Vascular Pattern Characterization in Colonoscopy Images

A dissertation submitted by **Joan M. Núñez Do Rio** at Universitat Autònoma de Barcelona to fulfill the degree of **Doctor of Philosophy**.

Bellaterra, September 26, 2015

Director	**Dr. Fernando L. Vilariño Freire** Dept. Ciències de la Computació & Centre de Visió per Computador Universitat Autònoma de Barcelona
Thesis committee	**Dr. Mª Gloria Fernández Esparrach** Universitat de Barcelona Barcelona, Spain **Dr. Oriol Ramos Terrades** Universitat Autònoma de Barcelona Barcelona, Spain **Dr. Michal Drozdzal** Medtronic R&D Yoqneam, Israel

This document was typeset by the author using LaTeX 2ε.

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

ISBN: 978-84-943427-6-9

Printed by Ediciones Gráficas Rey, S.L.

Als meus pares.

We shall not cease from exploration
and the end of all our exploring
will be to arrive where we started
and know the place for the first time.
from *Little Gidding* (1942), T. S. Eliot (1888 - 1965),

Acknowledgements

Aquesta tesi mai hauria estat possible sense tota la gent que m'ha recolzat i m'ha acompanyat durant aquests anys. Ja sigui des de la col·laboració directa o simplement des de la vostra amistad o el vostre suport, tots vosaltres m'heu empès en el dia a dia per arribar fins aquí.

En primer lloc, vull agrair el meu director el Dr. Fernando Vilariño per donar-me la oportunitat de ficar-me en aquesta aventura, per haver confiat en mi, per tots aquests anys de feina junts i per creure-hi en tot moment.

Un agraïment molt especial per al Dr. Jorge Bernal pel seu suport, per la seva col·laboració i per la seva generositat. Aquests anys no haurien estat el mateix sense les estones que hem compartit, sense les nostres converses i sense el seu recolzament.

Vull donar les gràcies també al Dr. Xavi Sànchez per la seva implicació, per les seves idees i per fer-ho tan fàcil a l'hora de compartir hores de feina. El meu agraïment també al Dr Miquel Ferrer i a la Dra. Dèbora Gil per la seva col·laboració i el seu valuós punt de vista. Gràcies en general a tothom que ha col·laborat amb mi encara que fos responent a una simple pregunta o compartint la seva opinió.

Gràcies a tots els companys del CVC amb qui he compartit aquesta aventura i que ara ja són bons amics: el Carles, la Ivet, la Camp, el Lluís, el Toni, el Jon, el David, l'Anjan, la Nuvis, l'Alejandro i el Fran. Tantes hores, tantes converses, esquiades (colla de *palilleros*), alguna calçotada, i no moltes birres... En general, gràcies a tothom del CVC que fa possible que tot aquest engranatge segueixi girant.

Gràcies a tots els meus amics que, malgrat amb prou feines saben el que he estat fent durant tots aquests anys, comparteixen el dia a dia amb mi: el Mario, la Nuri, la Carla, el Manolo, el Sergi, la Maria, l'Ana, el Joanet, el Koki i el Txarli. Gràcies també al Javi, al Juano i al Jordi per tantes bones estones al *niu*. Moltes gràcies Xevi pel teu suport. I would also like to thank those friends who have always supported me from the distance. Thank you George and Dan.

Moltes gràcies a la meva familia. Senzillament perque són els millors, i si me'ls hagues d'imaginar, me'ls imaginaria tal com són. Grácies al meu germà, a la Conchi i a la Diana. I moltíssimes gràcies als meus pares, perquè res no hauria estat sense ells, i si ho hagués estat, de ben segur que hauria estat pitjor. Moltes gràcies per ser-hi.

Abstract

Colorectal cancer is the third most common cancer worldwide and the second most common malignant tumor in Europe. Screening tests have shown to be very effective in reducing the amount of deaths since they allow an early detection of polyps. Among the different screening techniques, colonoscopy is considered the gold standard although clinical studies mention several problems that have an impact in the quality of the procedure. The navigation through the rectum and colon track can be challenging for the physicians which can increase polyp miss rates. The thorough visualization of the colon track must be ensured so that the chances of missing lesions are minimized. The visual analysis of colonoscopy images can provide important information to the physicians and support their navigation during the procedure.

Blood vessels and their branching patterns can provide descriptive power to potentially develop biometric markers. Anatomical markers based on blood vessel patterns could be used to identify a particular scene in colonoscopy videos and to support endoscope navigation by generating a sequence of ordered scenes through the different colon sections. By verifying the presence of vascular content in the endoluminal scene it is also possible to certify a proper inspection of the colon mucosa and to improve polyp localization. Considering the potential uses of blood vessel description, this contribution studies the characterization of the vascular content and the analysis of the descriptive power of its branching patterns.

Blood vessel characterization in colonoscopy images is shown to be a challenging task. The endoluminal scene is conformed by several objects whose similar characteristics hinders the development of particular models for each of them. To overcome such difficulties we propose the use of the blood vessel branching characteristics as low-level features for pattern description. We created two data sets including manually labeled vessel information as well as manual ground truths of two types of keypoint landmarks: junctions and endpoints. We present a model to characterize junctions in binary patterns. The implementation of the junction model allows us to develop a junction localization method. The proposed method outperforms the available algorithms in the literature in experiments in both, our newly created colon vessel data set, and in DRIVE retinal fundus image data set. In the latter case, we created manual ground truth of junction coordinates. Since we want to explore the descriptive potential of junctions and vessels, we propose a graph-based approach to create anatomical markers. In the context of polyp localization, we present a new method to inhibit the influence of blood vessels in the extraction of low-level profile information. The results show that our methodology decreases vessel influence, increases polyp information and leads to an improvement in state-of-the-art polyp localization performance.

Resum

El càncer de còlon és el tercer amb més incidència al món i el segon tipus de tumor maligne més comú a Europa. Les tècniques d'exploració directa del còlon han demostrat la seva eficiència en la reducció del nombre de víctimes mortals, permetent la detecció de pòlips en estadis prematurs. Entre les diferents tècniques d'exploració, la colonoscòpia és considerada actualment l'estàndard clínic, tot i que diferents estudis revelen la incidència d'alguns factors en la qualitat de l'exploració. La navegació al llarg del còlon i el recte evidencia una sèrie de reptes per als endoscopistes que provoquen un augment en la taxa d'errors. L'acurada inspecció del còlon ha de ser certificada per tal de minimitzar les possibilitats que alguna lesió no sigui detectada. La inspecció de les imatges de colonoscòpia pot aportar informació crucial per als endoscopistes i donar suport a la navegació durant el procediment.

Els vasos sanguinis i els seus patrons de ramificació poden aportar potencial descriptiu per desenvolupar marcadors biomètrics. Els marcadors anatòmics basats en vasos sanguinis podrien ser utilitzats per identificar escenes en vídeos de colonoscòpia i donar suport per a la navegació generant una seqüència d'imatges ordenades en el recorregut de les seccions del colon. Verificant la presència de contingut vascular a l'escena endoluminal també és possible certificar una acurada inspecció de les mucoses i millorar la localització de pòlips. Considerant els usos potencials de la descripció dels vasos sanguinis, aquesta contribució estudia la caracterització del contingut vascular i l'anàlisi de la capacitat descriptiva dels seus patrons de ramificació.

La caracterització dels vasos sanguinis en imatges de colonoscòpia suposa reptes importants. L'escena endoluminal inclou diferents objectes amb característiques similars, fet que dificulta el desenvolupament de models diferents per a cadascun d'aquests objectes. Per afrontar aquestes dificultats, proposem l'ús dels patrons de ramificació dels vasos sanguinis com a trets a baix nivell per a la descripció de formes. Hem creat dues bases de dades d'imatges que inclouen la segmentació manual dels arbres vasculars, així com la localització manual de dos tipus de punts d'interès: encreuaments i punts finals. Presentem un model per a la caracterització dels punts d'encreuament en patrons binaris. La implementació del model ens permet desenvolupar un mètode de localització de punts d'encreuament. El mètode supera els algorismes existents a la literatura en experiments en dues bases de dades: una de creació pròpia i la base de dades DRIVE, d'imatges de fons d'ull. En el segon cas, hem creat una extensió amb la localització manual dels punts d'encreuament. Pel fet que volem explorar la capacitat descriptiva de patrons vasculars i punts d'encreuament, proposem una aproximació basada en grafs per crear marcadors anatòmics. En el context de la localització de pòlips, establim un nou mètode per inhibir la influència dels vasos sanguinis en l'extracció d'informació de baix nivell. Els resultats mostren que la nostra metodologia disminueix la influència dels vasos sanguinis, augmenta la informació als pòlips i millora els mètodes de localització de pòlips.

Resumen

El cáncer de colon es el tercero de mayor incidencia en el mundo y el segundo tipo de tumor maligno más común en Europa. Las técnicas de exploración directa del colon han demostrado su eficiencia en la reducción del número de víctimas mortales, permitiendo la detección de pólipos en estadios prematuros. Entre las diferentes técnicas de exploración, la colonoscopia se considera actualmente el estándar clínico, aunque diferentes estudios revelan la incidencia de algunos factores en la calidad de la exploración. La navegación a través del colon y el recto evidencia una serie de retos para los endoscopistas que provocan un aumento en la tasa de errores. La inspección minuciosa del colon debe ser certificada con el fin de minimizar las posibilidades de que alguna lesión no sea detectada. La inspección de imágenes de colonoscopia puede aportar información crucial para endoscopistas y proporcionar herramientas de apoyo durante el procedimiento.

Los vasos sanguíneos y sus patrones de ramificación pueden aportar potencial descriptivo para desarrollar marcadores biométricos. Los marcadores anatómicos basados en vasos sanguíneos podrían ser utilizados para identificar escenas en vídeos de colonoscopia y proporcionar apoyo a la navegación generando una secuencia de imágenes ordenadas a lo largo de las secciones del colon. Verificando la presencia de contenido vascular en la escena endoluminal también es posible certificar la inspección minuciosa de las mucosas y mejorar la localización de pólipos. Considerando los usos potenciales de la descripción de vasos sanguíneos, esta contribución estudia la caracterización del contenido vascular y el análisis de la capacidad descriptiva de sus patrones de ramificación.

La caracterización de vasos sanguíneos en imágenes de colonoscopia supone retos importantes. La escena endoluminal incluye diferentes objetos con características similares, dificultando el desarrollo de modelos diferentes para cada uno de esos objetos. Para afrontar estas dificultades, proponemos el uso de patrones de ramificación de vasos sanguíneos como características a bajo nivel para la descripción de formas. Hemos creado dos bases de datos de imágenes que incluyen la segmentación manual de los árboles vasculares, así como la localización manual de dos tipos de puntos de interés: cruces y puntos finales. Presentamos un modelo para la caracterización de puntos de cruce en patrones binarios. La implementación del modelo nos permite desarrollar un método de localización de puntos de cruce. El método supera los algoritmos existentes en la literatura en experimentos en dos bases de datos: una de creación propia y la base de datos DRIVE, de imágenes de fondo de ojo. En el segundo caso, hemos creado una extensión con la localización manual de los puntos de cruce. Dado que queremos explorar la capacidad descriptiva de patrones vasculares y puntos de cruce, proponemos una aproximación basada en grafos para crear marcadores anatómicos. En el contexto de la localización de pólipos, establecemos un nuevo método para inhibir la influencia de los vasos sanguíneos en la extracción de información a bajo nivel. Los resultados muestran que nuestra metodología disminuye la influencia de los vasos sanguíneos, aumenta la información en los pólipos y mejora los métodos de localización de pólipos.

Contents

List of Figures

List of Tables

Chapter 1

Introduction: colonoscopy quality improvement

1.1 Colon cancer

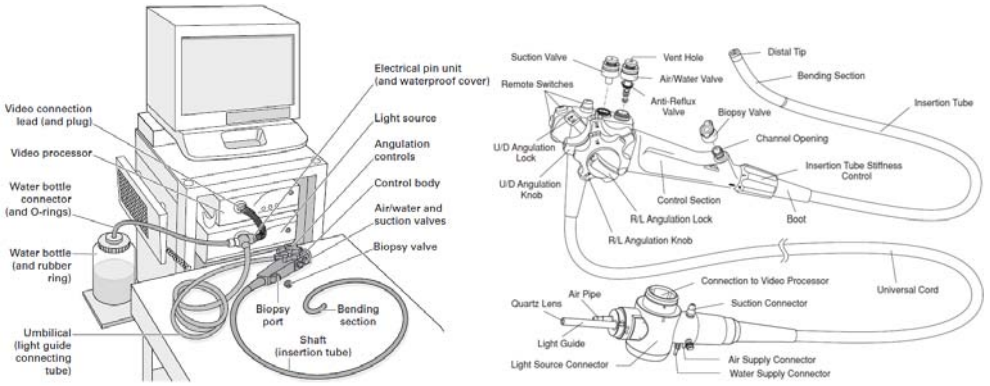
Colorectal cancer (CRC) is the cancer that develops in the tissues of the large intestine, which is part of the digestive system and is formed by the colon -the longest part of the large intestine- and the rectum -the part of the large intestine closest to the anus-. Cancer is more commonly found in the colon or rectum than in the small intestine. CRC is the third most common cancer worldwide and the second most common malignant tumor in Europe. Deaths related to CRC in Europe reach $230,000$ cases every year and estimations arise that colorectal cancer cases will reach the 9.7 % of the total global cancer cases by 2020 [1].

Besides prevention habits, such as a healthy diet [141], screening tests have shown to be very effective in reducing CRC deaths [90, 70]. Screening techniques allow physicians to find precancerous growths -called polyps- and cancer before the patient can feel any of its symptoms. Once cancer appears, it grows into the wall of the colon or rectum. This process is described in several stages. Polyps detected in its early stages are reported to have higher survival rates as well as allow a less extensive treatment and faster recovery [152].

1.2 Colon screening and colonoscopy

Different screening techniques for the inspection of the colon and rectum exist, such sigmoidoscopy, double-contrast barium enema, fecal occult blood test, colonoscopy or digital rectal exam. Colonoscopy is a procedure used to screen the colon and rectum which is considered the gold standard to detect and treat different intestinal pathologies. Different types of endoscopes are available to be used for colonoscopy. They allow endoscopists to screen and examine the colon so that different gastrointestinal pathologies can be detected and evaluated, such as inflamed tissue, ulcers, abnormal growths, tract bleeding or polyps. A single procedure permits clinical personnel to diagnose pathologies and perform therapeutic operations like tissue sample collection (biopsy) or polyp removal.

The equipment includes a colonoscope and a computer for monitorization (see Figure 1.1a). Although many colonoscopes are available, most of them share the main features. The colonoscope instrument has two different parts: the insertion tube and the control head (see Figure 1.1b). The insertion tube is a flexible shaft that contains the electrical wires



(a) (b)

Figure 1.1: (a) Endoscope system[1]. (b) Colonoscope components[2].

sending the images from the distal tip to the video processor as well as tubes for suction, air and water feeding and, in many cases, an extra tube for a forward water-jet. The characteristics of this tube are a important feature in a colonoscope. The tube diameter must be big enough to provide enough space for the different components while keeping the outer diameter as small as possible to maximize patient tolerance. The diameter can reach 15 mm. In addition, the tube design must combine flexibility to open its way through the intestine, with torqueability to be able to transmit accurately the torsion in the proximal tip of the tube to the distal tip. Usually, the flexibility of the tube is not the same throughout its length. The proximal portion of the insertion tube is more rigid to prevent loop formation whereas the distal part of the tube (40 cm) is remarkably more flexible to ease the insertion through the intestinal track. The most distal portion of the insertion tube -around 10cm- is known as the bending section. This portion of the tube is deflectable and can be fully angulated by the endoscopist. The control head holds the two angulation control wheels -left/right and up/down- and the suction, air/water and biopsy valves to control the air, water and suction channels. Endoscopists use the control head to command any action during the procedure, such as guiding the endoscope, cleaning the lens, pulling fluids up or performing any intervention by passing several types of devices through the biopsy channel. The design of any kind of endoscopy includes several safety features to prevent accidents in the air, water or suction channels which may lead to accidents such as overinsufflation of the patient.

1.3 Colonoscopy procedure

The whole procedure involves two steps: colon preparation and colon examination. The colon must be cleaned out prior to the examination so that colon mucosa inspection is not hindered by the presence of residual stool. Patients are given detailed instructions as the colon must be empty for the test. The second step consists in the colonoscopy itself.

The procedure can be split in two phases: insertion and withdrawal. The patient is usually placed in the left lateral position and the endoscope insertion tube is inserted in

[2]text1
[2]text2

the anus. The head control is operated by the endoscopist's left hand, and the right hand is mostly used to control the insertion tube. During the insertion phase the endoscope is guided through the different parts of the colon tract until the cecum is reached. Angulating and rotating the endoscope helps the endoscopist to find the lumen if any difficulty is found allowing the endoscope to move forward in the insertion. The insertion phase is the most uncomfortable for the patient. During this phase, the cecum, which is considered to be the beginning of the large intestine, must be reached as soon as possible. During the withdrawal phase the endoscopist pulls the endoscope out of the intestine. This phase is used to carry out the thorough inspection of the colon mucosa to clean the colon of neoplasia and perform colonoscopy polypectomy if needed (surgical excision of a polyp).

The endoscope and the colonoscopy procedure are designed to service the purpose of screening the colon and rectum. The final aim of the screening is to detect and remove polyps. Colonoscopy is still considered the gold standard for colon screening [126] and the cornerstone of current prevention guidelines [141]. It reduces drastically the risk of death of screened population [90, 70] and allows the longest interval for rescreening among all the forms of testing. However, several reasons can affect the quality of interventions increasing polyp missdetection -miss rate values can reach 20%, or even 10% in the case of large or advanced adenomas (larger than 5 mm) [62]-. The missdetection of polyps have been found to be related to different issues involved in the screening process that are liable to be exploited by Computer Vision techniques. Providing computerized knowledge of the endoluminal scene by means of an accurate analysis of the scene objects can provide a important support for endoscopists. That standard knowledge can be used to improve polyp localization as well as to improve the intervention itself or the post planification, by supporting the navigation through the colon and rectum and developing standard scales to measure the quality of an intervention and improve the planification of future screenings. In all this possible applications, the accurate characterization of the vascular content in the endoluminal scene and the accurate knowledge of its branching pattern can be crucial. The next section explores several matters affecting the quality of colonoscopy.

1.4 Colonoscopy quality

The effectiveness of colonoscopy in reducing colon cancer incidence depends on several reasons. This fact has prompted researchers to clarify these reasons so that colonoscopy quality indicators could be developed to standardize the reliability of the procedure. The proper visualization of the colon tract, the thoroughness in mucosa examination and the acceptance of the procedure by the patient are just some examples of elements that can affect the quality of the procedure [120]. Among the different indicators that have been proposed in the literature, we will focus on those which may be exploited from the point of view of computer vision.

Colon preparation

The patient must follow preprocedure instructions to clean the bowel from stool and allow a better examination of the colon mucosa [120]. An adequate bowel preparation reduces sedation requirements, increases polyp detection rates [60, 52] and decreases insertion and withdrawal time [79] and cecal intubation time [22] and increases intervals (<10 years) for repeated colonoscopies [118], which also results in a higher economical cost [119]. Commonly, categorical scales are used after any examination to classify separately the different parts of the colon among a variable number of categories to report bowel preparation [11, 128]. Standardized techniques to quantify the level of proper colon preparation would improve

the evaluation of interventions and the planning of future screening. Identifying colonoscopy frames with a certain amount of solid or liquid stool could quantify the frames containing stool to develop bowel cleansing metrics.

Center and endoscopist variability

Several works affirm that there is a clear evidence of the substantial variability of colonoscopy performance among different centers and endoscopists [156, 95]. The lack of measurable standard methods to provide the degree of quality of an intervention increases the ambiguity of the screening outcome. For instance, some studies suggest that the description of too many cleansing levels results on ambiguity and disagreement between experts [95].

Cecal intubation

Cecal intubation is defined as the complete intubation into the cecum with the tip of the endoscope. It is considered the prove that the whole colon tract is examined, including the proximal colon. Colonoscopy recommendations prompt endoscopists to report the identification of the cecum as a quality indicator of the colonoscopy procedure. If the cecum visualization is not documented the colonoscopy efficiency drops drastically. The report of cecal intubation can be based on different landmarks. The appendiceal orifice and the ileocecal valve are considered the most important landmarks [119]. Other possibilities are the lips of the ileocecal valve or intubation of the terminal ileum. The certification of cecum visualization guarantees the complete screening of the colon and rectum. Appendiceal orifice detection could provide an important clue to propose automated systems aiming to assess cecal intubation rates.

Withdrawal time

The withdrawal time is the time between the cecum was reached and the colonoscope was completely removed from the anus. Assuming that a certain variability may be possible given that the length of colons may differ, the withdrawal phase in patients without previous surgical resection should last at least 6 minutes [120, 15, 107]. Some studies suggest that the quality of colonoscopy withdrawal technique should also be considered [85], regarding fold examination and adequacy of colonic cleansing and distension.

Navigation and lesion tracking

The wall of colon and rectum is made up of several layers. In between the inner mucosa and the outermost connective tissue layer -with the exception of the rectum-, we found a thin and a thick muscle layer forcing the contents of intestines along. To navigate through this folding tube with the colonoscope can be challenging for the endoscopist as well as to orient himself in a specific colon section. Computer-assisted tools to provide guiding clues through the intervention can represent long step forward in the reduction of polyp missdetection.

1.5 Scope of the thesis: Blood vessels, junctions and colonoscopy quality

In the context of the improvement of colon quality, we focus in the development of computer-assisted tools to support a better knowledge of the endoluminal scene. We propose the

characterization of blood vessels that are visible in the colon inner layer and, as its most remarkable structural elements, the localization of bifurcations and branch crossings.

Up to our knowledge, there is no existing work that has paid attention to the role of blood vessels in polyp localization. The accurate characterization of the vessel patterns in the endoluminal scene is crucial to improve the knowledge of the other elements in the scene. Vessel branching patterns merge into these other elements of the scene and can lead to inaccurate polyp localization or even to complete missdetection. Moreover, the nature of branching patterns such as vascular trees can be used as anatomical markers in a navigation framework to support endoscopists in their exploration of the colon and rectum.

A proper knowledge and characterization of vessel branching patterns in colonoscopy images sets the foundations for important advances in the quality and reliability of the colonoscopy procedure. The analysis of vessel patterns arises a wide variability. The complexity of the branching tree described by the vessels when visible in the inner layer of the colon track makes its modeling a challenging task. Nevertheless, it also demonstrates its potential as anatomical markers which would be clue in the development of computer-assisted tools to improve the navigation of the colon track and the quality of a colonoscopy intervention:

- Anatomical markers based on the blood vessel patterns can be used to identify a particular scene in colonoscopy videos. Endoluminal scene identification can improve the navigation through the rectum and the colon track by generating a sequence of ordered scenes through the different colon sections. The scene matching based on vessel pattern anatomical patterns leads to a sequentially consistent map of the colon track to support endoscopist navigation.
- The endoscopist must try to reach the cecum in the insertion phase in order to perform the thorough inspection in the withdrawal phase. However, the endoscopist may notice a suspicious area in the colon mucosa during the insertion phase which would like to inspect carefully in the withdrawal phase. An anatomical marker based on the blood vessel content of the endoluminal scene would allow endoscopists to set up alarms. An alarm could be set up in the introduction phase and then be raised by the support computer system to notify the endoscopist when the region previously labeled is revisited in the withdrawal phase.
- The characterization of blood vessel patterns allows the development of support tools to analyze the visibility of colon wall mucosa. An standard quality measure quantifying the presence of vessel content in the course of an screening intervention can be used to score the adequate cleansing level.
- The identification of the vascular content in a given endoluminal scene can help to improve polyp localization methods. Several objects in the colon track and rectum appear as similar visual objects in the endoluminal scene as visualized in colonoscopy videos. A better knowledge of the blood vessel patterns in a given scene allows a better knowledge of the other objects in the scene, such as polyps or adenomas.

1.6 Main contributions of the thesis

This thesis focuses on the improvement of the computerized knowledge of the endoluminal scene focusing on the blood vessel content, the junction keypoints on their patterns as well as the usage of that information for the development of anatomical markers and the improvement of polyp localization and segmentation techniques.

- Analysis of the complexity of blood vessel characterization in colonoscopy images. The coexistence of different elements in the endoluminal scene with similar characteristics makes it difficult to develop models for each of them, particularly for blood vessels.
- Creation of two data sets of colonoscopy images including a selection of images from several colonoscopy videos and the corresponding manually created ground truth of the blood vessel content. Two manually created ground truths of keypoint landmarks are also included for one of them: junctions and endpoints.
- Creation of an extension for the DRIVE retinal fundus imaga data set that includes the coordinates of the junction in the vascular patterns.
- Definition of a general model of junctions and intersections in binary images. An implementation of the model is also proposed which leads to an automatic tool for the localization of junctions in a given binary pattern. The new method -GRowing Algorithm for Intersection Detection, GRAID- outperforms existing methods in the bibliography, as showed by experiments on blood vessel patterns in both colonoscopy and retinal images.
- Introduction of a new skeletonization based on the foundation of our novel junction localization method. This new skeletonization algorithm prioritizes the extraction of a skeleton that keeps the branching structure of the shape.
- Verification of the potential of blood vessels and junctions as anatomical markers based on a graph-based branching pattern characterization.
- Study of the impact of the influence of blood vessel mitigation in the state-of-the-art polyp localization method caused by the reduction of valley-shaped geometric information in the image.
- Development of a polyp segmentation algorithm based on state-of-the-art polyp localization method.

1.7 Structure of the thesis

The thesis is organized in three parts. Throughout the first part, which spans Chapters 2 and 3, we introduce the endoluminal scene and the objects found in colonoscopy videos and images. This analysis of the image formation and context let us introduce the challenges of the characterization of objects in colonoscopy images. We also present the data sets involved in the experiments exposed in the following chapters. Among those data sets, a new set of colonoscopy images including the corresponding newly manually created Ground Truth of vascular content and keypoints -junctions and endpoints-. The second part, spanning from Chapters 4 and 6, introduces a novel method for the characterization and detection of junctions in binary images containing vessel patterns. The performance of this new method is analyzed quantitatively and compared to the other state-of-the-art methods. Based on the novel junction localization method, we introduce a new skeletonization algorithm design to describe shapes including branching structures that must be extracted. Then, a graph strategy is proposed to create branching pattern markers and show the importance of junctions and endpoints as branching pattern descriptors. The third part, which includes Chapter 7, explores the endoluminal scene and presents a methodology to minimize the interference of vascular pattern information in the extraction of information for polyp localization. The methodology inhibits blood vessel influence and maximizes polyp information. A novel polyp segmentation algorithm is also presented and compared to general segmentation methods. Finally, Chapter 10 exposes the overall conclusions of the thesis and highlights the future lines of research.

Part I

CONTEXT & TOOLS

Chapter 2

Endoluminal Scene

2.1 Introduction to Colonoscopy images

The colonoscope is provided with a CCD (charge-coupled device) chip which transmits the images from the distal tip of the insertion tub to the video monitor. Color information is captured by color chipset cameras. However, many endoscopes use monochrome chipsets which obtain color information from monochrome chipsets in successive extractions with different color filters. These different streams are then merged into one single color video stream. The imaging system made up by the camera chipsets with a light source solid to it together with the nature of the colon track and rectum conform the endoluminal scene captured by colonoscopy videos. Colonoscopy images are a consequence of the 2D projections of the objects in the colon and rectum captured by the system. As a result, the analysis of the colon scene and the modeling of endoluminal objects present a wide range of challenges that will be introduced in this chapter.

2.2 Endoluminal scene objects

The navigation of the endoscope during the colonoscopy procedure can find several objects that can be more or less frequent. The main objects among the content of the common endoluminal scene can include lumen, wrinkles and folds, blood vessels, polyps, specular highlights, bubbles and specular content (see Figure 2.1). It is important to be able to provide an accurate model of each of the visual features that come out as a consequence of those endoluminal objects in order to provide a proper description of the endoluminal scene. The different objects may or may not be found in a single scene. Hence the model for every single object must cope with the problem of the separability of the different objects.

- Lumen. The inner cavity within the colon tubular cavity is known as the lumen. In the context of colonoscopy the camera may be pointing at the colon cavity -when moving through the colon- or may be focusing on the inspection of the colon walls and mucosa -commonly in the withdrawal phase-. Consequently, the lumen can be found in any position in the colonoscopy video procedure or even not be found at all. The region described as the lumen goes from dark to medium-dark, given the variability of ilumination in the endoluminal scene.
- Folds. Wrinkles and folds are a consequence of the muscular nature of the inner colon

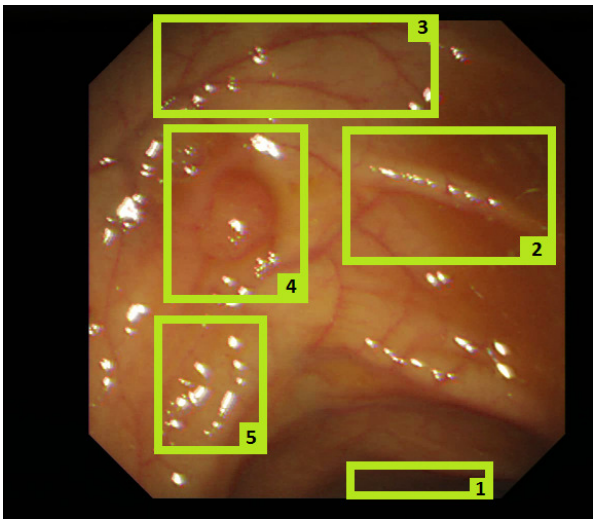


Figure 2.1: Graphical example of a typical endoluminal scene from a colonoscopy video: 1) Lumen; 2) Wrinkles and folds; 3) Blood vessels; 4) A polyp; 5) Specular highlights

layer. The existence of wrinkles and folds in the colon walls creates visual objects induced by occlusions and illumination changes.

- Polyps. Many kinds of polyps can be present in the endoluminal scene. There is a high variability is terms of shape and texture as well as the position of the polyp itself in scene regarding the lumen, the colon walls and the relative position to the camera.
- Specular highlights. The colon wall humidity and the proximity of the camera source light causes specular highlights to appear as brightest regions in the scene.
- Fecal content. Although part of the colonoscopy procedure includes colon preparation and cleansing, the presence of fecal content is still high. A high degree of fecal content hinders the inspection of the colon mucosa and decreases the reliability of the procedure.
- Blood vessels. Vascular patterns can be found in the endoluminal scene when vessels are close enough to the inner colon surface. The analysis of blood vessels will be tackled in subsequent sections in this chapter.

2.3 Bibliography on computer-assisted colonoscopy

Developing standardized systems to assess the quality of the colonoscopy procedure involves several tasks. Several visual objects or regions in a colonoscopy image can be the aim of an automated system, such as lumen, polyps, stool, vessels or folds. However, the final goal of a proper intervention is to guarantee the complete screening of the colon tract and the localization of all possible polyps in the colon walls. Existing polyp localization intelligent systems could be divided into shape-based polyp localization [18, 167] and texture-based polyp localization [151, 6].

A good knowledge of the elements of the scene must support the tasks of a proper polyp localization as well as an accurate screening of the colon and rectum. The lumen in a tubular

structure is the inner cavity within it. As a visual object in colonoscopy images, it is the darkest region of the image which points out the path to follow to move forward in the colon since it represents the deepest obstacle-free part in the colon. Consequently, lumen segmentation can be an interesting tool to develop computer-aimed systems to assist the advance of the colonoscope in the colon. Morever, the presence of the lumen region in the center of an image may indicate that the endoscopist is focusing on the examination of the colon mucosa. In this line we find the work of Duncan Gillies et al. which describes an autonomous guidance system to navigate automatically in the colon [116]. They propose a lumen segmentation system based on a pyramid-extraction technique where the image is represented by a successive subdivision of an the image into quadrants. The original image is at the bottom of the pyramid and each upper level is built by applying a transformation to the previous level quadrants in the pixel-level. They use pixel intensity transformations to detect the seed of the darkest area in the image followed by a merging process based on variance and average gray level criteria. Tan et al. proposed another technique for lumen detection based on a fuzzy-directional edge-detection method and a region-growing algorithm [116]. Tian et al. faced the problem of lumen segmentation with a two step algorithm [150]. In the first step an adaptative progressive thresholding technique is used to segment preliminary region of interest. Afterwards, and adaptive Iris filter is used to determine the actual region from the preliminary region of interest.

Hwang et al. presented a computer-based method which produces several objective metrics to reflect the quality of a colonoscopic procedure by analyzing its digitized video [66]. Oh et al. presented subsequently an extension to that work [105]. The analysis of the procedure video is based on frame classification and camera motion. Each frame is classified identifying different patterns in its frequency spectrum so that *non-informative* frames - blurred or containing blocking material- are discarded [106]. An affine camera model is used to provide a video segmentation based on camera motion that allows to estimate the boundary between insertion and withdrawal phases. The technique proposed to achieve camera motion estimation uses MPEG motion vectors. They use basically insertion and withdrawal time to compute several single quality scores and validate them qualitatively on a database made up of seven colonoscopy videos. Other proposal for metrics in [66] included lumen identification based on JSEG [48] and a technique to determine whether a planar region is convex or concave [44].They also included in [105] the detection of therapeutic shot detection [38]. The authors did not proposed any experimental setup to assess the possibilities of the metrics the proposed to discern the quality of colonoscopy videos.

Several kinds of frames can be expected to be found in a colonoscopy video, as for instance frames where the appendiceal orifice is visualized. The presence of the appendiceal orifice has been described as the main clue to obtain cecal intubation rates. Cao et al. proposed an algorithm for automated detection of the presence of the appendiceal orifice [37]. Their method exploits the idea that, when the appendix is closely inspected, the colon lumen is not visible. They use the segmentation method in [66] to segment colonoscopy frames and propose several new features. Finally, k-means is used to classify colonoscopy frames as *appendix images* or *non-appendix images*.

The presence of solid or liquid stool in the frames of colonoscopy provide information about the degree of bowel cleansing. We already introduced the algorithm proposed in [106] to classify frames as *non-informative*. Hwang et al. proposed a stool segmentation algorithm using color features and and support vector machines [67]. Arnold et al. introduced an algorithm to detect *indistinct* frames based on the wavelet analysis [10]. They define *indistinct* frames as those which do not carry any useful clinical information, which are not clinically relevant for further analysis and do not include clear views of the mucosa or the lumen.

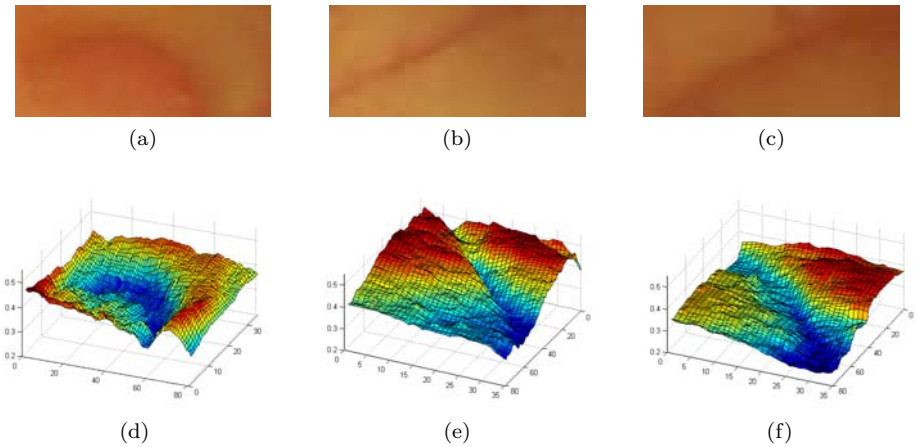


Figure 2.2: Image example and surface: (a)(d) polyp. (b)(e) vessel. (c)(f) fold.

2.4 Image formation and endoluminal object appearance

Colonoscopy videos are the result of the endoscope navigation through the patient colon track and large intestine with a camera and a light source. Endoluminal objects and light source interaction cause the visual objects that are finally found in the colonoscopy video. All the endoluminal objects in the scene have an impact in the endoluminal image. That impact is what we call the visual object, which is the consequence of an endoluminal object. Despite the different nature of the visual objects in a colonoscopy image, their impact as visual object in the scene always has the shape of valley objects in the image surface. This valley-shape of the visual objects can be used as the basis for the characterization of different endoluminal objects in the scene. That is the case of polyps [18] and blood vessels, but also folds, which leads to a high degree of ambiguity in valley-based models, as seen in Figure 2.2. The presence of specular highlights is also an important source of valley profiles in the images. Sharp profiles appear as a consequence of the presence of specular highlights and overshadow medium to small scale profiles due to their strong response to state-of-the-art valley detectors. Fecal content also increases the valley information in the scene. In this case, low scale valley information is added, which increases the overall valley information in the image. The similarities between visual objects since they share valley-based profiles hinder the development of univocal models.

The characterization of polyps is based on a model which results from both a model of the polyp itself and a model of the illumination in the colonoscopy videos, as described by state-of-the-art works [18]. A general model considers polyps as semi-spherical shapes. The interaction between the polyp shape and the colonoscope light source results on shadings -areas of darkness in reference to a source of light- around the polyp, that are valley profiles in the intensity image.

In the case of folds and wrinkles, their presence depends on the scene captured by the colonoscopy camera and the spectral properties of the light source as well as its position and orientation with respect to the colon walls wrinkles. If we assume that the properties of the intestinal wall remain unchanged in wall folds, or the difference is negligible, the visual

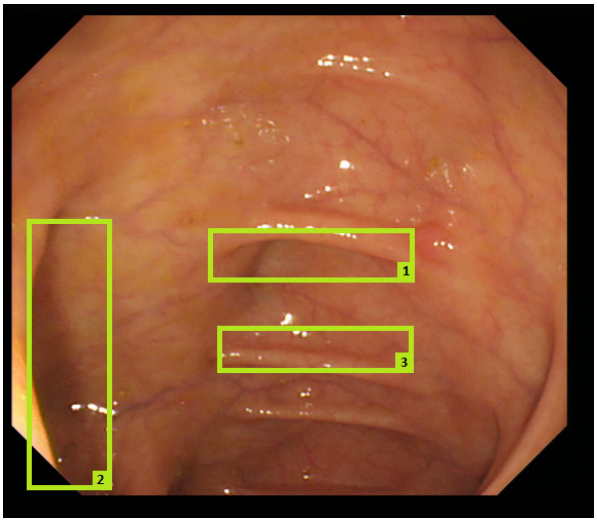


Figure 2.3: Wrinkle & fold features: 1) Shading. 2) Shadow 3) Edge.

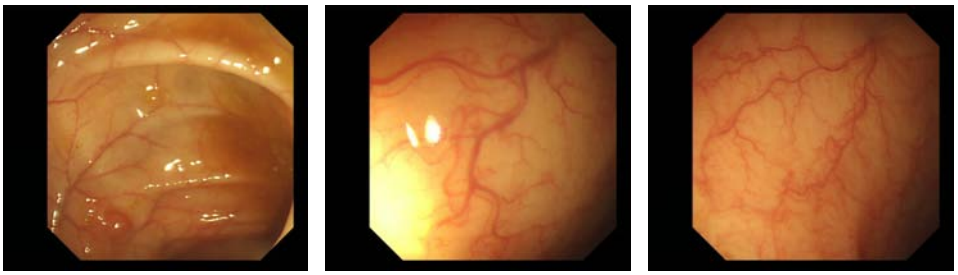
features caused by wrinkles and folds are a consequence of the illumination of the scene. Fold/wrinkle features have a 3D nature cause while blood vessels are 2D features.

Figure 2.3 exemplifies this nature of wrinkles and folds in the scene. The variability of folds and wrinkles in the scene in terms of shape and position in relation to the light source results in visual features with different nature. Folds and wrinkles can produce several visual features. An abrupt fold with the source light in a determined position and orientation can cause an occlusion that can be more or less prominent. In that case an edge appears in the image. That edge can be the only visual feature caused by the fold (see Figure 2.3)). However, depending on the illumination and the fold/wrinkle itself, the occlusion can also result in a shadow (see Figure 2.3)). A shadow is an area of darkness relative to the object blocking the source of light. Refers to a shape cast by an object that block the light source. A third case can be still described when the fold/wrinkle is not abrupt enough. In that event the visual feature is a consequence of the shading caused by the shape variation of the intestinal wall (see Figure 2.3)).

Unlike wrinkles and folds, the presence of blood vessels is not just a consequence of the light source orientation. The appearance of vessels in the scene is just caused by the nature of the intestinal wall that is modified by the presence of blood vessels close to its surface. The visual features caused by blood vessels in a colonoscopy image are just a consequence of the illumination of the scene and the reflectivity properties of the intestinal wall when there is a blood vessel, which modifies the reflectivity properties of the intestinal wall. However, despite the different nature of the objects conforming the endoluminal scene, modeling and characterizing them particularly is a challenging task.

2.5 Vessel branching pattern characterization

Vascular patterns describe an elongated path which branches off and crosses over itself. These elongated paths present a high variability in terms of shape, thickness and, although less evident, color. Unlike other branching patterns -such as road networks-, vascular patterns



(a) (b) (c)

Figure 2.4: Examples of blood vessel patterns.

-as in in retinal fundus images or colonoscopy videos- do not hold any symmetry or pattern periodicity. In the case of colonoscopy vessel patterns, they may not be connected in a fully tree-like way and they are usually not single-root tree structures. Concavity/convexity and tortuosity, defined as the property of a curve being twisted or tortuous, or having many turns, can be found in endless combinations. The thickness of the piecewise elongated structure which conforms the branching patterns can also vary among a high range of values even in a single continuous pattern. Branching and bifurcation frequency can also be remarkably different. Figure 2.4 shows three colon vascular patterns. This examples include different degrees of thickness, branching frequency and tortuosity.

Color information in colonoscopy videos can also be exploited. Colon images show little variance in terms of color. The high variability in terms of brightness also affects the visual objects in the scene. It all results in images with low contrast between the wall -considered as the background-, and the visual objects in the scene, such as polyps, vessels or folds. Color in the endoluminal scene expands from black and white in terms of intensity, while keeping very concentrated along a range of very few colors.

2.6 Junctions in vessel branching patterns

Characteristic visual objects that are easily recognizable by humans appear when blood vessel patterns branch off or cross over: junctions and intersections, respectively. Apparently, this definition states a clear rule. However, the characterization of junctions is not straightforward either in binary, gray-level or color images.

The human ability to analyze branching patterns such as the vascular content in colonoscopy image is not easily achieved automatically. This is due to the high variability of all the visual objects most people would agree to recognize as junctions. Modeling junctions represents a challenging task given the variability in their shape. The suitable approach should take into account 3 considerations we will introduce in the following sections: junction topology, junctions are not corners and junctions are an image object. Junctions are important image interest points with a clear shape or topology although they have an ambiguous geometry. Junctions have not been considered in depth as topological shapes. They have rather been tackled as a corner problem in the interest point literature. Such approach offers limitations and junctions as shapes should be considered as separate task. The complex characterization must also cope with the problems of shape information extraction from images.

2.6.1 Junction shape

Junctions or intersections appear in a branching pattern where three or more paths merge. Strictly, we name the points of a linear pattern as bifurcations when three paths converge and intersections when more than three paths converge. We will often refer to all of them as junctions, no matter the number of branches. This apparent simplicity of the definition entails sort of paradoxical outcomes. On one hand, it represents a simple description and, therefore, a clear statement for a human observer to discern junctions from non-junctions. On the other hand, the apparently simple definition involves a high level of ambiguity which must be utterly faced in a proper and accurate description of the object shape.

Given that a junction appears when several elongated branches converge in a common region, the shape will always be made up of a central region and 3 or more branches converging to it. This approach seems sufficient to describe the topology of a vascular pattern although it lacks of necessary accuracy to describe the geometry and shape of junctions. The variability in shape and size of the central region as well as the relative width and tortuosity of the branches result on a broad amount of shapes comprised in the general junction definition. The geometrical variability of a junction can be grouped in several aspects to consider:

- The central region shape and width and the relative width and length of the branches. The consideration of elongated structures can become a vague notion as it deals with the relative widths of the branches and the central region. The width of the branches can make a difference to be considered as a branch or as concavity in a given region.
- The tortuosity of the branches when reaching the central regions. The more tortuose the branches are when reaching the central region, the less circular the junction is the and less radial its topology.
- The lack of any kind of symmetry and the variability of the angle of incidence of branches to the central region. Adding the fact that the topology does not offer any structural reference, the development of models verifying invariance to rotation becomes a challenging goal.

All the previous considerations interact together resulting in multitude of combinations. The shapes that are more accessible and easier to analyze are isolated cases with straight branches whose length and width have similar values to the central region. Figures 2.5c, 2.5a, 2.5b and 2.5d exemplify this kind of shapes. This case make it easier to verify the convergence of long enough branches to a single region since it can be locally isolated.

The variations of central region size and branch length and width hinders the possibility to isolate the structure and validate it as a junction. This is the case of junctions depicted in Figures 2.5e, 2.5f, 2.5g and 2.5h. This cases make more difficult to catch and describe the shape locally and increase the ambiguity of the location of the single-pixel intersection landmark.

The tortuosity and variability of branch incidence angle increases the difficulties since it eliminates the most intuitive radial shape of a junction. The junction shapes in Figures 2.5i and 2.5j show clear examples on this matter. Central region shape adds more difficulties in the same sense as well as in the location of the intersection landmark, as shown in examples in Figures 2.5k and 2.5l.

The previous situations are usually combined in a broad number of possibilities and aggravated by the fact that junctions are usually not isolated structures. Junctions can appear in a high density in branching structures, such as vascular patterns, which hinders even more their characterization and location, as depicted by the examples in Figures 2.5m, 2.5n, 2.5o and 2.5p

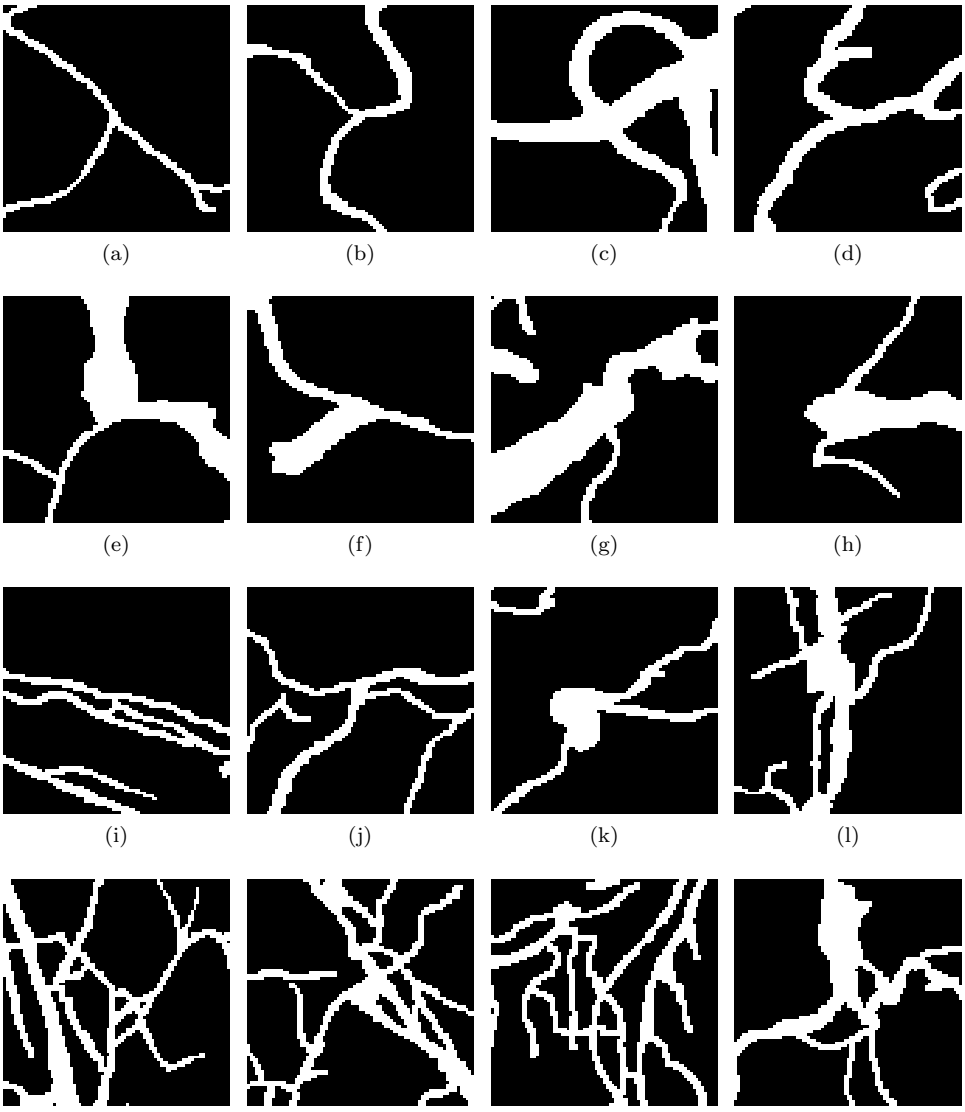


Figure 2.5: Junction shape examples.

2.6.2 Junctions versus corners

Interest points are very attractive in computer vision because they are useful in many problems. For that reason, many interest point detectors exist. The general notion of interest point refers to points in an image that are distinct and geometrically stable [135]. Both of these conditions are verified by junctions and intersections in vascular patterns. However, the traditional approach to interest point extraction focuses on points for which the image

suffers some kind of two-dimensional change. This definition usually includes corners of several kinds and even isolated extremum values or edge endings [12, 61, 94].

In the case of junction shapes as described above, any general approach aiming at curvature significant points or two-dimensional changes is too permissible. As already shown in the previous Section, junction shapes have a clear topological structure although a broad range of geometrical shapes are possible. As far as we know, the characterization of junction shapes can not be achieve by any of the existing low-level feature detectors in the literature . For this reason, unlike many works in the literature, we will make a difference between junction structures and corners although they can share some of their characteristics.

2.6.3 Junctions in images

Modeling junctions must also cope with the adjusted feature extraction from images. The nature of the images must be considered by any possible characterization of junction topology and shape variability. The information we can gather from images to tackle the problem of junction localization can basically be grouped in two categories: color and geometry. In terms of color, the problem will remain in the same terms as the blood vessel pattern itself. In terms of geometry, the shapes and structures must be extracted from the image surface information, which would then allow the analysis of shape topology as already introduced.

Given that junctions are shapes created at the convergence of elongated structures, the shape is the consequence of the intensity variations of all the converging branches and the central region. As show before, the branches converging to the junction describe valleys on the image surface. Therefore, junction localization may combine the challenges of a proper analysis of valley-profile structures in the image while adding the junction scale combination.

The extraction of geometric information leading to junction modeling must consider the variability of branch scales and profiles -valley scales and profiles- and the lack of a strictly defined converging region shape. A local analysis of the images

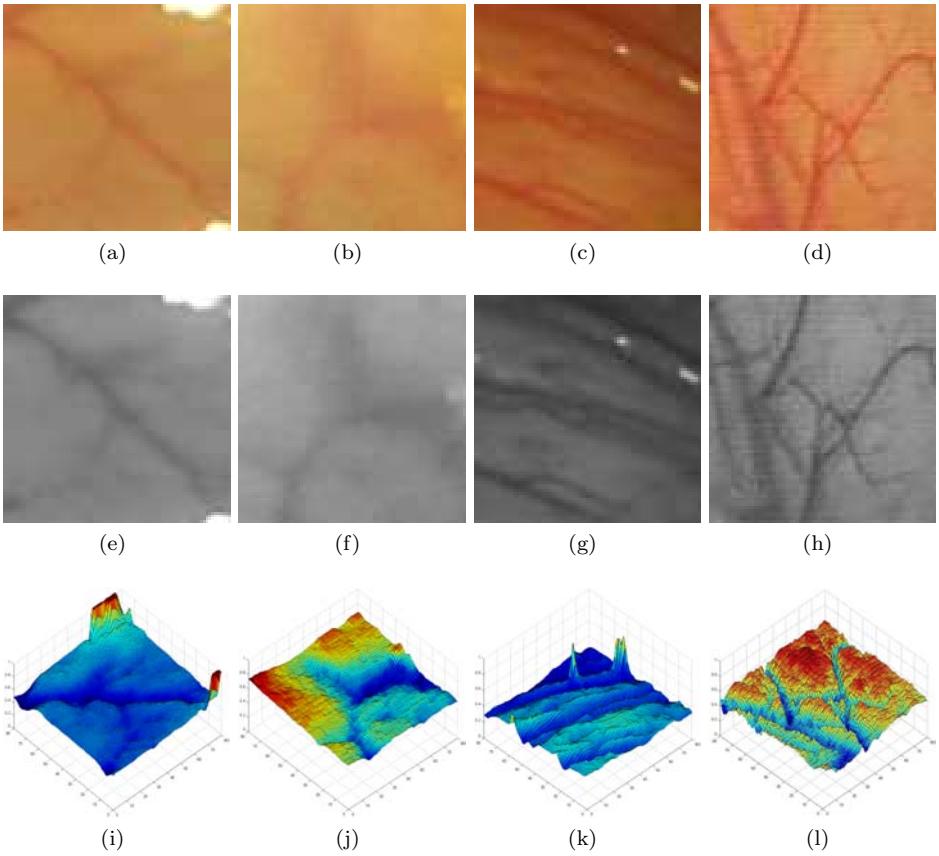


Figure 2.6: Junction image examples: color image and green component image and surface.

Chapter 3

Data sets and Grount Truths

Several data sets have been used in the different quantitative and qualitative experiments performed through this thesis. The new data sets created along this work are introduced in this chapter as well as different existing data sets used in some experiments.

In the context of colonoscopy, we present COLON-V-TRACE and COLON-VESSEL data sets , new data sets of colonoscopy images to study the vascular content in colonoscopy videos, its structure and landmarks -junctions and endpoints-, and its interaction with the other elements of the seen. Some experiments regarding the use of vacular content influence in the improvement of state-of-the-art polyp localization methods are tested on the only existing public colonoscopy polyp data set.

Name	# images	Content
COLON-VESSEL*	40	Original image. Vessel manual segmentation. Vessel manual junction landmark GT. Vessel manual endpoint landmark GT.
COLON-V-TRACE*	29	Original image. Vessel trace manual segmentation.
CVC-ColonDB	300	Original image. Polyp manual segmentation. Polyp manual segmentation contour. Non-informative region manual segm.

* Newly created data sets and ground truths.

Table 3.1: Colonoscopy data sets and ground truths.

In the context of retinal fundus images, we create a new ground truth for the images in a data set already available in the bibligraphy to perform experiments regarding vascular patterns. We present *NunGT*, a manually created intersection ground truth add-on for the well-known DRIVE retinal fundus image data set. *NunGT* together with the already existing intersection ground truth *AzzoGT* allows us to study vascular patterns in a different context and validate the influence and variability of several observers in the location of intersection landmarks.

Name	# images	Content
DRIVE	40	Original image. Vessel manual segmentation.
NunGT* (DRIVE add-on)	40	Vessel manual junction landmark GT.
AzzoGT (DRIVE add-on)	40	Vessel manual junction landmark GT.

* Newly created ground truth.

Table 3.2: Retinal data sets and ground truths.

3.1 Colonoscopy images

3.1.1 COLON-VESSEL data set

As fas as we know, there was no existing data set of colonoscopy images for the study of the blood vessels in the endoluminal scene. In order to carry out experiments on the vascular patterns in colonoscopy videos, we created a new data set and provided it with manual vessel segmentations and ground truths of the landmark points in the segmented structures.

COLON-VESSEL data set has been conformed by selecting frames extracted from 15 different colonoscopy videos created at St. Vincent's Hospital and Beaumont Hospital in Dublin, Ireland. An expert selected 40 frames whose resolution is 574×500. Considering the variability in the presence of vascular patterns in colonoscopy, the selected frames present different degrees of vascular content. The selected images also vary regarding the other elements in the scene, ranging from frames where only the colon wall and a vessel pattern is visible to images where several of the possible endoluminal object appear and interact. The images do not include any patient information and present a blank frame in the borders as commonly found in endoscopy videos.

The data set includes a ground truth consisting of a mask of the blood vessels present in of the 40 images. An expert segmented manually the vascular patterns appearing at every image. The mask has been created at pixel level and provides an accurate segmentation of blood vessels. Vascular structures in colonoscopy image do not show any regular pattern. The vascular content can be linked in a single branching structure or split in several structures. Figure 3.1 shows examples of COLON-VESSEL data set.

A ground truth of landmark keypoints in the blood vessel patterns is also included. The manual location of landmarks was achieved on the previously segmented vascular structures and it includes junctions and intersections with any number of branches. A single pixel was labeled as an intersection if it was identified as the point where at least three branches meet together. Given the ambiguity of the task of locating a single intersection pixel, the expert was given the rule of placing the landmark in the intersection of the imaginary axis of the branches as long as the length of the branches is proportionally high enough. Come examples can be seen in Figure 3.1.

The intersection ground truth includes 1516 landmarks and the number of intersections per image ranges from 3 to 124. In the case of endpoints, 2110 landmarks have been labeled and the number of endpoints per image ranges from 12 to 119. In both cases, the landmark density or concentration varies between the images. Figure 3.2 shows the number of intersections and endpoints for each image.

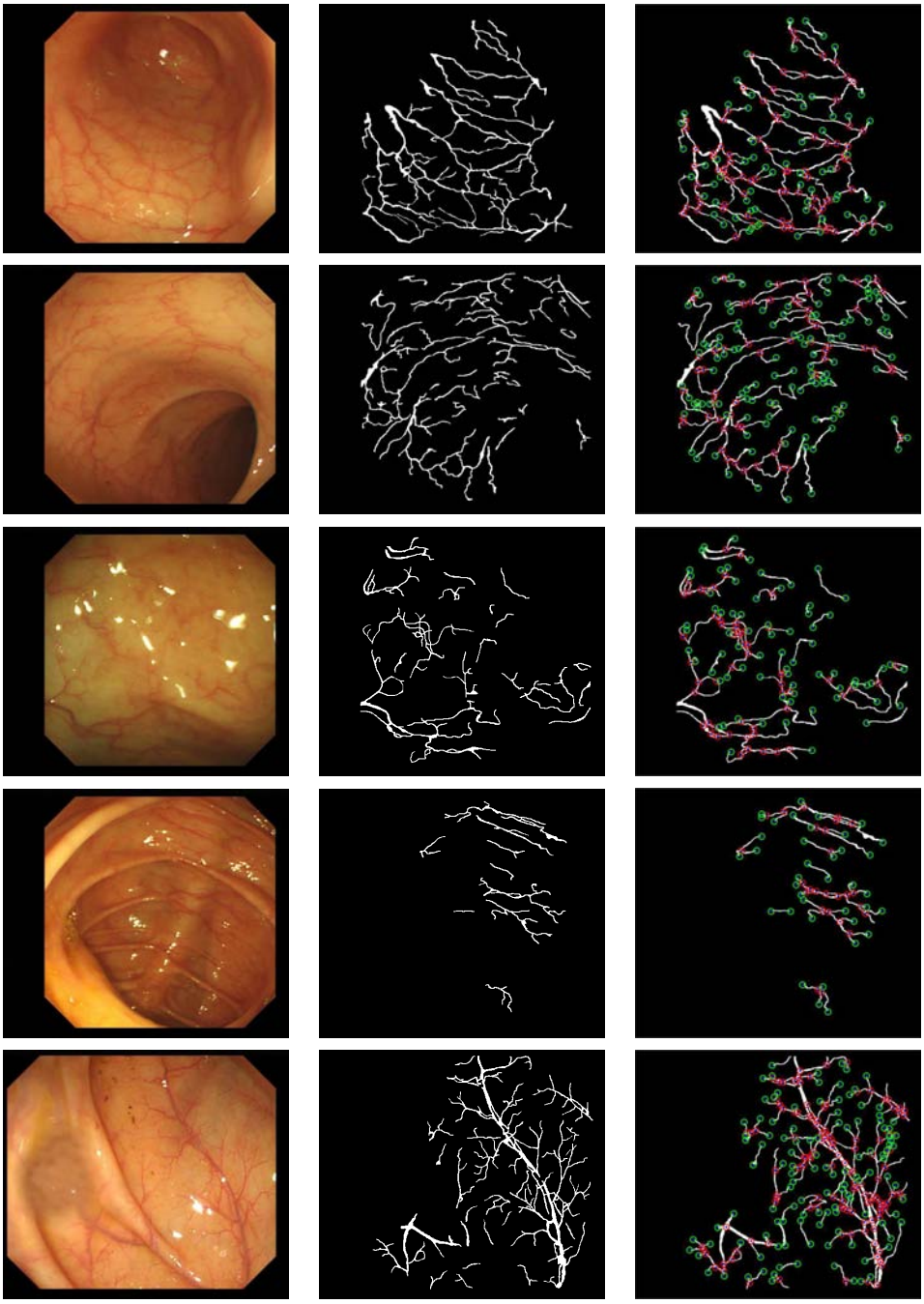
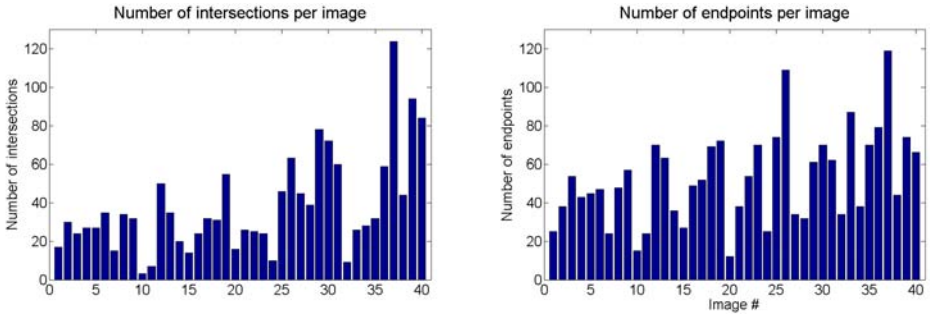


Figure 3.1: COLON-VESSEL data set examples. First row: Original image. Second row: Vessel manual segmentation. Third row: Junction (red) & endpoint (green) landmarks.



(a) (b)

Figure 3.2: COLON-VESSEL data set statistics. (a) Number of junctions. (b) Number of endpoints.

3.1.2 COLON-V-TRACE data set

The main goal of colon and track screening and colonoscopy is the search for lesions, adenomas and polyps. The previous data set includes blood vessels in colonoscopy but polyps are not present in all the imagse. The study of the interaction of polyps and the vascular patterns in the endoluminal scene demands for a data set that includes both objects in every single. This is the case of COLON-V-TRACE data set, which includes 29 colonoscopy images including both polyps and blood vessels (see Figure 3.3 for some examples).

COLON-V-TRACE data set is composed by a selection of images from CVC-ColonDB, the only public data set of colonoscopy images [18]. CVC-ColonDB includes 300 different images selected randomly by physician experts from the same 15 colonoscopy videos used in the creation of COLON-VESSEL data set. In this case, 20 frames per video sequence were obtained, whose image resolution is also 574×500 pixels. The frames were selected by experts so that they show endoluminal scenes containing polyps and adenomas and the points of view of the scene were significantly different. The database includes the original frame image as well as the manual segmentation of the polyps -and their contours- and the non-informative regions. Figure 3.4 includes some examples of images in the database and shows the variability in polyp appearance.

The images in CVC-ColonDB data set have been selected by the experts to provide a polyp database and maximize the variability in that sense. Different types of polyps, with different sizes and presenting different positions are covered by the data set. Nevertheless, the presence of other endoluminal objects -such as blood vessels, folds or highlights- in the selected colonoscopy frames is not assured since the frame selection is an expression of the variability aimed by the authors regarding polyps. That is why the creation of COLON-V-TRACE data set arises as a necessity for the exploration of polyp and vascular interaction.

COLON-V-TRACE is conformed by a subselection of 29 images from CVC-ColonDB data set following a selection criteria based on the presence of vascular content. By selecting images which contain blood vessels among those included in CVC-ColonDB, we create a set of images containing both polyps and vascular patterns.

The original colonoscopy image frames are paired with a ground truth of blood vessel traces manually labeled by experts. The vascular content in the images have been labeled without considering the width of the vessels. Therefore, only the vessel trace information is considered.

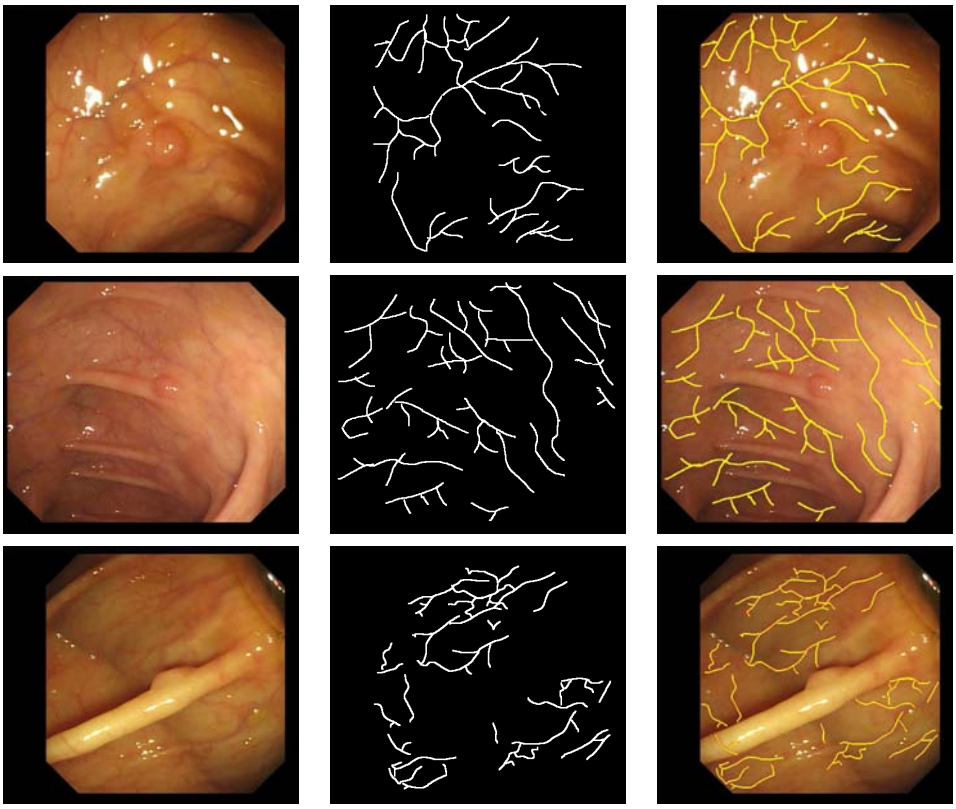


Figure 3.3: COLON-V-TRACE data set examples. First row: Original image. Second row: Vessel trace manual segmentation. Third row: Vessel trace mask superimposed on image.

3.2 Retinal fundus images

3.2.1 DRIVE data set, NunGT & AzzoGT

Branching structures or vascular patterns are present in different kinds of images besides colonoscopy. That is the case of retinal fundus images (see Figure 3.5a). In order to broaden our experiments to other contexts, we also experimented with DRIVE, a well-known retinal fundus data set. This data set has also been used in previous studies of intersections and a public ground truth of junctions and intersections is already available. We contributed with newly created ground truth. Together with the existing ground truth, they give us the chance to analyze the challenging task of intersection characterization regarding 2 observers. The location of intersections of any type is not a simple question and can be challenging for expert human observers or even a matter of disagreement. That is why it is important to have the possibility of different observers for the same data.

DRIVE is a public data set of retinal fundus images published in [144]. This data set has been commonly used in comparative studies on segmentation of blood vessels in retinal

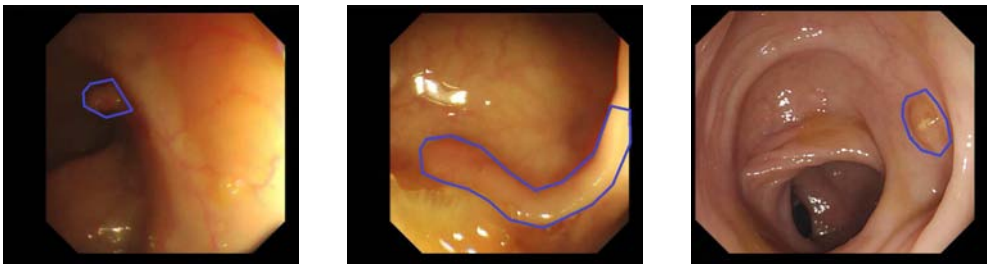
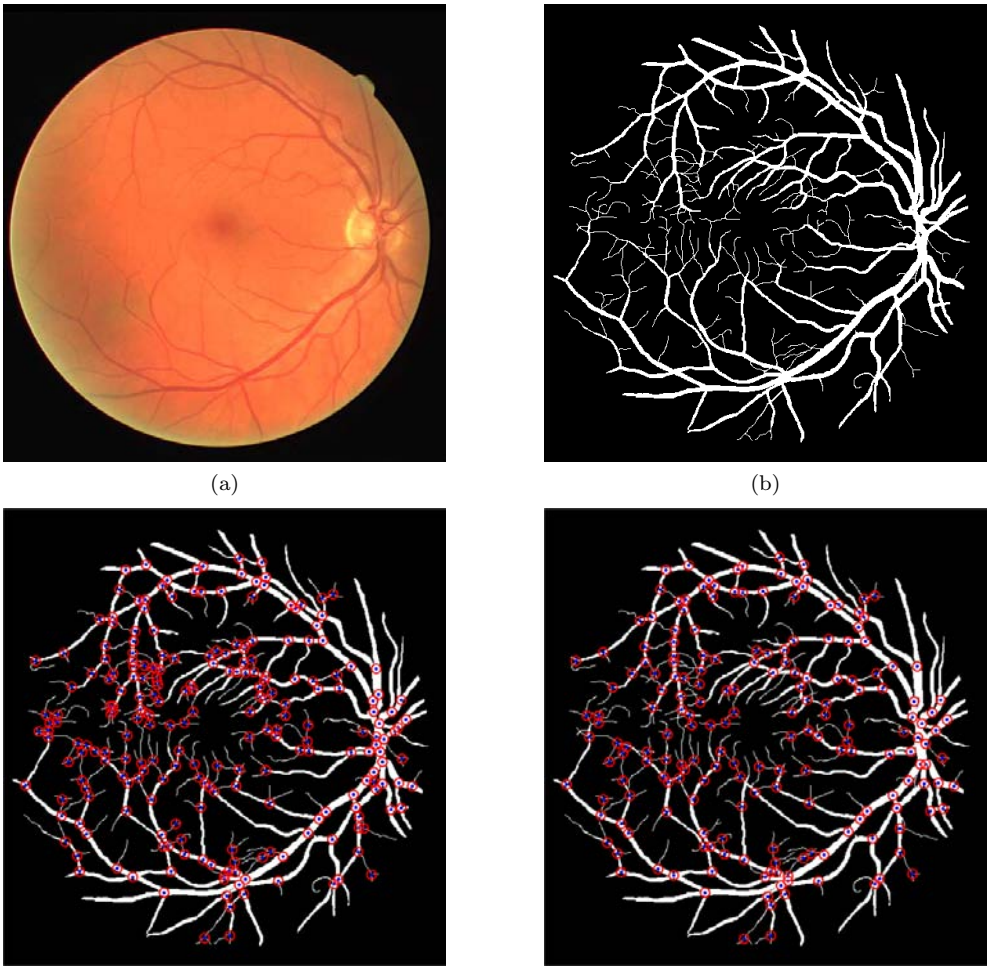


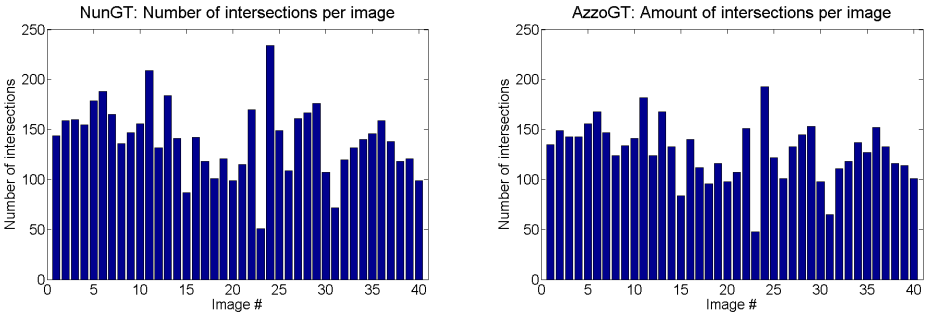
Figure 3.4: CVC-ColonDB data set examples (polyps in blue) (as in [18]).



(a) (b)

(c) (d)

Figure 3.5: DRIVE data set examples. (a) Original image. (b) Manual segmentation. (c) *NunGT*. (d) *AzzoGT*.



(a) (b)

Figure 3.6: Number of junctions in DRIVE ground truths: (a) *NunGT*. (b) *AzzoGT*.

images. It was obtained from a diabetic retinopathy screening program on a population of 400 diabetic subjects between 25 and 90 years old and 40 images of size 546×584 pixels were selected. The whole set of images is divided into a training and a test set, both containing 20 images. A single manual segmentation is provided for the training images. However, two manual segmentations are available for the test cases. As stated by the authors, one segmentation is the gold standard and the other one can be used as a reference for comparisons to computer generated segmentations. In the experiments performed in this work we will always refer to the 40 image data set including the gold standard manual segmentation in the case of the test set. Figure 3.5 shows examples of DRIVE data set.

We contributed with a new ground truth of manually annotated intersection landmarks for DRIVE data set, *NunGT*, that includes 5607 landmarks. As in the case of DRIVE-VESSEL landmark ground truths, the manual location of landmarks was achieved on the previously segmented vascular structures, which in this are case available in DRIVE data set. The expert was given the same rule of placing the landmark in the intersection of the imaginary axis of the branches.

There is an existing ground truth of manually annotated intersections for the DRIVE data set, *AzzoGT* [13], that has been used as a second observer. It includes 5118 bifurcations and crossroads, and it is publicly available[1]. Examples of both *NunGT* and *NunGT* ground truths are showed in Figure 3.5.

[1]http://goo.glMAKuPd

Part II

JUNCTIONS & BRANCHING PATTERNS

Chapter 4

Junction Characterization in Vessel Patterns

4.1 Branching patterns and junctions

Vascular structures are relevant in the analysis of colonoscopy scenes. The analysis of the blood vessels in colonoscopy images helps to improve the knowledge of the endoluminal scene. The better knowledge of the vascular content in the scene would represent an important step forward in the improvement of techniques for polyp and adenoma localization, and the development of endoscopy navigation support systems or colon wall mucosa visibility evaluation standards.

From a broader point of view, branching patterns are certainly common structures in nature. Patterns such as vascular trees, road networks, palm prints or topographical structures like rivers are just some examples of common branching patterns. Therefore, the analysis of branching patterns represents a very important task not only in the case of blood vessels and medical diagnosis contexts, but also in fields such as road detection or biometrics. The presence of branching structures in medicine and biology has been widely reported [3, 131]. Retinal blood vessel morphology has been described as an important indicator of hypertension, diabetes, arteriosclerosis or other cardiovascular diseases [40, 154]. Retinal vascular trees or palm prints have also been reported as reliable biometrics for personal identification tasks [117] or registration systems [35].

The detection of branching structures in every different context, their characterization and the measurement of the network properties are crucial tasks for subsequent applications. In every different context, the branching patters have particular characteristics. Their segmentation in every single context can be highly dependable on the nature of the problem and the images.

As a common valid notion, branching patterns can be defined as piecewise elongated structures that cross over each other or branch off more or less frequently. This definition is verified by branching patterns of any nature and certifies the importance of the branching points as crucial landmarks in this kind of structures. Detecting intersection keypoints becomes a necessary task when capturing the structure of branching patterns in any context. The knowledge of the branching points becomes an important step for any application that intends to characterize the original pattern and can be recognized as task of interest itself.

In this chapter a novel and very simple geometrical model to describe intersections in

(a) (b)

Figure 4.1: Usual thinning artifacts. (a) Original patterns. (b) Thinning results with artifacts (Necking: green, Tailing: red, Spurs: orange).

branching structures is presented. The model based on two conditions: Bounded Tangency condition (BT) and Shortest Branch condition (SB). The proposed model precisely sets a geometrical characterization of intersections and allows the introduction of a new unsupervised operator for intersection extraction. Ann implementation is also proposed, which handles the consequences of digital domain operation that, unlike existing approaches, is not restricted to a particular scale and does not require the computation of the thinned pattern.

The new proposal, as well as other existing approaches in the bibliography, are evaluated in a common framework for the first time. Although the new method is applied here in the context of medical 2D structures, it can be straightforwardly generalized to 3D or any kind of branching pattern. Performance results are showed on COLON-VESSEL and DRIVE data sets [144]. In the latter case results are validated by using both the existing *AzzoGT* intersection landmark ground truth [13] and the newly created *NunGT* ground truth.

4.2 Literature review

The existing approaches for vascular intersection detection, fundamentally proposed in the field of retinal imaging, can be separated into three categories [13, 25]: geometrical-feature based methods and model based methods.

4.2.1 Geometrical-feature based methods

Geometrical-feature based approaches usually perform a pixel-level processing stage followed by different kinds of post-processing analysis. These approaches usually involve adaptive filtering and branch analysis based on thinned structures. They are often computationally costly since they involve the processing of each pixel independently. An important step of the methods in this category usually consists of a thinning algorithm leading to compute the so-called skeleton of the structure, as in [25, 39, 74, 114]. These methods claim that it is desirable to reduce the original structures to one-pixel wide vascular trees. A skeleton, which has not a unified definition for the different implementations, is generated by a process of thinning. This process starts from the original structure and must identify the pixels belonging to it that are essential to keep the original structure shape [110]. Skeletonized shapes are usually affected by thinning artifacts like necking, tailing, spurs or staircase artifacts (see Figure 4.1), which landmark detectors will have to handle [110]. This topic is tackled in depth in Chapter 5.

Martinez-Perez et al. [114] proposed a characterization of retinal vascular content based on the one-pixel wide tree computed from the vessel pattern. Skeleton pixels are scanned in a 3×3 neighborhood so that bifurcation and crossover candidates are extracted by selecting skeleton pixels with 3 or 4 neighbors respectively. They propose a semiautomatic method to

overcome the fact that close bifurcations are usually joint into a crossover. Chanwimaluang et al. [39] proposal performs a similar first candidate selection procedure followed by a second processing step that removes small intersections by using the boundary pixels of an 11×11 window. Jung et al. [74] detector of vascular landmarks is also use the skeleton to detect crossroads as cross perpendicular structures with four connections and bifurcations as Y-type structures.

Bhuiyan et al. [25] method extracts vascular landmarks from the centerline image by using 3×3 rotational invariant masks to select potential candidates. The candidates are analyzed to find geometrical and topological properties that are used to classify landmark candidates as bifurcations or crossovers. Ardizzone et al. work [7] included vascular landmark extraction again based on the connectivity of the one-pixel wide vascular tree without any further candidate selection.

Another approach called combined cross-point number (CNN) method is introduced in [2]. This is a hybrid method of two intersection detection techniques: the simple cross-point number (SCN) [24] and the modified cross-point number (MCN) method. The former is based on a 3×3 window that is placed in the considered pixel to compute its so-called cross-point number (cpn), which basically counts the number of converging branches to the pixel. Bifurcation points must hold 3 transitions ($cpn = 3$). This method follows the same idea as the previous approaches. However, the authors propose a solution to the problem of turning a crossroad into a pair of bifurcations. The solution is based on MCN, a new operator based on a 5×5 which also computes the number of converging branches to the pixel but, in this case, in a 5-side window parameter.

The work proposed by Calvo et al. [33] also reduces the vascular structure to its skeleton, which is filtered to reduce spurious projections. The skeleton pixels are then classified by using their intersection number, equivalent to the already mentioned SCN, followed by post-processing techniques to solve crossover detection problems based on the intersections between a circumference of a given radius and the thinned pattern tree. The authors propose a voting system which involves three different radii. Finally, the classification is refined by merging two bifurcations into a crossroad if they are close enough (represented by a radius parameter) and connected by a single segment.

Saha et al. [132] method also takes skeleton tree extracted from the vascular structure and does not detect crossroads. They consider a window centered in the candidate pixel and each connected-component is uniquely labelled. The algorithm makes an anti-clockwise round-trip along the perimeter of the window. A pixel is classified as a bifurcation point if the cyclic path length is 3 and does not have any repetition.

4.2.2 Model based methods

These group of methods is based on a vectorial tracing of the desired structure. Seed points are usually placed as initial locations so that the vascular structures in the image can be tracked from them recursively. These methods usually have lower computational complexity than the methods in the previous category as they do not need to process every pixel in the image so they are usually proposed for real-time applications.

The method introduced by Can et al. [34] is based on an antiparallel edges model of the linear portions of the vascular pattern. The algorithm keeps relevant tracing information in two data structures as the tracking of the branching pattern proceeds, the so-called "centerline image" and "centerline network". The former is an array which keeps non-zero values for the already traced centerlines and increments a variable called the "segment number" when each new segment in the vascular structure is tracked. The latter consists of a linked list of segments so that every single segment is a linked list of connected pixels which rep-

resent the already traced centerline of that segment. The centerline image is checked from the current tracing point to label it as a bifurcation candidate if non-zero values are found in three different small line searches. At the same time, the centerline network is searched every time a previously detected vessel is intersected and the intersection point is updated. When multiple close intersections are detected they are replaced by their centroid.

Tsai et al. [153] presented an exploratory or tracking approach named exclusion region and position refinement (ERPR). This approach is also based on the antiparallel model. Nevertheless, this work considers this model is valuable for the tracing algorithm itself but it is no longer valid when approaching intersection or branching points. As a consequence, the authors claim that the estimation of vascular landmarks is clearly affected. They propose a model for intersections based on the landmark location, the set of vessel orientations that meet in the intersection and a circular exclusion region where the antiparallel model is violated. The landmark extraction algorithm starts at an endpoint of the trace, either when it intersects another vessel or when it meets at least two other trace endpoints. They launch an iterative process from those endpoints that re-estimates the traces when outside exclusion regions and re-estimates the landmark position otherwise.

4.2.3 Hybrid approaches

Azzopardi et al. [13] introduced a different approach proposing the use of so-called COSFIRE (Combination Of Shifted FIlter REsponses) filters [14]. COSFIRE filters are keypoint detection operators that must be trained to extract given local patterns. These filters are made up of Gabor filters that are combined so that the response of a given pixel is computed as a combination of the shifted responses of the Gabor filters. The final output includes the local maxima from the outputs of all trained filters.

4.3 GRAID: GRowing Algorithm for Intersection Detection

GRowing Algorithm for Intersection Detection (GRAID) is a hybrid approach based on the definition of a precise intersection model that operates at pixel level. An intersection model allows us to define the landmark which represents the location of an intersection. The model is defined by two conditions: Bounded Tangency (BT) condition, and Shortest Branch (SB) condition. The algorithmic implementation has one single parameter that states the leverage between the geometrical proportions of the branches and the intersection. The method is not restricted to the computation of the thinned pattern nor conditioned by a sliding window size. For these reasons the algorithm is independent from drawbacks of thinning methods and it is not restricted to any particular scale. The outcome is a straightforward and precise intersection detector which does not need to go through a training process and that is able to classify separately the intersections regarding its number of branches.

We propose a method to extract bifurcations and crossroads from branching patterns in binary images based on a general intersection model. A precise model to allow the definition of the landmarks representing the location of intersections is stated. Given that model, an algorithm which handles the consequences of working on a digital domain, such as the approximation of the Euclidian distance and the lack of resolution to reach maximal ball tangencies, is proposed. The first part of this section introduces our proposed model and the second part proposes the corresponding implementation of the detector.

4.4 Intersection model in binary patterns

An intersection candidate is defined by the center of a maximal circumference inscribed in the branching pattern. The candidates are extracted as intersections if and only if they hold the following two conditions:

- Bounded Tangency (BT): the maximal inscribed circumference and the pattern contour must have 3 or more tangencies.
- Shortest Branch (SB): the relation between the shortest branch and the radius of the inscribed circumference must be higher than a given ratio.

Given a binary image containing a structure pattern, S, and a point, $x \in S$, we define the circumferences with a radius r, centered at x and inscribed in S as $C_S(x, r)$, where $0 < r \leq r_{max}$. When $r = r_{max}$ the circumference is maximally circumscribed. Then, a decision function for intersection extraction is defined as follows:

$$B(x) = |P_S \cap C_S(x, r_{max})| \tag{4.1}$$

where P_S is the contour of the structure, S.

The verification of BT condition is achieved through the analysis of $B(x)$ function. $B(x)$ describes the number of tangent points between the maximal inscribed circumference and the branch pattern contour. Every single point within S will be forwarded as an intersection candidate if the number of tangent points between the inscribed circumference and the structure contour is ≥ 3. Since the number of tangencies is equivalent to the number of branches, $B(x)$ also describes the number of branches converging at each intersection candidate. Regarding the usual terminology in the bibliography, those points verifying $B(x) = 3$ will be bifurcation candidates (3 branches) and those verifying $B(x) = 4$ will be crossroad candidates (4 branches). Our model allows in this way to separately extract intersections with a particular number of branches, although we will focus in this work in general intersection extraction by simply allowing $B(x) \geq 3$.

After verifying BT condition each branch must be tracked to assess that SB condition imposed by our model is also held. SB condition is mathematically defined as:

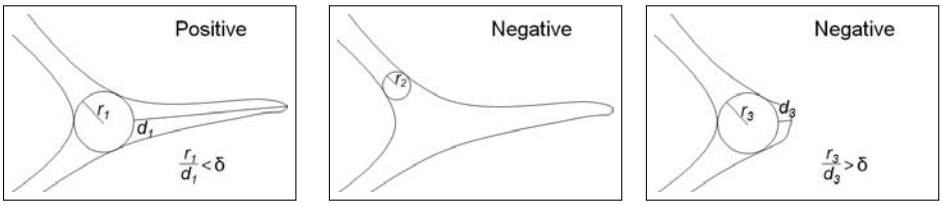
$$\frac{r_{max}}{\min_i (d_i)} < \delta \tag{4.2}$$

where δ is the geometrical parameter balancing the size of the circumscribed circumference and the length of the branches converg d_i, $0 < d_i \leq B(x)$, are the lengths of all branches from point x. Thus, the SB condition assures a minimal length for all the branches converging in the intersection landmark.

Figure 4.2a illustrates a positive candidate as it verifies both conditions imposed by our model. However, Figure 4.2b shows a negative candidate as there are only two tangent points so BT condition is not validated. A different negative candidate situation is illustrated by Figure 4.2c since BT condition is verified but SB condition is not.

4.5 Algorithm description and implementation

The input to the intersection detector is a binary image containing the branching pattern the output is the set of pixels representing the location of intersections. The input image is first pre-processed by a filtering stage that tackles the problem of isolated or spurred pixels inside the branching region by setting them to background value. This can be simply achieved by



(a) (b) (c)

Figure 4.2: Intersection model: candidate examples. (a) Positive (verifies both BT and SB conditions). (b) Negative (does not verify BT condition). (c) Negative (verifies BT condition but does not verify SB condition).

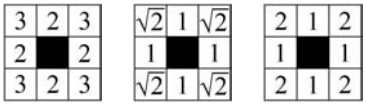
bringing to foreground those pixels whose 8-connectivity neighbors are at least 7 of them set to foreground.

Algorithm 1 shows the basis of the proposed method. The algorithm takes as input the binary image containing the branching structure and the branching ratio δ. We set $\delta = 1.5$ so that the targeted intersections can be unequivocally accepted, not allowing spurs to be taken as branches.

The first step of the algorithm is based on the Danielsson distance algorithm (line 1) [46], which computes the distance map -*dMap*- of the complementary of the input image based on the distance function *Cost*. Based on our model, we would ideally use the Euclidean distance. However, dealing with a discrete domain and the necessity of a geodesic distance in the moment of expanding branches prompts us to define an approximation. We define a cost function describing the distance to 8 neighbors starting from upper left corner as *Cost(n)*. We selected the cost function as shown in Figure 4.3a. Other possibilities can be: a float approximation to the Euclidean distance (Figure 4.3b) -which provides similar results although increasing computational cost- or the block distance (Figure 4.3c) -which misses some center ball candidates-.

Algorithm 1: Algorithm

```
Input: image: Input binary image, δ: minimum branching factor
Output: output: Intersection binary image
dMap = DanielssonDistance(Cost,!image);
for pix in image do
    if IsBallCenter(pix,dMap) then
        frontier = ExpandCenter(pix,dMap,Cost);
        len = dMap(pix)*δ;
        nbranch = ExpandFrontier(frontier,Cost,len);
        if nbranch > 2 then
            output(pix) = true;
        end
    end
end
return output;
```



(a) (b) (c)

Figure 4.3: Cost function examples. (a) Integer approximation of Euclidean distance. (b) Float approximation of Euclidean distance. (c) Block distance.

Next, each pixel in the image is processed separately. There are three basic functions in the algorithm implementation: **IsBallCenter** and **ExpandCenter**, both used to check BT condition, and **ExpandFrontier**, which checks SB condition.

IsBallCenter function (line 3) checks the branching pattern to select those pixels, x, that are centers of a maximal ball:

$$\{x \in S \mid \exists C_S(x, r_{max})\} \tag{4.3}$$

$$r_{max} = \arg\min_r \{P_S \cap C_S(x, r) \neq 0\} \tag{4.4}$$

A pixel will be extracted as a maximal ball center in any of the following cases:

1. The maximal ball radius r_{max} from the candidate pixel x is higher than the difference between the maximum radius from each neighbor n, r^n_{max}, and the distance to that neighbor, $Cost(n)$:

$$r_{max} > \max_n \{r^n_{max} - Cost(n)\}, \forall n \tag{4.5}$$

2. The difference between the maximum radius from each neighbor and the distance to that neighbor is positive, and the maximal ball radius from the candidate pixel is higher than the minimum cost to reach a neighbor:

$$\{r^n_{max} - Cost(n) > 0\} \vee \{r_{max} > \min_n \{Cost(n)\}\}, \forall n \tag{4.6}$$

The inequality in 1) would be enough if a discrete domain. The implementation in a digital domain forces the introduction of the or-condition in 2). Figure 4.4b shows an example of maximal ball centers by **IsBallCenter**.

Every ball center is then analysed further so that the second part of BD condition is tested. We must select those maximal ball centers which have at least 3 tangencies to the pattern contour, i.e. at least 3 branches. To asses the number of branches we must expand the ball from its center to its radius. This task is achieved by **ExpandCenter** function (line 4). The discrete domain can cause the maximal circumference radii we already computed to be too short to reach all the expected tangent points. We handle this problem by adding an offset equivalent to the maximum value of the *Cost* function to the radius of the maximal ball:

$$r\prime_{max} = r_{max} + \max_n \{Cost(n)\} \tag{4.7}$$

With this offset we make sure that the algorithm reaches the contour and the right amount of tangencies are identified. This improvement saves us from missing maximal ball centers with 3 or more branches (see Figure 4.4c and 4.4d for an example).

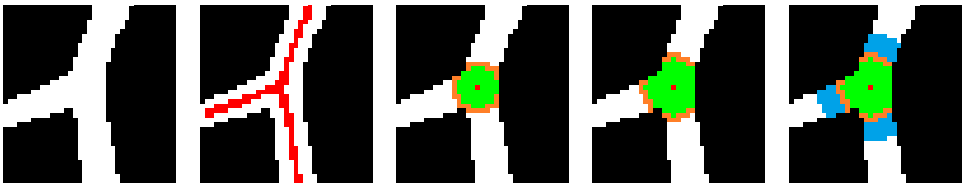


Figure 4.4: Algorithm samples. (a) Intersection pattern. (b) Maximal ball candidates. (c) Maximal ball (green) and 2 outer branch frontiers (orange). (d) Extended maximal ball (green) and 3 outer branch frontiers (orange). (e) Frontier expansion (blue)

ExpandCenter function (line 4) expands every center pixel to its maximal ball contour. The pixels that are part of that contour are tested to isolate those that have at least one neighbor that is part of the structure (foreground) and out of the maximal ball. Those isolated pixels are then grouped in connected blobs that we call *branch frontiers* (see Figure 4.4d).

Finally, branch frontiers need to be expanded as shown by Figure 4.4e (line 6) to assess the verification of SB condition. The corresponding ball center pixel will be labelled as an intersection candidate if and only if at least 3 branches verify SB condition expressed by Equation 4.2 (line 7).

Algorithm 2 shows the explicit pseudocode implementation of **ExpandCenter** and **ExpandFrontier**. The algorithm expands a given pixel based on the distance map by prioritizing the expansion of those pixels with a lower distance cost until the corresponding *StopCondition* is reached (line 17). In the case of **ExpandCenter**, the stop condition is to reach a frontier pixel. In the case of **ExpandFrontier** the condition would consist of reaching the branch distances that assess SB condition (Equation 4.2).

The final output of the algorithm are the centroids of the landmark candidate blobs since several candidates may be selected for a given intersection due to the discrete working domain.

4.6 Results

4.6.1 Validation framework

We implemented our operator in C/C++ and all the experiments were run in a Personal Computer with a 2.67 GHz processor. In order to validate or method, we use two different data sets of vascular images related to two different anatomical problems: 1) The DRIVE data set, for retinal fundus images, and 2) the COLON-VESSEL for colon vessels from colonoscopy images. In the case of DRIVE data set we use both the already existing $AzzoGT$ intersection ground truth and the newly created $NunGT$ ground truth, which allows us to compare the variability in intersection landmark placement. In the case of COLON-VESSEL data set we will use the only existing intersection landmark ground truth for colonoscopy vascular patterns.

In the case of our new $NunGT$ and COLON-VESSEL junction ground truths, as already introduced in Chapter 3 the landmark was placed in the intersection of the imaginary axis

Algorithm 2: Expand

```
Input: pix: pixel to be expanded dMap: Danielsson Distance map to
       background CostFunc: 8 connectivity cost function
Output: output: number of frontiers/branches
Make-Queue: queue;
queue.Push(pix);
Cost = ascendingSort(CostFunc);
temp = dMap;
while !queue.isEmpty do
    if queue.First! = NULL then
        queue.Push(NULL);
        dinc = min(Cost);
        forall the n neighbors in Cost do
            if Cost(n) < dinc then
                queue.Push(NULL);
                dinc = Cost(n);
            end
            forall the q in queue do
                if Pixelisforeground then
                    d = temp(n) + Cost(n);
                    if StopCondition then
                        [...];
                    else
                        temp(n) = d;
                        queue.Push(n);
                    end
                end
            end
        end
    else
        queue.Pop();
    end
end
return output;
```

of each branch as long as the branch length is proportionally high enough. We know this procedure verifies both BT and SB conditions.

The different performance results have been compared in terms of precision, sensitivity and their harmonic mean (F_1 score), which are defined as follows:

$$Precision = \frac{TP}{TP + FP} \tag{4.8}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.9}$$

$$H.mean = 2 \cdot \frac{Prec \cdot Sens}{Prec + Sens} \tag{4.10}$$

where TP (True Positives) are the number of landmarks extracted correctly, FP (False Positives) are the number incorrectly extracted landmarks and FN (False Negatives) are the landmarks that were not detected. Any detected landmark is considered correctly extracted (TP) if the distance to the corresponding landmark in the ground truth is smaller than an evaluation parameter ϵ. The value of ϵ has been set to 5 pixels for the evaluation of the different approaches. The impact of ϵ in the validation will be discussed in Section 4.7.

4.6.2 Experimental results

The performance of our method has been compared to previous approaches. We implemented several methods among those introduced in Section 4.2 that have never been compared in the same framework: Filter Based Junction detector (FBJ), Aibinu approach [2] and Saha et al. [132] proposal with some modifications. We call FBJ the basic idea used in intersection extraction methods such as Martinez-Perez et al. [114] and Chanwimaluang et al. [39]. These algorithm selects from the skeletonized structure those pixels which have at least 3 neighbors considering 8 connectivity. Saha et al. [132] algorithm is designed to extract bifurcations -3 branch intersections- by processing the cyclic path of a sliding window (see Section 4.2) whose length must be 3. We allow the length to be 3 or higher to widen the algorithm target to intersections with any number of branches. Since the authors did not clarify what window size should be used, after extensive tests we determined to use a window size of 10×10 as the optimal trade off to avoid missing intersections and not to join those that are closer.

The method published by Azzopardi et al. is also considered in the comparison although we just took the performance results published by the authors [13]. The method is based on COSFIRE filters, which must go through a training process. We performed different trainings following the authors directions which showed a large variety in the outcome and did not get to approach the performance published by the authors -96.60 % precision, 97.81 % recall, with no reference to ϵ-.

FBJ, Saha and Aibinu include a thinning step. The selection of a thinning algorithm has consequences in the performance of an intersection detector. The selection of a thinning algorithm should mind the problems described in Section 4.2 (see Figure 4.1). Aibinu is the only method, among those which use skeletonized structures, that explicitly proposes to use a particular thinning algorithm for its intersection detector: Kwon et al. algorithm [83]. For this reason we decided to use Kwon thinning method for FBJ, Modified Saha and Aibinu, although we also tested other standard methods without remarkable performance changes. In this way performance can be compared considering exactly the same advantages or drawbacks offered by the same single thinning algorithm.

Regarding GRAID, as introduced in Section 4.5, we defined $\delta = 1.5$ so that the geometrics of the targeted intersection are more inclusive, which just depends on the nature of the problem. The bigger the value of δ, the most restrictive SB condition is. GRAID does not need to go through a training stage and, give an input image, its performance is completely repeatable.

We first present the performance results for $\epsilon = 5$ and then we assess the impact of ϵ in the final performance. Three experiments are carried out: 1) *AzzoGT* on DRIVE data set, 2) *NunGT* on DRIVE data set, and 3) *ColonVesselGT* on COLON-VESSEL data set:

1. Table 4.1 shows the performance metrics for the different approaches using *AzzoGT* as the ground truth -we also include the 2nd observer results represented by *NunGT*-

Table 4.1: Exp. 1.1: Detector comparison on *AzzoGT* data set

Method	**Prec.** [%]	**Sens.** [%]	**H. mean** [%]
GRAID	90.60	93.22	91.89
Aibinu et al.	80.99	93.73	86.90
Modified Saha et al.	85.70	91.79	88.64
FBJ	53.24	89.55	66.78
2nd observer (*NunGT*)	89.23	96.64	92.78

Table 4.2: Exp. 1.2: Detector comparison on *NunGT* data set

Method	**Prec.** [%]	**Sens.** [%]	**H. mean** [%]
GRAID	96.67	93.12	94.86
Aibinu et al.	89.15	93.95	91.49
Modified Saha et al.	90.69	90.71	90.70
FBJ	56.84	88.26	69.15
2nd observer (*AzzoGT*)	96.68	89.35	92.87

Table 4.3: Exp 1.3: Detector comparison on COLON-VESSEL data set

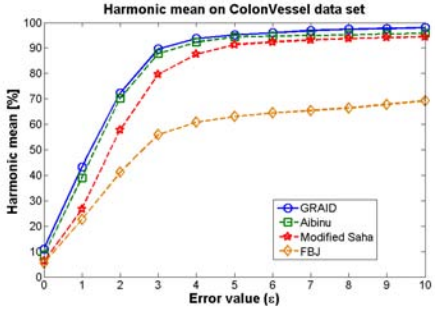
Method	**Prec.** [%]	**Sens.** [%]	**H. mean** [%]
GRAID	96.65	93.58	95.09
Aibinu et al.	91.64	95.76	93.65
Modified Saha et al.	87.93	94.97	91.31
Skeleton	47.72	92.97	63.07

. Our approach outperforms all the approaches considered and implemented in this study which have been compared in a common framework. The performance values published by Azzopardi et al. still remain higher -96.60 % precision, 97.81 % recall- . Nevertheless, as already mentioned, the evaluation conditions of the COSFIRE method are not clearly stated in the original work and our experiments showed a high performance variability when different training patterns are selected.

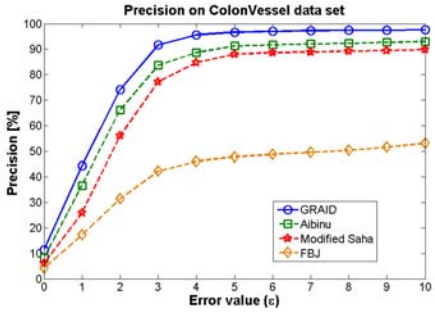
2. Table 4.2 shows the performance values for the same methods when considering *NunGT*. In this case we verify that again our proposal reaches values much higher than the other algorithms.

3. Table 4.3 shows results achieved for the images in COLON-VESSEL data set with *ColonVesselGT*. Again our method outperforms the state of the art.

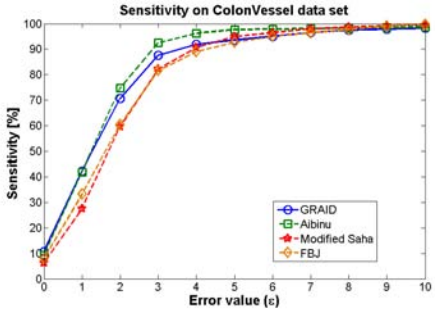
Finally , in order to clarify the importance of a common framework to achieve a comparison of different intersection detectors, the previous experiments were repeated modifying the value of $\epsilon = 5$. Similarly to the previous group of experiments, several tests were carried out. Figure 4.5 shows the results for COLON-VESSEL data set. Figure 4.6 shows the results for DRIVE data set and both *AzzoGT* and *NunGT*. The different plots show the variation of precision, sensitivity and harmonic mean when modifying the value of ϵ. These results demonstrate the important variation that performance metrics suffer when increasing ϵ value.



(a)



(b)



(c)

Figure 4.5: ϵ value influence on performance metrics for COLON-VESSEL data set. (a) Harmonic mean. (b) Precision. (c) Sensitivity.

4.7 Discussion

The experiments exposed above clarify that in all cases our algorithm reaches higher performance values than the other implemented methods. In the results shown in Table 4.1 and Table 4.2 the values of sensitivity keep close for the cases of GRAID, Aibinu and Modified Saha although GRAID reaches higher values of precision. The output of these two former methods is highly conditioned by the sizes of the windows they use since it varies the tar-

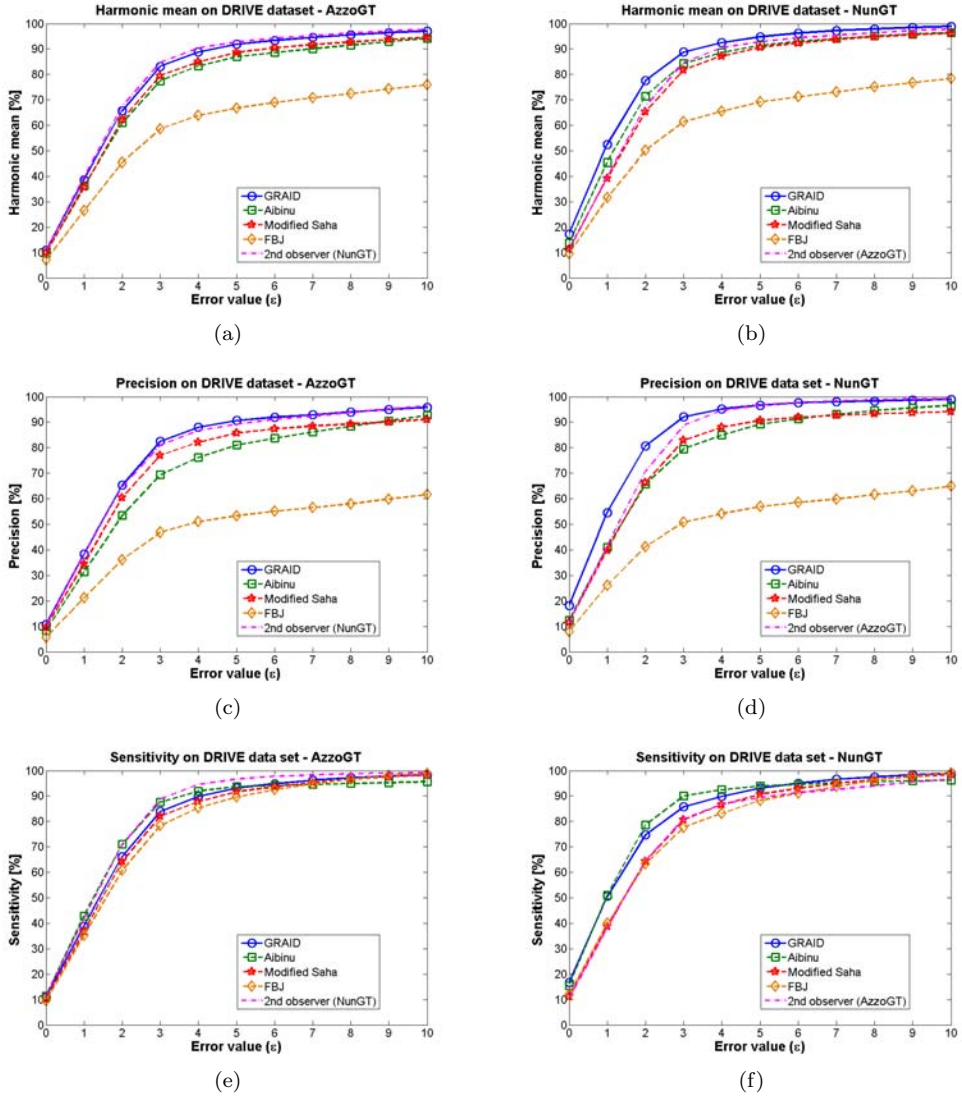


Figure 4.6: ϵ value influence on performance metrics: (a) Harmonic mean for *AzzoGT*. (b) Harmonic mean for *NunGT*. (c) Sensitivity for *AzzoGT*. (d) Sensitivity for *NunGT*. (e) Precision for *AzzoGT*. (f) Precision for *NunGT*.

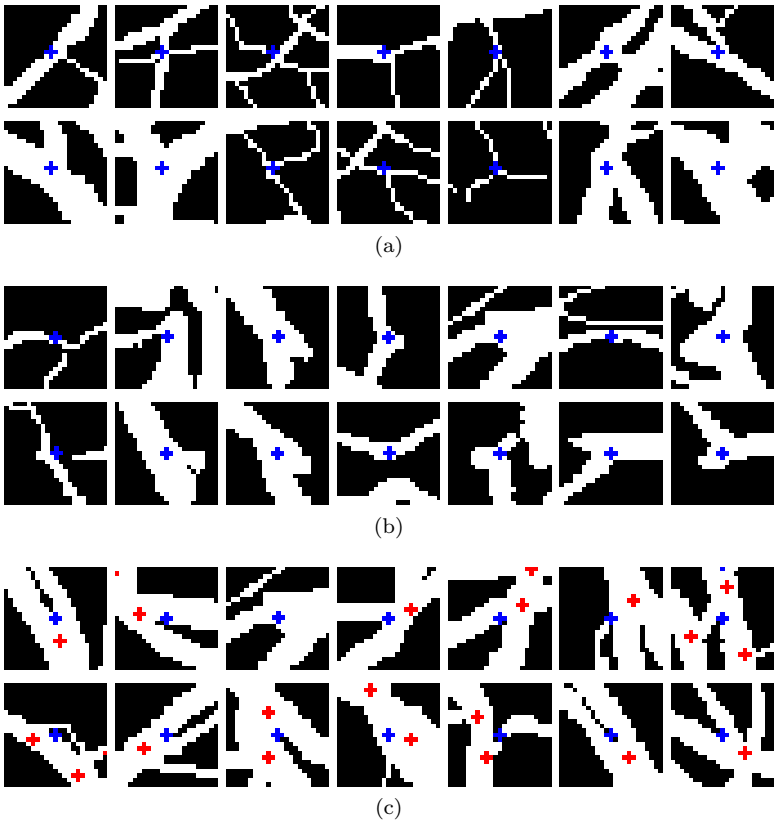
geted intersection size. As Aibinu states, we used 3×3 and 5×5 windows. As we explained above, 10×10 windows were selected in Saha algorithm. Conversely, our algorithm is not scale dependent so it is suitable to a wide range of images. FBJ algorithm extracts a high amount of False Positives decreasing to 53.24% of precision due to its basic approach based on a thinning process, suffering from the usual skeleton artifacts (see Figure 4.1). On the contrary, our proposal does not suffer from the problems caused by a thinning step.

The comparison between both ground truths -$AzzoGT$ and $NunGT$- shows an increase in the precision for all the methods when using $NunGT$ as the ground truth. To clarify these results we carried out a qualitative analysis of the extracted intersections when $NunGT$ expert is tested against $AzzoGT$ ground truth. We manually identified 501 out of 598 of the False Positives as actual True Positive intersections which were not considered in $AzzoGT$. Figure 4.7a shows some examples. Regarding False Negatives, 33 out of 177 resulted to be intersections that would not meet the formal criteria defined in 4.6.1 (see some examples in Figure 4.7b). In both cases the remaining intersections are caused by a shift in the pixel selected as the keypoint for each intersection. Some of these can be accepted as a consequence of different criteria. In this sense, we point out that our intersection model states a clear and concise criteria to select the representative keypoint for each structure. Some other cases, however, would not be accepted as good keypoints in our ground truth as they appear too shifted or they are not representative of the structure they should describe (Figure 4.7c shows some examples).

The experiment showed in Table 4.3 on COLON-VESSEL database points similar trends to the previous experiment. GRAID is still providing the higher values of harmonic mean although in this case the difference in terms of precision and sensitivity is lightly wider. Aibinu and Modified Saha reach higher levels of sensitivity but GRAID is much more precise.

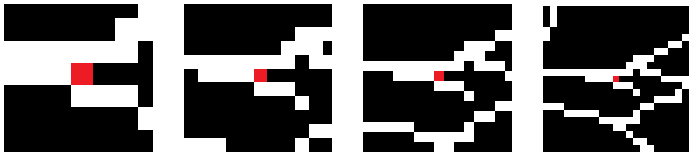
Experiments on ϵ value let us know about the algorithm accuracy as well as the influence of ϵ in the performance metrics. Plots in Figure 4.6 show that our value of $\epsilon = 5$ is big enough to be away from the sloppiest regions of the plots, which make it less prone to be influenced by small displacements of the landmark in the ground truth. At the same time, $\epsilon = 5$ is small enough to avoid the bias provided by random detections. In addition, the plots highlight the higher accuracy and performance of GRAID algorithm. For the particular case of Aibinu, sensitivity reaches higher values for the particular case of low ϵ, however showing lower values of precision for the same ϵ.

The results published by Azzopardi et al. -96.60 % precision, 97.81 % recall- are still higher than our tested method. Nevertheless, the values reached by GRAID keep considerably close. This is an important outcome since the validation conditions used by Azzopardi et al. are not completely clarified and present intrinsic problems for repeatability. The method they propose is based on COSFIRE filters, which need to go through a training process. Tools that need to go through a training process, and that are sensitive to the particular patterns chosen for the training phase, are less repeatable. The training process must consider the heterogeneity and redundancy of the training data -or patterns- to carry out a generalized implementation which is able to predict the correct output. For this reason, setting up the training process can be complex and demands for a deeper knowledge of the methodology. Moreover, the selection of the training samples, as well as their size, becomes crucial to reach a repeatable implementation. As seen in Figure 4.8, a difference in only a few pixels in the training pattern size entails including closer structures that will cause important differences in the resulting trained filter. Therefore, even though the final COSFIRE filter is not scale dependant, the final implementation is highly dependant on the shapes included in each selected pattern and, particularly, on its size. Differently, our method can be directly applied to a given binary pattern as it just requires a geometrical ratio to describe the targeted intersection proportions.



(a) (b) (c)

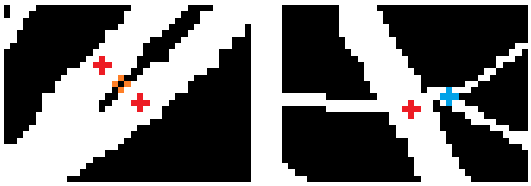
Figure 4.7: AzzoGT examples. (a) Not labelled intersections. (b) Labelled intersection not meeting our formal criteria. (c) Divergence in landmark placement (NunGT: red; AzzoGT: blue).



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

Figure 4.8: Branching patterns on different window sizes: influence on resulting training patterns. (a) 7 pixel side. (b) 11 pixel side. (c) 15 pixel side. (d) 21 pixel side.

This paper assumes the binary branching pattern is given as input to all methods. The binary pattern can be obtained in several ways regarding the nature of the images in a given problem -such as vascular tree segmentation in retinal images, which has been largely studied-. Geometrical-feature based methods outcome depend on the connectivity of the



(a) (b)

Figure 4.9: Error samples. (a) Caused by connectivity (orange: pixels causing connectivity). (b) Caused by proximity (blue: missing intersection).

given pattern. Model based methods does not rely on a given binary pattern. Their performance depend on the reliability of the tracking process, which is based on image surface gradient information. Since most of the methods used in branching pattern segmentation are also based on gradient information, the lack of connectivity will affect to the tracking process in the same way as it affects to most branching pattern segmentation approaches. Regarding hybrid approaches, Azzopardi et al. approach can manage a lack of connectivity although, for the same reason, False Positives will be extracted when closer branches are found. GRAID performance is based on the connectivity of the branching pattern. The lack of connectivity can be tackled by adding a previous morphological operation although False Positive intersection may be also extracted.

The main difference between the two considered hybrid approaches rely on the training process needed by the method proposed by Azzopardi et. al. The COSFIRE-based approach must go through a training process which highly conditions the performance of the trained detector and its repeatability. GRAID is applied as an operator based on a simple geometrical model whose complexity dealing with binary domain implementation is transparent to the user.

GRAID performance reaches state-of-the-art values when detecting intersection landmarks. The computation time of the operator depends on the nature of the branching pattern since it determines the number of times the conditions imposed by the model must be assessed. Our implementation of the methodology in C/C++ takes on average $60.8ms$ for each COLON-VESSEL image and $110ms$ for each DRIVE image.

The analysis of the intersections extracted by GRAID arises some error sources caused by: 1) the definition of the connectivity in the input image, and 2) the proximity of intersections. The former leads to erroneous extraction of intersections. Our method is based on the expansion from single pixels inside a given pattern, which we assume to be defined using 8-connectivity. The problem appears when 8-connectivity happens between parallel branches connected by a single pixel. This pattern verifies our model whereas the expert did not label it as an intersection (see Figure 4.9a for an example). The proximity of intersections causes our method to miss some landmarks due to the fact that our output integrates close intersections into a single one. This is caused by the approximations we make to implement our model in a digital domain (see Figure 4.9b for an example).

Another remarkable aspect on the performance of GRAID is related to one-pixel wide patterns. In such cases the area of the maximally inscribed circle is just one pixel. Considering our cost function -see Figure 4.3c-, the maximal ball candidates can be extracted by Equation 4.5. In these cases, the computation of the number of branches could be problematic though. Introducing an offset to the maximal ball candidate radius succeeded in making sure there will not be any branch missed.

4.8 Conclusions

We have proposed a new approach for precise intersection landmark extraction from binary branching structures based on a novel intersection model. The model states that junctions are those landmarks in the input branching pattern where a maximal inscribed circumference can be placed that has more than 2 tangent points with the pattern contour. The number of tangent points is equivalent to the number of branches, which allows our method to classify separately bifurcations -3 branches- and different kinds of crossroads -4 or more branches-. Given the radius of that circumference, the branches from that landmark must have a minimum length. The ratio between the circumference radius and branch lengths can be selected by the user regarding the targeted intersections. We have successfully overcome the implementation problems of this kind of approach given the digital domain of images providing a robust and simple interpretation of its performance. Moreover, our method can be naturally extended to 3-dimensional input data or branching patterns of any nature, such as vessels, roads, palm prints or topographical structures.

We have compared our algorithm with previously published works in order to provide the first evaluation of several approaches in a single evaluation framework. For that purpose, we have assessed the performance of our proposal in the a existing ground truth for DRIVE retinal data set and we have contributed with a second intersection landmark ground truth to the retinal DRIVE data set to provide a reliable interpretation of results. We also created a new data set of colonoscopy frames and the corresponding intersection ground truth.

The performance values reached in terms of precision and sensitivity place our method in the best performance level for those approaches implemented in this work. The performance of our method remains in lower levels than the cited values by Azzopardi et al. However, we have showed that the impact of evaluation conditions on the the final performance is high enough to make that difference less remarkable as well as the necessity of a training process and a complicated parametrizing process, which have a direct impact on results and overfitting. Conversely, the novel method we propose is simple, highly repeatable and does not need neither a parameter tuning step nor a training stage.

Chapter 5

Vessel Pattern Analysis and Skeletonization

Shape representation and description is an important task in several areas of Computer Vision, such as image retrieval or object categorization. Shape recognition is also key in the interpretation of medical images as well as in the development of biometric tools. The description of branching patterns, like blood vessels in colonoscopy images, must focus on the representation of topological information. In this chapter we propose the description of binary branching patters by means of a novel skeleton computation method based on GRADE junction localization algorithm. This approach keeps the advantages of skeletons in terms of simplicity and improves traditional approaches with an improvement on topology representation and by reducing sensitivity to boundary noise.

5.1 Related Work

Shape description has been a common research topic in the last decades in the context of different recognition applications. In the different contexts, shape representation has been approached in many ways. Shape description can be based on the shape contour information or on the whole region shape. This difference allows the classification of the shape description approaches into two categories [163].

Contour-based methods usually compute a feature descriptor by exploiting only shape boundary information. The comparison between different shapes is achieved by measuring some kind of metric distance between descriptor vectors. Shapes can be represented as simple global features computed from the shape -such as area, eccentricity, convexity or elliptic variance [161, 115]-, as point sets [54, 129, 16, 17], or by one dimensional functions [164]. Other countour-based approaches explored shape representation in other domains, such as Fourier or wavelet descriptors [42, 78, 130], or descriptors that break the contours down into smaller segments [51, 56, 23, 143].

Region-based methods use all the pixels within the region shape to compute its description. Basically, three approaches are found within this category: moment descriptors, scan-based descriptors and skeletons. Moment descriptors use nonlinear combinations of the image moments to describe shapes [65, 149, 148, 80]. Scan-based descriptors use a grid to scan the shape and acquire information, either a rectangular grid [88] or a circular raster sampling [55, 111]. Skeleton methods exploit the idea of eliminating redundant information

by reducing shapes to its medial axis.

Skeletons are one-pixel structures that presume to describe shapes by minimizing the number of necessary pixels. The definition of a skeleton is very ambiguous and usually has not a unified description for the different processes followed to generate it. The process to reduce a region to those minimum number of pixels that are essential to represent its shape is know as thinning. The definition of a skeleton and the description of the first thinning process is commonly taken at Blum's medial axis function (MAF) [28]. The medial axis is the set of points for which there exists more than one closest point in the shape's boundary. One way to implement this approach to thinning is to compute the region distances to the boundaries. The selection of the distance function has an impact on the skeleton result and arises several problems as, for example, the missing of pixels when using Euclidean distance.

The many different approaches in skeleton computation can be grouped in 4 categories: iterative methods, contour-based methods, polygon-based methods and force-based methods [110].

Iterative methods are based on a process of peeling away the outer pixel layers of a shape until any layer can be remove anymore. These techniques state a set of rules to define the process to remove pixels. The iterative process marks and deletes pixels and usually includes some kind of template matching [146, 160, 165]. The rules must ensure endpoint preserving and connectedness. These methods usually suggest the use of preprocessing stages to minimize some of the usual artifacts in skeletons [63]. Contour-based methods locate the entire shape contour and delete all pixels that are not necessary to ensure connectedness [112, 82]. These algorithms are supposed to provide faster execution speeds than iterative methods. Polygon-based methods consider the shapes as polygons and use their geometric properties to locate the skeleton [92]. In this category we find triangulation methods, which take advantage of the so-called computer graphics approach and coverts the polygon into triangles to extract its skeleton [104, 96]. These group of methods can have some difficulties related to the starting point taken when considering the polygon. Force-based methods exploit the idea of considering a repulsive *force* exerted on the object pixels. The nearer the pixel to the boundary, the greater the force it suffers. The skeleton lies on the pixels suffering forces from opposite directions [30].

5.2 GRAID-based skeletonization

The use of skeletons to represent branching patterns like blood vessels seems to arise as good possibility. Given their topological structure, branching patterns are clearly described by their junction distribution, which is basically a description of their topological structure. We propose a novel skeleton method which ensures the preservation of the branching characteristics. This skeleton method is a straightforward consequence of GRADE junction localization algorithm (chapter 4) and exploits Bounded Tangency condition (BT) and Shortest Branch condition (SB) for its implementation. Consequently, the resulting skeleton will minimize the loss of branching information.

5.2.1 Methodology

As mentioned in the previous section, skeletons can be defined by MAF introduced by Blum [28]. However, the implementation of such approach must achieve some approximations that may cause some undesired result in the computed skeleton. The MAF defines a function to guide the expansion of a wave front that starts at the boundaries. The wave front expands to neighbor pixels until two waves meet in a single pixel. The meeting points describe the

medial axis and conform the skeleton of the original shape. However, the implementation of this approach have been traditionally pointed out as having some drawbacks. The resulting skeletons can not ensure connectivity, spurious branches and junctions may appear and endpoints can be modified.

We present a methodology to compute skeletons from binary structures that is based on GRAID junction algorithm. This new algorithm prioritizes the preservation of the topology of the branching structure is prioritized. To accomplish that, the resulting skeleton must preserve the junctions and the endpoints of the original shape.

Our new methodology can described in 3 main steps: 1) distance computation, 2) maximal inscribed circumference extraction, and 3) branch expansion. Distance computation is based on a two scan Fast Euclidian distance transformation (EDT) [139]. This approach to distance transformation allows to compute the Euclidean distance function efficiently by recording the relative x- and y-coordinates in only two raster scans of image under a 3×3 neighborhood.

The second stage extracts the pixels where a circumference maximally inscribed in the shape can be located, given the distance function. This stage is based on the BT condition already introduced in Section 4.4. As defined by Eq. 4.1, the number of tangencies between the maximal inscribed circumference and the pattern contour is equivalent to the amount of branch candidates. In the case of junction extraction, candidates are selected when the amount of tangencies is higher than 3. In the case of skeleton extraction, all candidates with 2 or more tangencies will be forwarded to next stages. Therefore, the implementation in the case of the skeleton computation is the same as already described in Section 4.5 by Eq's 4.3 and 4.4.

The last stage grows conforms the skeleton by analyzing selected candidates on the basis of SB condition, defined by Eq. 4.2. As introduced in Section 4.5, this task is achieved in two steps. The first step expands the candidates to its maximal circumference and the second step expands the circumference boundaries along the branches to verify SB condition. Both cases follow the implementation described by Algorithm 2. However, in the case of maximal circumference expansion we use EDT since no geodesic expansion within the shape is needed.

5.3 Results

Figures 5.1 and 5.2 show some results on some colonoscopy images from the COLON-VESSEL data set and on some retinal fundus images from the DRIVE dataset, respectively.

5.4 Conclusions

We have presented a novel methodology to compute skeletons from binary shapes. Our method prioritizes the description of a shape in terms of its topology by accurately preserving the junctions and endpoints in the original structure. To achieve that goal, the algorithm is based on the GRAID junction localization algorithm and exploits Bounded Tangency and Shortest Branch conditions.

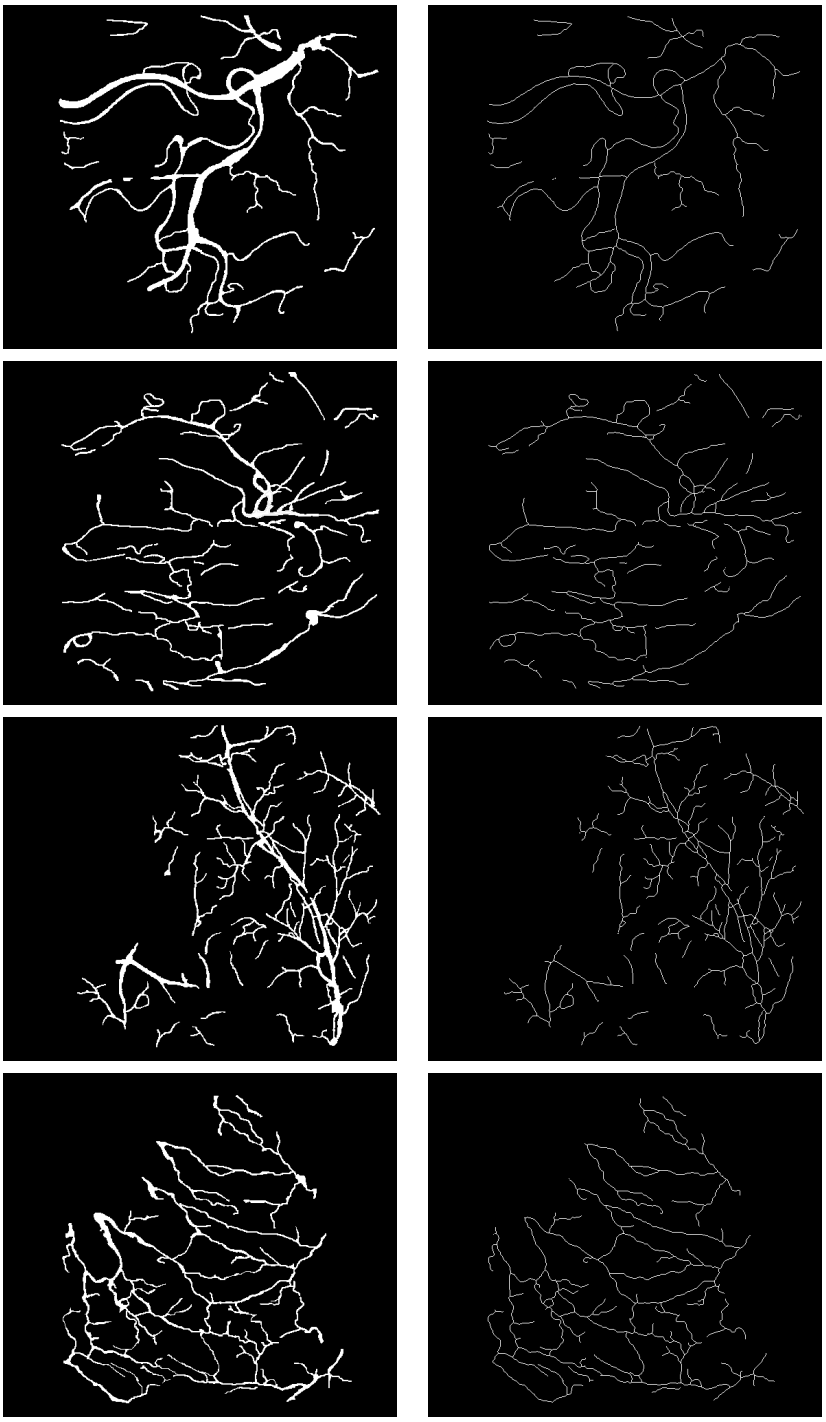


Figure 5.1: Skeleton examples on colonoscopy images. First row: Original image. Second row: Skeleton.

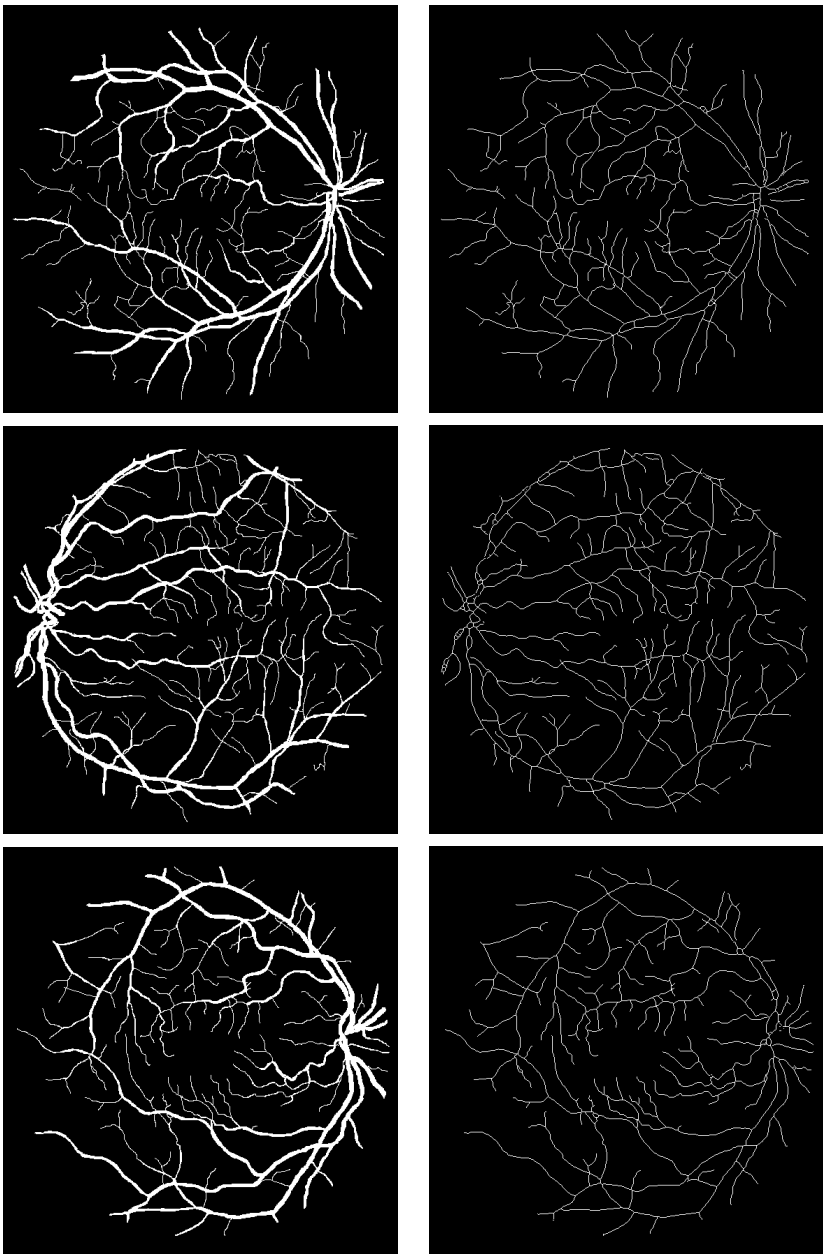


Figure 5.2: Skeleton examples on fundus retinal images. First row: Original image. Second row: Skeleton.

Chapter 6

Graph-based characterization of vessel patterns from keypoints

6.1 Introduction

One of the potential applications of computer-aided diagnosis methods (CAD) for colonoscopy is the development of patient's follow-up methods to allow the recognition of a single area of the colon containing a lesion when that area is revisited. Consequently we need to define and characterize those anatomical structures that remain stable over time in order to use them as markers to guide these follow-up methods. The only elements of the endoluminal scene whose appearance tend to keep stable are blood vessels and polyps -if they are not removed during biopsy-, as the appearance of structures such as folds is more prone to change.

In this chapter we focus on the characterization of blood vessels as anatomical landmarks with potential to be used as part of the development of follow-up methods. Blood vessels can be seen as branching patterns. The characterization of branching patterns has been deeply studied in other research fields, since the presence of these kind of structures is very common either in nature -palm prints [117]- or in anatomical structures -vascular trees [13]-. An accurate detection of these patterns along with a proper characterization of the network properties play a key role for applications using this kind of information. The segmentation of vessel patterns can be a difficult task though, given the nature of the procedure or image quality issues, such as resolution. However, keypoints in the pattern can be used to unequivocally characterize branching structures without the necessity of computing an accurate segmentation of the vessel pattern. These keypoints can potentially be identified as anatomical landmarks to be used in image registration methods.

Therefore, an accurate detection of these keypoints appears as a key stage for a good performance of these methods. There are basically two kinds of keypoints: junctions/intersections and endpoints. The detection of vascular intersections has mainly been studied in the field of retinal imaging. Available methods have been separated in two categories: *geometrical feature-based* and *model-based* methods [25, 13]. The first category groups methods which commonly start by a pixel-level processing stage followed by post-processing analysis specific for each methods. Methods belonging to this group tend to involve adaptive filtering and branch analysis based on thinned structures -being thinning a common step in the majority of available methods and an important source of error-. Some examples of *geometrical feature-based* methods appear in the works of Bhuiyan et al. [25] or Saha et al. [132].

Regarding the second category, *model-based* methods are based on a vectorial tracing of the desired structure. These methods use seed points as initial locations so vascular structures that appear in the image can be recursively tracked from them. Examples of this can be found in [35, 153]. Finally, the method of [13] offers a completely different approach which is based on the definition of COSFIRE filters to detect intersections in retinal images.

At this point, one question arises: which is the minimal performance these methods have to achieve so that those keypoints can be used as reliable anatomical landmarks. In other words, if we characterize blood vessels in terms of intersections and/or endpoints, how many of them could be missed and still be able to recognize a posterior appearance of the same structure in a different frame. In this paper we: 1) Assess the suitability of vessels on the colon wall as anatomical markers, and 2) We study the amount of keypoints allowed to be missed in order to still have a correct matching for a given vascular pattern.

To do this, we represent blood vessels using graphs. Graphs consist of a finite set of nodes connected by edges and they are one of the most general data structure in computer science. Due to the ability of graphs to represent properties of entities and binary relations at the same time, a growing interest in graph-based object representation can be observed in various fields. In bio and chemoinformatics, for instance, graph based representations are intensively used [89]. Further areas of research where graph based representations draw attention are web content and data mining [134], image classification [59], object categorization [49], shape analysis [113, 137], and graphical symbol and character recognition [87], among others.

We use a graph representation in which nodes can be either junctions or endpoints. Our validation will assess the similarity -using graph edit distance- between the original graph and an altered version created by removing some nodes from the original graph (simulating that those keypoints are not detected). We do not intend to propose the best graph matching framework but to assess the descriptive power of vessels. The validation is performed on a database of 40 colonoscopy images specially rich in blood vessel content.

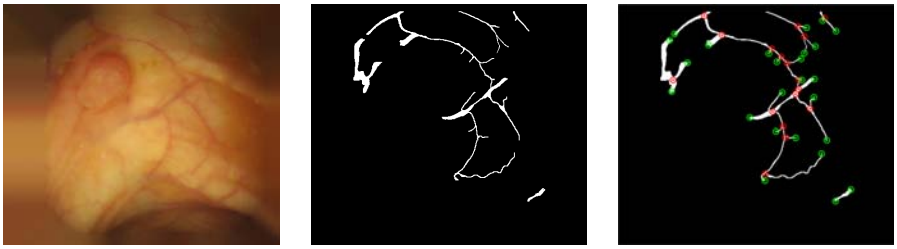
6.2 Vascular structures in colonoscopy images

Blood vessels appear as vascular structures in colonoscopy images. They can be seen as branching patterns and different keypoints can be used to help in their unequivocal characterization. In our case we define two different keypoints associated to blood vessels: junctions and endpoints. The definition of this keypoints can be better understood by looking at a binary segmentation of blood vessels. Considering this, a single pixel is labelled as an junction if it was identified as the point where at least three of the branches meet together. Conversely a pixel is labelled as an endpoint if it constitutes the end of a given branch and not reaching the boundary of informative area of the colonoscopy image.

Figure 6.1 shows an example of manually labelled keypoints in a vessel pattern. Once the keypoints are defined and characterized for each colonoscopy image with blood vessels we can proceed with the extraction of a graph.

6.3 Graph matching strategy

Graphs are used in many scientific fields as a powerful structural representation of objects. A graph is a mathematical structure for representing relationships which consists of a set of nodes connected by edges. This definition clarifies the great representative potential of graphs. A node could be a state, a condition or a location identified by features of any kind and the edges can represent the relations between nodes from a wide variety of interpretations.



(a) (b) (c)

Figure 6.1: Keypoint definition in colonoscopy images. (a) Original image. (b) Binary representation of blood vessels. (c) Keypoints: junctions (red), endpoints (green).

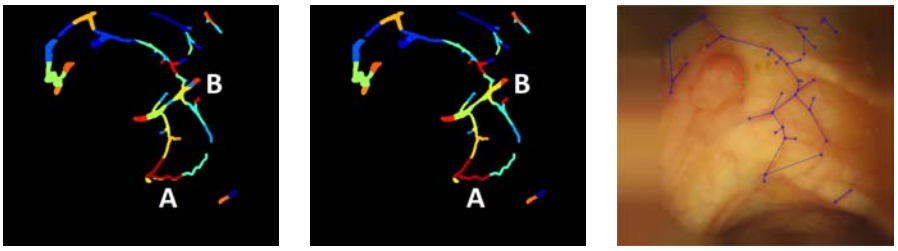
A graph matching strategy involves the construction of graph models of objects and the graph matching process. The graph models must be constructed from the given data and consist of a characterization of its nodes and a description of the relationships between them. The construction of object graphs is closely linked to the nature of the problem. The graph matching step must provide a matching environment which trusts on some kind of measure between graphs -similarity, dissimilarity, distance,...- able to provide a quantitative comparison rule between the graphs in the given data scenario. The graph matching process must find a correspondence between the nodes and edges of the graph while satisfying some constraints.

The kind of blood vessels we are dealing with contains high structural information. Junctions and endpoints and the way they relate to other points appear to be crucial for blood vessel characterization. For this reason, we use a graph matching framework to assess the impact of the selection of nodes in the robustness of a graph as a characterization of a vascular pattern. To accomplish this task, we first need to transform our images into graphs and define the attributes of both the nodes and the edges. Once the graphs are constructed, a similarity measure to compare such graphs is needed. In our case we use the graph edit distance. We remark that the purpose of this study is the assessment of the consequences of losing keypoint information in the descriptive power of the resulting vessel pattern graphs.

6.3.1 Graph extraction

We extract the graph given a set of keypoints and the binary pattern to be characterized. The keypoints will stand for the nodes and the binary pattern provides the structural information to create the adjacency matrix of the graph. The computation of the adjacency matrix from the binary pattern comprises the following steps (see example in Figure 6.2):

- In order to find adjacencies between branches, a first segmentation of the graph is achieved by grouping all the pixels behind the binary pattern into clusters represented by each keypoint detected. The criteria to assign a given pixel to a cluster is the Euclidean distance -Figure 6.2a-.
- We can observe from Figure 6.2a how the first segmentation may present some incoherences as some pixels and branches might be associated to keypoints which do not have connectivity. To solve this problem, all the regions which do not contain a keypoint are merged into any of the regions that are connected to them -Figure 6.2b-.
- The final graph is the region adjacency graph extracted from the segmented image -Figure 6.2c-.



(a) (b) (c)

Figure 6.2: Example of adjacency matrix calculation to create the final graph. (a) First segmentation. (b) Refined segmentation. (c) Final graph. Areas marked with A and B exemplify problems related with the first graph segmentation.

The distance of each node to the centroid of the vessel pattern is associated as the only attribute of each graph node. The distances are normalized with respect to the highest distance. This makes this simple characterization invariant to translation, rotation and scale. This decisions intend to lead to an experiment framework to validate the keypoint impact on vessel patterns as markers. The best configuration for the graph should be studied in future clinical works.

6.3.2 Graph edit distance

Graph edit distance [133, 31] is one of the most flexible and versatile approaches to error-tolerant graph matching. One of the major advantages of graph edit distance is that it is able to cope with directed and undirected graphs, as well as with labeled and unlabeled graphs. If there are labels on nodes, edges, or both, no constraints on the respective label alphabets have to be considered. Moreover, through the concept of cost functions, graph edit distance can be adopted and tailored to various applications such as fingerprint classification [99], diatom identification [4], or clustering of color images [125], just to mention a few.

The major drawback of graph edit distance is its high computational complexity that restricts its applicability to graphs of rather small size. In fact, graph edit distance belongs to the family of quadratic assignment problems (QAPs), which in turn belong to the class of NP-complete problems. That is, an exact and efficient algorithm for the graph edit distance problem can not be developed unless $P = NP$. Therefore, both the development of fast approximation algorithms and the gradual improvement of established approximation schemes are important and reasonable lines of research. In recent years, a number of methods addressing the high computational complexity of graph edit distance computation have been proposed [29, 142, 75, 102].

Given two graphs, the source graph g_1 and the target graph g_2, the basic idea of graph edit distance is to transform g_1 into g_2 using some edit operations. A standard set of distortion operations is given by insertions, deletions, and substitutions of both nodes and edges. A sequence of edit operations $e_1, e_2, ...e_k$ that transforms g_1 completely into g_2 is called an edit path between g_1 and g_2. To find the most suitable edit path out of all possible edit paths between two graphs g_1 and g_2, a cost is introduced for each edit operation, measuring the strength of the corresponding operation. The edit distance of two graphs is then defined by the minimum cost edit path between two graphs. As can be thought, the cost function is highly dependent on the attributes of the nodes and edges. A different cost, specific for each problem to be solved, is applied to each of the distortion operations.

The definition of the cost usually depends on the meaning of the graphs, and prior

knowledge on the graphs' labels is often inevitable for graph edit distance to be a suitable proximity measure. However, the possibility to parametrize graph edit distance by means of a cost function crucially amounts for the versatility of this particular dissimilarity model and automatic procedures for learning the edit cost can be used [97, 98, 100, 101] to calculate the optimal costs.

We propose the use of the algorithmic framework presented in [123] which allows the approximate computation of graph edit distance in a substantially faster way than traditional methods. The basic idea of this approach is to reduce the difficult problem of graph edit distance to a linear sum assignment problem (LSAP). LSAPs basically constitute the problem of finding an optimal assignment between two independent sets of entities. There is a large amount of available polynomial algorithms for LSAPs and an interested reader can find more information in [32].

In our case we are using the Euclidean Cost Function defined as follows: for two graphs $g_1 = (V_1, E_1, \mu_1, \nu_1)$ and $g_2 = (V_2, E_2, \mu_2, \nu_2)$, where μ_1, μ_2 and ν_1, ν_2 are the sets of node and edge labels, respectively, the Euclidean cost function is defined for all nodes $u \in V_1$, $v \in V_2$ and edges $p \in E_1$ and $q \in E_2$ by:

$$\begin{aligned} c(u \to \epsilon) &= \alpha \cdot \tau_{node} \\ c(\epsilon \to v) &= \alpha \cdot \tau_{node} \\ c(u \to v) &= \alpha \cdot \|\mu_1(u) - \mu_2(v)\| \\ c(p \to \epsilon) &= (1-\alpha) \cdot \tau_{edge} \\ c(\epsilon \to q) &= (1-\alpha) \cdot \tau_{edge} \\ c(p \to q) &= (1-\alpha) \cdot \|\nu_1(p) - \nu_2(q)\| \end{aligned} \tag{6.1}$$

where $\tau_{node}, \tau_{edge} \in \mathbb{R}^+$ are non-negative parameters representing the cost of a node and edge deletion/insertion, respectively. The weighting parameter $\alpha \in [0, 1]$ controls whether the edit operation cost on the nodes or on the edges is more important.

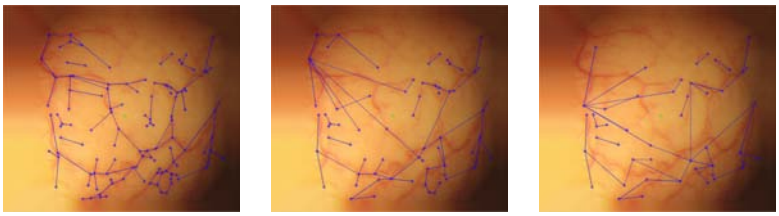
6.4 Results

6.4.1 Experimental setup

We run several experiments on COLON-VESSEL data set (see Section 3.1.1) to assess the degree of robustness of blood vessel representation using graphs. The graphs are created from the ground truth provided by experts. From the original image graph, we progressively and randomly eliminate junctions or endpoints -and the corresponding edges converging to them-. A query consisting of the identification of the altered graph over the set of the original graphs for each image is performed. The experiment removes a certain percentage of keypoints from the original graph (Figure 6.3 shows an example) before they are compared in terms of graph edit distance. This proposed graph matching framework will provide the closest graph among the data set for each query frame. Hence the system will always provide a matching -correct or not-. We do not address this issue as our goal is to assess the descriptive power of keypoint graphs rather than proposing a real solution to the frame matching problem, which should be addressed in the future.

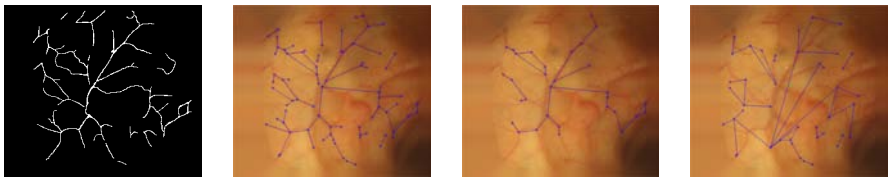
For the particular case of blood vessel structure representation using graphs we define the following cost function:

- Node deletion/insertion, $c(u \to \epsilon)$, $c(\epsilon \to v)$: cost is a constant value equal to 0.9.
- Node substitution, $c(u \to v)$: absolute value of the difference between distances to the centroid.
- Edge deletion/insertion, $c(p \to \epsilon)$, $c(\epsilon \to q)$: cost is a constant value equal to 1.7.



(a) (b) (c)

Figure 6.3: Graph modification examples (note that the removed keypoints are different as they are removed randomly). (a) Original graph. (b) Graph after 40 nodes removal. (c) Graph after 70 nodes removal.



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

Figure 6.4: Graph extraction examples. (a) Binary pattern. (b) Graph created with junctions and endpoints on original image. (c) Graph created with junctions on original image. (d) Graph created with endpoints on original image.

- Edge substitution, $c(p \rightarrow q)$: cost is 0.

For the parameter setting we apply the well documented values in the literature [122, 124], and which appear to provide good results for our particular the data set, being the deep study about parameter optimization out of the scope of this paper but a necessary step for a potential working prototype.

The removal of keypoints is carried out in a 5% step -which entails 20 intervals- and each step is repeated 10 times to provide statistically significant results regarding the impact of the removal of keypoints. We run three different experiments, regarding the keypoints we used -junctions, endpoints and both junctions and endpoints- to assess the descriptive power of each possibility (see examples in Figure 6.4).

6.4.2 Experimental results

Figure 6.5 shows two results for the three experiments introduced above.

Regarding junctions, we observe in Figure 6.5a that a 100% matching between incomplete graphs is possible if a given intersection detector is able to achieve less than 20% missing error. Matching success decreases when we remove more than 20% of the intersections in the image, reaching very low matching performance when removal reaches 70%. As mentioned above, we do not obtain a 0% matching because the experiment always provides the closest graphs. Hence when all the nodes are removed the similarity output will provide with a match with the graph with less nodes. Figure 6.5b presents a breakdown of the results for each image; this experiment confirms that for the majority of the images matching starts

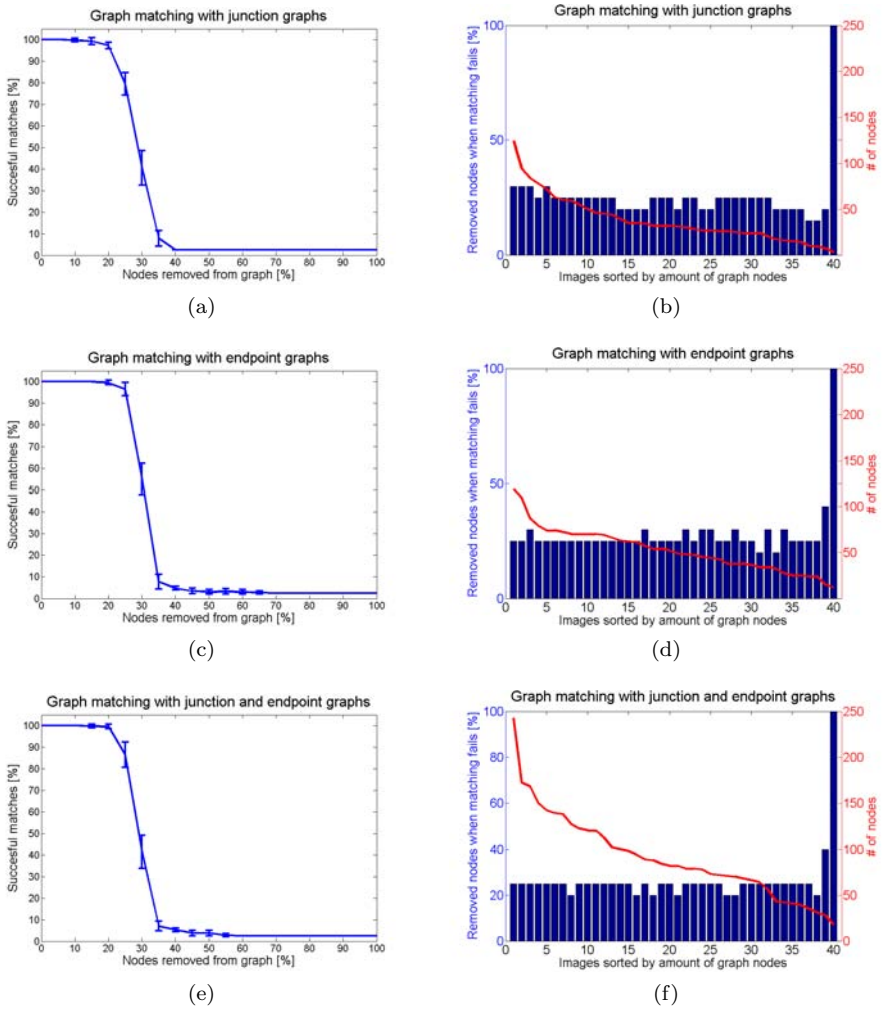


Figure 6.5: Impact of percentage of node removal in the number of correct matches. (a,b) Junctions. (c,d) Endpoints. (e,f) Junctions and endpoints.

to fail when we remove more around 25% of the junctions. Figure 6.5e shows the same behaviour when we remove randomly either a junction or a endpoint.

Regarding endpoint removal, we can observe a similar behaviour in Figure 6.5c although in this case the minimal error permitted for keypoint detectors reaches 25%. One possible reason behind this differences is that in this case nodes tend to have less edges reaching them. Results broken down per image -Figure 6.5d- show again the same trend, although in this case there seems to be a more direct relation between the number of endpoints in the image and the threshold percentage needed to lose matching precision.

6.5 Conclusions

One of the needs expressed by physicians is the ability of following-up lesions in colonoscopy procedures. One of the elements present in the endoluminal scene that can be used to help tracking these lesions are blood vessels, as their appearance tends to keep stable along different revisions of the same patient. Blood vessels can be seen as branching patterns and therefore they may be characterized by means of the position and number of branching points. Our approach for characterizing blood vessels consists of using graphs created from detected branching points in a way such their structure can be unequivocally recognized.

In this paper we present a study of the impact of the accuracy in keypoint detection in the way to develop graph matching based registration systems for colonoscopy frames. We propose a graph matching configuration whose only purpose is to provide a framework for our experiments. Experimental results show that there is a clear decrease in the success of the matching framework when a considerable amount of nodes is removed. We confirm that, when the removal of nodes keeps below certain percentages, blood vessels are still suitable as anatomical markers. This conclusion applies -with small differences- for the three kinds of graph characterization we have tested: using intersections/junctions, endpoints, or both intersections and endpoints.

Future work should consist of making further analysis regarding the impact of the connectivity of a given keypoint, that is, the relationship between missing the detection of a highly connected keypoints and the descriptive power of the associated graph. Given the trends concluded from this study, further research should be performed regarding bigger datasets. Different approaches to define the graph configuration and cost functions should be studied and experiments should also be extended to considering the several kind of image deformation that patterns can suffer in a colonoscopy procedure. As keypoints are the source of the graphs characterizing blood vessels, an accurate intersection and endpoint detector will be necessary. The degree of information in a query frame should also be evaluated to test the ability of a matching system to provide a matching for the query or to discard the request because of lack of information in the scene.

Another interesting research line would address the assessment of the descriptive power of graphs after a partial removal of node information in other kind of graphs from structured patterns.

Part III

ENDOLUMINAL SCENE ANALYSIS

Chapter 7

Blood Vessel Mitigation to improve polyp characterization

In this chapter we face the problem of blood vessels interfering in the analysis of the endoluminal scene. Vascular content hinders the extraction of scene information to locate polyps and adenomas. We will introduce a method for colonoscopy images which will allow us to separate information referring to blood vessels from scene objects related to the shape of the intestinal wall. By means of our approach we are able to make a difference between 2-dimensional objects, like blood vessels, and 3-dimensional objects, such as folds and polyps.

Our method is built on a previously published model of appearance for polyps. We follow the lines depicted in [18, 21] which pointed out the use of energy images, particularly the output of a valley detector, to make a first approach to endoluminal scene object boundaries detection. The valley information is used to generate accumulated energy maps to guide polyp localization methods. We provide a solution to mitigate the effect of blood vessels on the energy image, which shows to be useful to provide a more complete scene description. The mitigation step helps to improve the performance of the polyp localization algorithm.

We wil also present a novel segmentation method from Depth of Valley Accumulation (DOVA) Energy Maps (SDEM) algorithm for polyp localization in colonoscopy images. We work under the assumption that a faithful segmentation of the polyp region along with a exhaustive description of the polyp region could be potentially used to characterize polyps and will allow a posterior tracking of the lesion. Our segmentation method has been developed by considering the way the mentioned energy maps are calculated. We assess the performance of our method by comparing it with general and specific segmentation methods over a publicly annotated database.

7.1 Related work

As already mentioned, there are several bibliographic references devoted to the description of elements of the endoluminal scene. Regarding the scope of this paper, we can divide the published works into two different areas, namely: 1) Image enhancement and preprocessing; and 2) Polyp localization.

There are several types of artifacts associated to colonoscopy video that are a consequence of the image acquisition process, which fundamentally consists of color phantoms and specular highlights. Color phantoms are caused by a temporal misalignment of the color

channels implied by the use of monochrome CCD cameras in colonoscopy (see Section 1.1). A worsening on the quality of images is caused by the fact that RGB components are taken at different times [9, 45]. Specular highlights appear on the intestinal surface as an effect of frontal illumination, causing the apparition of highly saturated regions in the image. There are several approaches to detect and restore the surface below the specular highlights [8, 69].

Polyp localization concentrates the great majority of the bibliography devoted to intelligent systems for colonoscopy. One relevant issue, which has not received much attention, relates to the impact of the different elements of the endoluminal scene -such as folds, wrinkles and vessels- in the overall performance of the polyp localization methods.

7.1.1 Blood vessel segmentation

Up to our knowledge, besides the work presented in this chapter, the role of blood vessels in polyp localization has not received much attention in the literature. Therefore, there is not particular bibliography about vessel detection in colonoscopy videos. However, 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 [93].

- Pixel-based methods include different approaches: kernel-based methods, model-based techniques, classifier-based methods, and morphology-based strategies.
  - Kernel-based methods make use of a convolution operator with a particular kernel designed according to a model. 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 [41, 64]. These methods use Gaussian-shaped templates in different orientations and scales to identify vessel profiles.
  - Model-based techniques verify a model previously stated. An example of this approach [72] 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 what the authors called a primitive-based method [145]. 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 some examples a bayesian classifier is used after computing feature vectors obtained by Wavelet Gabor responses [140] or a neural network is used after computing a feature vector based on moment invariants-based features [91].
  - Morphology-based techniques use morphological operators to take advantage of shape characteristics of blood vessels. Morphological operators are usually combined with other techniques. Other authors used the extraction of vessel centerlines combined with local information as the vessel length is followed by an iterative vessel filling phase based on morphological filters [93]. Mathematical morphology can also be combined with curvature evaluation to differentiate vessels from other structures [162].
- Tracking-based methods aim to obtain the vasculature structure using local information to follow vessel centerlines. Tracking techniques trace vessels from selected

starting points which usually correspond to well known anatomical structures. 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 [50].

Many methods using techniques in different categories can also be found. For instance, some approaches combine a classification based on support vector machine followed by a tracking stage based on the Hessian matrix [159].

7.1.2 Polyp localization and segmentation

We propose a classification of the available works on polyp characterization according to their specific applications. The works in the literature can then be grouped into two categories: 1) polyp detection and localization, focused on determining whether there is or not a polyp in a given image and, in case of polyp presence, indicate the precise region in the image where the polyp is and 2) polyp segmentation, which aims at determining which pixels in the image correspond to the polyp.

Concerning the first category, we could divide the existing methods in two different groups according to the type of information used to characterize the polyps [19]: shape-based and texture and color-based. Shape-based localization methods propose the observation of specific cues in image contours that match to polyp usual outlines. Among this group, two main approaches are commonly used: detection by curvature analysis and detection by ellipse fitting. Polyp detection by curvature analysis intends to locate polyp-like shapes by analyzing curvature profiles of the contour information extracted from the image. Some examples belonging to this group can be found at [81, 71, 147, 19]. Polyp detection by ellipse fitting methods work under the assumption that polyps tend to present an elliptical shape and aim to fit deformable ellipses into the contour information found in the endoscopic image. Works within this group can be found at [76, 20]. There are also some methods that combine both curvature and ellipse fitting cues such as the one presented in [68].

Regarding texture and color-based methods, most of them are based on a classification framework to determine whether each pixel candidate belongs to a polyp based on the computation of a selected descriptor. Following this procedure, several state-of-the-art texture descriptors have been used such as wavelets [77], local binary patterns [158] or co-ocurrence matrices [5]. The work of [43] obtains polyp characterization by combining MPEG-7 texture and color descriptors. Other works, such as [109], propose a combination of texture and shape information to aid in polyp detection tasks.

Polyp segmentation methods in colonoscopy videos have been mainly applied for CT colonoscopy images [57] or chromoendoscopy [58]. Some simple segmentation methods have also been applied, although they are prone to be affected by noise and other image artifacts -specular highlights, image blurring- [121]. The only reported polyp segmentation method whose performance has been assessed in a public database is the one published in [18]. This paper compared the performance of a endoscopy image-focused method against general state-of-the-art segmentation methods such as Normalized Cuts [138] or Turbo Pixels [86] and proved that specific methods tend to outperform general approaches.

All the available works are focused on the characterization of polyps in the endoluminal scene but recent works [21] indicate the importance that the characterization of other elements (specular highlights, blood vessels, lumen) may have in polyp characterization. Consequently, this approach proposes to make efforts in the characterization of the whole

endoluminal scene to improve the performance of polyp detection and localization methods. Considering this, we will use the so-called MSA-DOVA localization method [21], to build up our polyp segmentation method in colonoscopy images. Our novel Segmentation from Energy Maps -SDEM- algorithm is based on the characterization of polyp boundaries in terms of valley information. SDEM also considers how MSA-DOVA energy maps integrate valley information to provide an initial segmentation of the polyp.

MSA-DOVA energy maps

MSA-DOVA energy maps are based on a model of appearance for polyps which was firstly described in [18]. This model combined information on how colonoscopy frames are acquired with the appearance of polyps in those colonoscopy frames. The model of appearance for polyps describes polyp boundaries by means of valley information.

The calculation of MSA-DOVA energy maps is based on the assumption that a pixel inside a polyp should be surrounded by valleys in several directions. The calculation of these maps is based on the use of a grid of radial sectors, S_i. Under each sector the maximum of DV image is extracted. MSA-DOVA offered an improvement over sum-based accumulation as presented in [21], using a median operator to calculate the final accumulation value. MSA-DOVA accumulation value for every image coordinates, $\vec{x}$, is calculated as follows:

$$MaxSec(\vec{x}) = \max_{S_i}\{DV\}, \quad i = 1, .., ns \tag{7.1}$$

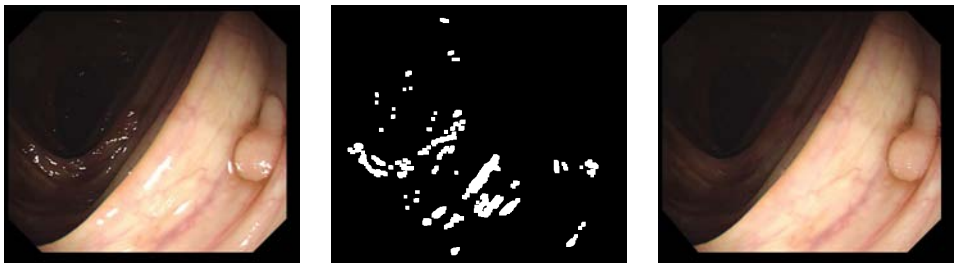
$$Acc(\vec{x}) = \underset{i}{\text{Median}}(MaxSec(\vec{x})) \tag{7.2}$$

where ns is the number of sectors in the grid. High energy regions in the accumulation map correspond with the polyp.

7.2 Feature image computation

As introduced in Section 2.5, blood vessels appear in 2-dimensional images as piecewise linear connected components. Unlike other image types, such as retinal images, the vascular structure in colonoscopy images is not connected in a fully tree-like way nor a single-root tree. The consequence of this is that spatial heuristics such as those mentioned above are not helpful in this case. Therefore, considering the previous definition, intensity valleys in a monochromatic image are a good starting point to detect the vascular structure, as confirmed by the existing related works. However, it becomes an overly broad model in the case of colonoscopy images since the endoluminal scene is made up of several objects of different nature. The problem with this simple blood vessel model is that it also matches other visual components of the endoluminal scene like boundaries of specular highlights, shades, bubble edges, colon wall folds or polyp contours.

We propose a framework adapted to MSA-DOVA localization method [21] to remove vessel information prior to polyp localization. In order to separate vessel information from the remaining anatomical structures we propose the following schema consisting of different stages, namely: 1) Image preprocessing; 2) Valley detection and 3) Vessel mitigation 4) MSA-DOVA. In the next sections we will present the proposed stage of the framework and we will also briefly introduce MSA-DOVA localization method.



(a) (b) (c)

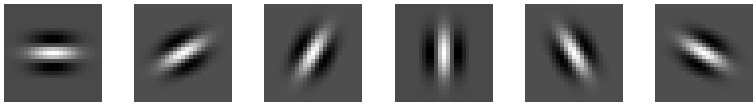
Figure 7.1: (a) Original image. (b) Specular highlights mask. (c) Output image.

7.2.1 Pre-processing

Our image preprocessing consists of two different stages: obtaining images from interlaced video and specular highlights detection and inpainting.

Interlaced video frame consists of two sub-fields taken in sequence, each sequentially scanned at odd and even lines of the image sensor [47]. All of the newer displays, contrary to previous analogic technology, are inherently digital. This means that the display comprises discrete pixels. Consequently the two fields need to be combined into a single frame, which leads to various visual defects which the deinterlacing process should try to minimise. In our case, we perform deinterlacing by taking one of each pair of horizontal lines of the image -the odd one- and then resizing it to half the original size. As mentioned in Section 7.1, specular highlights detection and inpainting is already covered in the literature. In our case, as the method is applied to the same type of images, we will use a method already available [8] to automatically segment and inpaint the specular highlights. The proposed specular highlights detection method consists of two different modules. The first one uses color balance adaptative 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). We have to take into account that specular highlights can appear only in one of the three RGB channels and, while it may suggest to apply a different threshold for each of the channels, the authors set one fixed gray scale threshold and compute the color channel thresholds using available image information. 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. After specular highlights detection is done, image inpainting is performed. Image inpainting is the process of restoring missing data in still images and usually refers to interpolation of the missing pixels using information of the surrounding neighborhood. One straightforward solution will be to replace each detected specular highlight by the average intensity on its contour. A problem with this approach is that the resulting hard transition between the inpainted regions and their surroundings may again lead to strong gradients. In order to prevent these artifacts, the inpainting is done in two levels. We use a filling technique where the image is modified by replacing all detected specular highlights by the centroid color of the pixels within a certain distance range of the outline. The second level implies the conversion of the binary mask that marks the specular regions in the image to a smooth weighting mask. The resulting integer valued weighting mask is used to blend between the original image and the smoothed filled image.

An example of specular highlight detection and inpainting can be seen in Figure 7.1.



(a) (b) (c) (d) (e) (f)

Figure 7.2: Second derivative of anisotropic gaussian (single scale σ and 6 orientations).

7.2.2 Valley detection

Our basic blood vessel model states that blood vessels appear as valleys in monochromatic images. Observation of colonoscopy images in RGB color space shows that the green component is the one that provides greater contrast between vessels and background, which agrees with the generalized idea regarding retinal images [93]. Therefore, the valley detection stage will have as input the preprocessed green component. Since vessels are described as piecewise linear connected components, different linear feature detectors appear as suitable candidates [108]. Among those detectors, we selected to use matched filters (see Appendix A for another example). It does not imply it to be the only possible solution, considering that designing a valley detector is not the aim of this study.

Blood vessels appear as darker line segments due to its lower reflectance with respect to colon walls. It prompted us to design our filter templates based on second derivatives of anisotropic Gaussian kernels. The kernel values are defined by the oriented Gaussian function described by:

$$G_{(\sigma_x,\sigma_y),\theta} = \frac{1}{(2\pi)\sigma_x\sigma_y} e^{-\left(\frac{\tilde{x}^2}{2\sigma_x^2}+\frac{\tilde{y}^2}{2\sigma_y^2}\right)} \tag{7.3}$$

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:

$$\begin{aligned} \tilde{x} &= x\cos\theta + y\sin\theta \\ \tilde{y} &= x\sin\theta - y\cos\theta \end{aligned} \tag{7.4}$$

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

$$G_{\sigma,\theta} = \frac{1}{(2\pi)2\sigma^2} e^{-\left(\frac{\tilde{x}^2}{2(2\sigma)^2}+\frac{\tilde{y}^2}{2\sigma^2}\right)} \tag{7.5}$$

Since we are modeling blood vessel profiles with second derivatives of anisotropic Gaussian kernels, the kernel will be defined as (see Figure 7.2):

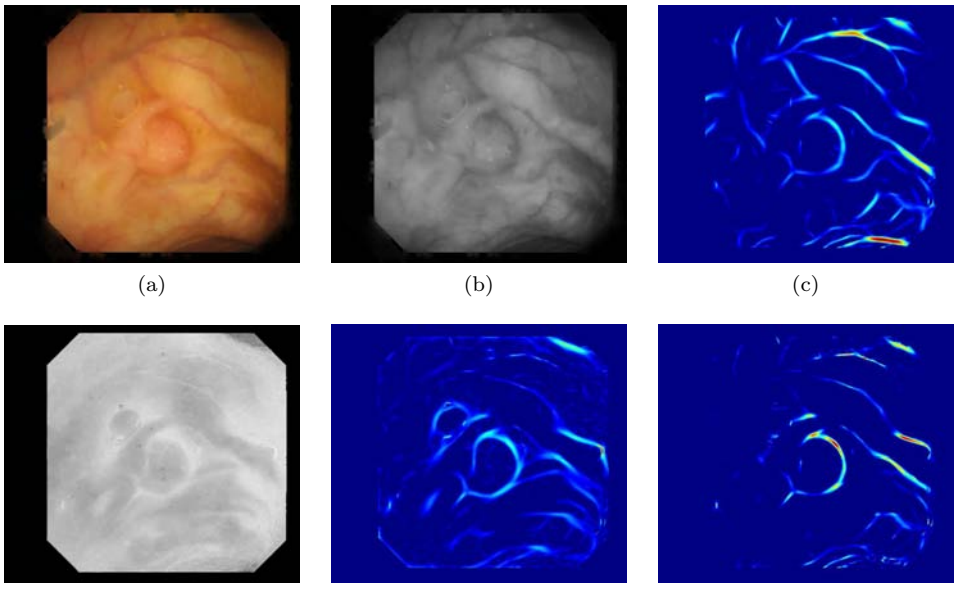
$$\partial_{\tilde{y}}^2 G_{\sigma,\theta} = \frac{\tilde{y}^2 - 1}{\sigma^4} G_{\sigma,\theta} \tag{7.6}$$

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

$$G_{\sigma,\theta}^N := \frac{\|\partial_{\tilde{y}}^2 G_{\sigma,\theta} * I\|}{\|\partial_{\tilde{y}}^2 G_{\sigma,\theta}\|\|I\|} \tag{7.7}$$

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 $\sigma = [2, 4, 6]$, which cover all vessels width in our test dataset. It all means we have 24 output images,



(a) (b) (c)

(d) (e) (f)

Figure 7.3: (a) Image after highlight removal. (b) Green component. (c) Valley energy image. (d) Saturation (HSV space). (e) Shadings energy image. (f) Final output energy image.

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:

$$I_{valleys} = \max_{i,j} \left(G^N_{\sigma^i,\theta^j} \right) \tag{7.8}$$

Prior to the valley detection method described above, structure preserving diffusion is applied in order to remove image surface irregularities while preserving image structure. This kind of feature-preserving filters are inspired in heat diffusion theory and apply iterative methods that use partial differential equations (PDEs). Diffusion is a physical process for balancing concentration changes with strong mathematical foundations. This way, smoothing filters based on diffusion interpret image intensity as that 'concentration'. Then, noise or artifacts are considered image inhomogeneities. The image intensity structure is adapted by a diffusion tensor. This tensor must be designed considering the diffusion values across structures to preserve.

Anisotropic diffusion proposed by Gil et. al. [53] will be used in this study. Non-linear anisotropic diffusion improves the adaptation of the tensor to image structure. This way diffusion values can be reduced on feature regions going around them. Anisotropic filtering will provide an output image so that the features are preserved and the noise (diffusion inhomogeneities) is reduced.

The output of this stage, $I_{valleys}$, is a gray level image in which the higher the value of a pixel, the higher the chances of that pixel to be part of a valley. See Figures 7.3a, 7.3b and 7.3c for an example of the process described so far.

7.2.3 Vessel mitigation

With regard to intensity values, both blood vessels and shadings from folds and wrinkles appear as elongated regions which are darker than intestinal walls in the background. Previous works showed that shadings from all the endoluminal structures can be approximated by the Phong's model [18], which includes ambient, diffused and specular components. Specular reflections have a spectral distribution nearly the same as the incident light but the diffuse component depends also on the object properties [136]. The ambient component is a non-directional source that groups environmental inter-reflections [27]. The resulting color of a given region in the endoluminal scene will depend on the orientation of the light source, which is coupled to the camera with its same orientation. In that sense, the dark areas created by folds are never oriented to the light source, and thus the nature of their color is conditioned by this orientation. Local variations of surface orientation in folds affect to the components in a different way. Specular reflection contribution decreases more quickly than the diffuse component in regions not oriented to the camera. Besides, since the diffuse component depends on the surface reflective properties and the surface orientations, regions which are not oriented to the camera, such as parts of folds and wrinkles, will appear as more saturated in color. In these regions the contribution of the specular component is lower, and the diffuse component will contribute to a higher saturation in color. Conversely, blood vessels are flat visual features that can be found in regions with any kind of orientation so that the nature of their color is not affected differently than the surrounding areas.

These considerations about the nature of the objects in the endoluminal scene based on its illumination led us to explore HSV color space [73], since it decouples the intensity of the image -which conveys no discriminative power between vessels and shadings- from its chromatic components. In HSV space H, S and V stand for hue, saturation and value, respectively. Hue is associated with the dominant wavelength in the color spectrum. Saturation refers to the amount of white light mixed with that dominant wavelength and it is defined as:

$$S = \frac{\max(r,g,b) - \min(r,g,b)}{\max(r,g,b)} = 1 - \frac{\min(r,g,b)}{\max(r,g,b)} \tag{7.9}$$

Assuming that colon wall properties remain unchanged at folds and wrinkles, the different color they show is to be related exclusively to changes in the illuminant contribution. Therefore, fold/wrinkle regions will have higher levels of saturation than the neighboring colon wall regions. An exhaustive test on our test dataset confirmed that color-saturation levels in vessel regions appear to be systematically closer to the levels of the adjoining background intestinal walls (see an example in Figure 7.3d).

Fold/wrinkle regions can then be described as piecewise linear connected regions in the saturation component image. An energy image describing the presence of folds and wrinkles in the scene, $I_{shadings}$, can be computed using the techniques exposed in Section **??** taking the complementary of saturation as input. Figure 7.3e shows an example of this result. The final output image,I_{out}, (see Figure 7.3f) will be computed as:

$$I_{out}(x,y) = I_{valleys}(x,y) I_{shadings}(x,y) \tag{7.10}$$

This resulting image is expected to enhance shadings from folds, wrinkles and polyps while mitigating blood vessels. When vessel mitigation is included in the polyp localization framework, DV image from Eq. 7.1.2 is substituted by $I_{out}(x,y)$, as obtained from Eq. 7.10.

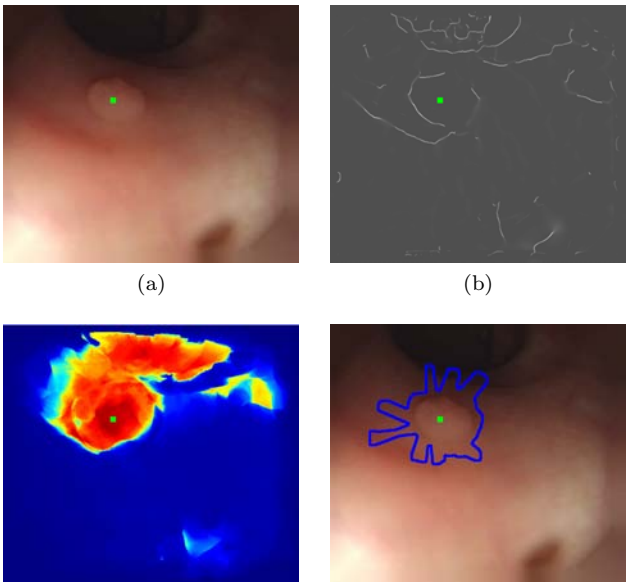


Figure 7.4: Examples of polyp segmentation from the output of an energy map: (a) Original image. (b) *DV* image. (c) *Acc* energy map. (d) Initial segmentation obtained by joining the position of the pixels that contributed to the maximum of MSA-DOVA accumulation image. Maximum of MSA-DOVA energy map is marked as a green square.

7.3 Polyp Segmentation method

Our polyp segmentation method -SDEM- uses information from both *DV* image and the way *Acc* energy maps from MSA-DOVA are calculated. Our method requires that maximum of MSA-DOVA maps falls within the polyp (an example of the output of MSA-DOVA energy maps is shown in Figure 7.4c). In that case we can obtain a first segmentation of the polyp by joining the position of the pixels that contributed to this maximum (Figures 7.4b and 7.4d).

This first segmentation may present irregularities due to several reasons, such as the presence of spurious valleys from other structures in the scene or the lack of contour information from the polyp (Figure 7.4b). These irregularities may cause positions of maxima in *DV* image for adjacent sectors not to be close to one another (Figure 7.5a).

Our goal is to eliminate the irregularities in order to have a continuous and locally circular boundary -typically associated to polyps- as the contour of the final segmentation. Our method locally explores distances from maxima under each sector to the coordinates of the maximum of accumulation, $\vec{c}^{max}$, extracted as the maximum from *Acc* energy map. This way we identify the sector maxima that are far from the circumference that represents the median of the distances from each maximum to the accumulation center (Figure 7.5b). We use the median distance as a way to correct irregular positions in favor of the most common distance value within a given neighborhood of positions. In this case the use of other options such as mean value is not suitable as the contribution of irregular positions have been proven to still have an strong influence in the performance. The positions of the pixels identified

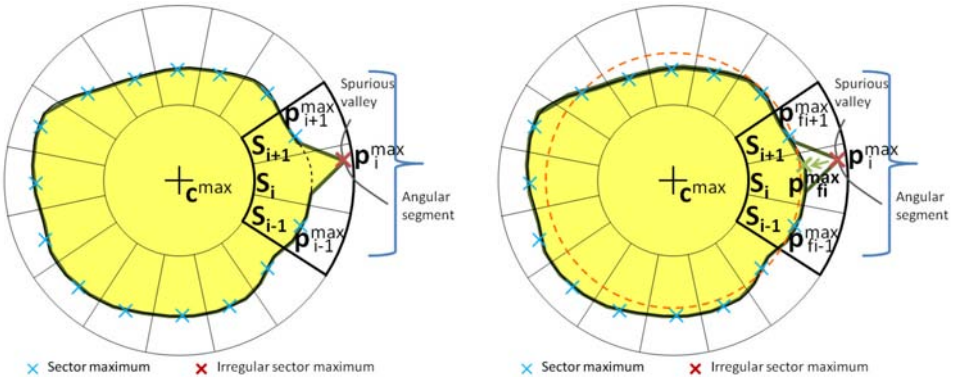


Figure 7.5: Graphical illustration of SDEM algorithm. Sector maxima are labeled as blue crosses. The original position showing irregularity and the corrected position are labeled as a red cross and a green cross, respectively. The median of distances to c^{max} is shown by the red discontinuous circumference.

as source of irregularities are corrected to have similar distances to $\vec{c}^{max}$. SDEM consists of the following steps:

1. Extraction of the position of the coordinates of the maximum of *Acc* energy map: $\vec{c}^{max} \in Image \mid Acc(\vec{c}^{max}) \geq Acc(\vec{x})\ \forall \vec{x} \in Image$.

2. Definition of a grid of ns radial sectors centered in $\vec{c}^{max}$.

3. Computation of $\vec{p}_i^{max}$, the position of the maximum of DV image under each sector S_i, as $\vec{p}_i^{max} \in S_i \mid \forall \vec{k} \in S_i, DV(\vec{p}_i^{max}) \geq DV(\vec{k})$, for $i = 1, .., ns$.

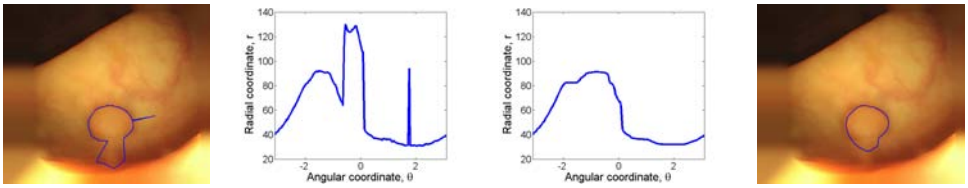
4. Conversion of $\vec{p}_i^{max}$ to polar domain: $\vec{q}_i^{max} = [r_i^{max}, \theta_i^{max}]$, where r stands for the radial coordinate and θ for the angular coordinate.

5. Computation of the new radial coordinate, r_i^{sdem}, as the median of the r_j values in an angular interval: $r_i^{sdem} = \underset{j}{\text{Median}}(r_j^{max}), j \in [i - \omega_s, i + \omega_s]$, where ω_s defines the integration angular interval centered on S_i.

6. Definition of the new polar coordinates as $\vec{q}_i^{sdem} = [r_i^{sdem}, \theta_i^{max}]$

7. Revert the conversion to cartesian coordinates to obtain the corrected position of the maximum under each sector, $\vec{p}_i^{sdem}$.

SDEM algorithm has one single parameter ω_s which sets up the size of the integration angular interval. MSA-DOVA parameters -minimum radii -*radmin*-, maximum radii -*radmax*- and the number of sectors -*ns*- are set to the values published in the original paper ($radmin = 25$,$radmax = 135$ and $ns = 180$). Figure 7.6 illustrates SDEM methodology with a qualitative example of segmentation.



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

Figure 7.6: SDEM illustrative example. (a) Preprocessed image and initial segmentation, (b) Polar representation of the initial segmentation, (c) Polar representation of corrected segmentation, (d) Preprocessed image and corrected segmentation.

7.4 Vessel mitigation experiments

7.4.1 Experimental Setup

Several experiments were developed to assess quantitatively the performance of our method on mitigating blood vessels. Both CVC-ColonDB and COLON-VESSEL-TRACE data sets have been used in the experiments (see Chapter 3.1.2). More specifically, we want to compare the energy corresponding to blood vessels in both the valley energy image and the valley energy image after blood vessel mitigation.

Since our vessel masks in COLON-VESSEL-TRACE data set have been created only as descriptors of its trace without any width information, we dilated the masks of blood vessels using morphological operators to provide us with a region of blood vessel influence. It allows us to separate the energy in blood vessel regions from the energy in non-vessel regions. Given L_v as the vessel mask and $\oplus$ as the dilation operator, vessel energy, E_v, in the considered energy image I is defined as:

$$E_v = \frac{\sum_{(x,y)\in I} I(x,y)(L_v \oplus S_r)(x,y)}{\sum_{(x,y)\in I} I(x,y)} * 100 \tag{7.11}$$

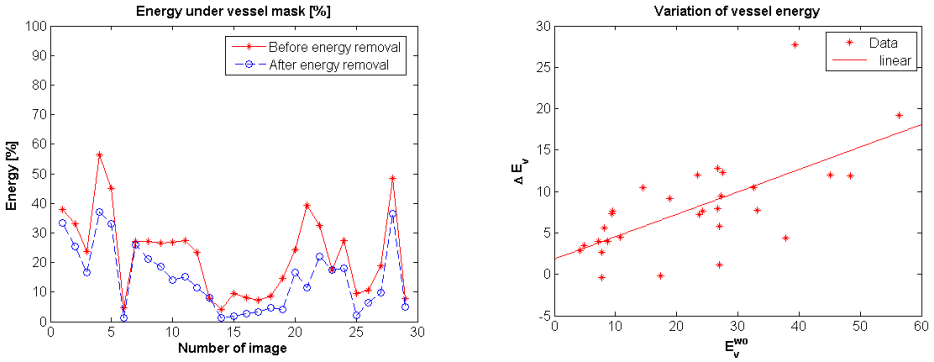
Consequently, the total energy in an image, E_{total}, will satisfy:

$$E_{total} = E_v + E_{nv} = 100; \tag{7.12}$$

which describes the balancing of energy between vessels and non-vessels as a percentage of contribution.

7.4.2 Vessel mitigation results

The proposed metrics have been computed for both the valley energy images and the valley energy images after vessel energy removal. Figure 7.7a shows E_v performance metric for each image in the whole COLON-VESSEL-TRACE already introduced. The figure allows us to verify the decrease of energy in areas previously identified as blood vessels as well as the variability of that decrease. The decrease of energy referred to vessels depends on the content of visual objects on the image. Images which had a high degree of vascular content prior to our processing and no folds interfering with them suffer an important decrease. Nevertheless, images whose vascular content was low or its trace is close or strongly crossed by folds do not show remarkable differences in terms of vessel energy, as expected. An example of both



(a) (b)

Figure 7.7: (a) Energy under vessel mask for each image. (b) Variation of vessel energy regarding vessel energy at input.

situations can be seen in Figure 7.8. First row in the figure shows an example where the input image has a large amount of vascular content in a clear surface not interfered with shades or folds. The example in the second row contains few vascular content and many clear folds. For this reason, folds and shades keep most of the image energy after vessel detection and vessel energy removal has less impact. Figure 7.7b plots the variation of energy under vessel masks regarding the energy under vessel mask prior to our removal step. The increase of energy, ΔE_v, is defined so that a positive value corresponds to vessel energy decrease:

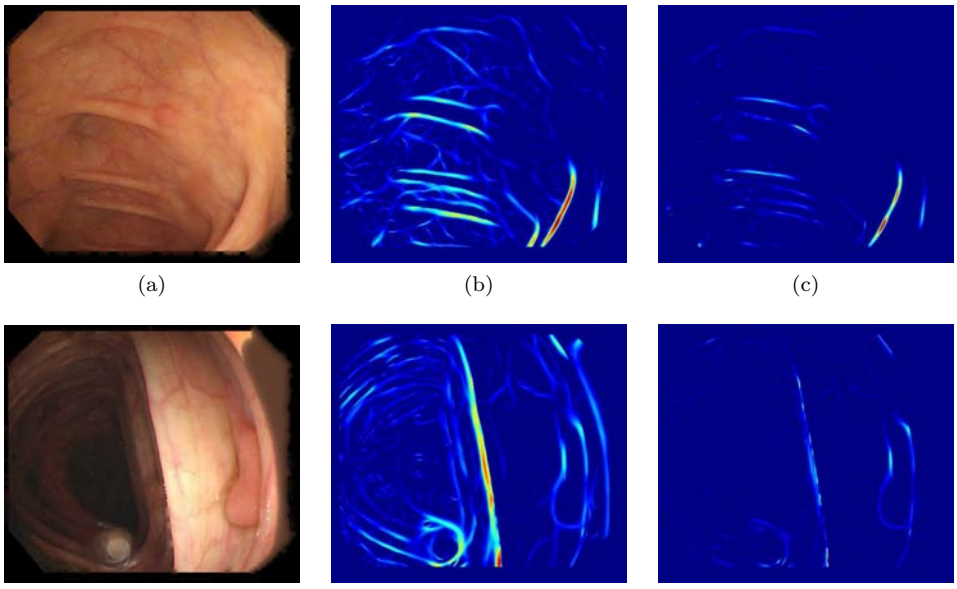
$$\Delta E_v = E_v^{wo} - E_v^{w} \tag{7.13}$$

where E_v^{wo} is the energy image without vessel mitigation and E_v^{w} is the energy image with vessel mitigation. We can see that the energy decrease is related to the energy at the beginning of the process (Pearson correlation coeficient of 0.65). The results regarding energy in non-vessel regions are the complementary of the ones presented in Figure 7.7a as stated in Equation 7.12. Therefore, we can also affirm that regions which has been manually identified as non-vessels does not suffer substantial energy decrease.

7.4.3 Application to polyp localization

Our aim is to provide a first approximation of a blood vessels characterization. This characterization could be useful to provide a better scene description, but it also shows to provide relevant information for some other applications such as helping in polyp localization. Polyp localization and the different approaches to this topic is further covered in Chapter **??**. In this section we will measure how the characterization of blood vessels could be useful to improve the polyp localization methods in the DOVA family [18, 21], which have already been tested in the CVC-ColonDB public data set (see Section 3.1.2). In our case, we will use the output of our processing scheme as the depth of valleys image, and we will measure the accumulation of energy by using the MSA-DOVA descriptor. MSA-DOVA descriptor defines an accumulation image by using data from a depth of valleys image.

Our hypothesis is that by identifying which parts of the energy image correspond to blood vessels information we could be able to mitigate their effect and check if the energy



(a) (b) (c)

(d) (e) (f)

Figure 7.8: Example 1: (a) Input image. (b) Valley energy image. (c) Energy image after energy removal). Example 2: (d) Input image. (e) Valley energy image. (f) Energy image after energy removal.

inside and outside the polyp changes. The metric that we will use in this experiments is:

$$E_p = \frac{\sum_{(x,y)\in I} I(x,y)L_p(x,y)}{\sum_{(x,y)\in I} I(x,y)} * 100 \tag{7.14}$$

where L_p is the polyp mask from the Polyp Dataset. We measure the percentage of energy inside the polyp mask whereas the energy outside the polyp will be the complementary. We expect that a polyp localization decision scheme based on the amount of energy concentrated on some area of the image will benefit from a blood vessel mitigation system which reduces the presence of vessel energy. We measure the increment of energy inside the polyp mask as:

$$\Delta E_p = E_p^w - E_p^{wo} \tag{7.15}$$

where E_p^{wo} stands for the energy image without vessel mitigation and E_p^w stands for the energy image with vessel mitigation. That difference of energy, ΔE_p, referred to E_p^{wo} is plotted in Figure 7.9. Table 7.1 shows that we improve the amount of energy inside the polyp mask in a large majority of images (217). This is true even considering that an increase or decrease lower than a 5% can be assumed as not significant (74 improved images).

To conclude with this section, we will show the direct impact that blood vessels mitigation has on polyp localization. In this case we will use the same polyp localization criteria than the one depicted in [18], that is, measuring if the maxima of the accumulation image is

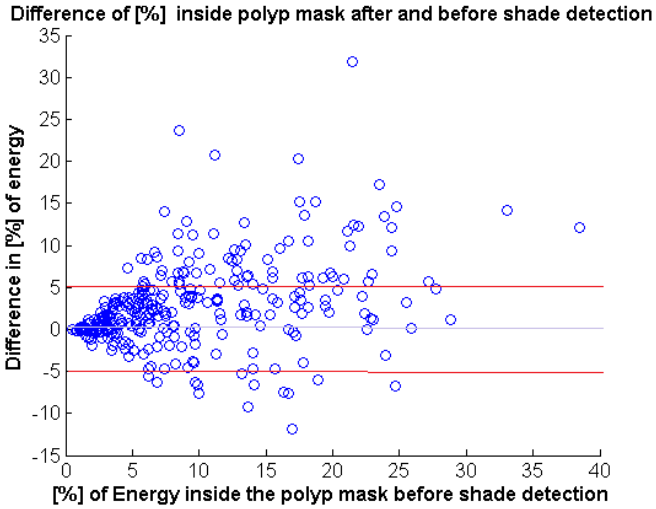


Figure 7.9: Difference of percentage of energy under polyp mask before and after blood vessel energy mitigation.

Table 7.1: Difference of percentage of energy under polyp mask with and without blood vessel mitigation.

ΔE_p	# of images
> 0	217
$> 5\%$	74
$< -5\%$	13

Table 7.2: Polyp localization results (placing accumulation maxima inside polyp mask): comparing results using vessel mitigation with no vessel mitigation.

	# of images	Polyp Dataset %
improved	59	19.67%
worse	17	5.67%
same	224	74.67%

placed inside the polyp mask. We can see a comparison between the results before and after applying our blood vessel energy mitigation in Table 7.2.

As we can see from Table 7.2, by applying our blood vessel mitigation algorithm the maxima of the accumulation image is placed inside the polyp mask in 59 more images (19.67%), the results were worse for 17 (5.67%), and no modification took place for 224 images (74.67%). This study shows that blood vessel mitigation can be a key part in the improvement of a polyp localization scheme, as it does have an impact on direct polyp localization results. Finally we show in Figure 7.10 some qualitative results of the comparison

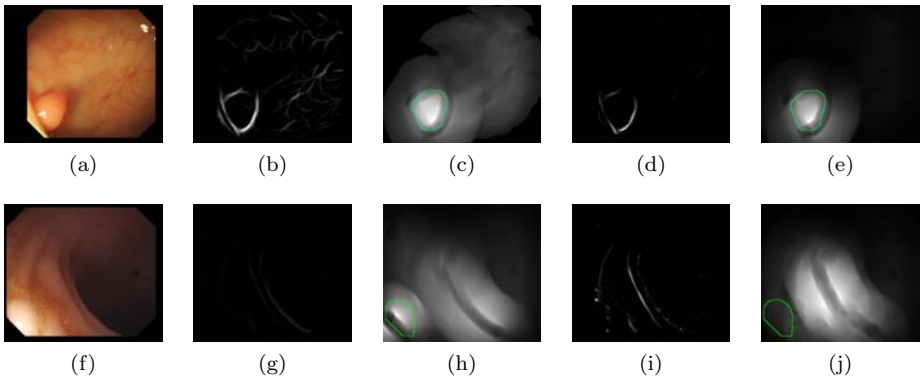


Figure 7.10: Example 1: (a) Original images. (b) Energy image before shade detection. (c) Accumulation image before shade detection. (d) Energy image after shade detection. (e) Accumulation image after shade detection (polyp region marked in green). Example 2: (f) Original images. (g) Energy image before shade detection. (h) Accumulation image before shade detection. (i) Energy image after shade detection. (j) Accumulation image after shade detection (polyp region marked in green).

of the accumulation images before and after applying our processing scheme. The first row shows a positive example, where the percentage of energy inside the polyp grows after applying vessel mitigation whereas the second row shows a negative example.

7.5 Polyp segmentation experiments

7.5.1 Experimental Setup

In order to assess the performance of our proposal, we will compare our method to the following both general and endoscopic-specific methods:

- Normalized Cuts (NCuts): The *normalized cuts* method [138] is a graph theoretic approach for solving the perceptual grouping problem in vision in which every set of points lying in the feature space is represented as a weighted, undirected graph. Segmentation is performed by disconnecting edges with small weights.
- Turbo pixels (TurPix): this algorithm [86] starts by computing a dense over segmentation of an image by means of a geometric-flow-based algorithm. This segmentation respects local image boundaries while limiting under segmentation by using a compactness constraint. Regions are refined by using criteria such as size uniformity, connectivity or compactness.
- Watershed with markers (WSM): watershed segmentation [18] considers a grayscale image as a topographic surface and achieves the segmentation by a process of "filling" of catchment basins from local minimums. Providing markers helps the algorithm to define the catchment basins that must be considered in the process of segmentation [166].
- Depth of Valleys (DoV)-based Region Merging Segmentation [18] (DV-RMS): this method assumes polyp boundaries to be described in terms of valley information.

Table 7.3: Segmentation results without -160 images- and with image preprocessing -203 images-.

Method	**Prec** [%]	**Sens** [%]	F_2	**Prec** [%]	**Sens** [%]	F_2
NCuts	20.29	80.27	0.50	18.02	83.84	0.48
TurPix	19.40	75.56	0.47	14.75	76.30	0.41
WSM	42.89	68.36	0.61	43.68	74.40	0.65
DoV-RMS	56.87	44.93	0.47	54.13	57.46	0.56
SDEM	69.93	69.32	**0.69**	65.07	81.22	**0.77**

The method starts from a first rough segmentation of the input image obtained by means of watershed. The segmented regions are merged using different criteria such as boundary strength and region content.

All the methods that have been compared were implemented under the same parameter configuration as described in [18]. We remark that both NCuts and TurPix need to be provided with a number of target regions nr to be extracted. After performing several segmentation tests we selected $nr = 3$ as the most representative result, considering that most colonoscopy images present three main regions, which are: 1) lumen; 2) polyp; 3) colon wall. For all the methods we used the position of maximum of MSA-DOVA to select the final polyp region. Regarding SDEM, we set $\omega_s = 20$ -corresponding to an angular segment of ± 40 deg- after a training state over 30 images not included in the validation data set.

We compare the performance of all methods over CVC-ColonDB data set (see Section 3.1.2) by means of three different metrics computed on the pixel-level segmentations. The metrics are the following:

$$Precision = 100 \cdot \frac{TP}{TP + FP} \quad Sensitivity = 100 \cdot \frac{TP}{TP + FN} \quad F_2 = \frac{5Prec \cdot Sens}{4Prec + Sens} \tag{7.16}$$

where TP, FP and FN stand for the number of True Positive, False Positive and False Negative pixels, respectively.

7.5.2 Results

The experiments were performed using as input both the original and the preprocessed image to compare its influence in the segmentation output. The preprocessing stage includes specular highlights detection and inpainting as introduced in Section 7.2.1. In both cases results are only considered for those images in which the polyp localization succeeded in order to focus on segmentation performance analysis.

Table 7.3 shows performance results for all methods. We can observe that our proposed methodology outperforms the rest of approaches, specially in terms of precision. Our method provides with regions with a higher amount of polyp content while adding less non-polyp areas. This result is confirmed by F_2-score. Our method provides with a segmentation that covers almost the 70% of the polyp region -much higher than the other methods- whereas it still keeps a reasonably high performance in terms of sensitivity. Our proposal also improves the results achieved by our most similar competitor -WSM-: segmentation guided by energy maps leads to obtain bigger final regions closer to the actual polyp region.

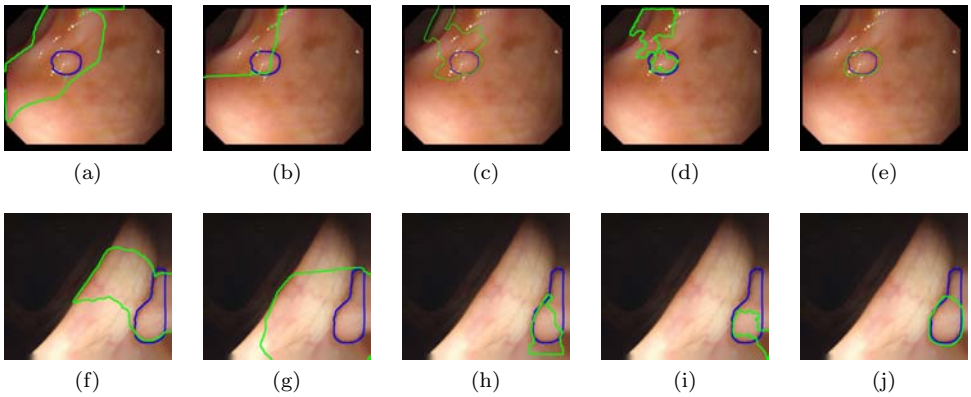


Figure 7.11: Examples of polyp segmentation results: (a-f) Normalized Cuts; (b-g) Turbo Pixels; (c-h) Watershed with MSA-DOVA markers; (d-i) PR and (e-j) Our proposal. Each image shows segmentation output (green) and polyp mask (blue). Top row shows results without image preprocessing, bottom row with image preprocessing operations applied.

Regarding the impact of image preprocessing, we also notice that our method outperforms the rest of approaches. In this case the final regions are much less prone to miss polyp regions -81.22% vs. 69.32% sensitivity values- although slightly less precise. The outcome is that preprocessing helps the procedure to minimize missing rates of polyp region extraction. Image preprocessing also has an impact in the performance of the rest of the methods. This impact follows the same trend in all cases, which means a considerable increase in terms of sensitivity and a light decrease in precision. WSM is the only case in which the precision value suffer and hardly noticeable variation. It must be remarked that preprocessing techniques have an impact in the polyp localization algorithm performance, which is reflected in an important decrease in the number of images considered in segmentation experiments. We can conclude that preprocessing has a positive impact in whole segmentation methodology.

Finally we present some qualitative results on polyp segmentation in Figure 7.11 some qualitative examples of polyp segmentation of several images before and after applying preprocessing operations.

7.6 Conclusions

In this chapter we introduced a characterization for blood vessels which allowed us to model them differently than other objects in a endoluminal scene, more specifically folds and wrinkles. We presented a procedure for mitigating blood vessels which consists of three stages: 1) Image preprocessing, to correct artifacts from the original image such as specular highlights; 2) Valley detection, to provide a first characterization of the objects in the image, and 3) Valley mitigation as a novel method which aims to discriminate between objects that have shades from objects that do not have them, such as blood vessels.

Our experiments show an encouraging trend, indicating that there is a decrease of energy on blood vessel areas. Quantitative results suggest that our method is able to achieve vessel mitigation successfully and that mitigation is more important on images with more blood vessel content. Our procedure was used to improve the only existing polyp localization

method that has been tested in a public database. As expected, the polyp localization decision scheme -based on the amount of energy concentrated on some area- benefited from a blood vessel mitigation system which reduces the presence of vessel energy. This is the first time that the impact of blood vessels in polyp localization has been measured quantitatively, proving that their presence makes it harder to identify 3-dimensional objects such as polyps.

Regarding future work, vessel characterization should be validated on a bigger manually labeled dataset. It should also involve the consideration of the superposition of blood vessels and other elements in the endoluminal scene. The appearance of vessels in folds must prompt us to add more information to the improved characterization presented in this work.

We have also presented a novel polyp segmentation method in colonoscopy videos, which is built on a general model of appearance for polyps which describes polyp boundaries using valley information. This information is integrated to generate energy maps linked with polyp presence in the image. Our method explores the way these maps are created to develop a polyp region segmentation algorithm, considering which pixels in the image contributed to the localization of the polyp. Our algorithm is able to improve an initial segmentation by adjusting the shape of the final region discarding some contributions prone to provide irregularity.

The results show that our method outperforms other general and specific segmentation methods in terms of Precision, Sensitivity and F2 measure. Our experiments also confirm the necessity of image preprocessing to improve the final segmentation of the polyp.

Chapter 8

Conclusions

Throughout the dissertation several tools and methods for the analysis of the endoluminal scene and the characterization of blood vessels have been presented. This chapter revisits the main contributions and summarizes the conclusions. Finally, future perspectives and lines of research are pointed out.

8.1 Conclusions and contributions

The following points summarize the structure and contributions of the thesis:

- In the context of colonoscopy procedure, the knowledge of the endoluminal scene can support technicians in their screening task. The impact of several aspects in the quality of colonoscopy procedure have been described in the medical literature. A better knowledge of the endoluminal scene may lead to the development of support tools for technicians. We focus on the characterization of blood vessels in the inner layer of the colon.
- Vessel characterization in colonoscopy images is a challenging task in the complex context of the endoluminal scene. The existence of several objects with similar characteristics and their interaction makes the segmentation of blood vessels in colonoscopy images a difficult
- We created two data sets of colonoscopy images: COLON-V-VESSEL and COLON-VESSEL. COLON-VESSEL includes a selection of 40 images and the corresponding manually created ground truth consisting of an accurate mask of blood vessel content. Two manually created ground truths of keypoint landmarks are also included: junctions and endpoints.
- We created a keypoint landmark ground truth for the DRIVE data set, an existing data set of retinal fundus images and the corresponding vascular content manual segmentation.
- A novel method for junction localization from binary branching patterns, such as blood vessels, have been presented. Our GRowing Algorithm for Intersection Detection (GRAID) is based on two simple geometric conditions: Bounded Tangency condition and Shortest Branch condition. The comparison of GRAID to other methods in the literature in a common framework verifies that our new method outperforms the other

approaches. We also tested the importance of localization accuracy and showed the high performance of GRAID on that point.

- We propose the use of junctions in the vascular content of the endoluminal scene, when present, to create anatomical markers to guide the navigation of endoscopists through the rectum and colon track. Junctions, representing the branching characteristics of the vascular pattern, show description abilities profitable for the development of biometric prints.
- Blood vessels in colonoscopy images can be modeled as piecewise elongated structures. Since this model also matches other elements in the endoluminal scene, such as polyp contours, we propose a methodology to inhibit the presence of blood vessel responses in valley energy images. We show the success of our methodology in both mitigating blood vessels presence in energy images and increasing the concentration of energy in polyp regions. Moreover, the methodology shows to have a positive impact in state-of-the-art polyp localization techniques.
- In the context of state-of-the-art polyp localization techniques, we propose a polyp-specific segmentation method that outperforms other general segmentation approaches. The development of an specific and local approach have been shown to be successful in providing a accurate segmentation of polyp contours.

8.2 Future perspective

This dissertation includes several contributions regarding branching patterns and, specifically, the vascular content in colonoscopy images. Blood vessel characterization as well as junction description and localization are important tools for the development of techniques to improve colonoscopy quality and provide support during the procedure. We list some ideas regarding the future perspective of this work:

- The junction localization algorithm we have introduced (GRAID) shows high performance results in terms of precision/sensitivity and location accuracy. Regarding the future use of this methodology in clinical frameworks it is important to be sure it is not computationally costly. We believe there is a margin of improvement around this matter.
- We propose the use of the vascular content in the endoluminal scene to create biometric markers. A new data set should be created to assess the graph-based markers in a framework including a higher variability of endoluminal scenes. The assessment should he assessment of vascular patterns as biometric markers should be tested on bigger data sets including more variability. The extraction of graph-based markers from junctions as representatives from the vascular content in the image offers a promising line of research to develop navigation support tools for the screening inspection.
- We have presented a method to minimize the impact of blood vessels in polyp localization techniques. However, our experiments to achieve an accurate segmentation of the vascular content in colonoscopy images have not been successful enough. An open challenge is to overcoming the difficulties -pointed out in this dissertation- of blood vessel segmentation regarding the similarities to other elements of the scene.
- The characterization and localization of binary junction structures should be extended to gray level or color images. Exist a large amount of studies in the literature regarding corner detection. The topological structures we described as junctions -including crossroad with any number of branches- are an interesting problem which would represent an important tool in several topics in Computer Vision.

List of publications

Journals

- Joan M. Núñez, Jorge Bernal, F. Javier Sánchez, and Fernando Vilariño. "GRowing Algorithm for Intersection Detection (GRAID) in branching patterns". In Machine Vision and Applications (MVAP),

International Conferences

- Joan M. Núñez, Jorge Bernal, Miquel Ferrer, and Fernando Vilariño. "Impact of Keypoint Detection on Graph-Based Characterization of Blood Vessels in Colonoscopy Videos". In the Proceedings of the 1st MICCAI International Workshop on Computer-Assisted and Robotic Endoscopy (CARE-MICCAI), pp. 22-33, 2014.
- Joan M. Núñez, Jorge Bernal, F. Javier Sánchez, and Fernando Vilariño. "Polyp Segmentation Method in Colonoscopy Videos by means of MSA-DOVA Energy Maps Calculation". In the Proceedings of the 3rd MICCAI International Workshop on Clinical Image-based Procedures: Translational Research in Medical Imaging (CLIP-MICCAI), pp. 41-49, 2014.
- Joan M. Núñez, Jorge Bernal, F. Javier Sánchez, and Fernando Vilariño. "Blood Vessel Characterization in Colonoscopy Images to Improve Polyp Localization". In the Proceedings of the 8th International Conference on Computer Vision Theory and Applications (VISAPP), Vol.1, pp. 162-171, 2013.
- Joan M. Núñez, Debora Gil, and Fernando Vilariño. "Finger Joint Characterization from X-Ray Images for Rheumatoid Arthritis Assessment". In the Proceeding of the 6th International Conference on Biomedical Electronics and Devices (BIODEVICES), pp. 288-292, 2013.

Appendix A

Valley-like feature characterization example: finger joints

A.1 Introduction to X-ray images for rheumatoid arthritis assessment

Rheumatoid arthritis (RA) is a chronic disease that causes joint dysfunction which results, among other symptoms, in the reduction of the joint space width (JSW) and the erosion of the joint bones. Since there is not any cure for RA, the assessment of the disease becomes very important and it has led to the emergence of several semi-quantitative assessment methods. Among them, the Sharp-Van der Heijde (SvdH) score, which provides separate discrete values for JSW and bone erosion based on X-ray image examination, is the most widely used nowadays [155]. The patient score is the sum of the scores for all the considered joints in hands and feet.

Three stages should be considered to face the automation of RA assessment: joint detection, joint characterization and joint measurements. As the main goal of this study is to propose and assess a JSW measure we will focus on the second and third stages taking hand joints as input (Figure A.1). This statement leads us to a hand joint modelling to describe the necessary features to compute JSW measures. Sclerosis and lower bone appear to be the main necessary features to obtain JSW measures and the corresponding detectors are proposed.

Few previous works have faced this problem. Bielecki et al. [26] developed the first study that faced the challenge of a fully automatic system for RA assessment. Vera et al. [157] provided a method which included all the different stages of the problem improving significantly the joint detection rates. Langs et al. [84] presented a new method whose scope is also the whole RA assessment process. However, these works faced the problem as a whole stage, not modular, and did not show a clear correlation among its proposed JSW score and Svdh.

Our proposal clearly separates joint detection, joint characterization and joint measurements in three independent modules. It also provides a framework for sclerosis and lower bone detection performance assessment, and proposes a JSW score which exhibits a close relation with SvdH score. This approach allows us to confirm for the first time a relation between an automatic score and the RA disease stage. Moreover, the modular orientation of our contribution, unlike previous approaches, enables future research to specifically focus

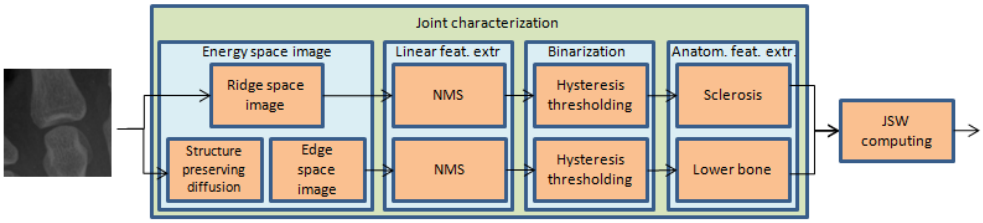


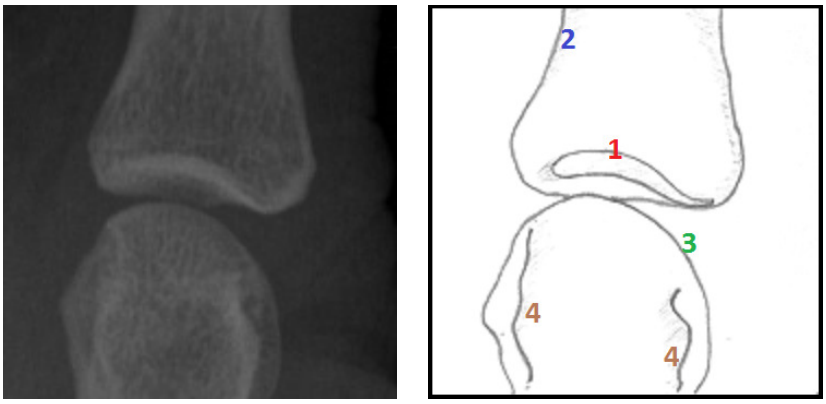
Figure A.1: Rheumatoid Arthritis characterization system.

on the improvement of the different open challenges independently.

A.2 Data analysis

This study is focused on the same digital database used by Vera et al. [157] enriched with our own manually segmented ground truth for the selected visual features necessary for JSW computation. Twenty X-ray images from different patients containing pairs of hands in posterioanterior view are available in DICOM format, 2828x2320 resolution and intensity range from 0 to 4095. Hand joints have been marked by an expert who was asked to spot the joint middle point along the finger axis and the joint orientation. A hand joint dataset containing 560 images of joints have been created (14 joints/hand). Among these samples, a total number of 320 joints were labelled with their SvdH score for inter-phalangeal distance (discrete values from 0 to 3). These 320 samples comprised 160 metacarpo-phalangeal joints (MCP) and 160 proximal inter-phalangeal joints (PIP) -distal inter-phalangeal joints (DIP) are not considered within SvdH score-. Erosion scoring was not collected because it is out of the scope of this this study.

Our proposed hand joint model does not consider thumb joints as they represent a different problem due to its specific profile. Since images are taken with the hands in posterioanterior view -from the hand back with the palm facing down- a profile view of the thumb is taken. Consequently, a frontal projection is obtained for the rest of the fingers whereas a lateral projection is obtained for the thumbs. This reason, as well as their specific skeletal structure, causes the visual features observed in thumbs to be substantially different and justifies that they are not included in this study. As far as the other finger joints are concerned, several features are distinguished within the model: sclerosis, upper bone contour, lower bone contour and lower bone inner edges, as depicted in Figure A.2 [103]. The sclerosis and the lower bone contour are the main necessary features to carry out measures on JSW. The sclerosis (feature 1 in Figure A.2b) appears as a consequence of the upper bone shape and the way an X-ray image is created. The X-ray beam have to pass through a higher density region due to the concavity of the lower part of the upper bone of the joint (feature 2 in Figure A.2b). As a consequence, the sclerosis appears as a prominent high intensity region. Regarding the lower bone, the visual features are also a consequence of the mentioned X-ray image acquisition mechanism and the lower bone shape (feature 3 in Figure A.2b). That shape can vary in the different fingers or the different hand joints and the inner edges may be present or not (feature 4 in Figure A.2b). However, the lower bone contour is defined by the outer part of the lower bone.



(a) (b)

Figure A.2: (a) Joint example. (b) Joint model schema (1: sclerosis, 2: upper bone, 3: lower bone, 4: inner edges).

A.3 Methodology

Our proposed sclerosis and lower bone detectors include several stages: energy space image, linear feature extraction, binarization and anatomical feature extraction.

The first stage aims to obtain the energy space image. The intensity levels of that image describe the chances of finding a feature in every pixel. Two separate processes are achieved in order to obtain the ridge space image and the edge space image. A second derivative of anisotropic Gaussian (DoG) method is used to compute the ridge space image.

On the other hand, the edge space image is obtained as the gradient of the input image after applying structure preserving diffusion [53]. Diffusion filtering proved its success in improving the quality of the edge detection by smoothing the image irregularities while keeping the main image structure.

The second stage performs the feature extraction using Non-maximum Suppresion algorithm [36], which only keeps pixels that are local maxima along the gradient direction. Gradients are computed using the structure tensor of the space image.

In the third stage hysteresis thresholding algorithm [36] is used to binarize the non-maximum suppressed images while preserving feature connectivity and removing weak responses.

Finally, the two images forwarded by the third stage must go through the fourth processing step in order to provide the final sclerosis and lower bone segmentations. This stage is different for the two thresholded images:

- The final sclerosis segmentation is the ridge in the binarized image that is closer to the center of the image following the finger orientation .
- As far as lower bone is concerned, the corresponding binarized image is processed to remove the edges in the upper part and the margins. Afterwards, the endpoints in the remaining processed edges are linked. The final lower bone segmentation is obtained by computing the convex hull of the linked-edge image.

A.4 Experimental setup

Two subsets were created from our dataset of 320 annotated images to perform our experiments: 1) Tune Dataset, with 40 randomly selected healthy joints (20 MCP and 20 PIP); 2) Test Dataset, with the remaining 280 joints (140 MCP and 140 PIP). The Tune Dataset was used to tune the parameters of the system for sclerosis and lower bone contour segmentation, and the Test Dataset was used to compute the output of our system and compare the proposed distance measure with the SvdH score. We added 20 DIP joints from the non-annotated dataset to enrich the variability of the Tune Dataset -these joints can be safely included because, although not having a SvdH score, they fit our model-. Both sclerosis and lower bone were manually segmented by an expert using OsiriX [127] software exclusively for the 60 images of the Tune Dataset.

Performance metrics were developed to evaluate the performance of the system, i.e. the quality of our detections, and tune the parameters. We based our metrics on the Average Surface Distance (ASD), defined as follows:

$$ASD(U,V) = \frac{1}{|S(U)|}\left(\sum_{s_U \in S(U)} d(s_U, S(V))\right) \tag{A.1}$$

where, given a pixel p and a region R conformed by a set of pixels $S(R)$, $d(p, S(R))$ is defined as:

$$d(p, S(R)) = \min_{s_R \in S(R)} \|p - S_R\| \tag{A.2}$$

where $\|.\|$ stands for the Euclidean distance.

Thus, if A denotes our automatic segmentation and M denotes the manual delineation, we define:

$$Caught = ASD(A, M) \tag{A.3}$$

$$Missed = ASD(M, A) \tag{A.4}$$

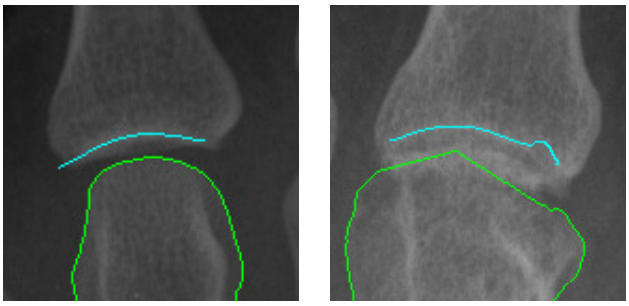
When both Caught and Missed metrics are zero the segmentation is perfect. Caught metric value is related to the quality of the detector at detecting valid pixels (true positives) and avoiding non-desired pixels (false positives). Analogously, the lower the value of Missed, the less desired information was missed (false negatives).

This experimental setup was used in order to train the detection systems with the Tune Dataset. The system providing the lowest values for the performance metrics was selected, giving priority to a lower Caught value so that the provided output is better although some parts may be missed. Afterwards, the selected detectors were applied to the Test Dataset. The final sclerosis and lower bone outputs were then visually tested in order to separate the wrong detections from those detections that can be useful to compute JSW scores. The final detections were accepted if they provided a good segmentation and there were not remaining spurs in the joint interspace.

Finally, the proposed JSW measurement is the minimum distance between the sclerosis and the lower bone. It is defined as:

$$d_{min} = \min_{p \in L} d(p, S) \tag{A.5}$$

where S and L stand for the sclerosis surface and the lower bone contour, respectively.



(a) (b)

Figure A.3: (a) MCP valid detection. (b) MCP wrong detection.

A.5 Results

Table A.1 shows the valid sclerosis and lower bone rates when applying our sclerosis detector to the Test Dataset. High detection rates have been obtained in the case of sclerosis detection, reaching higher rates in the case of healthy samples (SvdH 0), providing 96.1%, since they are closer to our joint model. However, these detection rates drop down to 58.3% in the case of joints in a more advanced RA stage, which makes clear the potential usefulness of the model in order to classify advanced stages of the disease. All in all, the detection rate reaches the remarkable value of 89.6% on the whole dataset. As far as MCP and PIP are concerned, the detection rates remain close although a light tendency is confirmed which shows that PIP joints present more detection problems.

On the other hand, lower bone detection arises as a very difficult problem. The overall detection rate drops to 58.9% although the healthy samples reach 69.7%. The lack of an stardard practice in X-ray image techniques causes an important variability in the samples which makes lower bone segmentation difficult to achieve. Figure A.3 shows two detection examples of wrong and valid segmentations.

The samples where both lower bone and sclerosis segmentations were labelled as valid were considered for further analysis. Table A.2 contains the corresponding results. The rates when the two segmentations are considered altogether are just slightly below the lower bone detection rates, reaching 67.0% for healthy samples and 54.2% for the whole dataset.

Afterwards, the proposed JSW measure was computed and tested in relation to SvdH score. In the case of MCP samples (Figure A.4a), the JSW estimation shows a light decreasing trend as the value of SvdH score increases. Nevertheless, that decreasing trend is clear in the case of PIP samples (Figure A.4b). It is an encouraging result taking into account that our system is only considering JSW values. Standard clinical assessment considers more information as the symmetry between hand joints or erosion information.

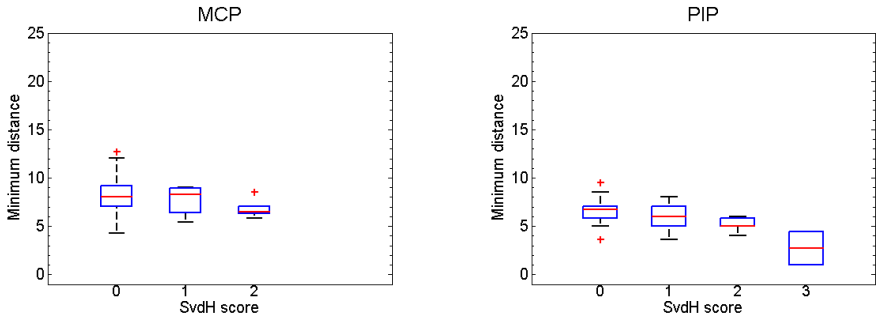
Finally, the reliability of the proposed system to provide a proper JSW estimation independently of its relation to the RA stage was tested. The JSW measure was computed both for the manual and automatic segmentations of the Tune Dataset, providing a final error of 1 ± 0.7 pixels. This result confirms the robustness of the JSW estimation.

Table A.1: Valid sclerosis and lower bone detection rates.

	J. type	SvdH 0	SvdH 1	SvdH 2	SvdH 3	Total
Sclerosis	**MCP**	111/115 (96.5%)	10/11 (90.9%)	10/12 (83.3%)	2/2 (100.0%)	**133/140 (95.0%)**
	PIP	64/67 (95.5%)	27/35 (77.1%)	22/28 (78.5%)	5/10 (50.0%)	**118/140 (84.2%)**
	Total	**175/182 (96.1%)**	37/46 (80.4%)	32/40 (80.0%)	7/12 (58.3%)	**251/280 (89.4%)**
L. bone	**MCP**	85/115 (73.9%)	5/11 (45.4%)	8/12 (66.6%)	0/2 (0.0%)	**98/140 (70.0%)**
	PIP	42/67 (62.6%)	14/35 (40.0%)	8/28 (28.5%)	3/10 (30.0%)	**67/140 (47.8%)**
	Total	**127/182 (69.7%)**	19/46 (41.3%)	16/40 (40.0%)	3/12 (25.0%)	**165/280 (58.9%)**

Table A.2: Valid sclerosis and lower bone detection rates (both in the same image).

J. type	SvdH 0	SvdH 1	SvdH 2	SvdH 3	Total
MCP	83/115 (72.1%)	5/11 (45.4%)	6/12 (50.0%)	0/2 (0.0%)	**94/140 (67.1%)**
PIP	39/67 (58.2%)	11/35 (31.4%)	6/28 (21.4%)	2/10 (20.0%)	**58/140 (41.4%)**
Total	**122/182 (67.0%)**	16/46 (34.7%)	12/40 (30.0%)	2/12 (16.6%)	**152/280 (54.2%)**



(a) (b)

Figure A.4: Distance measure and SvdH score. (a) MCP joints. (b) PIP joints.

A.6 Conclusions

The proposed system sets the foundations of a modularized RA assessment system. We created a hand joint sample dataset and introduced a hand joint model based on the skeletal structure of hand bones. Our sclerosis detector achieves remarkably good results. However, lower bone segmentation appears to be a harder task when faced in a local way. The JSW measure we propose was compared with the gold standard score for RA assessment. We showed for the first time that an automatic measure for JSW can be computed so that its value has a clear relation to the SvdH manual measure assessed by clinical personnel and, therefore, to RA disease stage. Future work should involve the confirmation of this trend with a dataset which should include a larger study with samples from multiple sources. Higher SvdH score samples, particularly scarce in number should be also considered.

Bibliography

[1] Colorectal cancer facts and figures, in support of the spectacolor biobank project. Technical report, European Organisation for Research and Treatment of Cancer (EORTC), 2014.

[2] A. M. Aibinu, M. I. Iqbal, A. A. Shafie, M. J. E. Salami, and M. Nilsson. Vascular intersection detection in retina fundus images using a new hybrid approach. *Computers in Biology and Medicine*, 40(1):81–89, 2010.

[3] K. A. Al-Kofahi, S. Lasek, D. H. Szarowski, C. J. Pace, G. Nagy, J. N. Turner, and B. Roysam. Rapid automated three-dimensional tracing of neurons from confocal image stacks. *Information Technology in Biomedicine, IEEE Transactions on*, 6(2):171–187, 2002.

[4] R. Ambauen, S. Fischer, and H. Bunke. Graph edit distance with node splitting and merging, and its application to diatom identification. In *Graph Based Representations in Pattern Recognition*, pages 95–106. Springer, 2003.

[5] S. Ameling et al. Texture-based polyp detection in colonoscopy. *Bildverarbeitung für die Medizin 2009*, pages 346–350, 2009.

[6] 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.

[7] E. Ardizzone, R. Pirrone, O. Gambino, and S. Radosta. Blood vessels and feature points detection on retinal images. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 2246–2249. IEEE, 2008.

[8] M. Arnold et al. Automatic segmentation and inpainting of specular highlights for endoscopic imaging. *Journal on Image and Video Processing*, 2010:9, 2010.

[9] M. Arnold et al. Quality Improvement of Endoscopy Videos. In *Proceedings of the 8th IASTED International Conference on Biomedical Engineering*, Insbruck, Austria, February 2011.

[10] M. Arnold, A. Ghosh, G. Lacey, S. Patchett, and H. Mulcahy. Indistinct frame detection in colonoscopy videos. In *Machine Vision and Image Processing Conference, 2009. IMVIP'09. 13th International*, pages 47–52. IEEE, 2009.

[11] C. Aronchick, W. Lipshutz, S. Wright, F. DuFrayne, and G. Bergman. Validation of an instrument to assess colon cleansing. *Am J Gastroenterol*, 94(9):2667, 1999.

[12] H. Asada and M. Brady. The curvature primal sketch. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (1):2–14, 1986.

[13] G. Azzopardi and N. Petkov. Automatic detection of vascular bifurcations in segmented retinal images using trainable cosfire filters. *Pattern Recognition Letters*, 34(8):922–933, 2013.

[14] G. Azzopardi and N. Petkov. Trainable cosfire filters for keypoint detection and pattern recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(2):490–503, 2013.

[15] R. Barclay, J. Vicari, A. Doughty, J. Johanson, and R. Greenlaw. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. *New England Journal of Medicine*, 355(24):2533–2541, 2006.

[16] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):509–522, 2002.

[17] S. Belongie, G. Mori, and J. Malik. Matching with shape contexts. In *Statistics and Analysis of Shapes*, pages 81–105. Springer, 2006.

[18] J. Bernal et al. Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition*, 45(9):3166 – 3182, 2012.

[19] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño. Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43:99–111, 2015.

[20] J. Bernal, F. J. Sánchez, and F. Vilariño. Depth of valleys accumulation algorithm for object detection. In *CCIA*, pages 71–80, 2011.

[21] J. Bernal, J. Sánchez, and F. Vilariño. Impact of image preprocessing methods on polyp localization in colonoscopy frames. In *Proceedings of the 35th International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, Japan, July 2013. (in press).

[22] C. Bernstein, M. Thorn, K. Monsees, R. Spell, and J. O'Connor. A prospective study of factors that determine cecal intubation time at colonoscopy. *Gastrointestinal endoscopy*, 61(1):72–75, 2005.

[23] S. Berretti, A. Del Bimbo, and P. Pala. Retrieval by shape similarity with perceptual distance and effective indexing. *Multimedia, IEEE Transactions on*, 2(4):225–239, 2000.

[24] V. Bevilacqua, S. Cambò, L. Cariello, and G. Mastronardi. A combined method to detect retinal fundus features. In *Proceedings of IEEE European Conference on Emergent Aspects in Clinical Data Analysis*, 2005.

[25] A. Bhuiyan, B. Nath, J. Chua, and K. Ramamohanarao. Automatic detection of vascular bifurcations and crossovers from color retinal fundus images. In *Signal-Image Technologies and Internet-Based System, 2007. SITIS'07. Third International IEEE Conference on*, pages 711–718. IEEE, 2007.

[26] A. Bielecki, M. Korkosz, and B. Zielinski. Hand radiographs preprocessing, image representation in the finger regions and joint space width measurements for image interpretation. *PRn*, 41(12):3786–3798, 2008.

[27] J. Blinn. Models of light reflection for computer synthesized pictures. In *ACM SIGGRAPH Computer Graphics*, volume 11, pages 192–198. ACM, 1977.

[28] H. Blum. A transformation for extracting new descriptors of shape. *Symposium on Models for Perception of Speech and Visual Form*, 1976.

[29] M. C. Boeres, C. C. Ribeiro, and I. Bloch. A randomized heuristic for scene recognition by graph matching. In *Experimental and Efficient Algorithms*, pages 100–113. Springer, 2004.

[30] D. Brunner and G. Brunnett. Fast force field approximation and its application to skeletonization of discrete 3d objects. In *Computer Graphics Forum*, volume 27, pages 261–270. Wiley Online Library, 2008.

[31] H. Bunke and G. Allermann. Inexact graph matching for structural pattern recognition. *Pattern Recognition Letters*, 1(4):245–253, 1983.

[32] R. Burkard, M. Dell'Amico, and S. Martello. *Assignment Problems, Revised Reprint*, volume 106. SIAM, 2012.

[33] D. Calvo, M. Ortega, M. G. Penedo, and J. Rouco. Automatic detection and characterisation of retinal vessel tree bifurcations and crossovers in eye fundus images. *Computer methods and programs in biomedicine*, 103(1):28–38, 2011.

[34] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam. Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms. *Information Technology in Biomedicine, IEEE Transactions on*, 3(2):125–138, 1999.

[35] A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(3):347–364, 2002.

[36] J. Canny. A computational approach to edge detection. *PAMI*, (6):679–698, 1986.

[37] Y. Cao, D. Liu, W. Tavanapong, J. Wong, J. Oh, and P. de Groen. Automatic classification of images with appendiceal orifice in colonoscopy videos. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 2349–2352. IEEE, 2006.

[38] Y. Cao, D. Liu, W. Tavanapong, J. Wong, J. Oh, and P. de Groen. Computer-aided detection of diagnostic and therapeutic operations in colonoscopy videos. *Biomedical Engineering, IEEE Transactions on*, 54(7):1268–1279, 2007.

[39] T. Chanwimaluang and G. Fan. An efficient blood vessel detection algorithm for retinal images using local entropy thresholding. In *Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium on*, volume 5, pages V–21. IEEE, 2003.

[40] N. Chapman, G. Dell'Omo, M. Sartini, N. Witt, A. Hughes, S. Thom, and R. Pedrinelli. Peripheral vascular disease is associated with abnormal arteriolar diameter relationships at bifurcations in the human retina. *Clinical Science*, 103(2):111–116, 2002.

[41] S. Chaudhuri et al. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transactions on medical imaging*, 8(3):263–269, 1989.

[42] R. Chellappa and R. Bagdazian. Fourier coding of image boundaries. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (1):102–105, 1984.

[43] M. T. Coimbra and J. P. S. Cunha. Mpeg-7 visual descriptors—contributions for automated feature extraction in capsule endoscopy. *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(5):628–637, 2006.

[44] T. Cormen, C. Leiserson, R. Rivest, and C. Stein. *Introduction to algorithms*. MIT press, 2001.

[45] R. Dahyot, F. Vilariño, and G. Lacey. Improving the quality of color colonoscopy videos. *Journal on Image and Video Processing*, 2008:1–7, 2008.

[46] P.-E. Danielsson. Euclidean distance mapping. *Computer Graphics and image processing*, 14(3):227–248, 1980.

[47] G. De Haan and E. Bellers. Deinterlacing-an overview. *Proceedings of the IEEE*, 86(9):1839–1857, 1998.

[48] Y. Deng and B. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(8):800–810, 2001.

[49] O. Duchenne, A. Joulin, and J. Ponce. A graph-matching kernel for object categorization. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1792–1799. IEEE, 2011.

[50] L. Espona et al. A snake for retinal vessel segmentation. *Pattern Recognition and Image Analysis*, pages 178–185, 2007.

[51] H. Freeman and A. Saghri. Generalized chain codes for planar curves. In *Proceedings of the 4th International joint conference on Pattern Recognition*, pages 701–703, 1978.

[52] F. Froehlich, V. Wietlisbach, J. Gonvers, B. Burnand, and J. Vader. Impact of colonic cleansing on quality and diagnostic yield of colonoscopy: the european panel of appropriateness of gastrointestinal endoscopy european multicenter study. *Gastrointestinal endoscopy*, 61(3):378–384, 2005.

[53] D. Gil et al. Structure-preserving smoothing of biomedical images. In *Computer Analysis of Images and Patterns*, pages 427–434. Springer, 2009.

[54] S. Gold and A. Rangarajan. A graduated assignment algorithm for graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(4):377–388, 1996.

[55] A. Goshtasby. Description and discrimination of planar shapes using shape matrices. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):738–743, 1985.

[56] W. I. Grosky and R. Mehrotra. Index-based object recognition in pictorial data management. *Computer Vision, Graphics, and Image Processing*, 52(3):416–436, 1990.

[57] H. Wang, L.C. Li et al. A novel colonic polyp volume segmentation method for computer tomographic colonography. In *SPIE Medical Imaging*, pages 90352W–90352W. International Society for Optics and Photonics, 2014.

[58] M. Häfner, A. Uhl, and G. Wimmer. A novel shape feature descriptor for the classification of polyps in hd colonoscopy. In *Medical Computer Vision. Large Data in Medical Imaging*, pages 205–213. Springer, 2014.

[59] Z. Harchaoui and F. Bach. Image classification with segmentation graph kernels. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[60] G. Harewood, V. Sharma, and P. de Garmo. Impact of colonoscopy preparation quality on detection of suspected colonic neoplasia. *Gastrointestinal endoscopy*, 58(1):76–79, 2003.

[61] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Citeseer, 1988.

[62] D. Heresbach, T. Barrioz, M. Lapalus, D. Coumaros, P. Bauret, P. Potier, D. Sautereau, C. Boustière, J. Grimaud, C. Barthélémy, et al. Miss rate for colorectal neoplastic polyps: a prospective multicenter study of back-to-back video colonoscopies. *Endoscopy*, 40(4):284–290, 2008.

[63] C. M. Holt, A. Stewart, M. Clint, and R. H. Perrott. An improved parallel thinning algorithm. *Communications of the ACM*, 30(2):156–160, 1987.

[64] A. Hoover et al. 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.

[65] M.-K. Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2):179–187, 1962.

[66] S. Hwang, J. Oh, J. Lee, Y. Cao, W. Tavanapong, D. Liu, J. Wong, and P. de Groen. Automatic measurement of quality metrics for colonoscopy videos. In *Proceedings of the 13th annual ACM international conference on Multimedia*, pages 912–921. ACM, 2005.

[67] S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. de Groen. Stool detection in colonoscopy videos. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 3004–3007. IEEE, 2008.

[68] S. Hwang, J. Oh, W. Tavanapong, J. Wong, and P. C. De Groen. 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.

[69] Y. Imai et al. Estimation of multiple illuminants based on specular highlight detection. *Computational Color Imaging*, pages 85–98, 2011.

[70] T. Imperiale, D. Wagner, C. Lin, G. Larkin, J. Rogge, and D. Ransohoff. Results of screening colonoscopy among persons 40 to 49 years of age. *New England Journal of Medicine*, 346(23):1781–1785, 2002.

[71] Y. Iwahori, T. Shinohara, A. Hattori, R. J. Woodham, S. Fukui, M. Bhuyan, and K. Kasugai. Automatic polyp detection in endoscope images using a hessian filter. *Proceedings of MVA*, pages 21–24, 2013.

[72] X. Jiang et al. Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(1):131–137, 2003.

[73] G. Joblove and D. Greenberg. Color spaces for computer graphics. *ACM SIGGRAPH Computer Graphics*, 12(3):20–25, 1978.

[74] E. Jung and K. Hong. Automatic retinal vasculature structure tracing and vascular landmark extraction from human eye image. In *Hybrid Information Technology, 2006. ICHIT'06. International Conference on*, volume 2, pages 161–167. IEEE, 2006.

[75] D. Justice and A. Hero. A binary linear programming formulation of the graph edit distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(8):1200–1214, 2006.

[76] J. Kang and R. Doraiswami. Real-time image processing system for endoscopic applications. In *Electrical and Computer Engineering, 2003. IEEE CCECE 2003. Canadian Conference on*, volume 3, pages 1469–1472. IEEE, 2003.

[77] S. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. Karras, M. Tzivras, et al. Computer-aided tumor detection in endoscopic video using color wavelet features. *Information Technology in Biomedicine, IEEE Transactions on*, 7(3):141–152, 2003.

[78] H. Kauppinen, T. Seppänen, and M. Pietikäinen. An experimental comparison of autoregressive and fourier-based descriptors in 2d shape classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(2):201–207, 1995.

[79] W. Kim, Y. Cho, J. Park, P. Min, J. Kang, and I. Park. Factors affecting insertion time and patient discomfort during colonoscopy. *Gastrointestinal endoscopy*, 52(5):600–605, 2000.

[80] W.-Y. Kim and Y.-S. Kim. A region-based shape descriptor using zernike moments. *Signal processing: Image communication*, 16(1):95–102, 2000.

[81] S. Krishnan, X. Yang, K. Chan, S. Kumar, and P. 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. IEEE, 1998.

[82] P. Kwok. A thinning algorithm by contour generation. *Communications of the ACM*, 31(11):1314–1324, 1988.

[83] J.-S. Kwon, J.-W. Gi, and E.-K. Kang. An enhanced thinning algorithm using parallel processing. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 3, pages 752–755. IEEE, 2001.

[84] G. Langs, P. Peloschek, H. Bischof, and F. Kainberger. Automatic quantification of joint space narrowing and erosions in rheumatoid arthritis. *Medical Imaging, IEEE Transactions on*, 28(1):151–164, 2009.

[85] R. Lee, R. Tang, V. Muthusamy, S. Ho, N. Shah, L. Wetzel, A. Bain, E. Mackintosh, A. Paek, A. Crissien, et al. Quality of colonoscopy withdrawal technique and variability in adenoma detection rates (with videos). *Gastrointestinal endoscopy*, 74(1):128–134, 2011.

[86] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson, and K. Siddiqi. Turbopixels: Fast superpixels using geometric flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(12):2290–2297, 2009.

[87] J. Llados and G. Sanchez. Graph matching versus graph parsing in graphics recognition—a combined approach. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(03):455–473, 2004.

[88] G. Lu and A. Sajjanhar. Region-based shape representation and similarity measure suitable for content-based image retrieval. *Multimedia Systems*, 7(2):165–174, 1999.

[89] P. Mahé, N. Ueda, T. Akutsu, J.-L. Perret, and J.-P. Vert. Graph kernels for molecular structure-activity relationship analysis with support vector machines. *Journal of chemical information and modeling*, 45(4):939–951, 2005.

[90] J. Mandel, J. Bond, T. Church, D. Snover, G. Bradley, L. Schuman, and F. Ederer. Reducing mortality from colorectal cancer by screening for fecal occult blood. *New England Journal of Medicine*, 328(19):1365–1371, 1993.

[91] D. Marín et al. 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.

[92] M. P. Martínez-Pérez, J. Jiménez, and J. L. Navalón. A thinning algorithm based on contours. *Computer Vision, Graphics, and Image Processing*, 39(2):186–201, 1987.

[93] A. 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.

[94] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International journal of computer vision*, 65(1-2):43–72, 2005.

[95] S. Morán Sánchez, E. Torrella, P. Esteban Delgado, R. Baños Madrid, A. García, A. Ono, E. Pérez Cuadrado, P. Parra, J. Cruzado Quevedo, F. Pérez Riquelme, et al. Colonoscopy quality assessment. *Revista Espanola de Enfermedades Digestivas*, 101(2):107, 2009.

[96] P. Morrison and J. J. Zou. Skeletonization based on error reduction. *Pattern recognition*, 39(6):1099–1109, 2006.

[97] M. Neuhaus and H. Bunke. Self-organizing graph edit distance. In *Graph Based Representations in Pattern Recognition*, pages 83–94. Springer, 2003.

[98] M. Neuhaus and H. Bunke. A probabilistic approach to learning costs for graph edit distance. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3, pages 389–393. IEEE, 2004.

[99] M. Neuhaus and H. Bunke. A graph matching based approach to fingerprint classification using directional variance. In *Audio-and Video-Based Biometric Person Authentication*, pages 191–200. Springer, 2005.

[100] M. Neuhaus and H. Bunke. Self-organizing maps for learning the edit costs in graph matching. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 35(3):503–514, 2005.

[101] M. Neuhaus and H. Bunke. Automatic learning of cost functions for graph edit distance. *Information Sciences*, 177(1):239–247, 2007.

[102] M. Neuhaus, K. Riesen, and H. Bunke. Fast suboptimal algorithms for the computation of graph edit distance. In *Structural, Syntactic, and Statistical Pattern Recognition*, pages 163–172. Springer, 2006.

[103] J. Núñez, V. F., and G. D. Computer vision techniques for characterization of finger joints in x-ray images. Technical report, Centre de Visió per Computador, 2011.

[104] R. L. Ogniewicz and O. Kübler. Hierarchic voronoi skeletons. *Pattern recognition*, 28(3):343–359, 1995.

[105] J. Oh, S. Hwang, Y. Cao, W. Tavanapong, D. Liu, J. Wong, and P. de Groen. Measuring objective quality of colonoscopy. *Biomedical Engineering, IEEE Transactions on*, 56(9):2190–2196, 2009.

[106] J. Oh, S. Hwang, J. Lee, W. Tavanapong, J. Wong, and P. de Groen. Informative frame classification for endoscopy video. *Medical Image Analysis*, 11(2):110–127, 2007.

[107] B. Overholt, L. Brooks-Belli, M. Grace, K. Rankin, R. Harrell, M. Turyk, F. Rosenberg, R. Barish, N. Gilinsky, et al. Withdrawal times and associated factors in colonoscopy: a quality assurance multicenter assessment. *Journal of clinical gastroenterology*, 44(4):e80, 2010.

[108] G. Papari and N. Petkov. Edge and line oriented contour detection: State of the art. *Image and Vision Computing*, 29(2-3):79–103, 2011.

[109] S. Y. Park, D. Sargent, I. Spofford, K. G. Vosburgh, et al. A colon video analysis framework for polyp detection. *Biomedical Engineering, IEEE Transactions on*, 59(5):1408–1418, 2012.

[110] J. R. Parker. *Algorithms for image processing and computer vision*. John Wiley & Sons, 2010.

[111] S. K. Parui, S. E. Sarma, and D. D. Majumder. How to discriminate shapes using the shape vector. *Pattern Recognition Letters*, 4(3):201–204, 1986.

[112] T. Pavlidis. *Algorithms for graphics and image processing*. Springer Science & Business Media, 2012.

[113] M. Pelillo, K. Siddiqi, and S. W. Zucker. Matching hierarchical structures using association graphs. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(11):1105–1120, 1999.

[114] M. Perez, A. Highes, A. V. Stanton, S. A. Thorn, N. Chapman, A. A. Bharath, and K. H. Parker. Retinal vascular tree morphology: a semi-automatic quantification. *Biomedical Engineering, IEEE Transactions on*, 49(8):912–917, 2002.

[115] M. Peura and J. Iivarinen. Efficiency of simple shape descriptors. In *Proceedings of the third international workshop on visual form*, volume 443, page 451. Citeseer, 1997.

[116] S. Phee, W. Ng, I. Chen, F. Seow-Choen, and B. Davies. Automation of colonoscopy. ii. visual control aspects. *Engineering in Medicine and Biology Magazine, IEEE*, 17(3):81–88, 1998.

[117] M. Pudzs, R. Fuksis, and M. Greitans. Palmprint image processing with non-halo complex matched filters for forensic data analysis. In *Biometrics and Forensics (IWBF), 2013 International Workshop on*, pages 1–4. IEEE, 2013.

[118] D. Rex, J. Bond, S. Winawer, T. Levin, R. Burt, D. Johnson, L. Kirk, S. Litlin, D. Lieberman, J. Waye, et al. Quality in the technical performance of colonoscopy and the continuous quality improvement process for colonoscopy: recommendations of the us multi-society task force on colorectal cancer. *The American journal of gastroenterology*, 97(6):1296–1308, 2002.

[119] D. Rex, T. Imperiale, D. Latinovich, and L. Bratcher. Impact of bowel preparation on efficiency and cost of colonoscopy. *The American journal of gastroenterology*, 97(7):1696–1700, 2002.

[120] D. Rex, J. Petrini, T. Baron, A. Chak, J. Cohen, S. Deal, B. Hoffman, B. Jacobson, K. Mergener, B. Petersen, et al. Quality indicators for colonoscopy. *The American journal of gastroenterology*, 101:873–885, 2006.

[121] F. Riaz, M. Ribeiro, and M. Coimbra. Quantitative comparison of segmentation methods for in-body images. In *Proceedings of EMBC 2009*, pages 5785 –5788, September 2009.

[122] K. Riesen and H. Bunke. Iam graph database repository for graph based pattern recognition and machine learning. In *Structural, Syntactic, and Statistical Pattern Recognition*, pages 287–297. Springer, 2008.

[123] K. Riesen and H. Bunke. Approximate graph edit distance computation by means of bipartite graph matching. *Image and Vision Computing*, 27(7):950–959, 2009.

[124] K. Riesen and H. Bunke. *Classification and clustering of vector space embedded graphs*. World Scientific, 2010.

[125] A. Robles-Kelly and E. R. Hancock. Graph edit distance from spectral seriation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(3):365–378, 2005.

[126] D. Rockey, E. Paulson, D. e. Niedzwiecki, W. Davis, H. Bosworth, L. Sanders, J. Yee, J. Henderson, P. Hatten, S. Burdick, et al. Analysis of air contrast barium enema, computed tomographic colonography, and colonoscopy: prospective comparison. *The Lancet*, 365(9456):305–311, 2005.

[127] A. Rosset, L. Spadola, and O. Ratib. Osirix: An open-source software for navigating in multidimensional dicom images. *Journal of Digital Imaging*, 17:205–216, 2004. 10.1007/s10278-004-1014-6.

[128] A. Rostom and E. Jolicoeur. Validation of a new scale for the assessment of bowel preparation quality. *Gastrointestinal endoscopy*, 59(4):482–486, 2004.

[129] W. J. Rucklidge. Efficiently locating objects using the hausdorff distance. *International Journal of computer vision*, 24(3):251–270, 1997.

[130] Y. Rui, A. C. She, and T. S. Huang. A modified fourier descriptor for shape matching in mars. *SERIES ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING*, 8:165–180, 1997.

[131] A. Saaristo, T. Karpanen, and K. Alitalo. Mechanisms of angiogenesis and their use in the inhibition of tumor growth and metastasis. *Oncogene*, 19(53), 2000.

[132] S. Saha and N. Dutta Roy. Automatic detection of bifurcation points in retinal fundus images. *Latest Research in Science and Technology. International Journal of*, 2(2):105–108, 2013.

[133] A. Sanfeliu and K.-S. Fu. A distance measure between attributed relational graphs for pattern recognition. *Systems, Man and Cybernetics, IEEE Transactions on*, 13(3):353–362, 1983.

[134] A. Schenker. *Graph-theoretic techniques for web content mining*, volume 62. World Scientific, 2005.

[135] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *International Journal of computer vision*, 37(2):151–172, 2000.

[136] S. Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4):210–218, 1985.

[137] A. Sharma, R. Horaud, J. Cech, and E. Boyer. Topologically-robust 3d shape matching based on diffusion geometry and seed growing. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 2481–2488. IEEE, 2011.

[138] J. Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, 2002.

[139] F. Y. Shih and Y.-T. Wu. Fast euclidean distance transformation in two scans using a 3× 3 neighborhood. *Computer Vision and Image Understanding*, 93(2):195–205, 2004.

[140] J. Soares et al. Retinal vessel segmentation using the 2-d gabor wavelet and supervised classification. *Medical Imaging, IEEE Transactions on*, 25(9):1214–1222, 2006.

[141] A. C. Society. *Colorectal Cancer Facts & Figures 2014-2016*. American Cancer Society, 2014.

[142] S. Sorlin and C. Solnon. Reactive tabu search for measuring graph similarity. In *Graph-Based Representations in Pattern Recognition*, pages 172–182. Springer, 2005.

[143] D. M. Squire and T. M. Caelli. Invariance signatures: Characterizing contours by their departures from invariance. *Computer Vision and Image Understanding*, 77(3):284–316, 2000.

[144] 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.

[145] J. Staal et al. Ridge-based vessel segmentation in color images of the retina. *Medical Imaging, IEEE Transactions on*, 23(4):501–509, 2004.

[146] F. W. M. Stentiford and R. G. Mortimer. Some new heuristics for thinning binary handprinted characters for ocr. *IEEE Transaction on Systems, Man, and Cybernetics*, 13(1):81–84, 1983.

[147] N. Tajbakhsh, S. R. Gurudu, and J. Liang. Automatic polyp detection using global geometric constraints and local intensity variation patterns. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pages 179–187. Springer, 2014.

[148] G. Taubin and D. B. Cooper. Recognition and positioning of rigid objects using algebraic moment invariants. In *San Diego,'91, San Diego, CA*, pages 175–186. International Society for Optics and Photonics, 1991.

[149] C.-H. Teh and R. T. Chin. On image analysis by the methods of moments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(4):496–513, 1988.

[150] 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.

[151] M. Tjoa and S. Krishnan. Feature extraction for the analysis of colon status from the endoscopic images. *BioMedical Engineering OnLine*, 2(9):1–17, 2003.

[152] Tresca, A. The Stages of Colon and Rectal Cancer. *New York Times (About.com)*, page 1, 2010.

[153] C.-L. Tsai, C. V. Stewart, H. L. Tanenbaum, and B. Roysam. Model-based method for improving the accuracy and repeatability of estimating vascular bifurcations and crossovers from retinal fundus images. *Information Technology in Biomedicine, IEEE Transactions on*, 8(2):122–130, 2004.

[154] M. O. Tso and L. M. Jampol. Pathophysiology of hypertensive retinopathy. *Ophthalmology*, 89(10):1132–1145, 1982.

[155] D. Van der Heijde. How to read radiographs according to the sharp/van der heijde radiological assessment in rheumatoid arthritis in long term studies. *Journal of Rheumatology*, (26):743–745, 1999.

[156] J. van Rijn, J. Reitsma, J. Stoker, P. Bossuyt, S. van Deventer, and E. Dekker. Polyp miss rate determined by tandem colonoscopy: a systematic review. *The American journal of gastroenterology*, 101(2):343–350, 2006.

[157] S. Vera. Finger joint modelling from hand x-ray images for assessing rheumatoid arthritis. Technical report, Centre de Visió per Computador, 2010.

[158] P. Wang, S. M. Krishnan, C. Kugean, and M. Tjoa. Classification of endoscopic images based on texture and neural network. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 4, pages 3691–3695. IEEE, 2001.

[159] L. Xu and S. Luo. A novel method for blood vessel detection from retinal images. *BioMedical Engineering OnLine*, 9(1):14, 2010.

[160] T. J. Yokoi, S. and T. Fukumura. Topological properties in digitized binary pictures. *Systems Computer Controls*, 4:32–39, 1973.

[161] I. T. Young, J. E. Walker, and J. E. Bowie. An analysis technique for biological shape. i. *Information and control*, 25(4):357–370, 1974.

[162] F. Zana and J. Klein. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *Image Processing, IEEE Transactions on*, 10(7):1010–1019, 2001.

[163] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern recognition*, 37(1):1–19, 2004.

[164] D. Zhang, G. Lu, et al. A comparative study of fourier descriptors for shape representation and retrieval. In *Proc. 5th Asian Conference on Computer Vision*. Citeseer, 2002.

[165] S. Zhang and K. Fu. A thinning algorithm for discrete binary images. In *Proc. Int. Conf. Computers and Applications*, pages 879–886, 1984.

[166] X. Zhang, F. Jia, S. Luo, G. Liu, and Q. Hu. A marker-based watershed method for x-ray image segmentation. *Computer methods and programs in biomedicine*, 2014.

[167] 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.

This work was supported in part by the research grant from Universitat Autonòma de Barcelona 471-01-2/2010, and by the Spanish Government through the founded projects "COLON-QA" ($TIN2009-10435$) and "FISIOLOGICA" ($TIN2012-33116$).