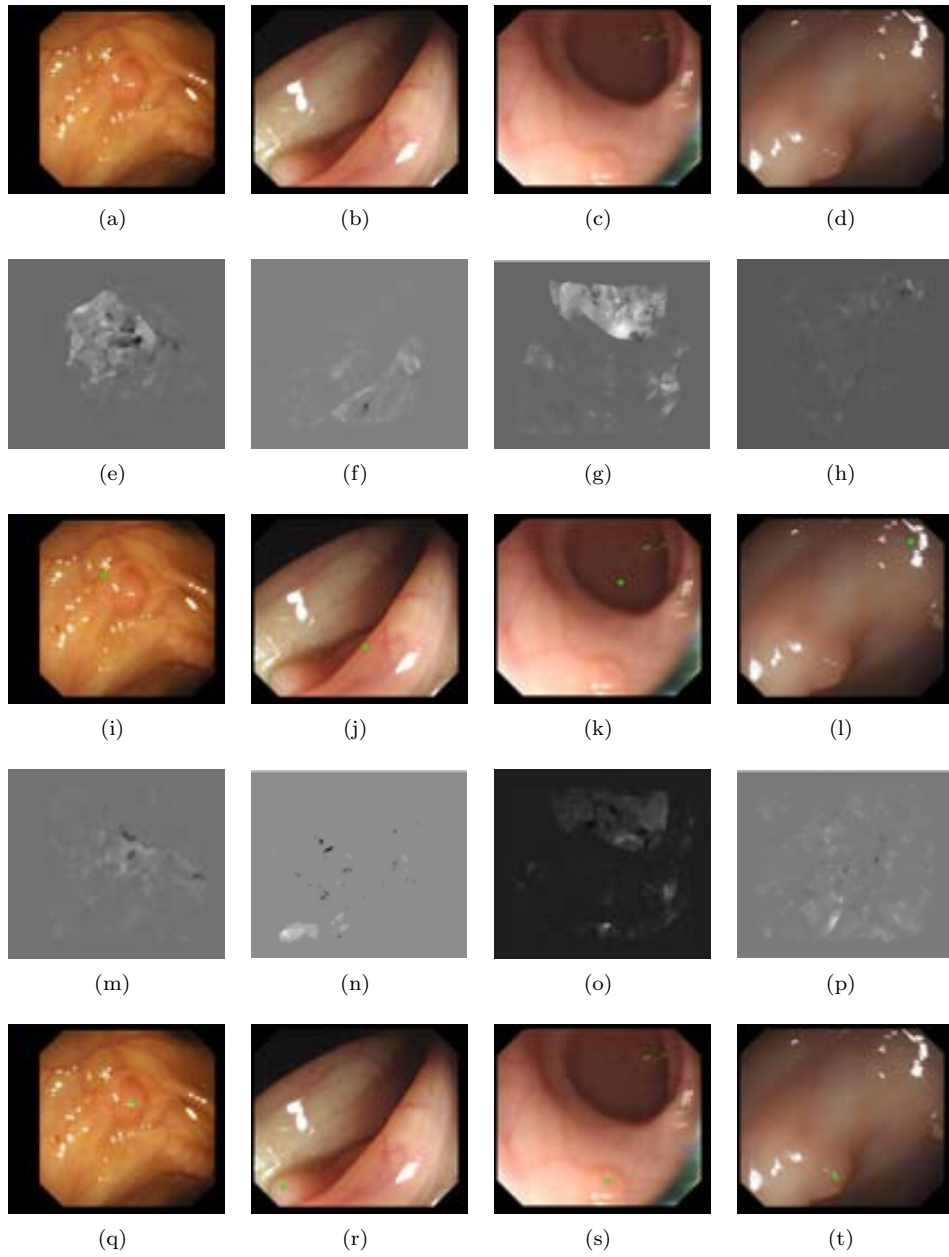


Figure 7.15: Improvement of polyp localization results by combining all the methods explained. (First row) Original images; (Second row) Original WM-DOVA energy map; (Third row) Polyp localization by means of WM-DOVA (maxima of WM-DOVA is painted as a green square); (Fourth row) WM-DOVA map after applying all methods; (Fifth row) New polyp localization by means of WM-DOVA.

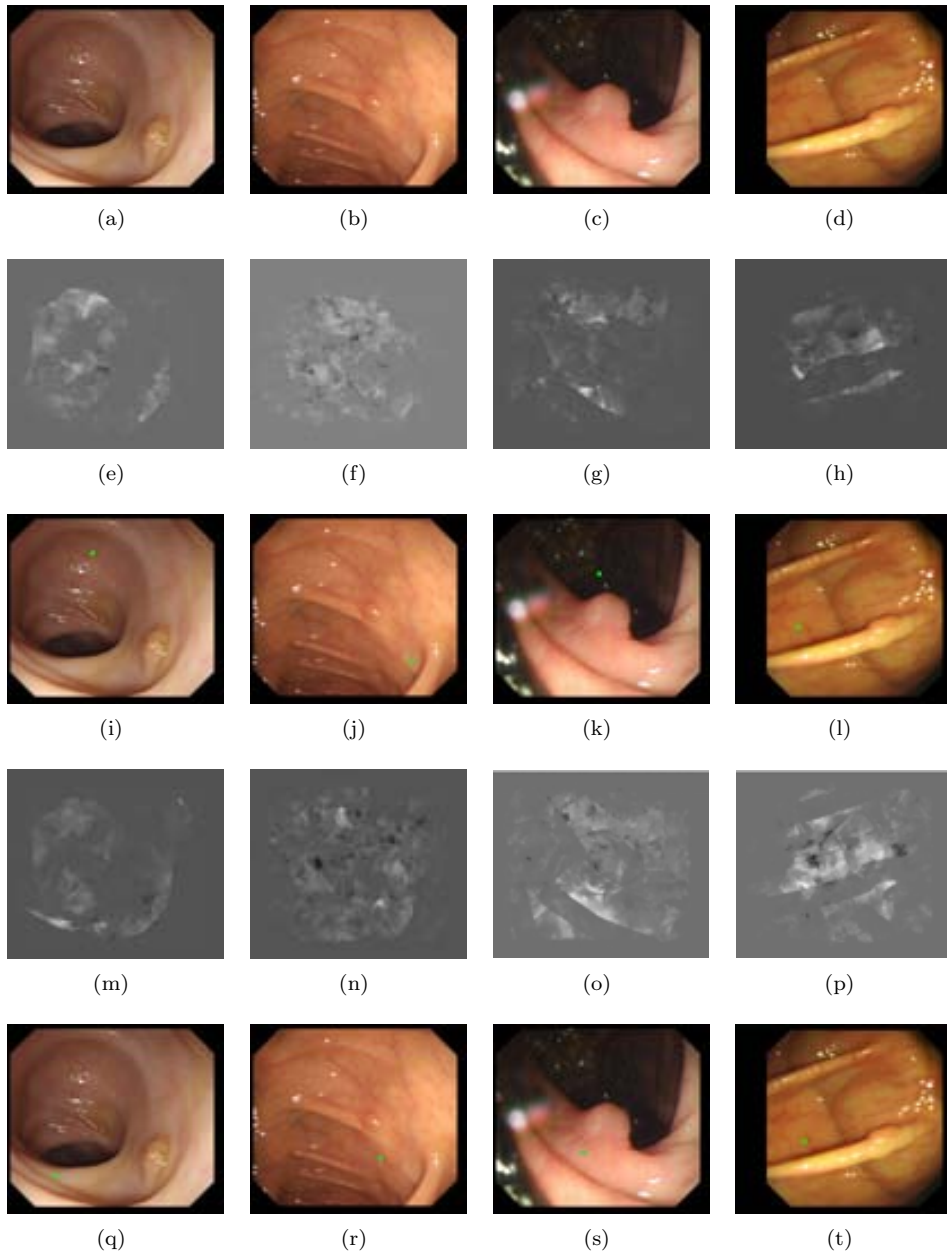


Figure 7.16: Errors of polyp localization results obtained even by combining all the methods explained. (First row) Original images; (Second row) Original WM-DOVA energy map; (Third row) Polyp localization by means of WM-DOVA (maxima of WM-DOVA is painted as a green square); (Fourth row) WM-DOVA map after applying all methods; (Fifth row) New polyp localization by means of WM-DOVA.

7.5 Polyp Segmentation results

We present in this chapter by exposing our polyp segmentation results. In this case we propose two different methods: polyp segmentation after region merging and polyp segmentation from the output of an energy map.

Objective of the experiment

The objective of this experiment is to check which polyp segmentation algorithm provides better results. In this case, a good result will be a final segmented region which covers the polyp in great extent. However we will not accept as a result a big region that covers a great part of the image and, by chance, the polyp. More precisely we want our final polyp segmentation to fit the original polyp's shape as much as possible.

We will make two types of analysis in order to compare the alternatives that we will present. The first analysis will be focused only on pure segmentation results via specific metrics. The second analysis will compare classification results, understanding in this case that, for every image, we will have only two regions: the polyp region and the background. The polyp region will be the one where the maxima of WM-DOVA energy map falls whereas the background region will be the rest of the image. As it can be seen, by using this scheme we balance the number of positive and negative examples, which in this case makes it possible to use general performance metrics, as it will be explained in the next subsection. Before entering with segmentation results, we will make a brief introduction to another segmentation techniques which we have compared our methods with.

Approximation to the segmentation problem

In general segmentation, which is one of the most difficult and critical tasks in computer vision, can be viewed as a perceptual grouping problem in which the image is divided into homogeneous regions, which can represent different features in the images depending on the methodology adopted. Some simple ways of segmentation exist however they prove to be over simplified for semantic grouping of image regions in more complex scenarios, as they are more sensitive to noise and other artifacts [91]. More sophisticated methods of image segmentation can be mainly divided into two different categories: segmentation by fitting and segmentation by clustering [39]. In the former, the problem of segmentation is viewed as an assertion that the pixels in an image conform to a model while, in the latter, the pixels are grouped according to some criteria such as gray level, color or texture. In order to perform efficiently, segmentation by fitting methods need strong gradient differences pertaining to the objects in the images which have to be segmented, which is not our case. Given that we want to segment informative regions containing polyps from clinically uninteresting areas, methods that segment by clustering seem well suited for our scenarios. Doing an extensive review of the state-of-the-art on segmentation is out of the scope of this

thesis and we will only make a brief review of some techniques whose performance is similar that the want our methods to have. Because of this, we have chosen three methods from this group to carry out our research and we will compare the output of our polyp segmentation methods with the output that some of this methods provide:

- Normalized Cuts: The *normalized cuts* method [98] is a graph theoretic approach for solving the perceptual grouping problem in vision. In *normalized cuts*, all the sets of points lying in the feature space are represented as a weighted, undirected graph. The weight of each arc is assigned using a set of pre-defined criteria. These can be based on the spatial distance among the pixels, their brightness values, etc. Usually the easiest way to perform segmentation in graph theoretic algorithms is to disconnect the edges having small weights usually known as the *minimum cut* [25].
- Watersheds: Watershed transformation [112] is one of the clustering based methods used as a tool for image segmentation. Watersheds operate on intensity gradients to perceive an image as a combination of catchment basins in a hilly area (a hill corresponds to high gradient) simulating the formation of image regions with projected flow of water. After identification of an intensity valley in an image, region growing algorithms are used to combine all the pixels which have similar intensities.
- Turbo pixels: this algorithm [62] start by computing a dense oversegmentation of an image by means of a geometric-flow-based algorithm. This first segmentation results on segments that respect local image boundaries while limiting undersegmentation by using a compactness constraint. The method is guided by five basic principles, namely: 1) Uniform size and coverage of the final segmented regions; 2) Connectivity: each final region should represent a simply connected set of pixels; 3) Compactness; 4) Smooth and edge-preserving flow and 5) No overlap between the final regions.

Metrics of the experiment

As mentioned before, we will make two different analysis of the results and each of them needs of different metrics. For the case of segmentation results, we will use AAC and DICE. We evaluate the performance of our segmentation methods by using two different measures: Annotated Area Covered (AAC) (Eq. 7.1 and Dice Similarity Coefficient (DICE) (Eq. 7.2) [91].

$$AAC = 100 \cdot \frac{\#oftruepolypixels}{\#oftotalpolypixels}; \quad (7.1)$$

$$DICE = 100 \cdot \frac{\#oftruepolypixelsinthefinalregion}{\#oftotalpixelsofthefinalregion}; \quad (7.2)$$

Both measures are complementary, as the former calculates the amount of annotated polyp area while the latter complements it with the amount of non-polyp

information that is kept in the region. We can see in Figure 7.17 two different segmentation results that can help to understand the decision about using these two concrete measures. The segmentation results for the first image (Figure 7.17 (a-c)) is very good in AAC but bad in terms of DICE, since the region that contains the polyp also contains lots of non-polyp information. In the other hand, segmentation results for the second image (Figure 7.17 (d-f)) may be a little worse in terms of accuracy but the final polyp region will contain more polyp than non-polyp information.

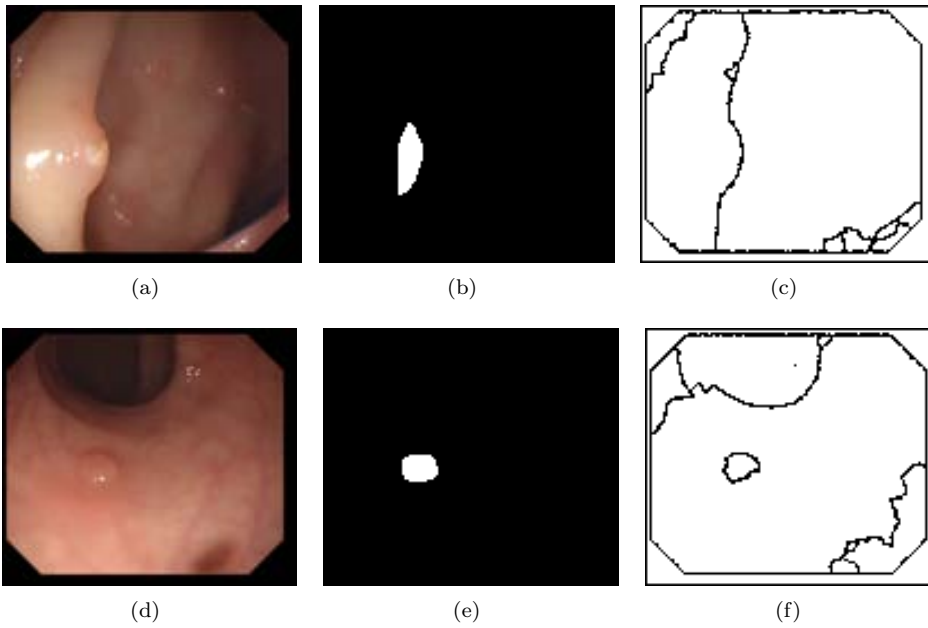


Figure 7.17: Examples of AAC and DICE results: (a) and (d) Original image; (b) and (e) polyp masks (the polyp is shown in white), and (c) and (f) segmentation results.

For the second type of analysis we will use the well-known classification metrics, which are:

$$Precision = \frac{TP}{TP + FP}; \quad (7.3)$$

$$Recall = \frac{TP}{TP + FN}; \quad (7.4)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}; \quad (7.5)$$

$$Specificity = \frac{TN}{TN + FP}; \quad (7.6)$$

$$Fallout = \frac{FP}{FP + TN}; \quad (7.7)$$

$$F2measure = 5 \cdot \frac{TP}{5 \cdot TP + 4 \cdot FN + FP}; \quad (7.8)$$

In this case a True Positive (TP) will correspond to a final segmented region which contains the polyp and on which the maxima of WM-DOVA falls. A False Positive (FP) will correspond to a final segmented region where the maxima of WM-DOVA has fallen but does not contain a polyp inside. A True Negative (TN) will consist of assigning correctly as non polyp containing region the background of the image and consequently a False Negative (FN) will correspond to an erroneous assignment of the polyp region by means of the maxima of WM-DOVA falling outside the polyp. Considering the domain of application of our algorithms, we will be interested in a high number of TP and TN and a low number of FP and FN but in this case, as it will be shown later, we will also take into account the size of the region with respect to the polyp size.

Experimental Results

Segmentation Results

In this subsection we will compare the results achieved by using the current state of the art on polyp segmentation with the ones achieved by several techniques, including our new contribution. We present a comparison of our polyp segmentation after region merging with the output of other segmentation methods in Table 7.23:

Measure / Method	Polyp Segmentation after Region Merging	NCuts	TurboPixels
AAC	70.29%	69.06%	69.2%
DICE	44.6%	37.75%	38.04%

Table 7.23: Comparison between the results obtained by our method, Normalized Cuts and TurboPixels with respect to the depth of valleys threshold.

As mentioned in [13], our first polyp segmentation method constitutes up to our knowledge the state-of-the-art on polyp segmentation and considering that it outperforms both normalized cuts and turbo pixels in terms of AAC and DICE (see Table 7.23), we will use its results as the starting point for the comparison with our new method. We can see that, by using our first method, we achieve high AAC results, which indicate that our final regions cover, in mean, a 70% of the polyp. But as we can see, DICE results are not that good. In our final regions the percentage of polyp content out of all the region is below 50%, that is, our final regions are composed by more non-polyp information than by polyp information. We show some qualitative comparison results in Figure 7.18.

Therefore our efforts should be concentrated on improving this DICE results. From this point we will compare the previous state-of-the-art results on polyp segmentation, achieved by our polyp segmentation after region merging method, with the ones

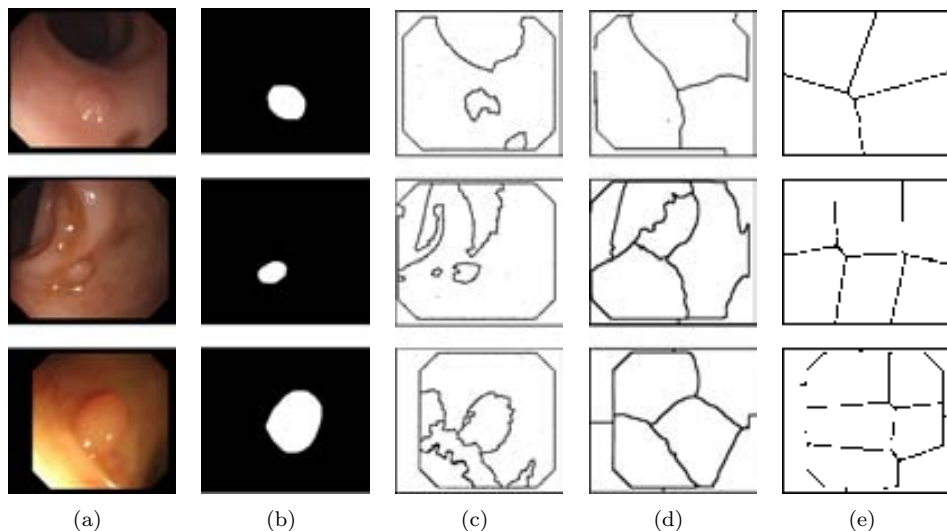


Figure 7.18: Comparison of segmentation results: (a) Original images; (b) polyp masks; (c) polyp segmentation after region merging output; (d) normalized cuts' output, and (e) turbo pixels' output [13].

obtained with our new proposal: polyp segmentation from the output of an energy map. Taking this into account, we offer a comparison of segmentation results between our two methods in Table 7.24.

Measure / Method	Polyp segmentation after region merging	Polyp segmentation from DOVA map
AAC	70.29%	63.98%
DICE	44.6%	52.53%

Table 7.24: Comparison between the results obtained by our two methods.

We can see from Table 7.24 that by using our new method we outperform current state-of-the-art in terms of DICE, although we are worse in terms of AAC. That means that our final regions contain more polyp information with the new method, but they fit worse the shape of the polyp. We show some comparison examples in Figure 7.19. As can be seen from the image, there are some cases where the first method offers better performance than the second (Figure 7.19 (b) and (c)) and some cases where we obtain the opposite result (Figure 7.19 (e) and (f)).

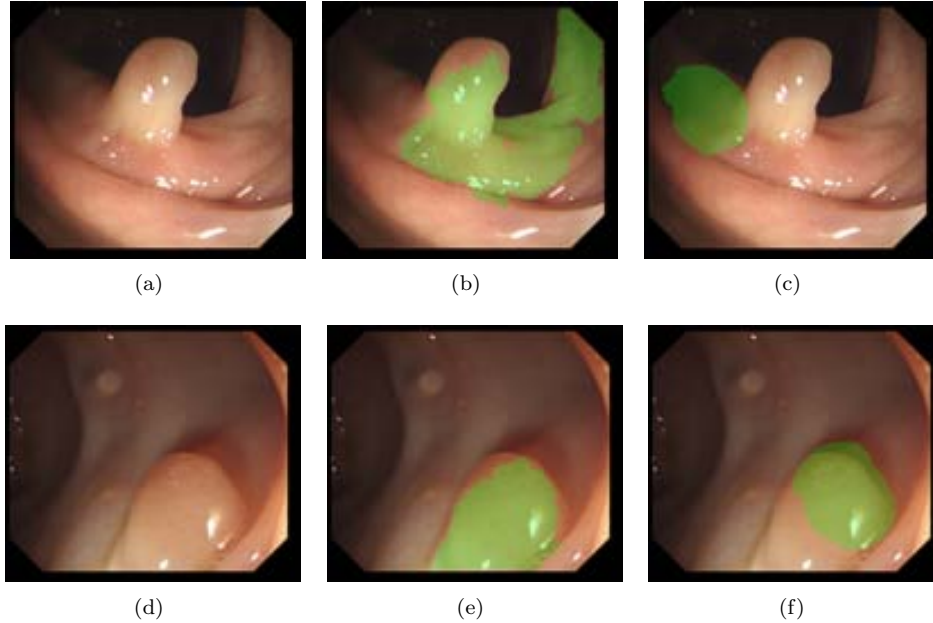


Figure 7.19: Comparison of segmentation results after applying our two segmentation methods: (a) and (d) Original image; (b) and (e) Segmentation results achieved with method proposed in [13], and (c) and (f) Segmentation results obtained with the new method.

But we have to take into account that both results are calculated according to the placing of DOVA maxima, which affects the segmentation results specially the ones obtained by the second method. In this case, if the placing of DOVA maxima is incorrect, the second method will find more difficulties to capture part of the polyp than the second, which only decides which is the final region by selecting the one where the maxima of DOVA falls. This can be better understood by taking a look at Figure 7.20.

We can see from Figure 7.20 (b) that in this case the maxima of WM-DOVA is placed outside the polyp. For the case of the second segmentation it means that the polyp region will contain hardly any part of the polyp, as it is calculated from the points that contributed to the maxima and in this case they are also placed outside the polyp. But by using the segmentation method proposed in [13] even considering the misplacing of the polyp, the final segmentation covers great part of it, despite of being a very rough segmentation that seems to cover incidentally the polyp.

In order to really assess the performance of both segmentation methods, we have decided to offer segmentation results only for the images where we place the maxima inside the polyp. By doing this we can really check if the final polyp regions obtained by both methods capture both the shape and content of the polyp. We show comparison results in Table 7.25.

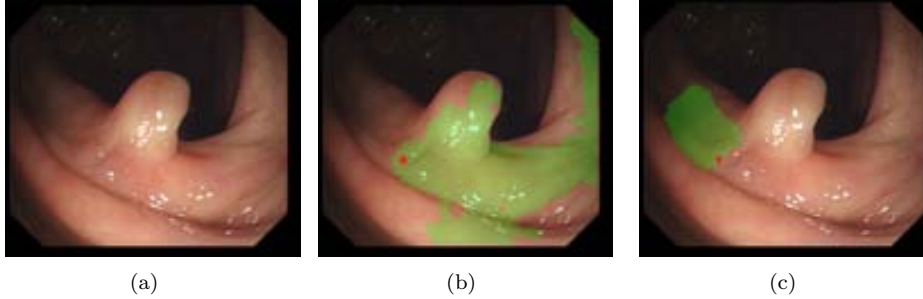


Figure 7.20: Comparison of segmentation results after applying our two segmentation methods: (a) Original image; (b) Segmentation results achieved with method proposed in [13], and (c) Segmentation results obtained with the new method.

	Mean AAC	SD AAC	Mean DICE	SD DICE
Polyp segmentation after region merging	59.95%	41.19	30.77%	21.33
Polyp segmentation from DOVA map	83.86%	19.24	74.81%	15.68

Table 7.25: Comparison between the results obtained by our two methods only for images with correct placing of maxima of DOVA.

As we can observe from Table 7.25 by only comparing segmentation results from the 208 images of our database where we place the maxima of WM-DOVA inside the polyp, our new method outperforms current state-of-the-art, specially in terms of DICE results, where the performance of the method exposed in [13] decreases as we do not consider as good results like the ones shown in Figure 7.20. We can also see from Table 7.25 that the results achieved by the new method do not only outperform the ones obtained by the first one in terms of mean value of both AAC and DICE but we can also see that the difference is also high in terms of standard deviation. As we have mentioned before, for the case of the first method we can have a final segmented region that covers almost all the frame and, by incidence, the polyp. This will result on a high AAC value, almost 100% for this particular image but the DICE value will be very low. We present some final segmentation results obtained with the final proposed method, polyp segmentation from the output of an energy map, in Figure 7.21. We can see how by means of the new segmentation method we are able to get good final polyp regions, no matter how the view of the polyp is. We can see examples from zenithal polyps (see Figure 7.21 (b)), images with high specular and blood vessel content (Figure 7.21 (c)), lateral views of polyps (Figure 7.21 (h), (n) and (p)) and even in poor quality images such as Figure 7.21 (l) we are able to provide with a polyp region that fits the area that the polyp occupies on the original image.

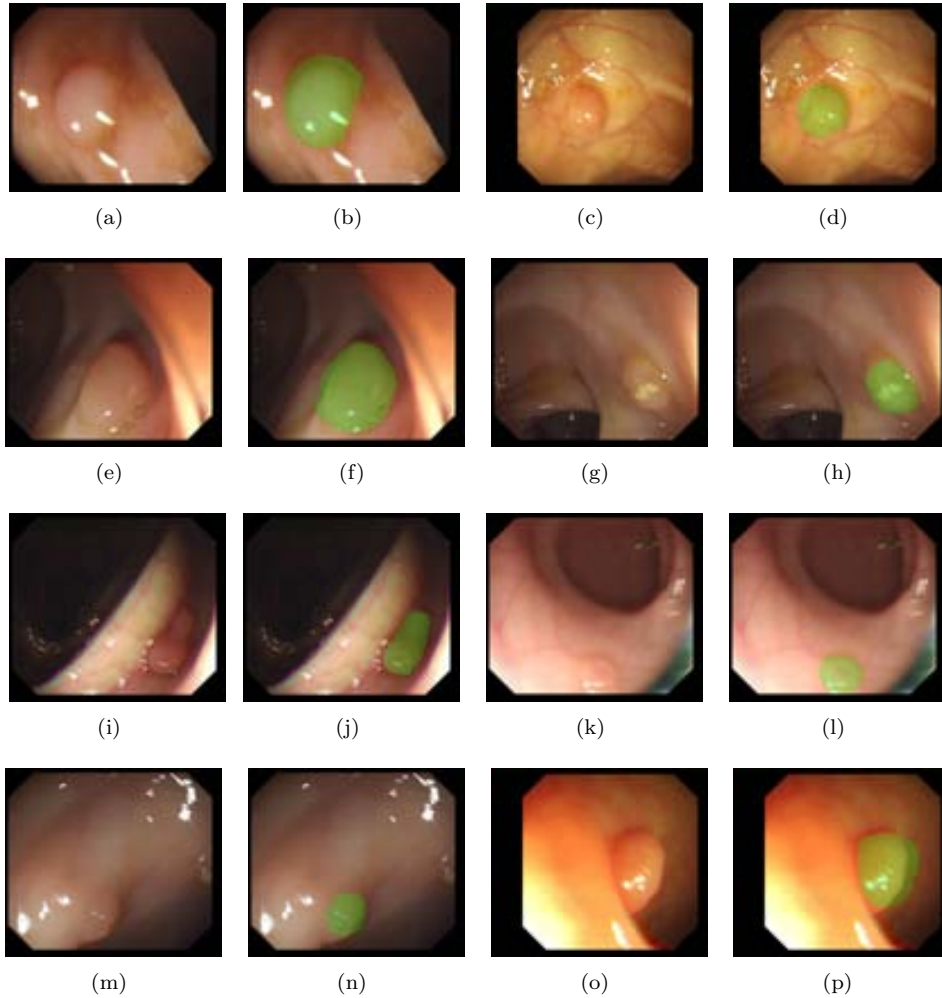


Figure 7.21: Segmentation results obtained by polyp segmentation from the output of an energy map.

Classification results

This final experiment treats the polyp segmentation as a classification problem in which, given an input image, we classify the final regions into polyp containing or not. In order to fit our experimental setup to this classification paradigm, we will assume the following:

- For every frame from the database, our polyp segmentation method will deliver only two regions as output: polyp region and background. For regions obtained by the first method the final region will be the one where the maxima of WM-

DOVA falls. For polyp segmentation from the output of an energy map the polyp region is the direct output of the algorithm.

- We will consider as True Positive when the polyp region contains fully or partially the polyp, a False Positive when the polyp region does not contain any polyp information. Following this criteria, we will label a True Negative when the background does not contain any polyp information and a False Negative when the background contains the polyp.

Taking into account these assumptions, we will expose in this subsection classification results using the metrics explained in a previous subsection of this chapter. Moreover, we will compare classification results achieved after performing polyp segmentation with our two methods with previous state-of-the-art classification results, which were presented in [13]. We will compare classification results from four experiments which are:

1. State-of-the-art results: achieved with the first polyp segmentation method and by means of SA-DOVA energy maps [13] (Baseline). In this case we choose as polyp region the one where the maxima of SA-DOVA falls, no matter if it is not the region with more polyp content out of all the final regions (Baseline).
2. State-of-the-art results improved by means of using WM-DOVA instead of SA-DOVA (Baseline with WM-DOVA).
3. Polyp Segmentation after region Merging and WM-DOVA (Method 1). Conversely to the previous case, we choose as final segmented region the one with more polyp content and we will check if the maxima of WM-DOVA falls on it or not.
4. Polyp Segmentation from the output of an energy map.

We present classification results in Table 7.26:

	Baseline	Baseline with WMDOVA	Method 1	Method 2
TP	265	280	252	249
FP	31	20	48	51
TN	269	280	252	249
FN	35	20	48	51
Precision	88%	93.3%	84%	83%
Recall	89%	93.3%	84%	83%
Accuracy	89%	93.3%	84%	83%
Specificity	88%	93.3%	84%	83%
F2Measure	0.89	0.9	0.77	0.765

Table 7.26: Comparison between the results obtained by our two methods.

The results shown in Table 7.25 can be analyzed in several ways. First of all, we can see that the improvements on polyp localization lead to an overall improvement

on every classification measure, as it can be seen in the table. But, as it has been mentioned before, these results have been obtained by considering a good classified polyp region even those where the amount of polyp information in it is very small.

In this case, we want to stress the importance of a good segmentation and, because of this, we have recalculated all the measures considering for the first method as final region the one with bigger AAC and, more importantly, DICE. As it can be seen in the Table, classification results are not the same, in fact, there is a global decrease in all the classification measures, leading to losing almost 30 polyps. The difference in this case is related to the fact that, by discarding regions with non significant polyp content, the maxima of WM-DOVA may fall in a big region that contains very small polyp information. In our case we prefer to deliver regions closer to how polyps appear on the image.

In our case, thinking about a possible application that our polyp segmentation could have, we prefer to offer as output regions obtained by our new method, even considering that, for some cases, we are losing polyps. We can see an example of a good result but a bad polyp region in Figure 7.22 (b) and (d). We can see that by using the second method the size of the polyp region is reduced. We can observe an example of a bad result but with a more reduced final region in Figure 7.22 (c) and (e).

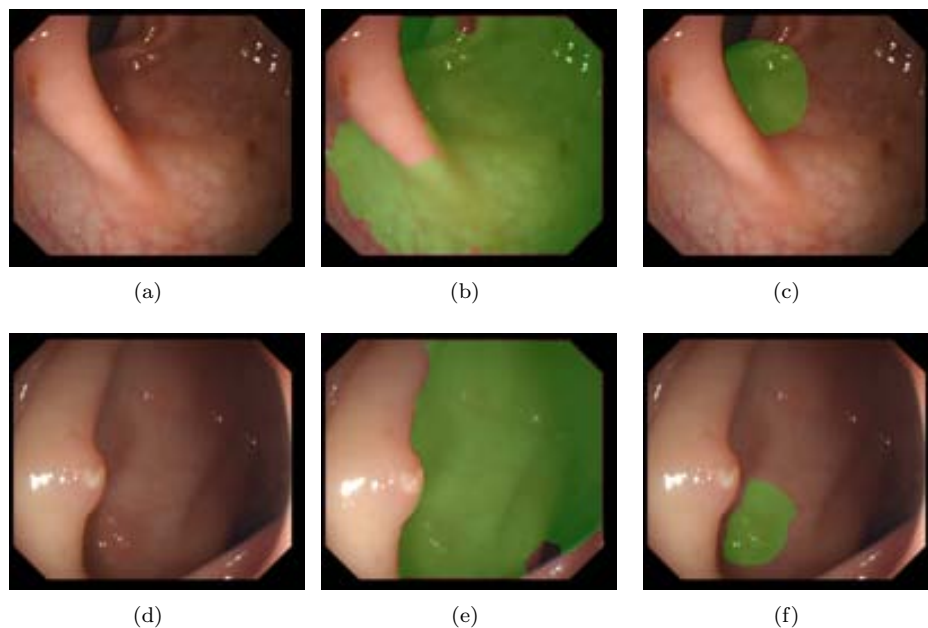


Figure 7.22: Comparison of classification results after applying our two segmentation methods: (a) and (d) Original image; (b) and (d) Classification results achieved with method proposed in [13], and (c) and (e) Classification results obtained with the new method.

As a way to confirm our preferences, we can observe the result shown in Figure 7.22 (b) and (c). In the first case, we would be indicating the physician that the interesting area is almost all the frame. In the second case, the indicated area will be smaller. In both Figure 7.22 (b) and (c) the polyp is captured but we are more precise with our new method and therefore, the results obtained by this method could be potentially more useful, although worse in region classification terms. Even when we do not capture the polyp, the final region obtained by means of the output of an energy map is smaller than the one achieved with polyp segmentation after region merging method.

Chapter 8

Use of a polyp localization method to model physicians' observations

8.1 Motivation

We have presented in the previous chapter the core of the thesis, which is to develop polyp localization and polyp segmentation methods in the scope of analysis of colonoscopy videos. These methods are based on a model of appearance for polyps, which has been created by observing how polyps appear on video frames but also by considering the presence of some other elements of the endoluminal scene. We have also expressed in both chapters 1 and 2 that the methods that we have developed could be enclosed into the field of Endoluminal Scene object description but, as it can be expected, these methods have been built in order to be incorporated in the future into intelligent systems for colonoscopy.

We have sketched in chapter 2 some of the domains of application of an intelligent system for colonoscopy, such as providing tools to assist on the diagnosis (via Computer Aided Diagnosis systems) or as a way to calculate performance metrics. Related to the first domain of application, we could think of incorporating our polyp localization methods as a way to indicate the physician which are the regions of interest of the image. But before thinking about the potential future of our methods, we must assess that they can provide key information to the process.

Taking this into account we will present in this chapter the first available study that compares the output of a computer vision method with the observations of the physicians during a colonoscopy procedure. More precisely, we will compare the output of our WM-DOVA energy maps with the fixations captured via an eye-tracker device, in order to check if our polyp localization method offers comparable results with the ones achieved by the physicians. In this case we will use the concept of saliency, understood as a way to point out which are the most interesting parts of the image. We will develop a little more about saliency in later sections but as an

introduction, the basic idea is to compare regions of interest calculated either by means of our WM-DOVA descriptor or by interpreting the physicians information.

8.2 Objective of the analysis

The objective of the analysis that we will present in this chapter is to create a framework that permits to establish correspondences between physicians' fixations (human regions of interest or hROIs) and the output of interest region detection algorithms (aROIs). We can think of three different scopes of analysis:

1. Comparison between the output of different hROIs.
2. Comparison between the output of different aROIs.
3. Comparison between the output of hROIs and aROIs.

We show a graphical representation of the three types of analysis in Figure 8.1, where we show the objective of our analysis which is to check for correspondences between computer vision algorithms (our polyp localization method) and visual attention models obtained via capturing physicians' observations with an eye-tracking device.

As mentioned before, there are three different analysis that can be performed but, as the objective of this chapter is to offer a first study on the potential of inclusion of some of our applications into real-life systems, we will concentrate on the first and third scope of analysis. We will first make a comparison between the observations of physicians, grouped according to their expertise and later we will compare the output of our polyp localization method with hROIs.

Our analysis will try to answer some of the following questions: Are the different hROIs attracted to the same regions of the image? Is there any difference in behaviour between experts and novices? Does our polyp localization method approximate physicians' observations? In order to tackle some of this questions, we need to define a framework of analysis. In this case, we will integrate our scopes of analysis into eye tracking and saliency theory domains, as we want to compare the output of hROIS (which could be analyzed in terms of fixations' types) with aROIS (which could be seen as saliency maps as they try to highlight which are the regions of interest of the image).

Before entering into the details of the experiments, it is necessary to make a brief theoretical introduction of the elements that constitute the basis of the analysis, which are: 1) Eye tracking devices; 2) Saliency theory and 3) Integration of eye tracking fixations into saliency maps. Therefore we will use the next section to perform this study and after this, we will continue with the experimental setup and the results.

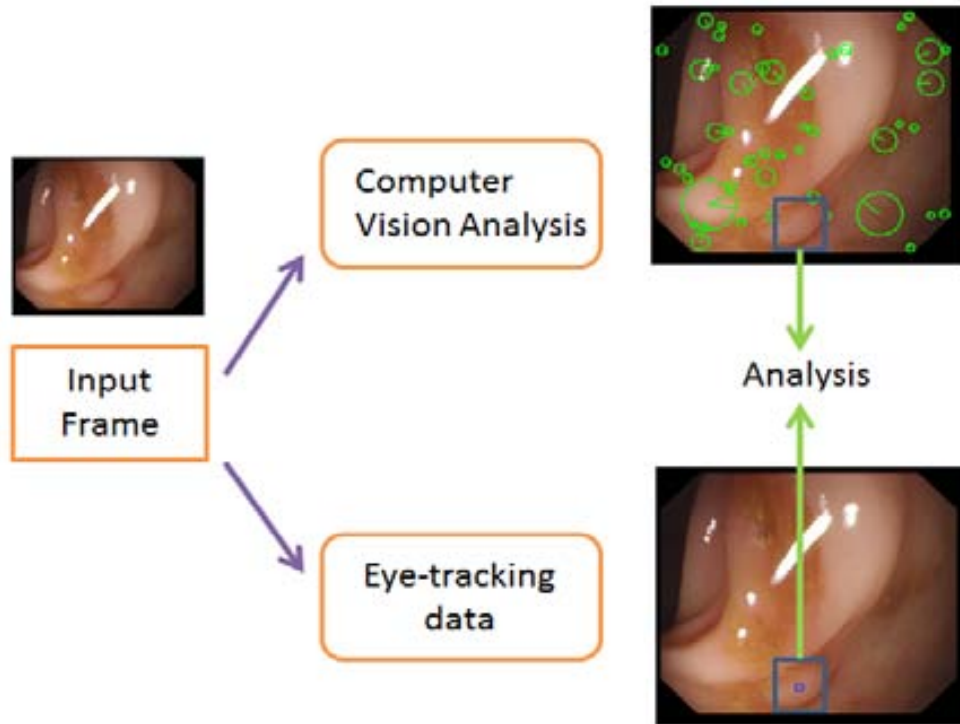


Figure 8.1: Graphical representation of the scope of the analysis.

8.3 From eye tracking to saliency theory. Integrated analysis of fixations via task-driven saliency maps.

In this section we will make a brief introduction of eye tracking although in our experiments we will only analyze the results that an eye tracker device provides, without entering into details such as how to configure correctly the eye tracker device. As mentioned in [37], when approaching the topic of eye tracking, we should start from considering the motivation for recording human eye movements. Taking this into account, when we are fixing our attention at a certain point of the field of view, we move our eyes to bring a particular portion of this view into high resolution so we can see in finer detail whatever the focus of our attention has been. In this sense, we can think of following the path of attention of a given observer by tracking his/her eye movements. In our case we will be interested on finding out which parts of the image attracts the visual attention of the physicians. This problem is tackled on the literature in two specific domains: analysis of fixations and saliency theory. In this section we will settle the basics of both eye tracking devices and saliency theory in order to provide a framework for our later study.

8.3.1 Analysis of the fixations

One important aspect that must be clarified is how to identify and label the different types of fixations that we may find [93]. More specifically we must distinguish between fixations and saccades. In order to understand its difference we must take a look at the physical structure of the eye itself. The fovea is the part of the eye located at the center of the macular region of the retina and it is responsible for sharp central vision, which is a characteristic in humans that enables us to read or watch television, coming into play in any human activity where visual detail is of primary importance. Taking this into account, a fixation occurs when our foveal attention stays on a particular object and the movement of our eyes from one fixation to another are known as saccades. Some works extends these definitions to another concepts such as smooth pursuit, blink and noise. In order to capture the observer's focus of attention an eye tracker device is used.

Eye tracker devices measure the rotation of the eyes in several ways, although the methods could be grouped into three categories. The first one uses an attachment to the eye, such as a special contact lens with an embedded mirror and the movement of this attachment devices is measured under the assumption that it does not slip along the eye rotation. The second group uses non-contact optical methods to measure eye motion by using the assumption that light (typically infrared) is reflected from the eye and sensed by a video camera or some other optical sensor. The information is later analyzed to extract eye rotation from changes in reflections. The last group of methods uses electric potentials measured with electrodes placed around the eyes, as they are the origin of a steady electric potential field which could also be detected in total darkness and with eyes closed. For our experiments we will use an eye tracker device from the second group, where a camera is mounted in a helmet-like device which is attached to the head of the person whose fixations we want to track.

In our case, as it will be explained in the Experimental Setup subsection, we have information from 22 different physicians. We will use this information in two different ways: first of all we will analyze separately the performance of each type of physicians and, after this analysis is done, we will compare physicians' performance with the output of our polyp localization method. This last comparison will be done in the framework of saliency theory, which fundamentals we will explain next.

8.3.2 Introduction to saliency theory

As mentioned before, we can define saliency of an item (object, person or pixel) as a state or quality which it stands out relative to its neighbours. More precisely, we will use the definition of saliency to identify which objects on the image are meant to attract more attention. Visual attention is a process that enables biological and machine vision systems to select the most relevant regions from a scene. Modeling visual saliency has attracted much interest recently and there are several frameworks and computational approaches available. Before enumerating some of the available alternatives, we must make a distinction between saliency and attention. Visual

attention is a broad concept that covers many topics such as bottom-up/top-down or spatial/spatio-temporal. Visual saliency has been recurrently referred as bottom-up process that delimits which image regions hinder different features from their neighbors [17].

In this sense, bottom-up saliency has been studied in search tasks such as finding an odd item among distractors as well as in eye movement prediction on free viewing of images or videos. In contrast to bottom-up, top-down attention deals with high-level cognitive factors that make certain areas of the image more relevant. In this concept we can include task demands or emotions. As it can be seen, our DOVA energy maps could be enclosed to this last group, as it models which areas of the image are meant to contain a polyp, which is the main objective to find for the case of physicians during a colonoscopy procedure.

Visual attention studies can also be categorized in other ways, depending on what they are interested in. For instance another division groups existing approaches into spatial (still images) or spatio-temporal models (video stimuli). Considering this classification, the analysis that we will perform is based on the individual analysis of frames extracted from a video and therefore we will concentrate on methods that cover still image analysis.

Another classification divides the models of attention into space-based or object-based. The first group has as goal to create saliency maps that may predict which locations have higher probability of attracting human attention, which is the aim of our polyp localization method. The later group aims at segmenting or detecting objects to predict salient regions. We could fit our polyp segmentation methods into this last group, as our aim is to provide with an accurate polyp region segmentation.

As it can be seen there are several classification criteria and there are many methods with are devoted to define visual attention maps and saliency maps. At this point of our research we are not interested on comparing our DOVA energy map to another saliency maps but, as we will depict in the last chapter of the thesis, it is a task that we would like to perform in the near future. As we have mentioned before, we could group our polyp localization methods into the field of top-down attention saliency. Referring to this concepts, there are several works in the literature that tackle the problem of task driven attention in videos, and we will make a brief review of some of them in the next section.

8.4 Linking eye tracking with saliency. Task-driven saliency maps

As we have mentioned before, with an eye tracking device we are able to capture physicians observations. At this point we can perform two different analysis: we can compare the individual observations of each physician to the output of our polyp localization methods or to group physicians according to the number of interventions and compare group results with the output of polyp localization methods.

In this case we aim at the second type of analysis and, in order to compare the output of a group of physicians with DOVA energy maps, we should find a way to integrate physicians information. Related to this, there are several approaches in the literature that provide with interesting methods to integrate this information. It has to be noted that there are very few works which deals with video studies, although some of them are very interesting.

For instance, we can find a preliminary study on how to extract visual features with eye tracking for saliency driven registration in [26]. Although the method presented is interesting the scope of the analysis is different from ours, because the creation of saliency maps from fixations is not task oriented. Another relevant work is the one presented in [86], where a comparison very similar to the one we will present (hROIs vs aROIs) is introduced. In this case, the analysis is performed on still images and, as the method presented before, it is not task oriented. What is interesting about this work is that it introduces a way to compare different ROIs such as using a distance measure after clustering results of aROIs.

More related with the scope of our analysis is the work presented in [43], where the concept of task driven saliency maps is introduced. In this case in order to compare the performance of difference saliency maps the authors propose a thresholding on saliency levels and then check if the higher values correspond to positions of the object of interest. Although under a different name, the work presented in [55] also builds on the idea of defining consistent locations (task driven objectives) and also provides with methods to create region of interest detector's saliency maps from the output of aROIs.

Among all the references that we have found, we will use for our experiments the process of creating task-driven saliency maps explained in [24]. In this work the attention of an image region is measured by fixation density, that is, the salience of an image is represented by a fixation density map. The fixation map is created from a set of discrete fixation points $(x_n^f, y_n^f), n = 1, \dots, N$ where N is the total number of fixation points found in a frame and (x_n^f, y_n^f) is the location of the n -th fixation point. Those fixation points are interpolated by a Gaussian function to generate a fixation density map $s(x, y)$:

$$s(\omega, l) = \frac{1}{N} \sum_{n=1}^N \frac{1}{2\pi\sigma_s^2} \cdot \exp\left(-\frac{(\omega - \omega_{f_n})^2 + (l - l_n^f)^2}{2\sigma_s^2}\right) \quad (8.1)$$

, where x and y denote, respectively, the horizontal and vertical positions of an observation pixel and s is the standard deviation of the Gaussian function, determined according to the visual angle accuracy of the eye tracking system. More precisely,

$$\sigma_s = L \times \tan \frac{0.5\pi}{180} \quad (8.2)$$

, where L is the viewing distance between the subject and the display (in this case, 60 cm). By using this method, the value of each fixation pixel is propagated to

its neighbor pixels. Therefore, a pixel in a densely populated fixation area is more attractive than a pixel in a sparsely populated fixation area. For the same reason, a region is more attractive if it is brighter on the fixation density map, as shown in Figure 8.2:

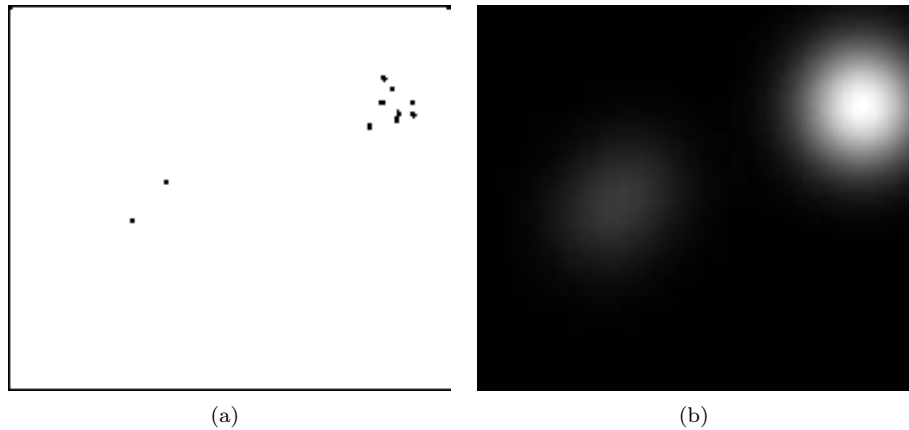


Figure 8.2: Creation of task-driven saliency maps: (a) Position of physician's fixations; (b) Task-driven saliency maps

We will use these task-driven saliency maps to integrate the physicians observation and, as it will be mentioned later, to create saliency maps from the output of our polyp localization methods. In the following sections we will present the experimental setup along with some definition of the metrics that we will use, before proceeding with the experimental results.

8.5 Experimental Setup

8.5.1 COLON-PHYS database

In order to measure how close computer vision algorithms are to physician's knowledge, a series of experiments have been run in association with physicians from Beaumont Hospital and St. Vincent's Hospital. The experiments consisted of showing, to each physician, a series of videos and, by using an eye tracking device, acquiring the gaze position for each frame. Before starting with the analysis of the results, we present the basic data of the experiments.

As mentioned before, these experiments have been made with the collaboration of 22 different physicians from Beaumont Hospital and St. Vincent's Hospital. Although their identities cannot be exposed here, we show in Table 8.1 a summary of the number of interventions done by each physician. It has to be noted that some of them have not done any intervention whereas some of them have participated in a great number of procedures. This fact leads us to distinguish between two classes of physicians,

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

Table 8.1: Number of interventions performed by each physician.

experts and novices. In this case, the threshold number of interventions used to classify the physicians has been set to 100, as it is the number of interventions each trainee should have performed in order to be considered as experts, following the guidelines depicted by the Joint Advisory Group on Gastrointestinal Endoscopy [5].

One of the objectives that we want to achieve with the analysis of eye tracking data is to measure if there is any difference in behaviour between experts and novices, taking into account that we have the number of interventions that each physician has made. This analysis can be done in two ways: we can create a 'mean expert' and 'mean novice', by integrating all the fixations from each group by using task-driven saliency maps or by analyzing separately the behavior of each physician.

Before starting with this analysis, we have to define the experimental setup. We count with 12 different videos and we have, for each of them, a ground truth that consists of a elliptical mask centred in the polyp's centre along with major and minor radii values. We can see one example from each video (along with the superimposed ground truth) in Figure 8.3.

8.5.2 Metrics used in the analysis

As it has been mentioned before, the objective of the analysis presented in this chapter is to compare the observations of the physicians', obtained via an eye-tracking device, with the output of our polyp localization methods. But before going to this comparison, we will make a first analysis of pure eye tracking information. This first analysis is done with the purpose of checking if the degree of expertise, defined from the number of procedures that each physician has performed, has an impact on the discovering of polyps. Therefore we must define a series of metrics in order to perform this analysis:

1. First Fixation: If we integrate each group's fixation information, we will refer as

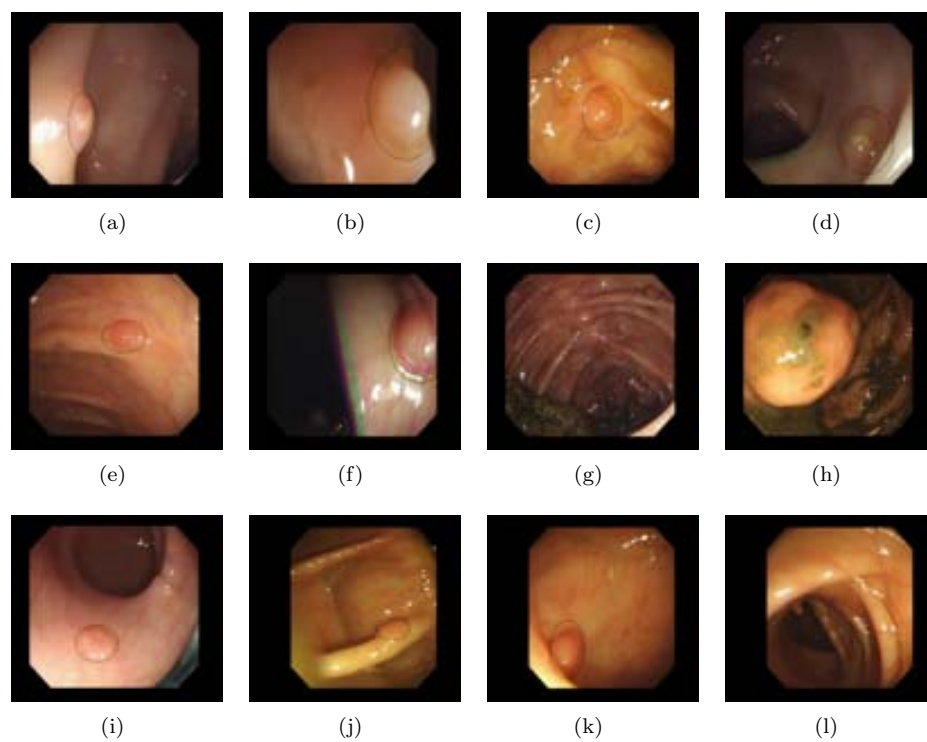


Figure 8.3: Ground-truth for Eye Tracking analysis. We show, for each video, one representative frame containing a polyp with the ground truth label superimposed: (a) Video 2; (b) Video 3; (c) Video 4; (d) Video 5; (e) Video 8; (f) Video 9; (g) Video 10; (h) Video 12; (i) Video 13; (j) Video 14; (k) Video 15; (l) Video 18;

first fixation the number of frame where the maxima of the task-driven saliency map falls over the polyp mask. In the case of individualized analysis of the eye tracking data, we will define as the first fixation the number of frame where the distance between the physician's fixation and the polyp mask is zero.

2. Reaction time: Reaction time is the elapsed time between the presentation of a sensory stimulus and the subsequent behavioral response. In this case, we will refer as reaction time as the difference in number of frames between the first apparition of the polyp on the video and the first frame where the maxima of the task-driven saliency maps fall over the polyp mask (first fixation).
3. Dwelling time: The total time spent looking at a stimulus. In our experiments, we will measure the dwelling time as the number of frames (non-necessarily consecutive) that the 'mean expert' and the 'mean novice' place the maxima of the task-driven saliency map over the polyp mask.
4. Concentration Rate: As we are integrating physician's information into a task-driven saliency map we have to define a measure that inform us about how concentrated are the fixations in a certain point on the image (in our case, inside the polyp's mask). We define as concentration rate the percentage of energy that falls inside a certain area of the image.

We will present in the next two sections the experimental results obtained by this analysis. Although it may seem a little obvious, we must mention that in this case the task that both physicians and DOVA energy maps perform is to find polyps in colonoscopy studies and we will compare the performance of each group of methods according to how fast they are on detecting a polyp or to how focused is the attention on the polyp once it is discovered.

8.6 Eye tracking data analysis

In this section we will present analysis of pure eye tracking data, that is, we will compare the performance of experts and novices according to the metrics presented in the previous section.

8.6.1 Reaction time

In order to calculate the reaction time we have taken the following steps:

1. Elaboration of the ground truth.
2. Calculation of the task-driven saliency maps.
3. For each frame that appears on the ground truth, calculation of the position of the maxima of the task-driven saliency map and check if the maxima falls inside the polyp's mask.

The results of the reaction time for experts can be seen in Table 8.2 whereas the results for novices can be consulted in Table 8.3. We have to mention that for the case of video 4 the polyp appears briefly from frames 252 to 260 and after disappears until frame 714, therefore for this case we will consider when calculating reaction time the video to be cut from frame 260 to 714.

Video	First Apparition of Polyp	1 st Fixation Experts	Reaction Time (#of frames)
2	266	272	6
3	393	393	0
4	252	252	0
5	382	387	0
8	370	378	0
9	144	193	49
10	553	660	107
12	1373	1373	0
13	126	182	0
14	235	235	0
15	195	198	3
18	369	369	0

Table 8.2: Experts' fixation results.

Video	First Apparition of Polyp	1 st Fixation Novices	Reaction Time (#of frames)
2	266	275	9
3	393	400	7
4	252	736	30
5	382	389	2
8	370	392	14
9	144	175	31
10	553	565	12
12	1373	1373	0
13	126	185	3
14	235	236	1
15	195	212	17
18	369	370	1

Table 8.3: Novices' fixation results.

The difference between Experts and Novices' reaction time is presented in Table 8.4.

We can see from Table 8.2 that experts tend to detect the polyp soon after its first apparition with the exceptions of videos 9 and 10. In the other hand, we can see from Table 8.3 that novices detect the polyp later, although the difference in number

Video	First Apparition of Polyp	Difference in Reaction Time
2	266	3
3	393	7
4	241	30
5	382	2
8	370	14
9	144	-18
10	553	-95
12	1373	0
13	126	3
14	235	1
15	195	14
18	369	1

Table 8.4: Difference Reaction time.

of frames is not critical (considering that the videos are codified at 25 frames per second).

We can observe a comparison between experts' and novices' fixation rate by consulting Table 8.4. The difference between both groups are not large, although we can see that for 9 of the 12 videos the experts detect the polyp earlier than novices, which may suggest that the difference in the number of procedures has a relative impact in the reaction time.

Finally, there are two cases where experts' detect the polyp later than novices: videos 9 and 10. We can observe in Figure 8.4 the frames where experts and novices place their maxima of the task-driven saliency map.

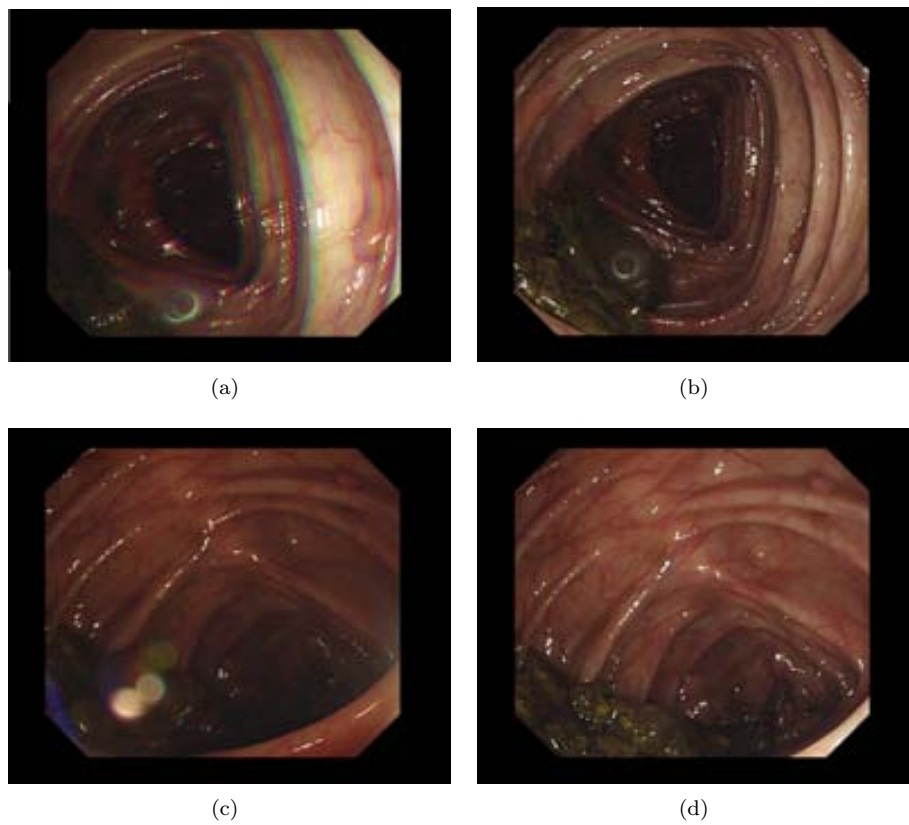


Figure 8.4: Difference in reaction time: cases of videos 9 and 10 where novices have their first fixation earlier than experts: (a) Video 9 Experts' first fixation; (b) Video 9 Novices' first fixation; (c) Video 10 Experts' first fixation; (d) Video 10 Novices' first fixation;

8.6.2 Dwelling time

In this case we compute the dwelling time by counting the number of frames where each group of physicians place the maxima of the task-driven saliency map over the polyp mask. We can see results on dwelling time on Table 8.5

Video	Dwelling time Experts	Dwelling time Novices	Difference Experts-Novices
2	85	60	25
3	70	56	14
4	146	103	43
5	78	73	5
8	106	80	26
9	71	70	1
10	3	10	-7
12	119	119	0
13	75	59	16
14	74	35	39
15	104	68	26
18	67	26	41

Table 8.5: Dwelling time results.

As it can be seen from Table 8.5 for 11 of the 12 videos the dwelling time for experts is higher than for novices. This can be seen in two different ways. On the one hand we can see that experts place their maxima of saliency on the polyp for more frames than novices. This can be interpreted as the experts have more confidence on where the polyp is on the image whereas novices do not and because of this the saliency map of novices is more scattered than experts', which seem to be more focused on polyp position. We present in Table 8.6 the number of frames with polyp for each video, in order to stress this last sentence's hypothesis.

In order to close this subsection we show in Table 8.7 the percentage of frames where polyps attract the attention of experts and novices, measured as placing the maxima of task driven saliency map on the polyp mask:

As it can be seen from this last experiment, there is a significative difference between experts and novices in terms of dwelling time. For 9 out of the 12 videos the differences is greater than 10% and, in this case, considering than once a polyp appears the physician that performed the intervention wanders around it, we can conclude that experts perform better than novices when the polyp is present, as they focus their fixation in the polyp on more frames.

Video	Total number of frames	Frames with polyp
2	1097	133
3	1500	138
4	2268	159
5	1688	133
8	1885	145
9	1733	149
10	1361	112
12	1500	120
13	1639	134
14	1501	132
15	946	137
18	923	101

Table 8.6: Number of frames with polyp for each video.

Video	Dwelling time Experts	Percentage Experts	Dwelling time Novices	Percentage Experts	Difference Experts - Novices
2	85	63.9%	60	45.11%	18.79%
3	70	50.72%	56	40.58%	10.14%
4	146	91.82%	103	64.78%	27.04%
5	78	58.65%	73	45.91%	12.74%
8	106	73.1%	80	55.17%	17.93%
9	71	47.65%	70	46.98%	0.67%
10	3	0.027%	10	0.09%	-0.06%
12	119	99.17%	119	99.17%	0%
13	75	55.97%	59	44.03%	11.94%
14	74	56.06%	35	26.52%	29.54%
15	104	75.91%	68	49.64%	26.27%
18	67	66.34%	26	25.74%	40.6%

Table 8.7: Dwelling time percentage results.

8.6.3 Energy concentration rate

The results obtained from the previous experiment indicate a slight difference in behaviour between experts and novices, we can not forget that we are basing our analysis on the creation of a 'mean expert' and a 'mean novice'. Because of that, we may have that the maxima of fixation falls under the polyp mask but the energy of the image is not concentrated on the polyp, hence the importance of having not only a strong decision on which frame results on the first fixation, but to also have a concentrated view.

Therefore, we have measured, once we have decided in which frame occurs the

first fixation, the concentration rate of the task-driven saliency map under the polyp mask, calculated as shown in Eq. 8.3.

$$CR = 100 \cdot \frac{E_p}{E_{all}} [\%] \quad (8.3)$$

, where CR stands for concentration rate, E_p stands for the energy under the polyp mask and E_{all} stands for the energy of all the image. The energy is calculated as the sum of the value of every pixel in the task driven saliency map.

Experimental results can be consulted in Tables 8.8 and 8.9. In this sense define as concentration rate the percentage of energy of the task driven saliency maps that falls under the polyp mask therefore a high concentration value will mean that nearly all the physicians for the particular group were looking inside the polyp whereas a low value will denote a random pattern of physicians fixations whose combined task driven saliency maps incidentally places the maxima inside the polyp mask.

Video	% Concentration Experts	% Concentration Novices	Difference Experts-Novices in %
2	13.59%	13.15%	0.44%
3	78.09%	65.39%	12.7%
4	22.27%	49.15%	-26.87%
5	52.47%	32.48%	19.98%
8	46.44%	17.53%	28.91%
9	75.6%	16.8%	58.8%
10	7.9%	14.82%	-4.09%
12	73.84%	70.79%	3.04%
13	53.28%	1.82%	51.45%
14	9.67%	9.15%	0.51%
15	34.43%	0%	34.43%
18	5.14%	2.7%	2.44%

Table 8.8: Difference in concentration rate in the experts' first fixation frame.

We can sketch some interesting results by observing Table 8.8 and Table 8.9 but, in order to organize better the information we will compile the results of the difference experts-novices in Table 8.10.

The before mentioned tables show that, for the majority of the cases, once the experts find the polyp, their concentration rate is higher than novices' (see Table 8.8) but, perhaps more interesting is the fact that, even when the novices find the polyp, the expert's concentration rate continues to be higher (see Table 8.9). We can observe from Table 8.10 that for a great number of videos (7 out of 12) even when novices find the polyp, experts' concentration rate is higher. Moreover, for some of the cases when this does not happen (videos 9 and 18) the difference is very small.

Taking into account these results we can conclude that not only experts localize

Video	% Concentration Novices	% Concentration Experts	Difference Novices-Experts in %
2	45.31%	23.37%	21.93%
3	67.26%	76.04%	-8.77%
4	13.31%	19.26%	-5.95%
5	29.75%	36.31%	-6.56%
8	17.53%	46.44%	-28.91%
9	93.57%	92.89%	0.68%
10	7.9%	14.82%	-6.92%
12	70.8%	73.84%	3.04%
13	27.18%	22.22%	4.95%
14	22.8%	7.9%	14.9%
15	10.3%	22.97%	-12.67%
18	4.76%	4.38%	0.37%

Table 8.9: Difference in concentration rate in the novices' first fixation frame.

Video	Difference Experts-Novices in % on First Fixation Experts	Difference Novices-Experts in % on First Fixation Novices
2	0.44%	21.93%
3	12.7%	-8.77%
4	-26.87%	-5.95%
5	19.98%	-6.56%
8	28.91%	-28.91%
9	58.8%	0.68%
10	-4.09%	-6.92%
12	3.04%	-3.04%
13	51.45%	4.95%
14	0.51%	14.9%
15	34.43%	-12.67%
18	2.44%	0.37%

Table 8.10: Difference in concentration rate in the experts' first fixation frame.

earlier the polyp than novices but, once they localize it, they are more focused on the polyp and this focus is kept for the rest of the frames. We can see in the Figure 8.5 an example of experts' and novices' concentration rates during all the frames with polyp for video 13. As it can be observed, the difference between experts and novices is higher than 0% for a big number of frames.

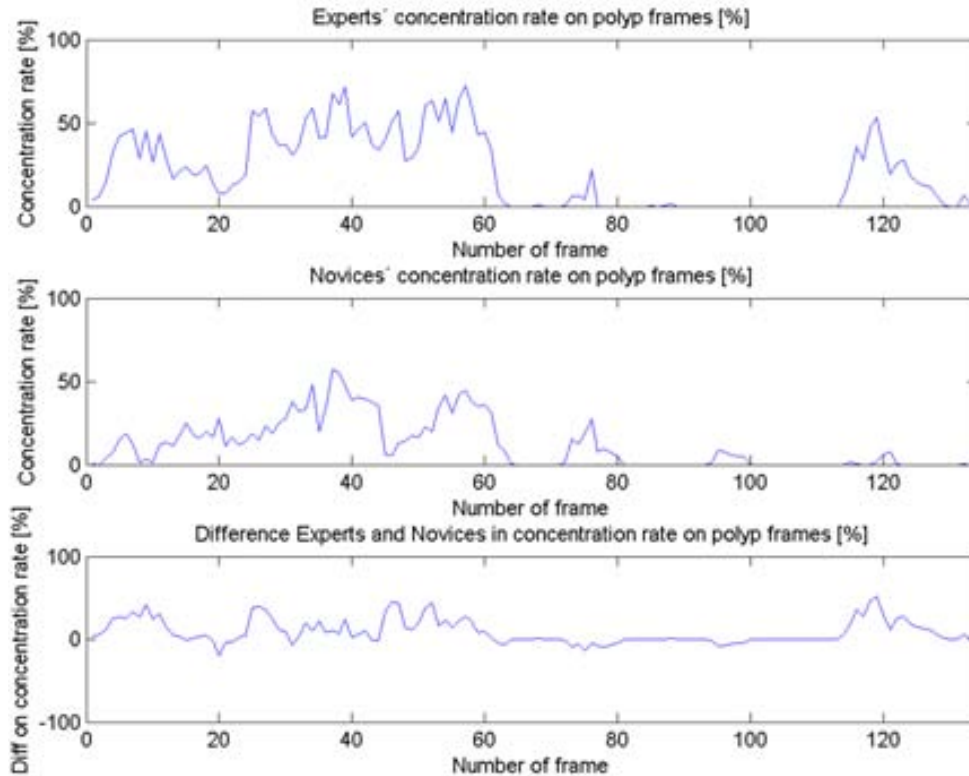


Figure 8.5: Difference in concentration rate between experts and novices during all the frames with polyp.

8.7 From polyp localization to saliency maps: DOVA saliency maps

After a preliminary study on physicians' fixations, we start with the second phase of the analysis which consists of correlating the output of hROIs with our polyp localization method. In this case we will use for polyp localization the method that has offered better performance, which is WM-DOVA and we will also perform specular highlights correction and blood vessels mitigation on the original images. As we want to compare the performance of DOVA with the physicians, we must transform one of the two members of the comparison in order to make a fair analysis.

Therefore, before presenting results of the comparison, we introduce in this section the new DOVA saliency maps, which are created from the output of our polyp localization method. In order to make comparable the output of our polyp localization methods and task-driven saliency maps, we will use energy map information in order to create 'virtual' physicians. This virtual physicians are placed in the position of the local maximas of the WM-DOVA energy maps. In this case we have imposed a

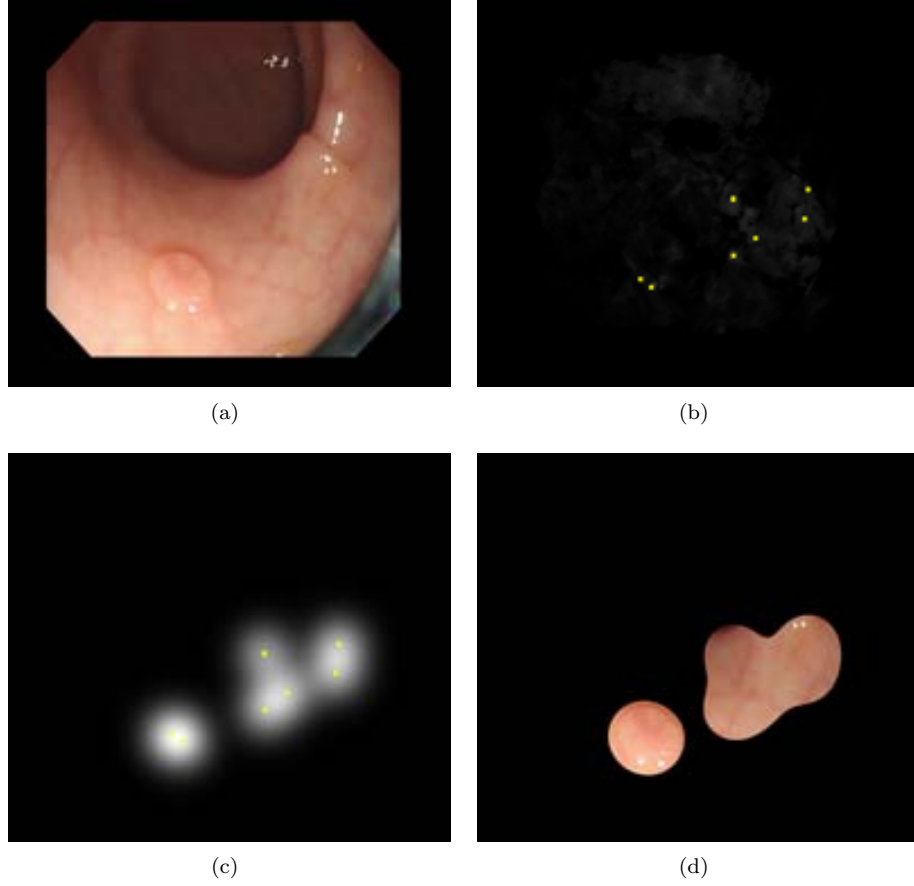


Figure 8.6: DOVA saliency maps from the output of polyp localization. (a) Original image; (b) WM-DOVA energy map; (c) WM-DOVA saliency map; (d) Saliency regions by means of WM-DOVA saliency map.

distance constraint of 50 pixels, which coincides with the mean radius of the polyps from our CVC 300 COLON DB database. Once we have calculated the position of the virtual physicians, we create the DOVA driven saliency map as depicted in Eq. 8.4.

$$s(\omega, l) = \frac{1}{N} \sum_{n=1}^N \frac{1}{2\pi\sigma_s^2} \cdot \exp\left(-\frac{(\omega - \omega f_n)^2 + (l - l_n^f)^2}{2\sigma_s^2}\right) \quad (8.4)$$

We explain graphically the process on how to create DOVA saliency maps from polyp localization in Figure 8.6.

As it can be seen from Figure 8.6 what we do is to extract the same number of local maxima of our DOVA energy maps and fit them to physicians so we can create

the task driven saliency map in the same way. It has to be noted that we will not accept any local maxima value, we will use only local maxima value up to a certain threshold from the maxima of DOVA map. In this case we have used as threshold value 0.9 in order to assess the performance of DOVA in a similar way that we have done before for polyp localization results.

8.8 Comparison of DOVA saliency maps with physicians' task driven saliency maps

As the objective of the analysis presented in this chapter is to check if our novel polyp localization method can approximate physicians' observations and therefore we will use the same metrics that we have introduced before in order to compare the results achieved via hROIs and aROIs.

8.8.1 Reaction time

We present in Table 8.11 the reaction time for our DOVA saliency maps. We can extract several conclusions from the results presented in the table:

- We localize the polyp for all the videos. That is, our method is robust enough to be useful for new studies.
- The reaction time is low for all the videos: only for the cases of videos 9 and 10 DOVA saliency maps last longer to localize the polyp. For the rest of the videos, considering that frames are acquired at 25 fps, the reaction time could be considered as good.

As it has been repeated during this chapter, the aim of this analysis is to compare DOVA saliency maps' performance with the output of hROIs therefore we provide comparison results in Table 8.12 and Table 8.13.

We can extract the following conclusions by observing the results from Table 8.12 and Table 8.13:

- On the one hand, we can observe at a glance that WM-DOVA obtains better results than novices for more than half of the videos. This difference is higher for video 4 although we can see that for videos 10 and 13 there is a big difference between novices and the performance of WM-DOVA. We will make a case-by-case review of this videos later on this section.
- On the other hand, we observe that WM-DOVA performs worse than experts for the great majority of the videos. It is true that for some of the cases, as it also happens for novices, that the difference in most of the cases is less or equal than 7 frames, which means that the delay is of about 0.25 seconds. We obtain better results than WM-DOVA for 2 videos, 9 and 15.

8.8. Comparison of DOVA saliency maps with physicians' task driven saliency maps183

Video	First Apparition of Polyp	1 st Fixation DOVA	Reaction Time
2	266	267	1
3	393	397	4
4	252	254	2
5	382	389	7
8	370	382	12
9	144	163	19
10	553	660	107
12	1373	1373	0
13	126	184	58
14	235	237	2
15	195	195	0
18	369	376	0

Table 8.11: DOVA saliency maps fixation results.

Video	Reaction Time DOVA	Reaction Time Novices	Difference DOVA - Novices
2	1	9	-8
3	4	7	-3
4	2	30	-28
5	7	2	5
8	12	14	-2
9	19	31	-12
10	107	12	95
12	0	0	0
13	58	3	55
14	2	1	1
15	0	17	-17
18	7	1	6

Table 8.12: Comparison of DOVA saliency maps' reaction time with novices.

- By taking a look at the tables, we can conclude that WM-DOVA performs better than novices when localizing first the polyp and worse than experts. In both cases the difference are not enough to consider them as significant but the numbers show that WM-DOVA performs comparably with the output of several hROIs.

Before closing with this section, we will make a review of those videos where DOVA's reaction time is different from zero in order to learn why we have not being able to localize the polyp in its first apparition.

Video	Reaction Time DOVA	Reaction Time Experts	Difference DOVA - Ex- perts
2	1	6	5
3	4	0	4
4	2	0	2
5	7	0	7
8	12	0	12
9	19	49	-30
10	107	107	0
12	0	0	0
13	58	0	58
14	2	0	2
15	0	3	-3
18	7	0	7

Table 8.13: Comparison of DOVA saliency maps' reaction time with experts.

Case by case analysis of videos

For the case of video 3 we can observe in Table 8.11 that the reaction time is of 4 frame. We can observe in Figure 8.7 (a) that in this case the reason behind this small delay may be related to the excessive blurring of the image, which may provide with less clear contours. Even considering that WM-DOVA places the maxima real close to the polyp mask, as it can be seen in the image.

The delay for video 4 is slightly smaller than for video 3. We can observe in Figure 8.7 (b) that in this case we can justify the delay of WM-DOVA because of the high density of wrinkles and folds and also because of the bad visibility of the polyp.

We obtain a delay of 7 frames for video 5. As we can see in Figure 8.8 (a), we can also associate the difference due to the presence of folds and wrinkles, although we have to mention that for this case the contours of the polyp are weaker than in later examples of this same video.

The delay for video 9 (Figure 8.8 (b)) is clearly caused by a profusion of elements of the endoluminal scene, specially folds and fecal content. Some other cause of error could be the small size of the polyp because, as we have mentioned in previous sections, we are applying directly our best polyp localization method and, in this image, the polyp may not fit even the minor radius size.

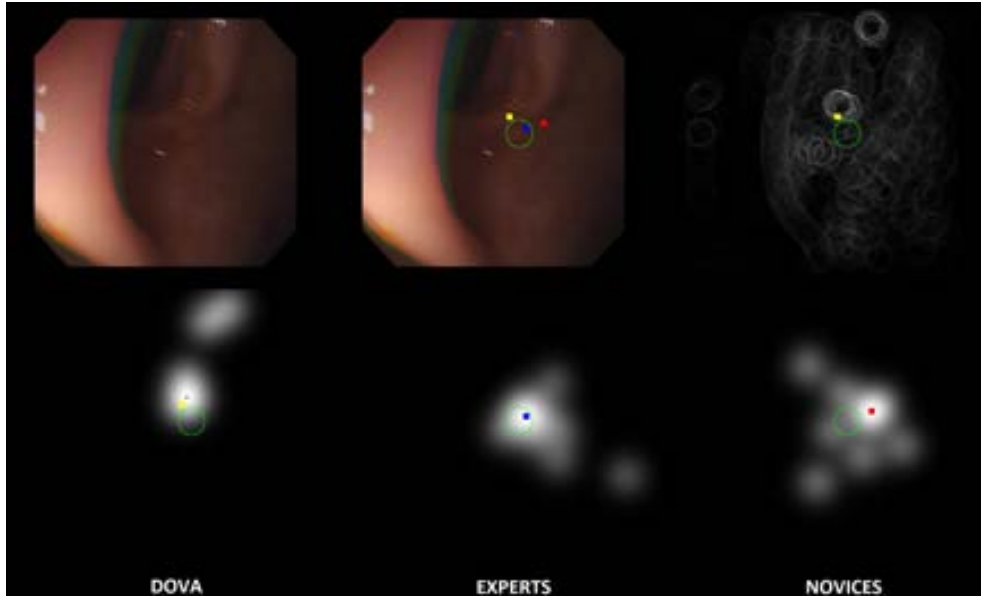
Similarly to the video 9, the delay for video 10 is caused by folds, fecal content and, in this case, blurring (see Figure 8.9 (a)).

We can see for the case of video 13 that the delay seems to be caused by the elements that we can consider to be part of the lumen or not clearly visible parts of the image, although we also have to consider that in this case fold contours are much clearer than polyps', as it can be seen in Figure 8.9 (b).

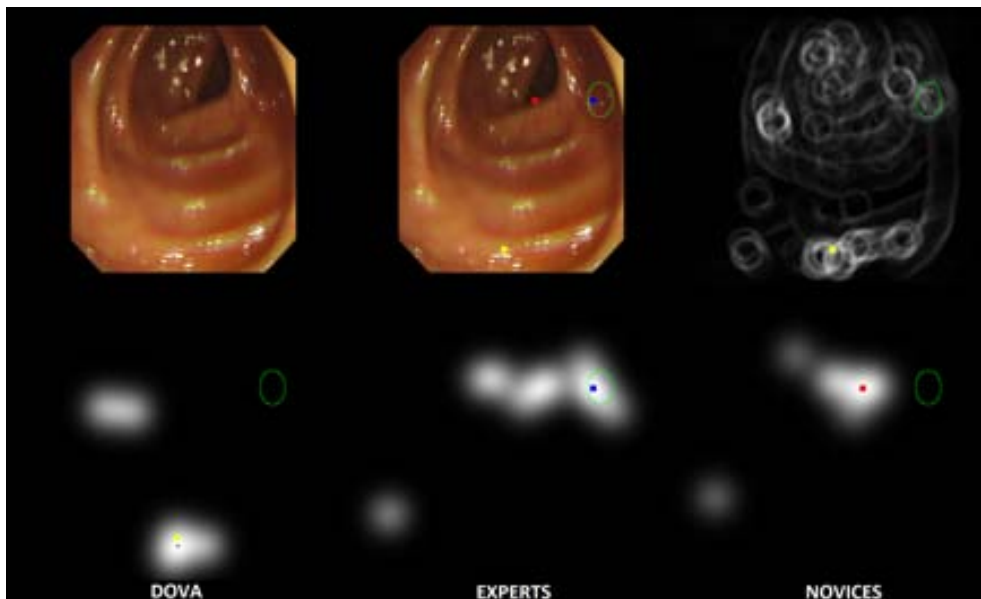
8.8. Comparison of DOVA saliency maps with physicians' task driven saliency maps¹⁸⁵

Video 14 present strong folds and a polyp which only present clear contours for one of its sides, as it can be observed in Figure 8.10 (a).

Finally the delay for video 18 is caused by folds and lumen but in this case we also observe that the mix of specular highlights and color channel misalignment deviates the maxima of WM-DOVA saliency map to an erroneous part of the image, as it can be seen in Figure 8.10 (b).



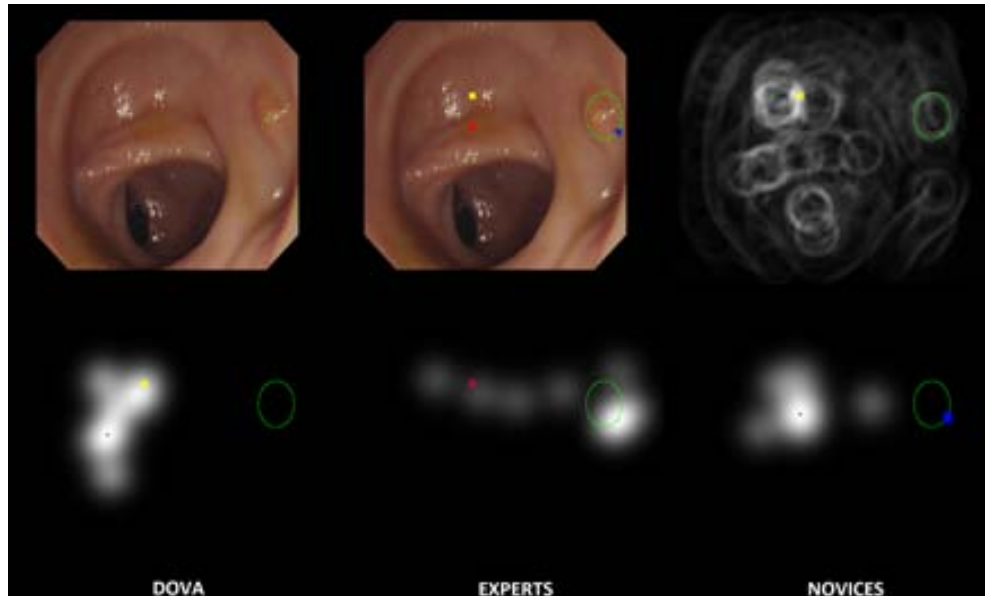
(a)



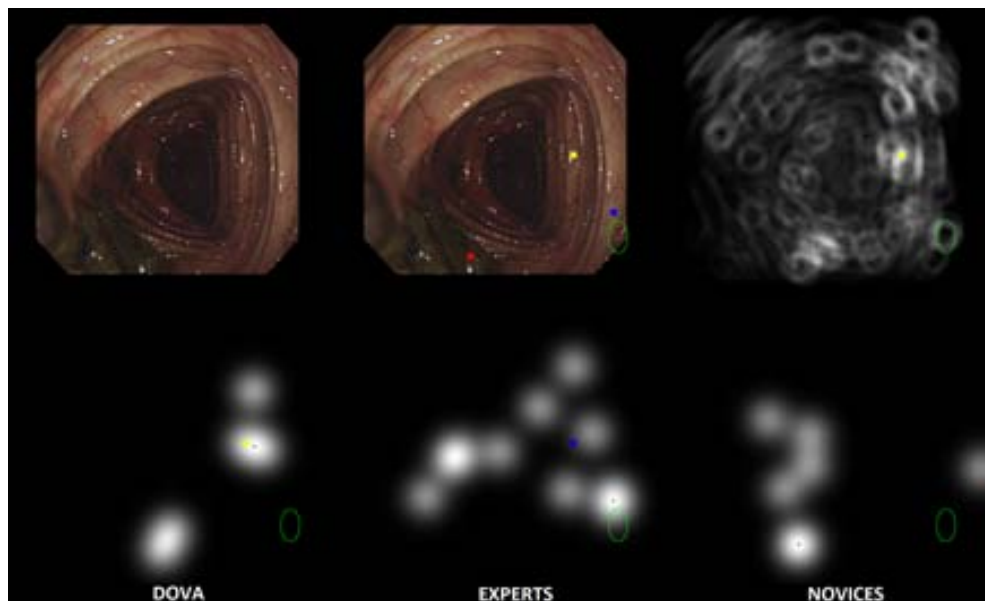
(b)

Figure 8.7: (a) Video 3: Comparison between WM-DOVA, experts and novices saliency maps. (b) Video 4: Comparison between WM-DOVA, experts and novices saliency maps. In both examples the maxima of WM-DOVA is painted by a yellow square, the maxima of experts is represented by a blue square and the maxima of novices is represented as a red square. Polyp mask is painted as a yellow ellipse.

8.8. Comparison of DOVA saliency maps with physicians' task driven saliency maps187

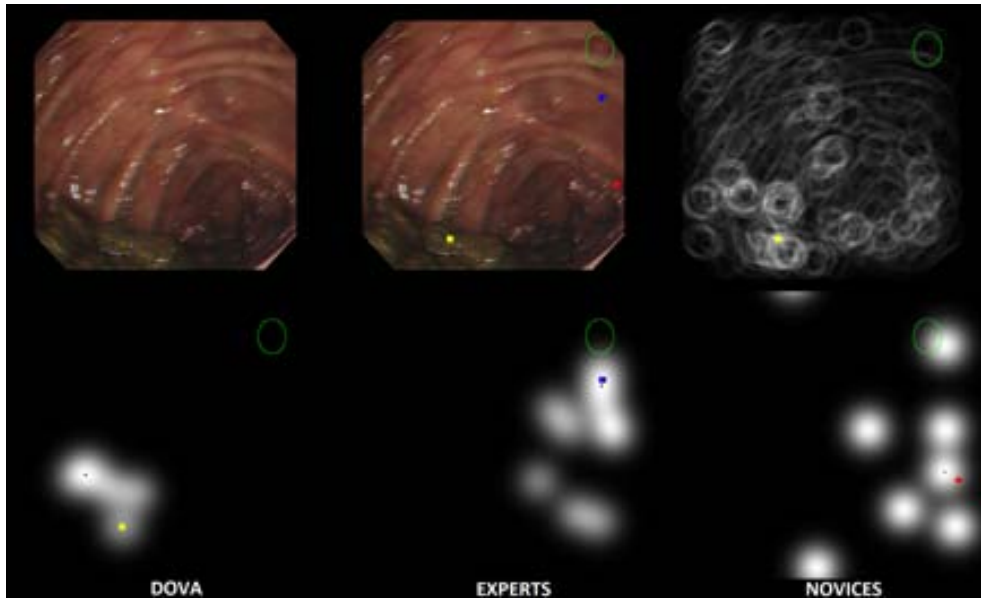


(a)

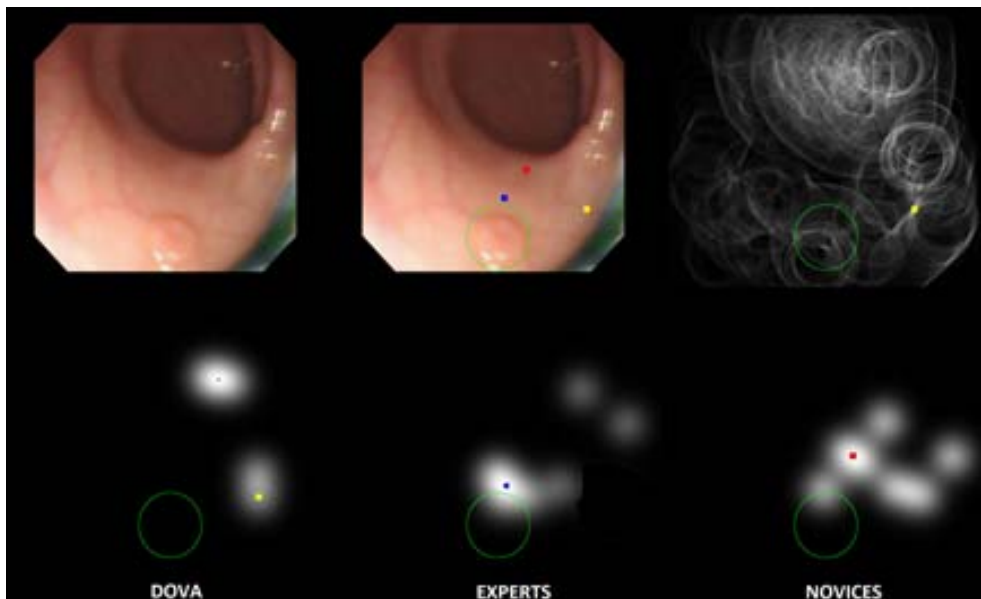


(b)

Figure 8.8: (a) Video 5: Comparison between WM-DOVA, experts and novices saliency maps. (b) Video 9: Comparison between WM-DOVA, experts and novices saliency maps. In both examples the maxima of WM-DOVA is painted by a yellow square, the maxima of experts is represented by a blue square and the maxima of novices is represented as a red square. Polyp mask is painted as a yellow ellipse.



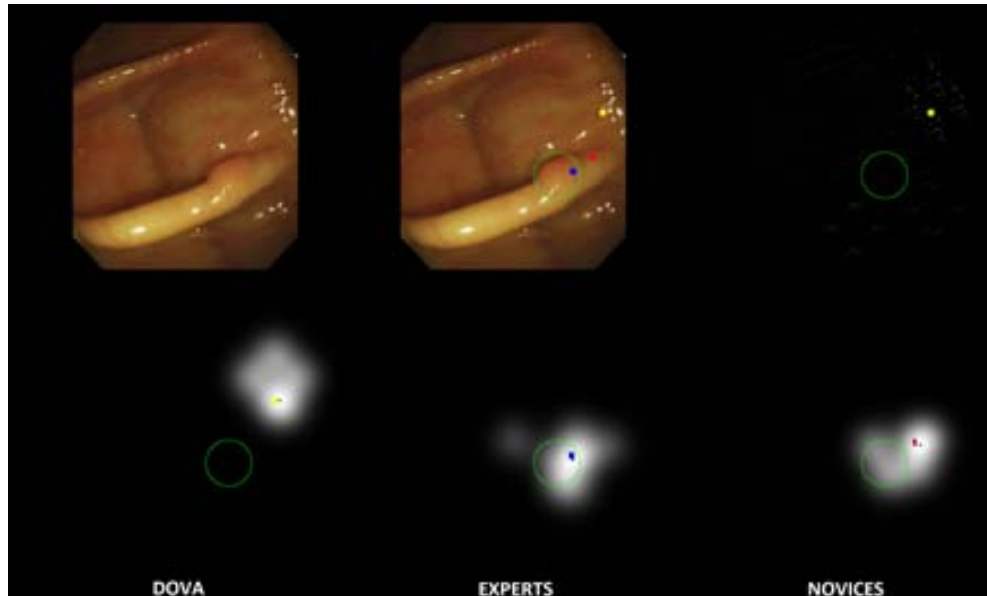
(a)



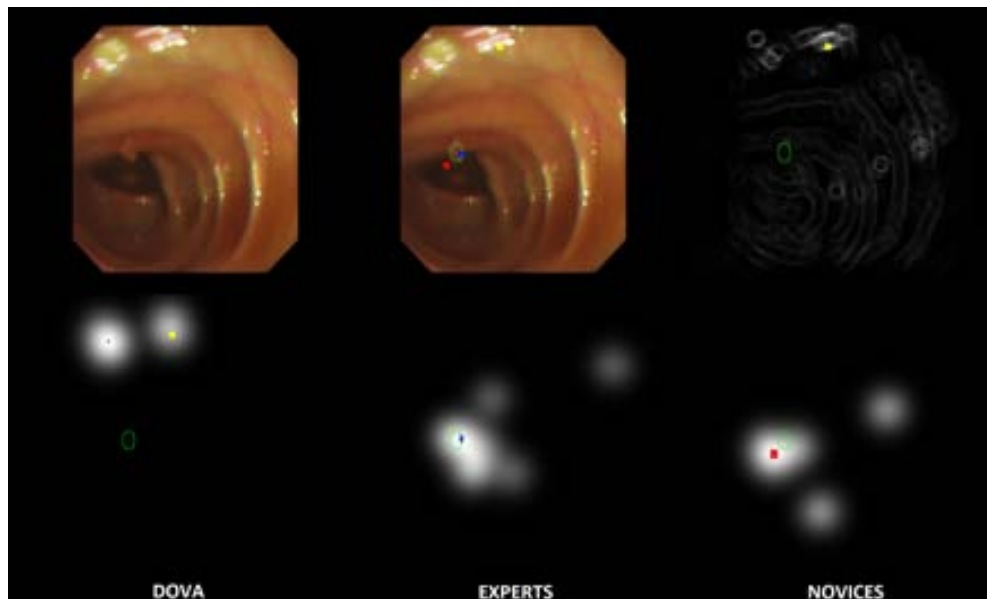
(b)

Figure 8.9: (a) Video 10: Comparison between WM-DOVA, experts and novices saliency maps. (b) Video 13: Comparison between WM-DOVA, experts and novices saliency maps. In both examples the maxima of WM-DOVA is painted by a yellow square, the maxima of experts is represented by a blue square and the maxima of novices is represented as a red square. Polyp mask is painted as a yellow ellipse.

8.8. Comparison of DOVA saliency maps with physicians' task driven saliency maps189



(a)



(b)

Figure 8.10: (a) Video 14: Comparison between WM-DOVA, experts and novices saliency maps. (b) Video 18: Comparison between WM-DOVA, experts and novices saliency maps. In both examples the maxima of WM-DOVA is painted by a yellow square, the maxima of experts is represented by a blue square and the maxima of novices is represented as a red square. Polyp mask is painted as a yellow ellipse.

8.8.2 Dwelling time

As we did for the case of task-driven saliency maps for experts and novices, we present here dwelling time results obtained by our DOVA saliency maps, along with a comparison with the data obtained from physicians. We can see in Table 8.14 the experimental setup:

Video	Dwelling time DOVA	Dwelling time Experts	Dwelling time Novices
2	40	85	60
3	44	70	56
4	32	146	103
5	13	78	73
8	29	106	80
9	28	71	70
10	1	3	10
12	120	119	119
13	5	75	59
14	10	74	35
15	44	104	68
18	6	67	26

Table 8.14: Dwelling time results.

As it can be seen from Table 8.14 for almost all the videos the dwelling time for DOVA saliency maps is lower than for experts' and novices'. There are several reasons for this difference:

- As expressed at the end of the previous subsection, WM-DOVA is still affected by some other elements of the endoluminal scene that have not been studied in this thesis such as folds, bubbles or fecal content.
- We have to consider that there are some cases where WM-DOVA will always fail because of the size of the polyp. If the polyp is too big (see Figure 8.11 (a)) or too small (Figure 8.11 (b)), the sectors that we use to accumulate will probably not catch the contours of the polyps.
- Finally, we have to mention that we have always used the maxima of WM-DOVA value in order to localize the polyp but maybe by lowering the threshold we could catch some polyps that we miss. We will sketch more about this in the Future Work.

8.8.3 Concentration rate

Finally in order to end the comparison between DOVA saliency maps and physicians' observations, we offer experimental results about concentration rate in Table 8.15 and Table 8.16:

8.8. Comparison of DOVA saliency maps with physicians' task driven saliency maps191

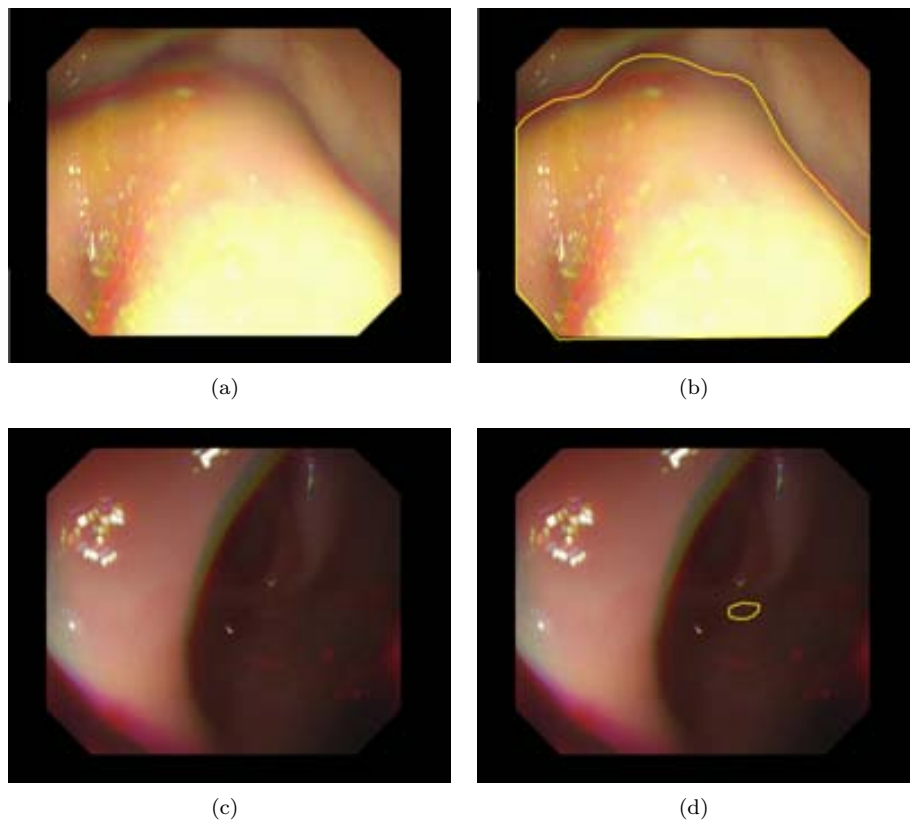


Figure 8.11: Some reasons for errors on locating polyps. (a) Original image with a big polyp; (b) Image with polyp mask superimposed; (c) Original image with a small polyp; (d) Image with polyp mask superimposed.

Video	% Concentration DOVA	% Concentration Experts	Difference DOVA-Experts
2	18.7%	26.66%	-7.96%
3	4.74%	3.22%	1.52%
4	15.21%	22.27%	-7.06%
5	13.02%	6.47%	6.55%
8	16.44%	22.44%	-6%
9	17.52%	25.6%	-8.08%
10	14.3%	17.9%	-3.6%
12	52.80%	72.84%	-20.04%
13	17.33%	23.28%	-5.95%
14	8.2%	9.47%	-1.27%
15	26.15%	34.43%	-8.28%
18	13.4%	5.34%	8.06%

Table 8.15: Difference in concentration rate in the DOVA saliency maps' first fixation frame.

Video	% Concentration DOVA	% Concentration Novices	% Difference DOVA-Novices
2	18.7%	24.14%	-5.44%
3	4.74%	2.39%	2.35%
4	15.21%	20.15%	-4.94%
5	13.02%	5.48%	7.54%
8	16.44%	17.53%	-1.09%
9	17.52%	23.8%	-6.28%
10	14.3%	15.82%	-1.52%
12	52.80%	69.78%	16.98%
13	17.33%	16.82%	0.51%
14	8.2%	7.45%	0.75%
15	26.15%	27.0%	-0.85%
18	13.4%	3.7%	9.7%

Table 8.16: Difference in concentration rate in the DOVA saliency maps' first fixation frame.

As it can be observed from Table 8.15 and Table 8.16, in this case our DOVA saliency maps offer worse concentration rate than experts in most of the cases but these results improve if we compare our DOVA saliency maps results with novices, where we get better or comparable results for half the videos. This results are clearly damaged by all WM-DOVA's original failures that have been expressed before. We also can observe that our concentration rates, except for some cases, are lower than 20% in mean which should indicate that although we place the maxima of DOVA saliency map inside the polyp, we are highlighting (in terms of giving saliency response) to another bigger parts of the image. This indicates that we should continue working on those challenges in order to reduce their contribution in terms of valley

information and, consequently, in terms of DOVA saliency maps.

8.9 Preliminary results of the use of DOVA saliency maps as polyp detectors

In the experiments that we have shown until this moment we have been working in frames with polyps in order to validate our methods, which are based on a model of appearance. In this final experimental section we show preliminary results on what happens in frames without polyps, in order to see if our current version of polyp localization method could also be used as a tool for a further polyp detection system or not.

We show in Table 8.17 results of the mean and standard deviation of WM-DOVA energy map for all the frames in each video.

Video	Mean WM-DOVA (all)	Mean WM-DOVA (polyp)	Mean WM-DOVA (no polyp frames)	Difference
2	1173.3	1209.3	1136.2	154.1
3	1217.7	1217.9	1217.2	0.67
4	44.7	6.05	59.09	-53.04
5	2739.8	2731.8	2740.7	-8.94
8	374.5	257.27	386.4	-129.12
9	1460.8	751.55	1556.1	-814.5
10	6265.6	9163.2	5964.6	3198.6
12	235.4	141	391.5	-249.5
13	1146.6	854.9	1184.6	-329.7
14	127.8	120.3	155.1	-35.1
15	4.78	4.0	6.6	-2.6
18	86.1	127.81	44.42	83.39

Table 8.17: Comparison of mean of WM-DOVA energy map for all the image, frames with and without polyp.

We can see from this table that there are half of the videos where the mean value of WM-DOVA energy map is higher for frames with polyps and without polyps but we can also see that there are videos such as video 8 and video 9 where the opposite difference is higher. One of the objectives of this final experiment was to test if we can obtain a threshold value that can determine if a frame contains a polyp or not. To do so we have done two different analysis, one considering all the frames of the video and the other one considering only what happens inside and outside the polyp mask in those frames that contain a polyp.

Concerning to the first analysis, we can observe in Table 8.18 that we cannot say by means of the maxima of WM-DOVA for the whole image that a frame contains a

polyp or not, because there are a few videos where it seems the maxima of WM-DOVA is higher for frames with polyp but the general trend is the contrary.

Video	Max WM-DOVA (frame with polyp)	Max WM-DOVA (frame without polyp)	Difference
2	1.6	1.16	0.44
3	1.12	5.9	-4.77
4	0.54	0.53	0.01
5	1.69	5.86	-4.17
8	0.51	2.15	-1.64
9	1.86	11.45	-9.58
10	0.92	2.62	-1.7
12	0.3	0.11	0.19
13	1.4	0.945	0.462
14	0.14	0.89	-0.75
15	0.12	0.34	-0.22
18	0.339	0.479	-0.14

Table 8.18: Comparison of maxima of WM-DOVA energy map for frames with polyp and without polyp.

Finally, in order to find out if this trend keeps going for frames with polyp, we have calculated the maxima of WM-DOVA inside and outside the polyp mask for frames with polyp. We can observe results of this analysis in Table 8.19, where again the conclusion is that, at its current state, we cannot use only the maxima of WM-DOVA energy map to determine if a frame contains a polyp or no because for half of the videos we obtain good results (maxima of WM-DOVA inside the polyp is higher than outside the polyp) but for the other half we obtain the opposite result.

As we can see from the experiments of this section, at this stage of development we cannot use the direct result of WM-DOVA to decide if a frame contains a polyp or not. As mentioned before, there are several reasons that could have led us to these results, such as not considering some elements of the endoluminal scene or just problems of the algorithm itself such as the parameter value dependance. Nevertheless these results encourage us to keep on improving our algorithms in order to provide in the future better polyp detection results.

8.10 Discussion

Our polyp localization and polyp segmentation methods perform well in our database, which consist of 380 examples of polyp appearance. Once these methods have been tested in our experimental setup, we want to check if their performance is at least comparable to physicians' in order to potentially incorporate some of our tools in future intelligent systems for colonoscopy.

Video	Max WM-DOVA (inside polyp)	Max WM-DOVA (outside polyp)	Difference
2	0.6852	1.33	-0.65
3	0.57	0.5	0.075
4	0.4	0.36	0.04
5	1.3	1.01	0.29
8	0.24	0.48	-0.24
9	1.56	1.63	-0.07
10	0.14	0.98	-0.84
12	0.3	0.22	0.08
13	1.4	0.945	0.462
14	0.14	0.19	-0.05
15	0.43	0.34	0.09
18	0.151	0.327	-0.17

Table 8.19: Comparison of maxima of WM-DOVA energy map for frames with polyp, inside and outside the polyp mask.

In order to do so, we were provided with several videos where the physicians' observations were captured by means of an eye-tracking device. The objective of the experiments shown in this chapter is to provide a first framework where to compare the performance of humans against computer vision algorithms. Another of the objectives planned for this experiment was to check if our polyp localization method can be used as a way to measure the expertise of a physician, in order to incorporate our methods in training systems.

Therefore, the first step was to determine the expertise of the physicians that participated on the experiments according to the number of procedures they have done and then we have integrated their fixations into task driven saliency maps, which highlight for a given frame the area of the images where the group of physicians has paid more attention. We have also calculated metrics related to eye tracking in order to compare factors such as reaction time or concentration rate, which have shown a difference between experts and novices.

Once the comparison between the different types of physicians have been done, we continue our analysis by incorporating our polyp localization method, which has to be transformed into a saliency map in order to make a fair comparison. Our polyp localization methods perform comparably to physicians in terms of reaction time but suffers from some limitations of the method (such size constraints or no description of some elements of the endoluminal scene) that make dwelling time and concentration rate results worse.

Considering the limitations of our method, we have finally performed an experiment in order to find out if our DOVA energy maps could be used in some way to indicate polyp presence in a general frame, which may or not contain a polyp. Preliminary results show that we get comparable results in reaction time but worse for other measures. These results encourage us to improve our current methods in order

to provide with better polyp detection results.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

We have proposed in this thesis our polyp localization and segmentation methods based on a model of appearance for polyps. The scope of the research included in this thesis is the building of intelligent systems for colonoscopy, which aim at providing additional information to the colonoscopy procedure. In this section we expose the main conclusions that can be extracted from this work:

- Polyp appearance varies a lot within a colonoscopy study and between different colonoscopy studies. Instead of relying on a model based on the general shape that polyps may have, we have developed our model of appearance for polyps based on how they do appear in colonoscopy frames. In order to develop this model we have taken into account how colonoscopy frames are acquired and generated. The aim of modelling the polyp appearance has been to find common elements that may lead to automatize its localization in further examples. After an exhaustive observation of several studies, we have found that intensity valleys appear to surround polyps therefore we use as a first cue for our algorithms the presence of valleys.
- Our model of appearance defines polyps as surrounded by valleys but polyps are not the only source of valley information in the endoluminal scene. There are other elements such as specular highlights, blood vessels or folds that also generate a response in terms of valley information. We state that by tackling the effect of these elements our polyp localization and segmentation methods will improve their performance, as was shown in Chapter 7. Therefore we have defined a novel depth of valleys image which aggregates and acknowledge the effect of the before mentioned elements.
- We have developed our polyp localization method by taking into account that protruding objects such as polyps should be enclosed by valleys. Our localization

methods integrate this depth of valleys information in order to assess if a given point is interior to an object. We have introduced several alternatives, from a more shape-specific approach such as EF-DOVA to more general approaches such as SA-DOVA or WM-DOVA. We have proven that as polyp appearance varies a lot it, shape-specific approaches tend to work well only in very specific cases. This led us to discard these type of approaches and develop more general algorithms. In this case WM-DOVA has proved to perform better than the other approaches, being more discriminative in terms of polyp localization.

- We have presented our two approaches for polyp segmentation and by comparing them we have shown that including polyp-specific constraints from the beginning such as the information that polyp localization provides leads to an improvement of the final polyp region. We have shown that our method outperforms more general segmentation methods such as normalized cuts [97] or TurboPixels [62]. We have also presented promising results on polyp region classification based on the output of our polyp localization methods.
- In general our DOVA energy maps have been proved as a successful method to localize polyps in images with polyp presence but they have yet to be proved as polyp detectors. DOVA maps also suffer from a worsening of the quality of the image. Although a first approach to solve the problem that lateral views may represent was presented via our Depth of Valleys image, there are still some cases where this approach fails to perform well. The lack of information about other elements of the endoluminal scene, such as lumen or folds, may also have an impact in our method. In order to go towards the development of polyp detection method we should be able to overcome these issues.
- We have also shown promising results on the correlation between DOVA energy maps and physicians' observation, which may indicate a potential use of our methods in application such as computer aided diagnosis or as a part of training systems. It is true that our preliminary results indicate that we should improve our methods in certain areas such as temporal coherence between consecutive frames or solving scale-based issues.

9.2 Future Work

Although we have presented promising results on both polyp localization and polyp segmentation, there is room for improvement in all the areas that have been covered in this thesis. Following the same structure of the thesis, we introduce some of the future lines of research that could be followed in the future:

- As mentioned before we should be able to find a better solution for lateral views of polyps because our first approximation by means of the use of the morphological gradient still does not work well for all the cases. By detecting these lateral contours better we should be able to differentiate between polyp-generated contours and the rest.

- As we have expressed in the experimental results for polyp localization and polyp segmentation, there are elements of the endoluminal scene that still have an impact on the performance of our system, such as lumen, folds or fecal content. We believe that addressing these elements may provide with better localization results and, consequently, with better defined polyp segmentation regions. We also plan to extend the work done about blood vessels, as we think the mitigation process can be improved by studying some other color spaces or different metrics.
- We have shown experimental results obtained by the best parameter combination of each algorithm but, in the case of DOVA energy maps there are some parameters that could be automatized such as the minimum and maximum radius. We plan to explore this in the near future and we expect to improve by catching some either too small or too big polyps that we could have missed.
- There are some methods that still can be improved such as VO-DOVA polyp localization where, by means of better valley orientations, we could discard some non-polyp boundaries.
- All our polyp localization and polyp segmentation results have been obtained only by means of the maxima of accumulation as we wanted to assess the degree of validity of our models but we can think of adding more complex classification systems that could yield better results.
- In order to be incorporated in future intelligent systems, our methods need to be faster. Although we only last less than 5 seconds in the whole process of localization and segmentation, images needed of a preprocessing that also affects performance cost. This cost should be reduced in order to reach real-time performance levels.
- We have presented a preliminary study on how to compare physicians' observations with the output of computer vision methods. It is clear that this comparison could be extended by adding more players to the comparison, such as saliency models (Itti, GVBS) or feature detectors (such as SIFT or SURF), which should be done in the near future. This comparison will also help us to improve our polyp localization and segmentation methods by learning where our systems do not perform well and others do.
- In order to progress towards a polyp detection method we should be able to discard automatically those frames with bad quality, following the trend that was shown in Chapter 2 about non-informative frame identification. We should also include temporal and spatial coherence to improve polyp localization in videos. This could be done by considering for a given frame polyp localization results from anterior and posterior frames. This could lead to an improvement of polyp localization in bad quality frames.

Chapter 10

List of Publications

10.1 Journals, Book chapters

- Jorge Bernal, Javier Sánchez and Fernando Vilariño, "Towards Automatic Polyp Detection with a Polyp Appearance Model", Pattern Recognition, Volume 45, Issue 9, Pages 3047-3582 (September 2012) (Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011)).
- Jorge Bernal, Fernando Vilariño and Javier Sánchez . "Towards Intelligent Systems for Colonoscopy", Colonoscopy, Paul Miskovitz (Ed.), ISBN: 978-953-307-568-6, InTech, Available from: <http://www.intechopen.com/books/colonoscopy/towards-intelligent-systems-for-colonoscopy> (2011).

10.2 International Conferences

- Jorge Bernal, Javier Sánchez and Fernando Vilariño, "Integration of Valley Orientation Distribution for Polyp Region Identification in Colonoscopy", Proceedings of the 2011 MICCAI Abdominal Imaging Workshop. Computational and Clinical Applications, Lecture Notes in Computer Science Volume 7029, pp. 76-83, Toronto, Canada, Springer, Heidelberg (2012).
- Jorge Bernal, Javier Sánchez and Fernando Vilariño, " A Region Segmentation Method for Colonoscopy Images Using a Model of Polyp Appearance", Proceedings of the 2011 Iberian Conference on Pattern Recognition and Image Analysis IbPRIA 2011. Lecture Notes in Computer Science Volume 7029, vol. 6669, pp. 134-142. Las Palmas de Gran Canaria, Spain. Springer, Heidelberg (2011).
- Jorge Bernal, Javier Sánchez and Fernando Vilariño, " Current Challenges on Polyp Detection in Colonoscopy Videos: From Region Segmentation to Region Classification. a Pattern Recognition-based Approach", Proceedings of the 2nd

International Workshop on Medical Image Analysis and Description for Diagnosis Systems MIAD 2011, pp. 62-71. Rome, Italy (2011).

10.3 National Conferences

- Jorge Bernal, Javier Sánchez and Fernando Vilariño, "Depth of Valleys Accumulation Algorithm for Object Detection", Proceedings of the 14th International Conference of the Catalan Association of Artificial Intelligence CCIA 2011, pp. 71-80. Lleida, Spain (2011).
- Jorge Bernal, Javier Sánchez and Fernando Vilariño, "Reduction of Pattern Search Area in Colonoscopy Images by Merging Non-Informative Regions", Proceedings of the XXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica Caseib 2010, pp. 88-96. Madrid, Spain (2010).

10.4 Internal Conferences and Technical Reports

- Jorge Bernal, Fernando Vilariño and Javier Sánchez . "Feature Detectors and Feature Descriptors: Where We Are Now, Colonoscopy", CVC Technical Report 154 (2010).
- Jorge Bernal, Javier Sánchez and Fernando Vilariño, "Effect of Frontier Weakness Measures on the Accurate Identification of Non-Informative Regions", Proceedings of the 5th CVC Workshop on the Progress of Research and Development, pp. 167-170. Barcelona, Spain (2010).
- Jorge Bernal, Javier Sánchez and Fernando Vilariño, "Towards Automatic Polyp Detection with a Polyp Appearance Model", Proceedings of the 6th CVC Workshop on the Progress of Research and Development, pp. 115-119. Barcelona, Spain (2011).
- Jorge Bernal, Javier Sánchez and Fernando Vilariño, "Towards Polyp Localization and Polyp Segmentation by means of DOVA energy maps", Proceedings of the 7th CVC Workshop on the Progress of Research and Development, Barcelona, Spain (2012).

Apart from this publications, there is another contribution that has been sent to conferences:

- Joan Manel Nuñez, Jorge Bernal, Javier Sánchez and Fernando Vilariño, "Blood Vessel Characterization in Colonoscopy Images to Improve Polyp Localization", 8th International Conference on Computer Vision Theory and Applications. Barcelona, Spain (2013).

Appendix A

Technical data

Although we have provided some technical results when comparing the several polyp localization methods, we will use this appendix to measure the time each stage of the processing scheme, in order to provide with more data for comparison purposes.

A.1 Technical equipment

All the results have been obtained by a PC with an Intel Core i7-3930 K twelve-core processor with 8 GB of RAM Memory. In order to develop the different algorithms we have used Matlab scripts and also MEX functions to incorporate functions of the GIPL library. All the processing times that will be exposed along the next sections are calculated per image.

A.2 Image preprocessing and Valley Detection

In both polyp localization and polyp segmentation methods we have explained the reason why we use several image preprocessing techniques, which impact in processing time we show in Table A.1.

A.3 Polyp localization

We present again in Table A.2 the processing time for each of the four DOVA energy maps calculation methods.

As it was mentioned in the corresponding chapter, polyp localization needs of three stages namely: image preprocessing, valley detection and DOVA energy map.

Technique	Chapter	Time (seconds)
Image deinterlacing	4	0.1
Black mask inpainting (along with specular highlights)	4	8.53
Specular highlights detection	4	15.57
Specular highlights inpainting (along with black mask)	4	8.53
Calculation of the DoV image	4	2.57
Blood vessels mitigation	4	0
Combined image preprocessing time	4	30.77

Table A.1: Computation time of preprocessing methods.

Measure	EF-DOVA	SA-DOVA	WM-DOVA	VO-DOVA
Number of experiments	9600	72900	218700	218700
Mean processing time per experiment (mins)	1521	15	30	80
Mean processing time per image (secs)	304.2	3	6	16

Table A.2: DOVA energy map processing times.

We offer in Table A.3 a summary of all the processing times for each of the stages along with the accumulative result.

A.4 Polyp segmentation

We have explained in Chapter 6 the two different approaches that we have developed to obtain polyp segmentation. We present processing time results in Table A.4

As it was also mentioned in Chapter 6, our polyp segmentation methods consists of different stages, which include image preprocessing and polyp localization. We offer total segmentation results in Figure A.5.

As it can be seen by using polyp segmentation from the contributors to the maxima of the DOVA energy map we are three times faster than with the previous method. In order to finish this appendix, we show in Figure A.1 some graphs to illustrate the

Technique	Chapter	Time (seconds)
Image preprocessing	4	28.2
Calculation of Depth of Valleys image	4	2.54
WM-DOVA energy map	5	6
Polyp localization	5	36.77

Table A.3: Computation time of preprocessing methods.

Technique	Chapter	Time (seconds)
Segmentation after Region Merging	6	81.38
Polyp Segmentation by means of maxima contributors	6	3.07

Table A.4: Computation time of preprocessing methods.

Technique	Chapter	Time (seconds)
Polyp segmentation after Region Merging	6	118.15
Polyp Segmentation by means of maxima contributors	6	39.74

Table A.5: Polyp segmentation computation time.

percentage of processing time that is spent on each stage.

Analyzing image preprocessing graph (Figure A.1 (a)) we can see that the biggest part of the computation time goes to specular highlights detection followed by the inpainting of both specular highlights and black mask. Image preprocessing constitutes more than the 75% of total processing time for polyp localization (Figure A.1 (b)) and around 70% for the case of polyp segmentation from the contributors of the maxima of DOVA map. On the other hand, as region merging process takes a lot of time, the incidence of image preprocessing or DOVA energy maps is minor for polyp segmentation after region merging, as it can be seen in Figure A.1 (c).

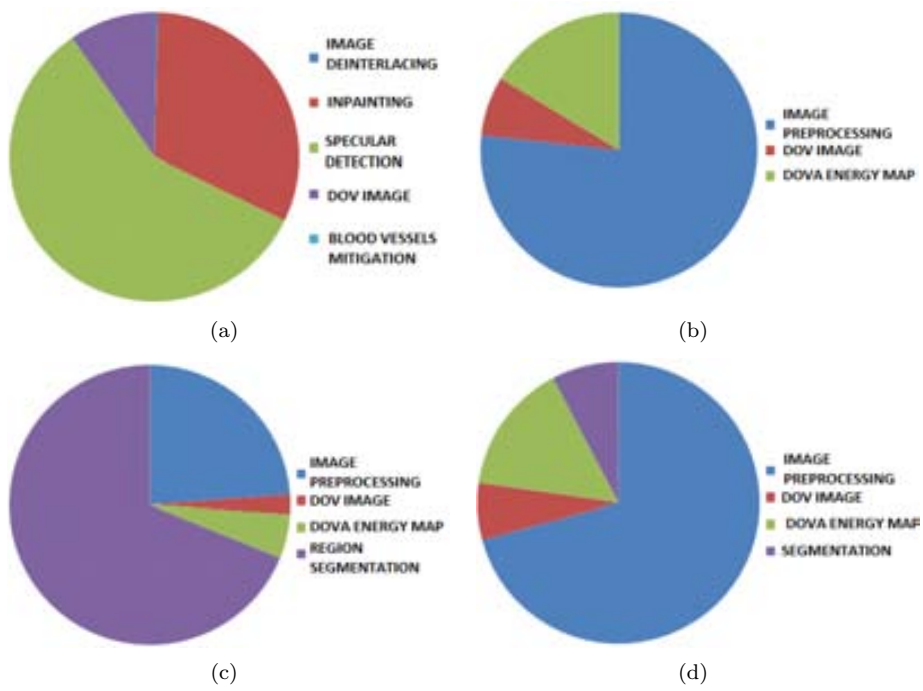


Figure A.1: Summary of processing time graphs: (a) Image preprocessing + DoV image; (b) Polyp localization; (c) Polyp segmentation after Region Merging and (d) Polyp segmentation by means of maxima contributors.

Bibliography

- [1] S. Ameling, S. Wirth, D. Paulus, G. Lacey, and F. Vilariño. Texture-based polyp detection in colonoscopy. *Bildverarbeitung für die Medizin 2009*, pages 346–350, 2009.
- [2] American Cancer Society. Colorectal cancer facts & figures 2011-2013, 2011. [Online; accessed 26-July-2012].
- [3] M. Arnold, A. Ghosh, S. Ameling, and G. Lacey. Automatic segmentation and inpainting of specular highlights for endoscopic imaging. *Journal on Image and Video Processing*, 2010:9, 2010.
- [4] M. Arnold, A. Ghosh, G. Lacey, S. Patchett, and H. Mulcahy. Indistinct frame detection in colonoscopy videos. In *2009 13th International Machine Vision and Image Processing Conference*, pages 47–52, 2009.
- [5] R. Barton. Accrediting competence in colonoscopy: validity and reliability of the uk joint advisory group/nhs bowel cancer screening programme accreditation assessment. *Gastrointestinal Endoscopy*, 67(5):AB77–AB77, 2008.
- [6] R.J. Basseri, B. Basseri, and K.A. Papadakis. Dysplasia and cancer in inflammatory bowel disease. *Expert Review of Gastroenterology and Hepatology*, 5(1):59–66, 2011.
- [7] Nancy N. Baxter, Meredith A. Goldwasser, Lawrence F. Paszat, Refik Saskin, David R. Urbach, and Linda Rabeneck. Association of colonoscopy and death from colorectal cancer. *Annals of Internal Medicine*, 150(1):1–8, 2009.
- [8] J. Bernal and F. Sánchez, J. andand Vilariño. *Colonoscopy Book 1: Towards Intelligent Systems for Colonoscopy*. In-Tech, 2011.
- [9] J. Bernal, J. Sánchez, and F. Vilariño. Feature Detectors and Feature Descriptors: Where We Are Now. Technical Report 154, Computer Vision Center, September 2010.
- [10] J. Bernal, J. Sánchez, and F. Vilariño. Reduction of Pattern Search Area in Colonoscopy Images by Merging Non-Informative Regions. In *Proceedings of the XXVIII Congreso Anual de la Sociedad Española de Ingeniería Biomédica*, pages 88–96, Madrid, Spain, November 2010.

- [11] J. Bernal, J. Sánchez, and F. Vilariño. Depth of valleys accumulation algorithm for object detection. In *CCIA*, pages 71–80, 2011.
- [12] J. Bernal, J. Sánchez, and F. Vilariño. Integration of valley orientation distribution for polyp region identification in colonoscopy. *Abdominal Imaging. Computational and Clinical Applications*, pages 76–83, 2012.
- [13] J. Bernal, J. Sánchez, and F. Vilarino. Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition*, 2012.
- [14] J. Bernal, F. Vilariño, and J. Sánchez. Feature detectors and feature descriptors: Where we are now. Technical Report 154, Computer Vision Center & Computer Science Department Universitat Autònoma de Barcelona, September 2010.
- [15] V. Bevilacqua, M. Cortellino, M. Piccinni, A. Scarpa, D. Taurino, G. Mastronardi, M. Moschetta, and G. Angelelli. Image processing framework for virtual colonoscopy. *Emerging Intelligent Computing Technology and Applications*, pages 965–974, 2009.
- [16] J.H. Bond. Polyp guideline: diagnosis, treatment, and surveillance for patients with colorectal polyps. *The American journal of gastroenterology*, 95(11):3053–3063, 2000.
- [17] A. Borji, DN Sihite, and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, 2012.
- [18] I. Bratko, I. Mozetič, and N. Lavrač. KARDIO: a study in deep and qualitative knowledge for expert systems. *MIT Press*, 1990.
- [19] G. Brelstaff and A. Blake. Detecting specular reflections using lambertian constraints. In *Computer Vision., Second International Conference on*, pages 297–302, dec 1988.
- [20] B. Bressler, L.F. Paszat, Z. Chen, D.M. Rothwell, C. Vinden, and L. Rabeneck. Rates of new or missed colorectal cancers after colonoscopy and their risk factors: a population-based analysis. *Gastroenterology*, 132(1):96–102, 2007.
- [21] C.S. Burrus, R.A. Gopinath, and H. Guo. *Introduction to wavelets and wavelet transforms: a primer*, volume 23. Prentice hall Upper Saddle River NJ, 1998.
- [22] Y. Cao, D. Liu, W. Tavanapong, J. Wong, J.H. Oh, and P.C. de Groen. Computer-aided detection of diagnostic and therapeutic operations in colonoscopy videos. *Biomedical Engineering, IEEE Transactions on*, 54(7):1268–1279, 2007.
- [23] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transactions on medical imaging*, 8(3):263–269, 1989.

- [24] H. Chen, WF Lee, TH Huang, and SL Yeh. Learning-based prediction of visual attention for video signals. *Image Processing, IEEE Transactions on*, 20(99):1–1, 2011.
- [25] Y. Cheng. Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(8):790–799, 2002.
- [26] A.J. Chung, F. Deligianni, X.P. Hu, and G.Z. Yang. Extraction of visual features with eye tracking for saliency driven 2d/3d registration. *Image and Vision Computing*, 23(11):999–1008, 2005.
- [27] J.M. Church. Clinical significance of small colorectal polyps. *Diseases of the colon & rectum*, 47(4):481–485, 2004.
- [28] M.T. Coimbra and J.P.S. Cunha. MPEG-7 visual descriptors and contributions for automated feature extraction in capsule endoscopy. *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(5):628–637, 2006.
- [29] L.C. Cummings and G.S. Cooper. Editorial: Detection of small polyps: Much ado about nothing&quest. *The American journal of gastroenterology*, 105(12):2586–2587, 2010.
- [30] A.H. Dachman, I. Sethi, and P. Lefere. Virtual colonoscopy. *Textbook of Clinical Gastroenterology and Hepatology, Second Edition*, pages 1027–1034, 2012.
- [31] R. Dahyot, F. Vilariño, and G. Lacey. Improving the quality of color colonoscopy videos. *Journal on Image and Video Processing*, 2008:1–7, 2008.
- [32] R.E. Davila, E. Rajan, D.G. Adler, J. Egan, W.K. Hirota, J.A. Leighton, W. Qureshi, M.J. Zuckerman, R. Fanelli, J. Wheeler-Harbaugh, et al. Asge guideline: the role of endoscopy in the patient with lower-gi bleeding. *Gastrointestinal endoscopy*, 62(5):656, 2005.
- [33] G. De Haan and E.B. Bellers. Deinterlacing-an overview. *Proceedings of the IEEE*, 86(9):1839 –1857, sep 1998.
- [34] BV Dhandra, R. Hegadi, M. Hangarge, and VS Malemath. Analysis of abnormality in endoscopic images using combined hsi color space and watershed segmentation. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4, pages 695–698, 2006.
- [35] Antipuesto D.J. Colonoscopy procedure. Nursingcrib website, September 2010.
- [36] K.A. Dobie and C. Burke. *Colonoscopy for Dummies*. Wiley Publishing Inc., 2011.
- [37] A.T. Duchowski. *Eye tracking methodology: Theory and practice*, volume 373. Springer-Verlag New York Incorporated, 2007.
- [38] L. Espona, M. Carreira, M. Ortega, and M. Penedo. A snake for retinal vessel segmentation. *Pattern Recognition and Image Analysis*, pages 178–185, 2007.

- [39] D.A. Forsyth and J. Ponce. *Computer vision: a modern approach*. Prentice Hall Professional Technical Reference, 2002.
- [40] Ron Gershon, Allan D. Jepson, and John K. Tsotsos. The use of color in highlight identification. In *Proceedings of the 10th international joint conference on Artificial intelligence - Volume 2*, pages 752–754, San Francisco, CA, USA, 1987. Morgan Kaufmann Publishers Inc.
- [41] D. Gil et al. Structure-preserving smoothing of biomedical images. In *Computer Analysis of Images and Patterns*, pages 427–434. Springer, 2009.
- [42] C. Gunduz-Demir, M. Kandemir, A.B. Tosun, and C. Sokmensuer. Automatic segmentation of colon glands using object-graphs. *Medical image analysis*, 14(1):1–12, 2010.
- [43] P. Harding and N. Robertson. A comparison of feature detectors with passive and task-based visual saliency. *Image Analysis*, pages 716–725, 2009.
- [44] J.P. Hassinger and S.D. et al. Holubar. Effectiveness of a Multimedia-Based Educational Intervention for Improving Colon Cancer Literacy in Screening Colonoscopy Patients. *Diseases of the Colon & Rectum*, 53(9):1301, 2010.
- [45] AD Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *Medical Imaging, IEEE Transactions on*, 19(3):203–210, 2000.
- [46] S. Hwang, J.H. Oh, J.K. Lee, Y. Cao, W. Tavanapong, D. Liu, J. Wong, and P.C. de Groen. Automatic measurement of quality metrics for colonoscopy videos. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 912–921, 2005.
- [47] S. Hwang, J.H. Oh, and W. et al. Tavanapong. Polyp detection in colonoscopy video using elliptical shape feature. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 2, pages II–465. IEEE, 2007.
- [48] ICT Programme Committee. Fp7. ict - information and communication technologies. work programme 2011-12. Community Research and Development Information Service, 2011.
- [49] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain. Wireless capsule endoscopy. *Nature*, 405:417, 2000.
- [50] Y. Imai, Y. Kato, H. Kadoi, T. Horiuchi, and S. Tominaga. Estimation of multiple illuminants based on specular highlight detection. *Computational Color Imaging*, pages 85–98, 2011.
- [51] X. Jiang and D. Mojon. Adaptive local thresholding by verification-based multi-threshold probing with application to vessel detection in retinal images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(1):131–137, 2003.
- [52] J. Jones. Haustral folds. 2012.

- [53] J. Kang and R. Doraiswami. Real-time image processing system for endoscopic applications. In *IEEE Canadian Conference on Electrical and Computer Engineering*, volume 3, pages 1469 – 1472 vol.3, May 2003.
- [54] S.A. Karkanis, D.K. Iakovidis, and D.E. et al. Maroulis. Computer-aided tumor detection in endoscopic video using color wavelet features. *Information Technology in Biomedicine, IEEE Transactions on*, 7(3):141–152, 2003.
- [55] W. Kienzle, F.A. Wichmann, B. Scholkopf, and M.O. Franz. Learning an interest operator from human eye movements. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 24–24. IEEE, 2006.
- [56] C.E. Kistler, K.A. Kirby, D. Lee, M.A. Casadei, and L.C. Walter. Long-term outcomes following positive fecal occult blood test results in older adults: benefits and burdens. *Archives of internal medicine*, pages archinternmed–2011, 2011.
- [57] Gudrun J. Klinker, Steven A. Shafer, and Takeo Kanade. Using a color reflection model to separate highlights from object color. In *Proc. ICCV*, pages 145–150. IEEE, 1987.
- [58] S.M. Krishnan and P.M.Y. Goh. Quantitative parametrization of colonoscopic images by applying fuzzy technique. In *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, volume 3, pages 1121 –1123 vol.3, oct-2 nov 1997.
- [59] SM Krishnan, X. Yang, KL Chan, S. Kumar, and PMY Goh. Intestinal abnormality detection from endoscopic images. In *Engineering in Medicine and Biology Society, 1998. Proceedings of the 20th Annual International Conference of the IEEE*, volume 2, pages 895–898, 1998.
- [60] K.E. Lasser, J. Murillo, S. Lisboa, A.N. Casimir, L. Valley-Shah, K.M. Emmons, R.H. Fletcher, and J.Z. Ayanian. Colorectal cancer screening among ethnically diverse, low-income patients: a randomized controlled trial. *Archives of internal medicine*, 171(10):906, 2011.
- [61] B. Levin, D.A. Lieberman, B. McFarland, R.A. Smith, D. Brooks, K.S. Andrews, C. Dash, F.M. Giardiello, S. Glick, T.R. Levin, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: A joint guideline from the american cancer society, the us multi-society task force on colorectal cancer, and the american college of radiology. *CA: a cancer journal for clinicians*, 58(3):130–160, 2008.
- [62] A. Levinshtein, A. Stere, K.N. Kutulakos, D.J. Fleet, S.J. Dickinson, and K. Siddiqi. Turbopixels: Fast superpixels using geometric flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2290–2297, 2009.
- [63] P. Li, K.L. Chan, and SM Krishnan. Learning a multi-size patch-based hybrid kernel machine ensemble for abnormal region detection in colonoscopic

- images. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.
- [64] J.S. Lin, E.M. Webber, T.L. Beil, K.A. Goddard, E.P. Whitlock, et al. Fecal dna testing in screening for colorectal cancer in average-risk adults, 2012.
- [65] D. Liu, Y. Cao, K.H. Kim, S. Stanek, B. Dounggratanaex-Chai, K. Lin, W. Tavanapong, J. Wong, J.H. Oh, and P.C. de Groen. Arthemis: Annotation software in an integrated capturing and analysis system for colonoscopy. *Computer methods and programs in biomedicine*, 88(2):152–163, 2007.
- [66] V. Lohsiriwat, S. Prapasrivorakul, and W. Suthikeeree. Colorectal cancer screening by double contrast barium enema in thai people. *Asian Pacific Journal of Cancer Prevention*, 13:1273–1276, 2012.
- [67] A.M. López and F. et al. Lumbreras. Evaluation of methods for ridge and valley detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4):327–335, 1999.
- [68] Machine Vision Group. Cvc-databasecolon: A database for assessment of polyp detection, 2012. [Online; accessed 24-July-2012].
- [69] A.R. Malik. *End of insertion detection in colonoscopy videos*. PhD thesis, University of North Texas, 2010.
- [70] P. Mamula, D.G. Adler, J.D. Conway, D.L. Diehl, F.A. Farraye, S.V. Kantsevov, V. Kaul, S.R. Kethu, R.S. Kwon, S.A. Rodriguez, et al. Colonoscopy preparation. *Gastrointestinal endoscopy*, 69(7):1201, 2009.
- [71] D. Marín, A. Aquino, M.E. Gegúndez-Arias, and J.M. Bravo. A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *Medical Imaging, IEEE Transactions on*, 30(1):146–158, 2011.
- [72] DE Maroulis, DK Iakovidis, SA Karkanis, and DA Karras. CoLD: a versatile detection system for colorectal lesions in endoscopy video-frames. *Computer Methods and Programs in Biomedicine*, 70(2):151–166, 2003.
- [73] A.M. Mendonca and A. Campilho. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *Medical Imaging, IEEE Transactions on*, 25(9):1200–1213, 2006.
- [74] L. Meylan, S. Daly, and S. Süssstrunk. Tone mapping for high dynamic range displays. In *Proceedings of SPIE*, volume 6492, page 649210, 2007.
- [75] S.S. Morán, E. Torrella, D.P. Esteban, M.R. Baños, A. García, A. Ono, C.E. Pérez, P. Parra, Q.J. Cruzado, R.F. Pérez, et al. Colonoscopy quality assessment. *Revista española de enfermedades digestivas: Órgano oficial de la Sociedad Española de Patología Digestiva*, 101(2):107, 2009.
- [76] National Cancer Institute. Image Archive Resources, 2011.

- [77] C.S. Nunes, T.F. Mendonça, P. Amorim, D.A. Ferreira, and L. Antunes. Comparison of Neural Networks, Fuzzy and Stochastic Prediction Models for return of consciousness after general anesthesia. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, pages 4827–4832, 2005.
- [78] J. Oh, S. Hwang, J. Lee, W. Tavanapong, J. Wong, and C. de Groen. Informative frame classification for endoscopy video. *Medical Image Analysis*, 11(2):110 – 127, 2007.
- [79] J.H. Oh, S. Hwang, Y. Cao, W. Tavanapong, D. Liu, J. Wong, and P.C. de Groen. Measuring objective quality of colonoscopy. *Biomedical Engineering, IEEE Transactions on*, 56(9):2190–2196, 2009.
- [80] J.H. Oh, S. Hwang, W. Tavanapong, P.C. de Groen, and J. Wong. Blurry-frame detection and shot segmentation in colonoscopy videos. In *Proceedings of SPIE*, volume 5307, page 531, 2003.
- [81] Manuel M. Oliveira, Brian Bowen, Richard Mckenna, and Yu sung Chang. Fast digital image inpainting. In *Proceedings of the International Conference on Visualization, Imaging and Image Processing (VIIP 2001*, pages 261–266. ACTA Press, 2001.
- [82] Olympus. Colonoscopes. Accessed: 30/08/2012.
- [83] Pentax. Colonoscopy. Accessed: 30/08/2012.
- [84] B. T. Phong. Illumination for computer generated pictures. *Communications of ACM*, 18(6):311–317, 1975.
- [85] P.J. Pickhardt, P.A. Nugent, P.A. Mysliwiec, J.R. Choi, and W.R. Schindler. Location of adenomas missed by optical colonoscopy. *Annals of internal medicine*, 141(5):352–359, 2004.
- [86] C.M. Privitera and L.W. Stark. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(9):970–982, 2000.
- [87] E. Quintero, A. Castells, L. Bujanda, J. Cubiella, D. Salas, Á. Lanás, M. Andreu, F. Carballo, J.D. Morillas, C. Hernández, et al. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. *New England Journal of Medicine*, 366(8):697–706, 2012.
- [88] L. Rabeneck, L.F. Paszat, R.J. Hilsden, R. Saskin, D. Leddin, E. Grunfeld, E. Wai, M. Goldwasser, R. Sutradhar, and T.A. Stukel. Bleeding and perforation after outpatient colonoscopy and their risk factors in usual clinical practice. *Gastroenterology*, 135(6):1899–1906, 2008.
- [89] D. Regge and S. Halligan. Cad: How it works, how to use it, performance. *European Journal of Radiology*, 2012.

- [90] D.K. Rex, T.F. Imperiale, D.R. Latinovich, and L.L. Bratcher. Impact of bowel preparation on efficiency and cost of colonoscopy. *The American journal of gastroenterology*, 97(7):1696–1700, 2002.
- [91] F. Riaz, M.D. Ribeiro, and M.T. Coimbra. Quantitative comparison of segmentation methods for in-body images. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 5785–5788, September 2009.
- [92] Ch. Saint-Pierre, J. Boisvert, G. Grimard, and F. Cheriet. Detection and correction of specular reflections for automatic surgical tool segmentation in thoroscopic images. *Machine Vision and Applications*, 22:171–180, 2011. 10.1007/s00138-007-0099-6.
- [93] D.D. Salvucci and J.H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.
- [94] R.E. Schoen, P.F. Pinsky, J.L. Weissfeld, L.A. Yokochi, T. Church, A.O. Laiyemo, R. Bresalier, G.L. Andriole, S.S. Buys, E.D. Crawford, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *New England Journal of Medicine*, 366(25):2345–2357, 2012.
- [95] N. Segnan, J. Patnick, and L. von Karsa. *European guidelines for quality assurance in colorectal cancer screening and diagnosis*. Luxembourg: Publications Office of the European Union, 2011.
- [96] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, Inc. Orlando, FL, USA, 1983.
- [97] J. Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2000.
- [98] J. Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2002.
- [99] J.V.B. Soares, J.J.G. Leandro, R.M. Cesar, H.F. Jelinek, and M.J. Cree. Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *Medical Imaging, IEEE Transactions on*, 25(9):1214–1222, 2006.
- [100] J. Staal, M.D. Abràmoff, M. Niemeijer, M.A. Viergever, and B. Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *Medical Imaging, IEEE Transactions on*, 23(4):501–509, 2004.
- [101] R.T. Tan and K. Ikeuchi. Separating reflection components of textured surfaces using a single image. *Digitally Archiving Cultural Objects*, pages 353–384, 2008.
- [102] R.J.M. Teixeira. Interfaces and visualization in clinical endoscopy. *Master Thesis*, 2009.

- [103] H. Tian, T. Srikanthan, and K. Vijayan Asari. Automatic segmentation algorithm for the extraction of lumen region and boundary from endoscopic images. *Medical and Biological Engineering and Computing*, 39(1):8–14, 2001.
- [104] M.P. Tjoa and S.M. Krishnan. Feature extraction for the analysis of colon status from the endoscopic images. *BioMedical Engineering OnLine*, 2(9):1–17, 2003.
- [105] MP Tjoa, SM Krishnan, and R. Doraiswami. Automated diagnosis for segmentation of colonoscopic images using chromatic features. In *Electrical and Computer Engineering, 2002. IEEE CCECE 2002. Canadian Conference on*, volume 2, pages 1177–1180, 2002.
- [106] Tresca, A. The Stages of Colon and Rectal Cancer. *New York Times (About.com)*, page 1, 2010.
- [107] C. van Wijk, V.F. van Ravesteijn, F.M. Vos, and L.J. van Vliet. Detection and segmentation of colonic polyps on implicit isosurfaces by second principal curvature flow. *Medical Imaging, IEEE Transactions on*, 29(3):688–698, 2010.
- [108] F. Vilariño. *A Machine Learning Approach for Intestinal Motility Assessment with Capsule Endoscopy*. PhD thesis, Universitat Autònoma de Barcelona and Computer Vision Center, 2006.
- [109] F. Vilariño, S. Ameling, G. Lacey, A. Ghosh, S. Patchett, and H. Mulcahy. Eye tracking search patterns in expert and trainee colonoscopist: A novel method of assessing endoscopic competency? In *Book of abstracts from the Digestive Disease Week*, Chicago, America, May-June 2009.
- [110] F. Vilariño, G. Lacey, J. Zhou, H. Mulcahy, and S. Patchett. Automatic labeling of colonoscopy video for cancer detection. *Pattern Recognition and Image Analysis*, pages 290–297, 2007.
- [111] F. Vilarino, P. Spyridonos, F. DeIorio, J. Vitrià, F. Azpiroz, and P. Radeva. Intestinal motility assessment with video capsule endoscopy: automatic annotation of phasic intestinal contractions. *Medical Imaging, IEEE Transactions on*, 29(2):246–259, 2010.
- [112] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE transactions on pattern analysis and machine intelligence*, 13(6):583–598, 1991.
- [113] S. Viswanath, D. Palumbo, J. Chappelow, P. Patel, B.N. Bloch, N. Rofsky, R. Lenkinski, E. Genega, and A. Madabhushi. Empirical evaluation of bias field correction algorithms for computer-aided detection of prostate cancer on T2w MRI. In *Proceedings of SPIE*, volume 7963, page 79630V, 2011.
- [114] Z. Wang, L. Li, J. Anderson, D. Harrington, and Z. Liang. Colonic polyp characterization and detection based on both morphological and texture features. In *International Congress Series*, volume 1268, pages 1004–1009. Elsevier, 2004.

- [115] J. Wei, H.P. Chan, C. Zhou, Y.T. Wu, B. Sahiner, L.M. Hadjiiski, M.A. Roubidoux, and M.A. Helvie. Computer-aided detection of breast masses: Four-view strategy for screening mammography. *Medical Physics*, 38:1867, 2011.
- [116] J.A. Wilschut, E.W. Steyerberg, M.E. Van Leerdam, I. Lansdorp-Vogelaar, J.D.F. Habbema, and M. Van Ballegooijen. How much colonoscopy screening should be recommended to individuals with various degrees of family history of colorectal cancer? *Cancer*, 117(18):4166–4174, 2011.
- [117] C.L. Wyatt, Y. Ge, and D.J. Vining. Automatic segmentation of the colon for virtual colonoscopy. *Computerized Medical Imaging and Graphics*, 24(1):1 – 9, 2000.
- [118] L. Xu and S. Luo. A novel method for blood vessel detection from retinal images. *BioMedical Engineering OnLine*, 9(1):14, 2010.
- [119] Q. Yang, S. Wang, and N. Ahuja. Real-time specular highlight removal using bilateral filtering. *Computer Vision–ECCV 2010*, pages 87–100, 2010.
- [120] K.J. Yoon and I.S. Kweon. Correspondence search in the presence of specular highlights using specular-free two-band images. *Computer Vision–ACCV 2006*, pages 761–770, 2006.
- [121] F. Zana and J.C. Klein. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *Image Processing, IEEE Transactions on*, 10(7):1010–1019, 2001.
- [122] H. Zhu, Y. Fan, and Z. Liang. Improved Curvature Estimation for Shape Analysis in Computer-Aided Detection of Colonic Polyps. *Beijing, China*, page 19, 2010.