Universitat
Autònoma
de Barcelona

A Machine Learning Approach for Intestinal Motility Assessment with Capsule Endoscopy

A dissertation submitted by **Fernando Luis Vilariño Freire** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor**.

Bellaterra, June 12, 2006

Supervisor: **Dr. Petia Ivanova Radeva**
Universitat Autònoma de Barcelona
Dept. Ciències de la Computació & Computer Vision Center

Centre de Visió
per Computador

This document was typeset by the author using $\LaTeX\,2_\varepsilon$.

The research described in this book was carried out at the Centre de Visió per Computador & Departament de Ciències de la Computació, Universitat Autònoma de Barcelona.

ISBN: 84-933652-7-0

Printed by Ediciones Gráficas Rey, S.L.

A Rosa, por su fuerza, su valentía
y porque ella sabe el porqué.

Acknowledgements

In the small Catalonian town of Sant Cugat del Vallès, there is a special place for me. There, you can walk around the stone cloister, staring for hours at the animal-shaped capitals which decorate the old columns. An ancient anthem is said to be codified within these capitals, since each different animal corresponds to a different note. The music is hidden among the stones of the Sant Cugat's monastery -as it is also hidden within the stones of the Galician city of Santiago de Compostela-. The disclosure of this symphony supposed a fascinating finding.

A few miles away, there is another beautiful cloister-shaped building, where the light which enters through the central garden is distributed to all the offices in which the people integrating the Computer Vision Center staff develop their research tasks. During the last years, I have walked this *cloister* around sharing comments, suggestions and solutions with these people. This is a list of acknowledgements to all of them.

First of all, I would like to thank Dr. Petia Radeva, my supervisor in this thesis, for her trust on my person when she offered me a proposal for developing my doctoral work within her team. Thanks also to Prof. Juan José Villanueva, head of the Computer Vision Center, for the warm welcome that he showed to me, which made my integration so easy in this dynamic and multi-disciplinary scientific community. In addition, I would like to express thanks to Dr. Fernando Azpiroz, who leads the medical research project that constitutes the fundamental reason of the advances which we present today in this work.

I would like to give a special acknowledgement to all the remaining members of the *intes-team* at CVC, headed by Dr. Jordi Vitrià, from whom I have obtained so many insightful observations about the world of classifiers. A special mention must be made to Dr. Panagiota Spyridonos, for her great collaboration task and her hours of dedication to the improvement of the results in this project, and to the rest of people who have joined our team: Dr. Joaquin Salas, Dra. Hella Khoufi -"de tout et de votre sourire"-, Eric Sommerlade, and Dr. David Guillamet. Regarding the Hospital Vall d'Hebron team, I would like to thank all the members who contributed to this work with their comments and suggestions: Dr. Fosca de Iorio and Dr. Carolina Malagelada.

Voldria expressar la meva gratitud a tots els companys amb els quals he tingut l'oportunitat de compartir despatx: l'Anna Salvatella, el Dr. Marco Bressan, el Dr. Xavier Otazu, la Dra. Dèbora Gil, en Xavier Jimènez, en David Rotger, el Dr. Misael Rosales -amigo Misael, cuando sea mayor, tal vez sea como tú-, l'Aura Hernàndez -que et saluda, i llavors el matí és inevitablement més alegre-, el Dr. Oriol Pujol -tantes coses per fer i tan poques hores de joventut-, en Joel Barajas, en Miquel Ferrer, i tots els membres del Centre de Visió per Computador i el Departament de Ciències de la Computació. Ha estat un plaer compartir comentaris, suggeriments, i tants i tants moments en les hores dedicades a la meva labor de recerca i docència a la Universitat Autònoma de Barcelona.

Esta lista de agradecementos non estaría completa se non fixera unha mención especial a todos os membros do Departamento de Electrónica e Computación da Universidade de Santiago de Compostela, no que comezou esta a miña andaina polos eidos da Intelixencia Artificial e as Ciencias da Computación.

Finally, I would like to close these acknowledgements with two special recalls: The first one is devoted to Anna, with whom I've shared insightful comments and moments around this thesis. The last one is for Lucy, to whom I own my gratitude for the professional scientific perspective I learnt from her, and for the unforgettable moments playing *The Beatles* in Bangor.

Al ĉiu, dankon.

Abstract

Intestinal motility assessment with video capsule endoscopy arises as a novel and challenging clinical fieldwork. This technique is based on the analysis of the patterns of intestinal contractions obtained by labelling all the motility events present in a video provided by a capsule with a wireless micro-camera, which is ingested by the patient. However, the visual analysis of these video sequences presents several important drawbacks, mainly related to both the large amount of time needed for the visualization process, and the low prevalence of intestinal contractions in video.

In this work we propose a machine learning system to automatically detect the intestinal contractions in video capsule endoscopy, driving a very useful but not feasible clinical routine into a feasible clinical procedure. Our proposal is divided into two different parts: The first part tackles the problem of the automatic detection of phasic contractions in capsule endoscopy videos. Phasic contractions are dynamic events spanning about 4-5 seconds, which show visual patterns with a high variability. Our proposal is based on a sequential design which involves the analysis of textural, color and blob features with powerful classifiers such as SVM. This approach appears to cope with two basic aims: the reduction of the imbalance rate of the data set, and the modular construction of the system, which adds the capability of including domain knowledge as new stages in the cascade. The second part of the current work tackles the problem of the automatic detection of tonic contractions. Tonic contractions manifest in capsule endoscopy as a sustained pattern of the folds and wrinkles of the intestine, which may be prolonged for an undetermined span of time. Our proposal is based on the analysis of the wrinkle patterns, presenting a comparative study of diverse features and classification methods, and providing a set of appropriate descriptors for their characterization. We provide a detailed analysis of the performance achieved by our system both in a qualitative and a quantitative way.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The analysis of intestinal motility activity is one of the main sources of information which gastroenterologists have in order to assess the presence of certain intestinal disfunctions. Currently, this motility assessment is mainly performed by means of highly invasive techniques, such as intestinal manometry, which carry multiple drawbacks, including the need of hospitalization, the introduction of a probe into the patients intestinal tract, the presence of qualified staff during the clinical intervention, etc.

Recently, a novel technique named Capsule Endoscopy has proved its efficiency as an alternative endoscopic technique. With capsule endoscopy, a pill with a micro-camera attached to it is swallowed by the patient. During several hours, the capsule emits a radio signal which is recorded into an external device, storing a video movie of the trip of the capsule throughout the gut. The application of this technique to intestinal motility assessment allows the specialist to overcome most of the difficulties associated to classical clinical procedures. However, capsule endoscopy carries a main drawback: the visualization analysis of the video frames is a tedious and difficult task, which deserves specifically trained staff, and which may last for more than one hour for each study. Thus, although the information provided by capsule endoscopy is unique and there is no other current technique which improves the reach and quality of the capsule images, the consequent procedure of analysis of the video material makes this clinical routine not feasible.

1.1 Providing a *novel* and *feasible* clinical routine by means of machine learning

In this scenario, the need of an alternative procedure for the obtention of the temporal pattern of intestinal contractions in video capsule endoscopy is mandatory. This urgent need boosted the collaboration between a group of gastroenterologists from Vall D'Hebron Hospital, Barcelona, and our group in the Computer Vision Center at the Universitat Autònoma de Barcelona, so as to evaluate the possibility of starting a new research line in this fieldwork. From that moment on, our efforts were focused on the development of a machine learning based system for the automatic detection of intestinal contractions in video capsule endoscopy. Our goal is to automatically provide the physicians with the temporal pattern of intestinal contractions for each study, i.e., to label the video frames which the system proposes as candidates for intestinal contractions in an automatic way without any expert interaction. The main exigence we demand from our system is to be able to catch the greater number of existing contractions (*high sensitivity*) minimizing at the same time the number of frames wrongly labelled as contractions (*high specificity* and *precision*). It is important to underline that the relevance of our outcome is based on the following fact: we are making it possible to turn *a highly useful but not feasible clinical routine* into a *feasible clinical routine*, allowing the experts to take profit of all the powerful source of motility information existing in video capsule endoscopy.

1.2 Main contributions of our work

The work presented in this thesis is the result of gathering a set of points of view, focusing on the main aim of providing the experts with a robust solution to the challenging problem of automatic detection of contractions. Figure 1.1 pictures a graphical representation of the different approaches in which we developed our research:

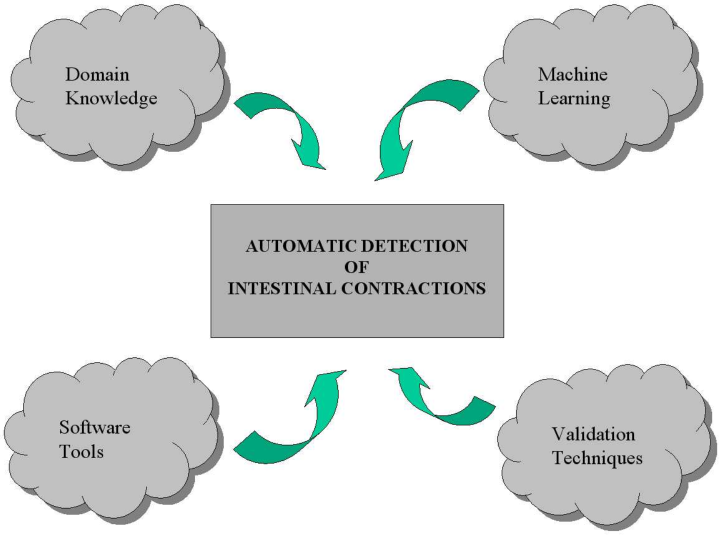


Figure 1.1: Multidisciplinary framework for the automatic detection of intestinal contractions in wireless capsule endoscopy videos

Our contribution: A general review of the intestinal motility framework

A preliminary background about the physiological phenomena associated to intestinal motility appears as an essential tool for understanding the multiple processes which yield to the generation of the diverse patterns of contractions. The assimilation of the basic foundations in which intestinal motility is sustained is an essential source of information in order to identify the traits which distinguish contractions from other general events. For this reason, our work is introduced by an extended study about the medical framework on which our research was developed. We specifically focused our attention on the physiological origin and nature of phasic and tonic intestinal contractions, their duration, their frequency and their typology, in order to use this domain knowledge for their further characterization.

Our contribution: An extended study of capsule endoscopy video imaging

This thesis presents a deep study of the capsule endoscopy imaging technology. Although several works have been published for diverse applications of capsule endoscopy, our contribution represents the first approach in the specific fieldwork of motility analysis, since, as far as we know, no previous work has been reported in this issue. We focused our research on the characterization of the diverse events which provide useful information regarding motility. Throughout our study, we propose the use of multiple image descriptors for the intestinal motility events, taking into account color information, textural features and blob analysis. We propose a first definition of the visual paradigms of both phasic and tonic intestinal contractions in video endoscopy, which is a completely new issue in the current literature.

Our contribution: Improving classification techniques for an imbalanced problem

The prevalence of intestinal contractions in video is very low, showing a ratio about $1 : 50$, which typically corresponds to less than 1,000 contractions within 50,000 video frames. This represents an imbalanced problem for classification. Imbalanced problems are characterized by a large number of false positives. In our case of study, this translates into a large number of non-contractions frames which are wrongly classified as contractions at the output stage of the system. With the aim of reducing this false positive rate, we focused our efforts on the research of classification techniques applied to imbalanced problems. A comparative analysis of multiple classifiers, including support vectors machines, AdaBoost, decision trees and others, were tested by including specific modifications for the optimization of the classification performance in this specific scenario. We propose several contributions:

- We show that the sequential design of the classification system provides good results in the reduction of the imbalance rate, yielding to an effective reduction in the false positives number.

- We introduce a novel method for improving the classification based on ROC curve analysis, which we prove to provide a substantial reduction in video visualization time.

- We present a detailed study and discussion of the effectiveness of the use of different stratified sampling techniques applied to standard classifiers, such as support vectors machines.

- Finally, we show our results in experiments which investigate different methods for the training stage of adaboost.

Our contribution: Automatic annotation of intestinal contractions

Our proposal is based on a machine learning system which automatically learns and classifies contractions from a capsule video source, providing the expert with the portion of the video which is highly likely to contain the intestinal contractions. This yields to a considerable reduction in visualization time, and the consequent reduction of stress, since most of the sequences to be analyzed are real contractions. One of the main advantages of our system is related to its ability to dynamically adapt itself to the different patterns of intestinal activity associated with intestinal contractions in a robust way. Furthermore, our implementation appears to be flexible and easily extensible, since the modular design of our approach allows the expert to include domain knowledge into the system by means of the addition of new modular stages. We present our results split into two different parts in this thesis.

- In Part II, we present a machine learning based procedure for the automatic detection of phasic intestinal contractions in capsule endoscopy videos. Our main contributions consist of a proposal for a set of image descriptors for the characterization of phasic intestinal contractions, and a proposal for a classification system based on a cascade of classifiers which identifies the diverse motility events, and automatically provides the expert with the suggested set of contractions.

- In Part III, we present a completely different machine learning approach for the automatic detection of tonic contractions. Our main contribution in this topic consists of a proposal for a set of image descriptors for the characterization of tonic contractions, and a proposal for a classification system for tonic contractions based on support vectors machines classifiers.

These two contributions provide the experts with the pattern of all the intestinal events in video in an automatic way, with good performance results. This allows the specialist to use capsule endoscopy as a helpful tool for the intestinal motility assessment, driving a useful but not feasible clinical technique into a feasible clinical technique.

Our contribution: Validation techniques

Validation of performance results is not a straightforward task in this scenario. On the one hand, time constraints force that only a small portion of the video frames are labelled by the specialists -namely, the intestinal contractions-. It makes it not possible to have access to an objective numerical performance analysis when other kind of events related to intestinal motility are to be detected. In this sense, we implemented and tested several graphical procedures to facilitate the qualitative analysis of preliminary results by means of visual techniques.

On the other hand, in order to numerically assess the actual performance of our system, we studied several performance metrics. In this sense, we propose the use of our own false alarm rate (FAR) definition, and the sensitivity-FAR curves, as useful tools for performance assessment which provide specific information about performance in imbalanced problems.

Our contribution: Software tools

The ultimate step of our project consists of providing the experts with the appropriate software tools which implement all the related advances. We present a general description of the software applications we developed. These applications are currently being used in a successful way by the specialists, and they have shown their performance as the basic source of data in recent publications. Some of these software tools are currently being tested for their inclusion into the clinical protocols which intestinal motility assessment involves.

1.3 How is this thesis organized?

The presentation of the contributions of this thesis is split in four parts: Throughout the first part, which spans Chapters 2 and 3, we introduce the medical framework of intestinal motility, we explain the technological issues related to capsule endoscopy and present our definition of the paradigms of intestinal contractions in capsule endoscopy videos. The explanation of the detection methods for intestinal contractions, and the performance results of our approach, are distributed into the following parts: Part II, which spans Chapters 4, 5, 6 and 7, is devoted to the exposition of the methodology and results about the automatic detection of phasic intestinal contractions, while Part III, which spans Chapters 8 and 9, is devoted to the analysis of the methodology and results about the automatic detection of tonic intestinal contractions. Finally, Part IV closes this thesis with the general discussion of our work in Chapter 10, and the exposition of the conclusions and highlighting of our proposals for future lines of research in Chapter 11.

Part I

MEDICAL AND TECHNOLOGICAL FRAMEWORK

Chapter 2

Medical framework

2.1 Introduction

Small intestine motility dysfunctions are shown to be related to certain gastrointestinal disorders which can be manifest in a varied symptomatology [50]. The analysis of the intestinal contractions of the small bowel, in terms of number, frequency and distribution along the intestinal tract, represents one of the methods with the highest clinical pathological significance [67], and has been successfully applied and reported in recent studies [68]. Several techniques have been developed and tested in a wide range of modalities to achieve this aim. Nevertheless, all of these techniques suffer from important drawbacks: they are highly invasive, they usually generate patient discomfort, they need hospitalization and specialized staff, etc. Wireless video capsule endoscopy appears as a novel technique which has been recently applied in several clinical scenarios related to gastroenterology [30, 70, 32]. Many of the drawbacks mentioned above are no longer present in capsule endoscopy, which makes it very interesting from a clinical point of view and highly useful for the specific objective of intestinal motility assessment.

In order to understand how intestinal motility works in the human being, an introduction to its main neurophysiological antecedents is needed. The aim of the following sections goes in this direction, providing an overview of what intestinal motility is and trying to solve questions about the origin of intestinal contractions from a physiological perspective, as well as a brief description of the different techniques which have been used by the scientific community for their detection. In the final section, we present a technical overview of the wireless capsule endoscopy video analysis technique, which will be especially useful to understand the nature of the acquisition of the video images that we used in our research.

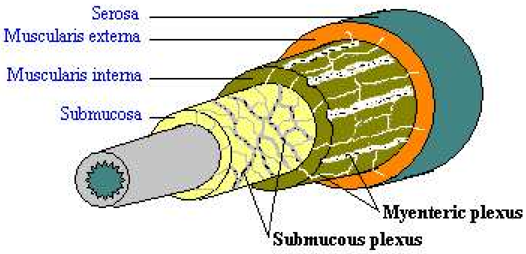


Figure 2.1: Main components of the enteric nervous system

2.2 Basic concepts about gastrointestinal motility

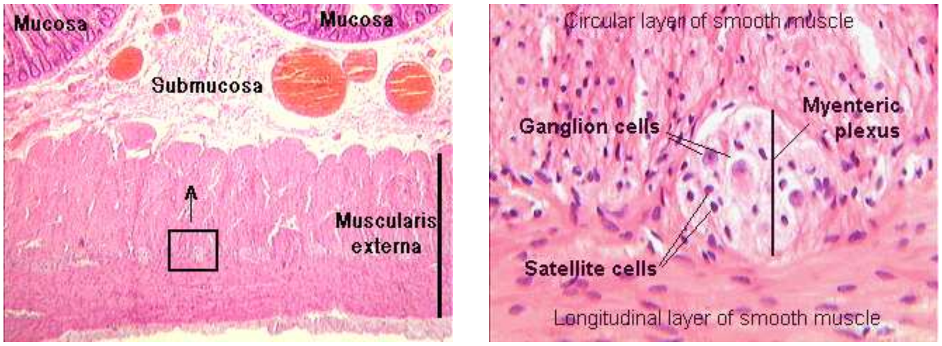
2.2.1 Physiological frame of gastrointestinal motility

The intrinsic nervous system

Muscle layers of the gut wall and their innervation are organized so as to provide the motor functions along the intestinal tract. This motor functions generate muscle contractions and motility that are integral parts of the digestive function. The interaction of the gut with the central nervous system is performed through either somatic or autonomic neurons [50]. Essentially, the digestive system is endowed with its own, local nervous system referred to as the *enteric* or *intrinsic nervous system*. This intrinsic nervous system, together with the somatic neurons belonging to the central system, provide the electrical signals that excite the contraction of the muscles surrounding the gut. The main components of the enteric nervous system consist of two networks or *plexuses* of neurons, both of which are embedded in the wall of the digestive tract and extended from the esophagus to the anus in the way represented in Figure 2.1. These two plexuses of neurons show the following characteristics:

- The *myenteric plexus* is located between the longitudinal and circular layers of muscle in the tunica muscularis and, appropriately, *exerts control primarily over digestive tract motility.*
- The *submucous plexus*, as its name implies, is buried in the submucosa. Its main role is *sensing the environment within the lumen*, so as to regulate gastrointestinal blood flow and to control epithelial cell function. In regions where these functions are minimal, such as the esophagus, the submucous plexus is sparse and may actually be missing.

The neurophysiological scheme of the gut matches the following pattern: Sensory neurons, mainly found in the submucous plexus, receive information from sensory receptors in the mucosa and muscle. At least five different sensory receptors have been identified in the mucosa, which respond to mechanical, thermal, osmotic and



(a) (b)

Figure 2.2: (a) Colon histology. (b) Focus on myenteric plexus

chemical stimuli. Sensory receptors in muscle respond to stretch and tension. Collectively, enteric sensory neurons compile a comprehensive battery of information on gut contents and the state of the gastrointestinal wall. Motor neurons, mainly in the myenteric plexus control gastrointestinal motility, secretion, and possibly absorption. In performing these functions, motor neurons act directly on a large number of effector cells, including smooth muscle secretory cells and gastrointestinal endocrine cells. Interneurons play the role of ”controllers”, thus being largely responsible for integrating information from sensory neurons and for providing it to motor neurons. The histology pictured in Figure 2.2 shows the distribution of the different layers and neurons described so far for a human colon.

Although the enteric nervous system can and does function autonomously, normal digestive function requires communication links between this intrinsic system and the central nervous system. Moreover, the connection to the central nervous system also implies that signals from outside of the digestive system can be relayed to the digestive system: for instance, the sight of appealing food stimulates secretion in the stomach.

Intestinal contractions

As a result of this muscular stimulation, a contractile activity and tone are produced, and intestinal contractions are generated. These intestinal contractions can be manifest in two different ways depending on their duration:

1. *Short contractions* (**phasic**), or
2. *Sustained contractions* (**tonic**)

It is well known that both the type and the spatial frequency of intestinal contractions depend on the region of the gastrointestinal tract where they are found (stomach, small intestine or colon), and the temporal patterns they present are different during *fasting* (before the ingestion of nutrients) and the *postpandrial* stage (after the ingestion of nutrients):

- On the one hand, stomach and small bowel present a common cyclical motility pattern during the fasting stage, known as the *migrating motor complex* (MMC), which has been the object of intense study over several decades. The MMC consists of three different phases: firstly, a period of motor steadiness; secondly, a period of apparently random and irregular contractions, and finally a sequence of regular phasic contractions, known as the *activity front* [50], reaching the maximum frequency of contractions in this final stage. The MMC cleanses the stomach and small intestine and so it has been termed as the gut ”housekeeper” [42], taking about 100 minutes in average (50-180 minutes) for this aim. Phase one lasts a 60% of the overall time, phase two lasts a 30%, and phase three is sustained for the remaining $5-10$ minutes. This pattern banishes if food is ingested, being replaced in the stomach by antral regular contractions and by irregular random contractions in the small bowel, sustaining this pattern during $2.5-8$ hours, depending on the size and type of food.

- On the other hand, the colon presents several forms of contractility, but the most representative patterns can be gathered into two different types: a) high amplitude wave propagated contractions, and b) irregular contractions. The former act as a propelling wave moving the intestinal content from the ascending colon to the transverse, descending and sigmoid colon and rectum; they accomplish the displacement function. The latter act by moving the bolus back and forth over short distances in a random way, providing a mixing function with the different enzymes and letting the bolus come in contact with the epithelial cells that absorb nutrients. Finally, several fluctuations in the muscular tone provide the basic scenario for good accommodation and storage.

The frequency of appearance of phasic contractions is strictly related to a consistent feature of gastrointestinal myoelectric activity consisting of an omnipresent, highly regular, and recurring electrical pattern called the *slow wave*: contractions are phase-locked to the slow wave frequency, i.e., they can only occur, although not in a mandatory way, at the crest of the slow wave. In this manner, the maximal frequency of contractile activity at a given site in the intestine is directly related to the slow wave frequency in that region. Thus, in the stomach, where slow waves occur at a frequency of three cycles per minute, the maximum frequency of phasic contractions is also three cycles per minute. Similarly the duodenal slow wave frequency and maximal frequency of phasic contractions are between 11 and 12 cycles per minute -these frequencies decline along the intestine to 9 cycles per minute in the distal ileum [67]-. A schematic view both of the contraction process for the propulsion functional pattern or peristalsis and the mixing functional pattern is rendered in Figure 2.3.

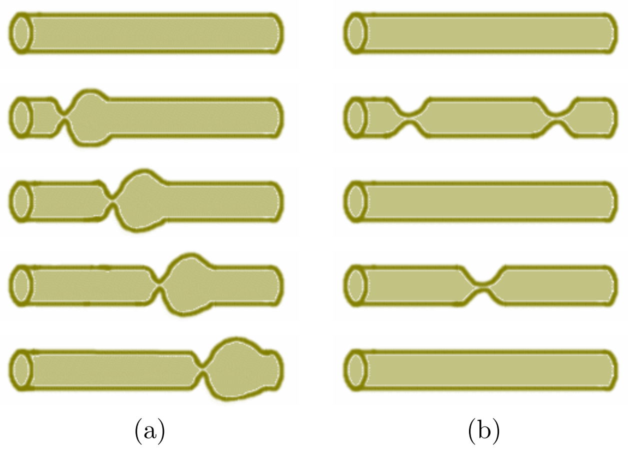


Figure 2.3: (a) Peristalsis functional pattern. (b) Mixing functional pattern.

2.2.2 Several pathologies associated with motility dysfunctions

Both the number, type and temporal distribution of intestinal contractions along the intestinal tract characterize motility patterns that are indicative of the presence of different malfunctions [42].

Gastric motor dysfunctions are essentially related to the emptying process, and *gastroparesis*, or delayed emptying, represents the most common clinical case. Its symptoms involve postprandial fullness, bloating, distention, nausea, and vomiting. Several postsurgical syndromes -such as the *early dumping syndrome*- affect acceleration of the early phase of liquid emptying, including an impairment of the motor response to feeding. It has been observed that gastroenterostomy may be associated with a specific clinical syndrome -*the Roux syndrome*- whose patients develop severe symptoms of postprandial abdominal pain, bloating, and nausea. *Diabetic gastroenteropathy* has been reported in diabetic patients showing a clinical frame of gastroperesis, involving dysphagia, early satiety, postprandial distress, constipation, diarrhea, and fecal incontinence. In these cases, fasting and postprandial antral hypomotility, and various but less consistent abnormalities of the proximal small intestinal MMC have been described. *Nonulcer dyspepsia*, *idiopathic gastroparesis*, and *dysmotility* are accompanied by postprandial fullness, nausea, and bloating in association with delayed gastric emptying. Unexplained vomiting has also been reported to be related to *the rumination syndrome*, that usually reflects a primary psychological disturbance. In addition, *accelerated gastric emptying* has been described in association with the *Zollinger-Ellison syndrome* and *duodenal ulcer disease*. Several gastrointestinal motor dysfunctions have been proved to be related to viral illnesses (*cytomegalovirus* and *herpes simplex virus gastritis*, among others) associated with the development of symptoms suggestive of gastric motor dysfunction and directly linked to disturbed emptying. Finally, gastroparesis and related symptoms may be a prominent feature of any disorder associated with autonomic neuropathy and may also be a component of both primary and secondary intestinal pseudoobstruction syndromes.

Small intestine motility disorders can be described in terms of ileus, pseudoobstruction, bacterial overgrowth, and the irritable bowel syndrome, mainly. *Ileus* is a consequence of surgery, in particular, of abdominal surgery, which presents a moderate, diffuse abdominal discomfort, abdominal distension, nausea and vomiting, especially after meals, lack of bowel movement and/or flatulence. These symptoms can be boosted when food administration fails to interrupt cyclic MMC activity, as in the case of post-vagotomy, which may contribute to the pathogenesis of *postvagotomy diarrhea*. *Mechanical obstruction* has been reported to produce patterns of nonpropagated prolonged contractions and nonpropagated clusters, as well as similar intense bursts of phasic contractions or clusters, which become more intense following meal administration. On the other hand *intestinal pseudo-obstruction* is a especially difficult and important clinical problem, whose patients present repeated episodes of nausea, vomiting, abdominal pain, distention and, on clinical grounds, are often initially suspected of having mechanical obstruction. As a result of stasis, bacterial colonization may occur with the resultant development of diarrhea, steatorrhea, weight loss, and nutritional problems. This syndrome may be the intestinal manifestation of a systemic disorder or may reflect a primary disorder of the intestinal musculature or its neural apparatus. *Bacterial overgrowth syndromes* are closely related to motility dysfunctions, since normal acid secretion and motility are the most important factors in preventing overgrowth of bacteria in the bowel, being related to certain abnormalities in the MMC. *Functional dyspepsia* is characterized by a constant pain in the stomach. It may be caused by conditions such as stomach ulcers and it is often aggravated by high acidity; it may also appear as a side-effect of drugs treating other illnesses such as arthritis and schizophrenia. Small intestinal dysmotility, with or without associated gastric dysmotility, has also been described in patients with functional dyspepsia. Finally, several motor abnormalities in patients with *irritable bowel syndrome* have been reported in the recent years. These have included abnormalities of the MMC, duodenal clusters, prominent ileal high pressure waves, and a disturbed postprandial motor response, as well as abnormal transit and impaired emptying.

2.2.3 Different ways of assessment of intestinal motility

Current techniques for assessment of small intestinal motility are multiple, complementary and specific for the area of the intestinal tract to be studied. Most of the medical imaging modalities have been used for diagnosis support purposes, including plain abdominal X-ray, computed tomography (CT), magnetic resonance imaging (MRI), functional MRI (f-MRI) and photon emission computed tomography (PET). The following enumeration holds some of the most extended techniques used so far. A deeper survey on this issue can be found in the references [67, 68, 42]:

The stomach motility disfunction has been faced in different ways. *Scintigraphic* studies with marker isotopes have resulted to be especially sensitive in demonstrating gastroparesis. *Ultrasonography* has shown to be a helpful tool for the evaluation of antropyloric function. *Magnetic resonance imaging*, *electrical impedance methods* and *electrogastrography* have been used in the evaluation of delayed gastric emptying. For the small intestine dysfunction assessment, *radiologic observation of the passage of barium* through the gut has been widely used as one of the main techniques, resulting especially useful in the detection of anatomic abnormalities. *Hydrogen breath tests*, *detection of sulfasalazine in blood*, and *scintigraphic studies with marker isotopes* have also been successfully applied in the evaluation of the time transit throughout the gut. Finally, *small intestine manometry* has shown to be highly useful for the analysis of the different parameters of the MMC, for the assessment of the presence and nature of the motor response to the meal, and for the detection of abnormal motility patterns. *Electromyography* and *intubation* have been also applied in clinical cases at the time of surgery.

All these techniques are currently in use, but the interpretation of their different recordings is difficult, so they are usually applied complementary. In some cases these techniques are highly invasive (radiation, fluor-markers, intubation, etc). In other cases, such as conventional endoscopic techniques, they are limited by the length of the visual sensors related to the length of the gut -typically 3.5-7.0 m.- and its complex looped configurations [30]. Small intestinal manometry is, nevertheless, widely accepted as the most reliable technique for intestinal motility assessment so far. Small intestinal manometry consists of the measurement of the pressure in certain points of the small intestine by means of multiple pressure sensors distributed along a thin tube that is introduced through the esophagus, giving as a result a graph with the contractile activity presented as variations in pressure detected by the sensors. Two main drawbacks are associated with this technique: On the one hand, it is an invasive test which carries discomfort problems for the patient, the presence of medical staff is needed throughout the whole process and hospitalization is required. On the other hand, its clinical value is limited to the examination of severe intestinal motor alterations, and it shows a lack of sensitivity over certain types of intestinal contractions that cannot be detected by means of this method.

In this scenario, wireless capsule endoscopy video analysis appears as a promising technique that overcomes most of the drawbacks previously described. The next section is devoted to the explanation of the technical framework of capsule endoscopy.

2.3 Wireless capsule endoscopy

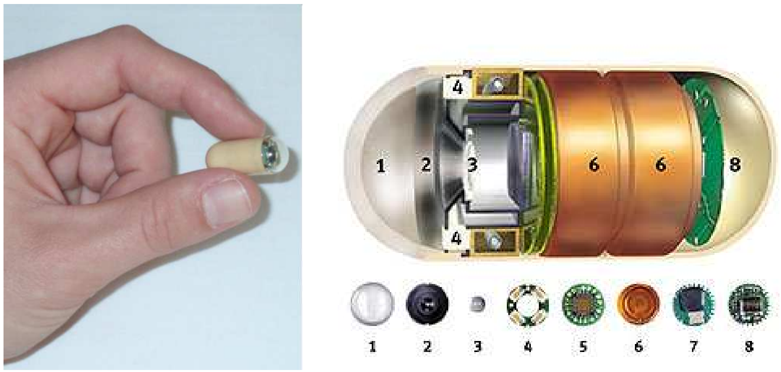
Wireless Capsule Endoscopy (WCE) [45] is a novel imaging technique which allows the visualization of the whole intestinal tract. The WCE procedure consists of the ingestion of a capsule, with a complete visualization device attached to it, which registers a video movie of its trip throughout the gut. This video is emitted by radio frequency and recorded into an external device carried by the patient. Once the study is finished, the final record can be easily downloaded into a PC with the appropriate software for its posterior analysis by the physicians.

Recently, several works have tested the performance of capsule endoscopy in multiple clinical studies. Some of these clinical scenarios include *intestinal polyposis* and the diagnosis of *small bowel tumors*, o*bscure digestive tract bleeding*, *Crohn's disease* and *small bowel transplant surveillance*. Some authors have analyzed the validity of capsule endoscopy for detection of *small bowel polyps* in hereditary polyposis syndromes, concluding the clinical value of the methodology for the *familial adenomatous polyposis* and that CE could be used as a first line surveillance procedure for *Peutz-Jeghers syndrome* [72]. Following a different line of research, other studies have focused their efforts on the evaluation of the clinical effectiveness of wireless capsule endoscopy in the management of patients with *obscure digestive tract bleeding* [2, 24], concluding that capsule endoscopy seems best suited for patients with obscure gastrointestinal bleeding who have undergone inconclusive standard evaluations and in whom the distal small bowel (the portion beyond the reach of a push enteroscope) needs to be visualized. In this direction, comparative studies have been published showing the main advantages and drawbacks of CE in comparison with push enteroscopy, sonde enteroscopy and intraoperative endoscopy in this kind of pathologies [5]. CE has also demonstrated to be highly effective for diagnosing several pathologies associated with *Crohn's disease* which are frequently missed by conventional tests [33, 49]. Other researchers have reported CE to be a helpful tool for post-transplant surveillance of small intestine. CE displayed post-transplant changes in the villi that ranged from blunted white villi seen at day 20 to normal villi observed at 6 months [26]. A more exhaustive review and summary about the current literature regarding wireless capsule endoscopy can be found in the following bibliographic references [30, 70, 32].

2.3.1 Device description

WCE was first developed and introduced by *Given Imaging Limited* under the trade mark of *M2A Given Diagnostic Imaging System* -M2A are the acronyms for "mouth to anus"-, being cleared for marketing for the first time through the U.S. government on August 1, 2001 [17]. This technology is performed by means of three main components: the capsule, the registration device and the proprietary data analysis software.

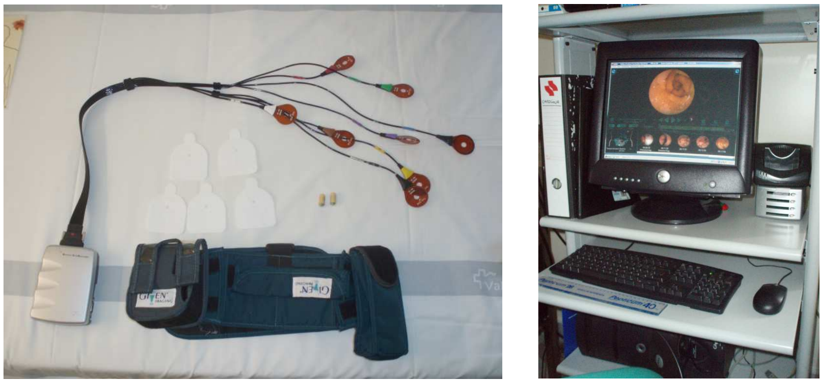
The capsule is an ingestible device equipped with all the suitable technology for image acquisition, including illumination lamps and radio frequency emission. Figure 2.4 shows a graphical scheme of the capsule together with the distribution of its components in scale. It consists of an external envelope with a transparent dome front sizing 11x30 mm (1), which contains a lens holder (2) with one lens (3), four illuminating leds (4), a complementary metal oxide silicon (CMOS) image sensor (5), a battery (6), an application-specific integrated circuit (ASIC) transmitter (7), and a micro-antenna (8) [45] . The field of view of the lens spans 140-degree, very similar to that of standard endoscopy. The illuminating lamps are low consume white-light emitting diodes (LED). The video images are transmitted using UHF-band radiotelemetry to aerials taped to the body which allow image capture, and the signal strength is used to calculate the position of the capsule in the body. Synchronous switching of the LEDs, the CMOS sensor and the ASIC transmitter minimize power consumption, which lets the emission of high-quality images at a frame ratio of 2 frames per second during 6 hours. The capsule is completely disposable and does not need to be recovered after use, being expelled by the body 10 to 72 hours after ingestion.



(a) (b)

Figure 2.4: (a) The M2A© camera. (b) Camera components.

The registration device consists of a set of aerial sensors for the RF signal reception, connected to a CPU with a hard disc for data storage. The registration device is carried by the patient fastened into a belt, altogether with a battery for power supply. The aerial sensors are taped to the body of the patient, forming an antenna array which collects the signal transmitted by the capsule and sends it to the receiver. The received data is subsequently processed and stored in the data storage by the CPU. Figure 2.5 shows a picture of (a) the external device, and (b) the workstation.



(a) (b)

Figure 2.5: (a) The external device, the belt and aerials, which are to be taped to the patient's body. (b) The workstation with the software installed.

The proprietary software is installed into a PC workstation. It allows the physicians to retrieve the data from the recorder and to transfer it to the workstation for additional processing and visualization on the display. The performed study can be stored independently on the workstation hard disk or be registered into a CD, a DVD or any other storage device, being ready for visualization and annotation on any computer in which the displaying software has been previously installed.

2.3.2 Pros and cons of capsule endoscopy

Capsule endoscopy overcomes most of the drawbacks related to manometry and other intestinal motility assessment techniques. It is much less invasive, since the patient simply has to swallow the capsule, which will be secreted in the normal cycle through the defections. Depending on the type of study, during the whole process of the capsule endoscopy video recording the patient may be laid on, sat down, walk or even lead an ordinary life. No hospitalization is needed nor special staff, since the video recording is performed without the need of any type of interaction. One of the main technological breakthroughs that this technology allows is the direct study and visualization of the entire small intestine, something that was not possible with the previous techniques so far [8]. In terms of cost, a US economic analysis in 2003, which was funded by Given Imaging Ltd., concluded that CEs per unit cost as a diagnostic tool for small intestine bleeding was comparable to that of other current endoscopic procedures (for example, the average direct cost was US$517 for CE and US$590 for enteroscopy)[17].

Nowadays,however, capsule endoscopy cannot replace any of the other procedures in a general and exclusive way, and the investigation of the small intestine should include capsule endoscopy together with the rest of technologies. Since the capsule has no therapeutic capabilities, any lesion discovered by capsule endoscopy must be further investigated using other standard techniques. In addition, the capsule use is contraindicated in patients with cardiac pacemakers, defibrillators or implanted electromechanical devices (due to the risk of radio-interference with the UHF signal), and in those patients with known or suspected obstruction or pseudo-obstruction (due to the risk of causing bowel obstruction) [2]. Nowadays, the number of capsule studies performed worldwide is still small and it is too soon to appreciate the sensitivity and specificity of this technique. The study of a capsule endoscopy video takes between 1 and 2 hours, which means a heavy load for the physicians. In this sense, the research on computer-aided and intelligent systems, such as the one presented in this work, results highly interesting for the development of this technology. The limited information currently available, however, is quite promising.

2.3.3 Common examples of capsule video images of the gut

Capsule endoscopy video images are quite similar to those acquired by classical endoscopic techniques. Each video frame consists of a 256x256 pixel image, rendering a circular field of view of 240 pixels of diameter, which spans 140-degrees, in which the gut wall and lumen are visualized -see Figure 2.6-.

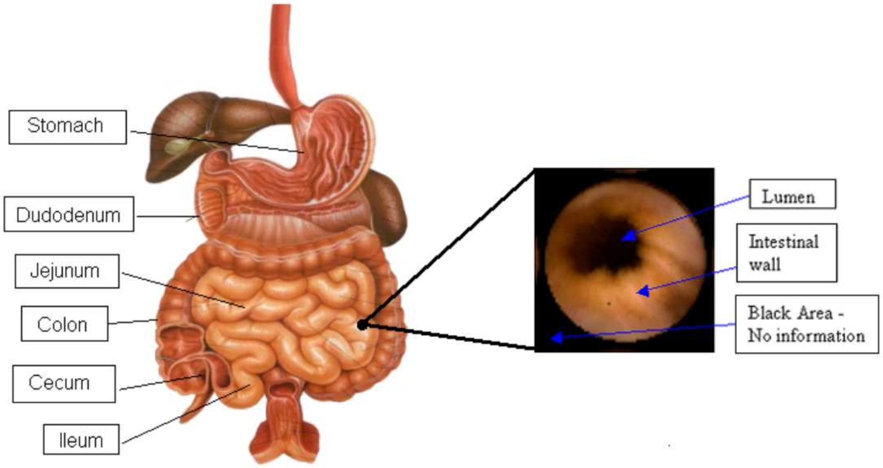


Figure 2.6: Appearance of a frame in capsule video endoscopy. The intestinal lumen and walls are rendered in a circular field of view.

Typically, the aspect shown by each part of the gastrointestinal tube presents differences in texture, shape and color, is patient-dependant and presents variability with several pathologies. For instance, gastric images appear with the common folded shape of the stomach wall and a pink tonality. This folded pattern is replaced in the jejunum by a plain pattern, describing star-wise distributed wrinkles when a contractile event occurs and showing a color tonality closer to orange. A hybrid pattern is shown in the duodenal zone, and almost no visibility is achieved in the cecal area. Figure 2.7 shows a set of sample images such as those described previously.

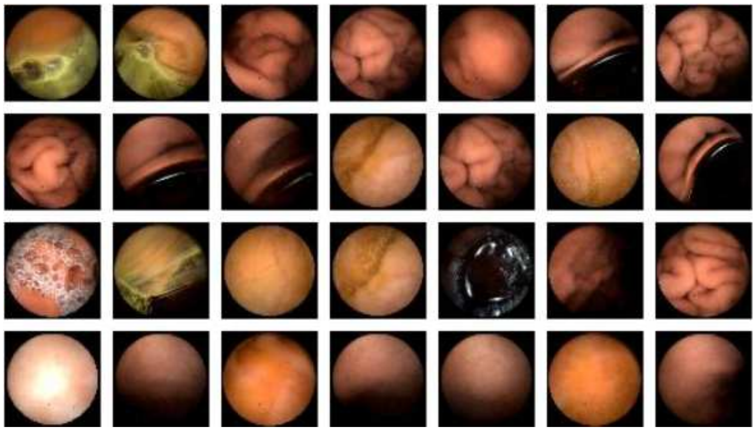


Figure 2.7: Different examples of capsule endoscopy video images.

The multiple patterns and appearances of the different images, and sequences of images, that could be found in intestinal motility studies will be the object of deep analysis in the next chapter.

2.3.4 Annotation of intestinal contractions in capsule endoscopy

Traditional procedure: Manual annotation

Video annotation is an essential characteristic of the capsule endoscopy technology. The general protocol is as follows: Once the study is downloaded to the workstation, the physician visualizes the whole video, selecting those frames where the object of interest is present (bleeding, polyp, wound, etc). Once the whole video is annotated, the expert can analyze, if necessary, each labelled frame in order to obtain information for clinical purposes, diagnosis, etc.

For the specific case of intestinal contractions, the expert visualizes the zone of interest where contractions are searched, and labels those frames where a contraction event is detected. Once this procedure is finished, the temporal pattern of intestinal contractions obtained is analyzed, inferring the corresponding conclusion about suitable intestinal motility dysfunctions. Figure 2.8 shows a snapshot from the visualization tool provided by Given Imaging: Rapid Viewer. On the left, a time line indicates the real time (time in the real life experiment) and relative position of the frame in the video. Together with the time line, a set of findings labelled by the expert are shown in their relative position. The main screen renders the video sequence showing the gut wall and lumen. In addition, the relative position of the camera in the human body and a graph of illuminance variation can be visualized, if desired, in the lower part.

The annotation process of intestinal contractions is, thus, not straightforward, time consuming and stressful. A typical study may contain up to $50,000$ images obtained at a frame ratio of 2 images per second (6-7 hours of capsule recording). The visualization time can be adjusted from 5 to 25 frames per second. At a typical visualization rate of 15 frames per second, the specialist needs at least one hour only for visualization purposes, without taking into account the time consumed in labelling the findings. In addition to this, the prevalence of contractions in video is very low. A typical ratio of 1:50 implies the presence of only 1000 findings in a whole video of 50,000 frames, making this procedure not feasible as a clinical routine. In other words: while the information provided by manual annotation of video capsule endoscopy images is enormously useful for the inference of important conclusions about motility in multiply clinical studies, and while CE has shown to be a contrastably helpful technique for motility assessment in patients in which no other technique would carry an improvement in analysis and/or diagnosis, the very procedure of manual annotation does not result feasible due to all the drawbacks previously exposed.

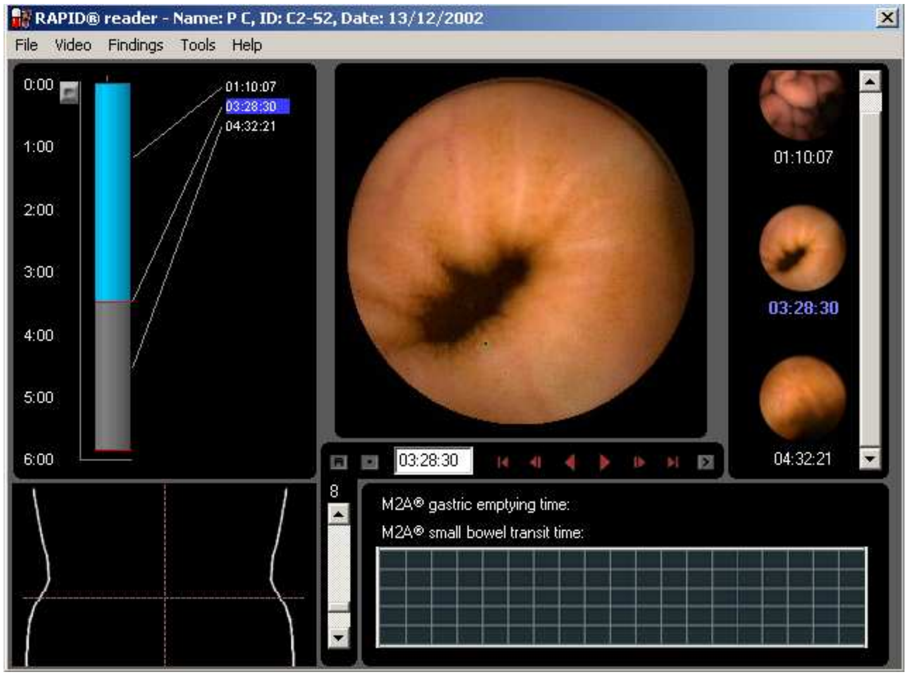

RAPID® reader - Name: P C, ID: C2-S2, Date: 13/12/2002
File
Video
Findings
Tools
Help
0:00
1:00
2:00
3:00
4:00
5:00
6:00
01:10:07
03:28:30
04:32:21
03:28:30
8
M2A® gastric emptying time:
M2A® small bowel transit time:

Figure 2.8: a) Annotation tool. The specialist saves into an external file the time of the frames of interest.

Requirements for an automatic annotation system

Regarding the specifications described in the previous sections, an automatic annotation process of intestinal contractions must accomplish the following constraints: the system and the trained expert must provide an equivalent pattern of intestinal contractions. This means that the system output must provide a good pattern of number of contractions per unit of time, in terms of inter-observer variability.

Chapter 3

Intestinal contractions in Video Capsule Endoscopy

The temporal pattern of intestinal contractions along the video is the main source of information for motility assessment in capsule endoscopy. In the previous chapter, we introduced the procedure of annotation the specialist must perform in order to obtain this pattern. We pointed out that the video annotation procedure consists of the labelling of the video frames corresponding to intestinal contractions. Consequently, the specialist is obliged to visualize the whole video in order to be able to label all the existing findings, in a tedious and time consuming procedure. The latter situation was stated as the main drawback of this technique, which turns it into a not feasible clinical routine. Our aim is to provide the specialist with the temporal pattern of intestinal contractions in an automatic way, allowing the specialist to take profit of the existing motility information, and making this clinical routine feasible.

In order to design a machine learning system for the automatic detection of intestinal contractions, two main pre-requisites must be achieved, namely: firstly, a good set of *findings* from which a generalizable pattern could be inferred, and secondly, a good set of *descriptors* for those patterns. In this chapter, we focus on the former, i.e., the introduction of the visual patterns associated with intestinal contractions, leaving the latter for a specific study in the next chapters. Section 1 introduces the main aspects regarding the fundamental elements which can be found in gut images from capsule endoscopy, in order to state the visual paradigm of intestinal contractions which is developed in section 2. The multiple sources of variability of these patterns are analyzed in section 3. We gather all this information in order to build up and present a taxonomy for intestinal contractions in section 4.

3.1 What is the gut like?

3.1.1 The intestinal walls

The gut can be seen as a long tube (about 4-7 meters) with its beginning in the mouth and its end in the anus. The constitution, function and visual appearance of each different part of the gut is multiple, and highly depends on the physiological task to which each part is devoted. In a general way, four main divisions can be enumerated (*proximal* -closer to the mouth- to *distal* -closer to the anus-): esophagus, stomach, small intestine and colon. The small intestine presents three different zones: duodenum, jejunum and ileum. The colon can be studied, in its turn, as ascending colon, transversal colon, descending colon and rectum. The typical visual aspect in capsule endoscopy of each one of these different parts is pictured in Figure 3.1.

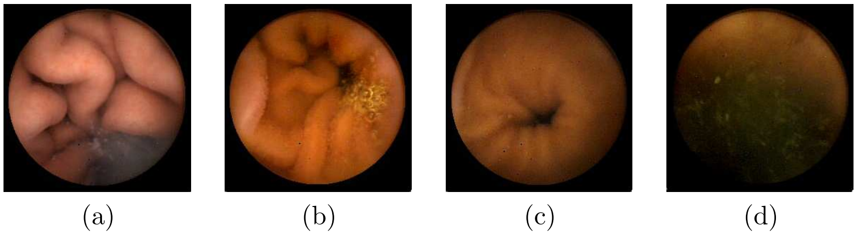


Figure 3.1: Typical visual aspect of (a) stomach, (b) duodenum, (c) jejunum, and (d) cecum in capsule endoscopy

The transit of the pill through the esophagus is very fast -typically two or three seconds- with no useful information present in this area. The stomach is the first zone from which the specialists can obtain clinical information. The stomach has the shape of a sac with folded walls; these folded walls increase the overall surface of the stomach, allowing a higher performance in the physiological processes involved. It usually presents a pale color close to pink. The typical aspect of the stomach walls in capsule endoscopy is shown in Figure 3.1 (a). The duodenum is the first part of the small intestine. Outside the stomach, the intestinal lumen is visualized as a tunnel delimited by the intestinal walls. The ovoid shape of the capsule lets itself keep pointing at the longitudinal direction in its trip throughout the gut, randomly in proximal or distal orientation -i,e., pointing towards the mouth or the rectum-. The intestinal walls are still folded in the duodenum, but with softer folds, presenting a color ranging from pink to orange. A frame showing the common appearance of the duodenal walls can be observed in Figure 3.1 (b). The jejunum and ileum present a similar appearance: the intestinal walls are plain in the relaxation state, but they contract creating folds during the contractile activity. The color appearance in these areas of the small intestine usually ranges from orange to red. A typical post-duodenal frame is shown in Figure 3.1 (c). Finally, the colon is the last part of the intestinal tube. The processes of assimilation of nutrients which take place in the colon are slow, in comparison with the previous stages. Moreover, since all the fecal content is released in the colon, the visual aspect is dark and the visualization quality is poor.

Figure 3.1 (d) exemplifies this situation.

3.1.2 Intestinal content

In addition to the former, capsule endoscopy allows the visualization of all the intestinal content through the gut. It may be classified in three main groups regarding their physiological origin and visual appearance:

- The *food in digestion* appears in small pieces through the gut. They are mixed with the intestinal juices, presenting a color ranging from brown to green.
- Intestinal juices are present as a *turbid liquid*, which performs a hindering effect in the visualization field. Its color presents a high variability, ranging from brown to yellow and green. Its density and opacity are also variable, and sometimes it can be dense enough to block the whole field of view of the camera.
- Intestinal juices may produce *bubbles* as well, which differ in shape and color from the turbid liquid defined above. These bubbles range from yellow to green, mainly centered in yellow.

Figure 3.2 renders typical examples of intestinal content.

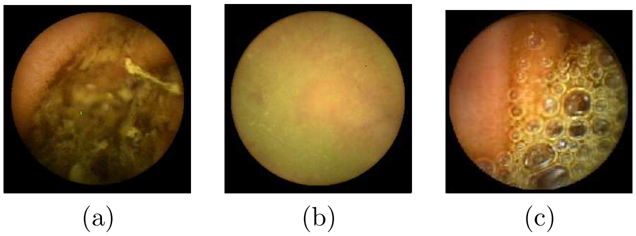


Figure 3.2: (a) Food in digestion. (b) Turbid liquid. (c) Bubbles

3.2 Visual paradigms of intestinal contractions in capsule endoscopy

3.2.1 Region of interest: post-duodenum to cecum

As we stated in the previous chapter, different motility patterns are linked to different motility disfunctions throughout the stomach, the small intestine and the colon. One of the first decisions to be taken in our research work concerned the scope of study, i.e., whether to take into account the whole gut or to restrict the analysis to a specific region. Since no previous work has been published so far about any research on intestinal motility with capsule endoscopy, being this Ph.D. thesis, to the best of our knowledge, the first approach in this fieldwork, we focused our efforts on a preliminary study of the problem in terms of the analysis of the different scenarios which may be present in motility analysis with capsule endoscopy. For this aim, the specialists provided us with a set of capsule endoscopy videos to be analyzed. After studying these videos, we accomplished a series of meetings in order to define the region of interest of our research. During these meetings, the experts exposed their main targets relative to the different motility patterns they expected to validate with the new technique of capsule endoscopy, receiving our feedback about the complexity of its analysis in terms of computer vision. Gathering the important background knowledge obtained through these multidisciplinary sessions, we decided to focus our efforts on the analysis of the intestinal motility in the post-duodenal region of the small intestine: *the region spanning jejunum and ileum.* This decision was underpinned both in the specific clinical interest of this region and its special suitability for capsule endoscopy, supported by the following reasons:

1. Regarding the clinical interest, small intestine motility is a widely spread technique which has been accepted as one of the most useful sources of information for motility assessment. Small intestine manometry is accepted as the most reliable technique in this field [42]. In this sense, the long experience in small intestinal manometry of the group of experts, with whom we have developed this research, allowed us to have direct access to the state-of-the-art information and knowledge in small intestine motility [50].

2. In addition to the former, the analysis of small intestine motility allowed the experts to accomplish an important full validation study of capsule endoscopy, comparing this novel and never used before technique for motility assessment with the gold-standard of manometry.

3. Moreover, the use of capsule endoscopy allows the experts to have access to motility information far from the duodenum and the proximal jejunum, the regions of study which manometry is restricted to, completing the validation study mentioned above by reaching the extension of the area of analysis just to the ileum-cecal valve.

4. In terms of the technique suitability, stomach, duodenum and colon present relevant complexities for motility assessment based on visual analysis, in comparison with the proposed region of study. On the one hand, the stomach is not rendered in capsule endoscopy as a tube, like the small intestine, and only the gastric walls are visualized, with no lumen. In this scenario, the motility activity is shown as a sudden movement of the gastric wall, and since the capsule is freely embedded in the gastric sac, its movement is apparently random, and the motility patterns related to it are not directly generalizable from a visual point of view. On the other hand, the analysis of the motility patterns in the colon is hindered by the poor visibility conditions due to the fact that all the fecal content is emptied in the cecum. Moreover, the digestive process in the colon may last 7 to 48 hours, 20 to 280 minutes in the stomach, and 45 to 140 minutes in the small intestine. Since the capsule's battery usually employed

in experiments with capsule endoscopy typically performs during 6 to 7 hours from its ingestion, this power supply should not be enough to track the whole colon. Finally, we decided to reduce the area of analysis excluding the duodenal section, due to its peculiar folded morphology which shows a high deviation of the visual pattern present in the rest of the small bowel. This exclusion allowed us to face a physiological scenario sharing a common morphology and appearance of the intestinal walls and contractile activity, with no loose of functional information regarding motility.

3.2.2 Visual paradigms

Intestinal contractions respond to two main categories from a physiological point of view, as described in the previous chapter, namely: phasic and sustained. Phasic contractions in the small intestine are visualized as a sudden closing of the intestinal lumen in a concentric way. The maximum frequency of these events has shown to be between 10-12 contractions per minute [50, 67], although this value varies depending on the region of the intestine and whether the patient has previously ingested or is fasting. Sustained contractions are produced by muscular tone, and can be visualized as a continuous closing of the intestinal lumen with a high variability in length. Figure 3.3 shows a characteristic example of a phasic contraction. Figure 3.4 shows a characteristic example of a sustained contraction.

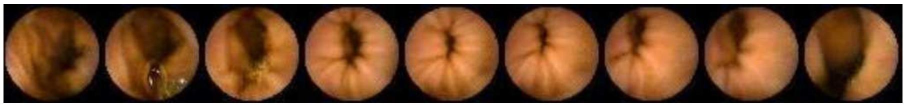


Figure 3.3: Phasic contraction.

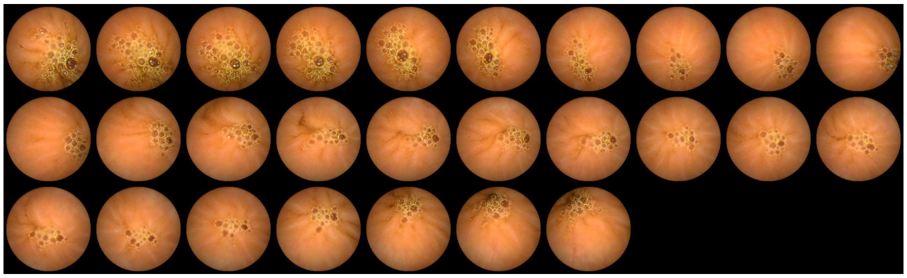


Figure 3.4: Sustained contraction.

With the aim of inferring information about their descriptive paradigms, we asked the specialist to label a set of 10 videos, annotating all the existing contractions of any kind in order to build up a knowledge database. Once these studies were finished, we gathered all the annotated frames in two separate and exclusive sets, namely, phasic and sustained. The further exhaustive analysis of this information led us to the definition of the paradigm of phasic and sustained contractions in the following way:

Phasic contractions

This kind of intestinal contractions are characterized by a sudden closing of the intestinal lumen, followed by a posterior opening. This open-closed-open scheme resulted to span 4-5 seconds -corresponding to 8-10 frames and providing an acquisition rate of 2 frames per second-. The specialist usually labelled the frame where the lumen was completely closed as a phasic contraction. Using this information, we finally defined a phasic contraction as *the central frame in a sequence of 9 frames where the pattern of an open-closed-open lumen is present.* If the closing of the intestinal lumen is complete, this event is categorized as an *occlusive* contraction, while, when the closing of the intestine is not complete, allowing a small portion of lumen at the central frame, it is categorized as a *non-occlusive* contraction. Figure 3.5 shows the referred paradigm of occlusive phasic contractions for three different sequences. Figure 3.6 shows the referred paradigm of non-occlusive phasic contractions.

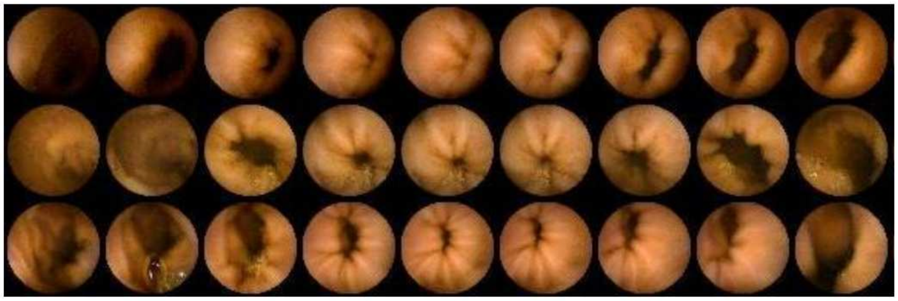


Figure 3.5: Three sequences of occlusive phasic contractions.

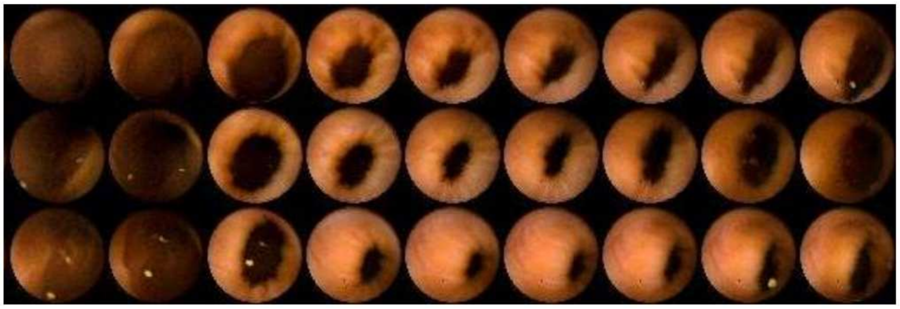


Figure 3.6: Three sequences of non-occlusive phasic contractions.

Sustained contractions

The pattern of sustained contractions corresponds to a sequence of a closed lumen in a undefined number of frames. This pattern is highly recognizable for the presence of the characteristic wrinkles which the continuous muscular tone produces when the intestinal walls are folded. Figure 3.7 shows the referred paradigm of sustained contractions.

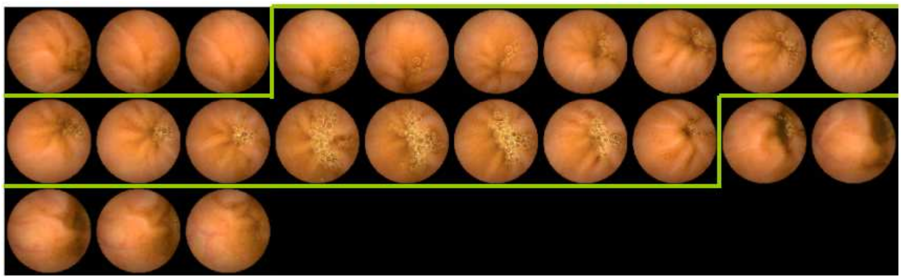


Figure 3.7: One sustained contraction spanning for 15 frames.

3.2.3 A protocol for video annotation of phasic and sustained contractions

Once the paradigms of phasic and sustained contraction were defined, the annotation protocol was reformulated in the following way:

- For *phasic contractions*, the experts must label the *central frame* of a contraction sequence. In order to make this protocol feasible, and assuming that in many cases it is impossible to define this central frame with precision, we stated that an intestinal contraction was present *within the sequence defined by* $+/-5$ *frames* around the labelled one. This threshold is trivially underpinned by the physiological and empirical reasons exposed in the previous sections (i.e., maximum frequency of contractions, statistical analysis of the annotated frames in the database, etc.)
- For *sustained contractions*, we asked the specialists to use a different file, indicating for each contraction its beginning and its end.

3.3 Elements with influence in the variability of the paradigms

The intestinal contractions shown in Figure 3.5 and 3.6 correspond to optimal examples that exactly match the paradigms defined above. Unfortunately, this pure pattern is perturbed by both the *free movement of the capsule* and the presence of *intestinal content*.

Impact of the free movement of the capsule within the gut

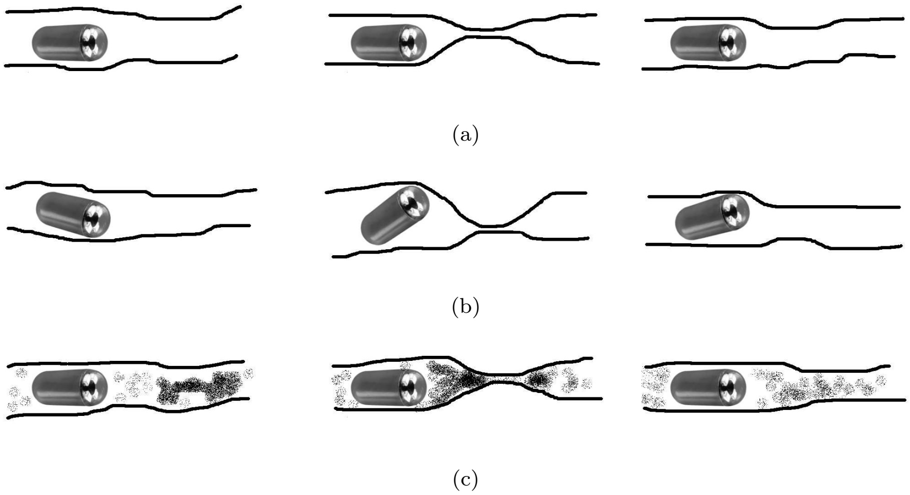
The position of the camera in the intestinal lumen during the contractile activity is not steady. Since the capsule is freely moving into the gut, multiple changes in direction (namely, focusing the intestinal lumen or the lateral intestinal wall) and orientation (i.e., facing the proximal or distal parts of the tract) are performed. As a result, the camera is not always focusing the central part of the lumen -see Figure 3.8 for a graphical representation-, and this yields to a high variability of the visual patterns obtained in the video sequences, which can be summarized in the following categories:

- *Incomplete contractions*: The central frame shows the intestinal lumen but it is not centered in the image. Part of the lumen lays out of plane.
- *Lateral contractions*: The first or the last part of the sequence is missed, but the central frame is present.
- *Out-of-plane contractions*: The central frame of the contraction is completely out of plane. The contraction event is deduced nevertheless by the remaining part of the sequence.

Figure 3.9 renders a representative set of incomplete, lateral and out-of-plane examples of this situation with three intestinal contractions labelled by the specialists.

Impact of the intestinal content

As was explained in the previous section, the presence of intestinal content, basically small pieces of food, turbid liquid or bubbles, has a negative effect in visualization. This impact has different weights depending on the type of intestinal content. While small pieces of food are not usually relevant enough to mislead the assessment of intestinal contractions by the experts, the presence of turbid liquid and bubbles may cause serious difficulties. If the turbid liquid is not very dense, or the region with bubbles is small, the intestinal lumen can be tracked and the contractions present in the video can still be labelled. But as far as the turbid liquid or the presence of bubbles are to cover the field of view of the camera, important information for assessment is lost. On the one hand, an analogous phenomenon to the angular deviation of the camera may be performed, since only a half of the sequence -incomplete-, part of the frames -lateral-, or the central part -out-of-plane- may be affected by the presence of turbid or bubbles. On the other hand, the presence of turbid or/and bubbles may affect the whole sequence; depending on the opacity of turbid and bubbles, the intestinal lumen may be still traceable, or completely lost. Figure 3.10 shows several examples of labelled intestinal contractions undergoing the described hindering elements.



(a)

(b)

(c)

Figure 3.8: Graphical representation in three steps (before, at the time of, and after the contraction event): (a) The paradigm of a phasic contraction, (b) the camera pointing towards the intestinal wall, and (c) the presence of turbid liquid hindering the visualization. These patterns match the sequences rendered in Figures 3.5, 3.9 and 3.10, respectively.

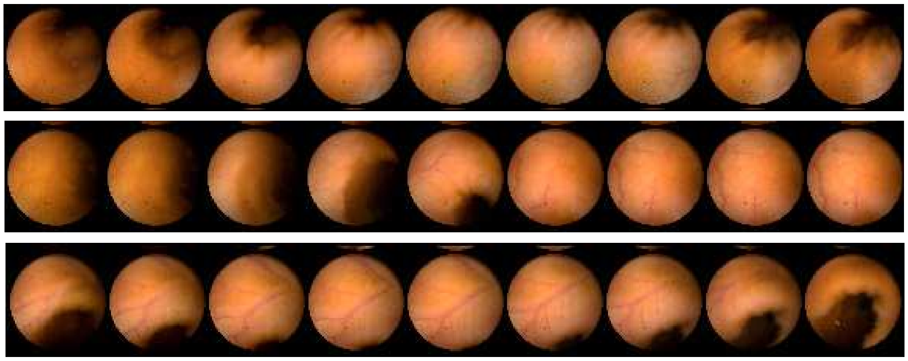


Figure 3.9: Three sequences of intestinal contractions showing the incomplete, lateral and out-of-plane patterns due to the random camera orientation.

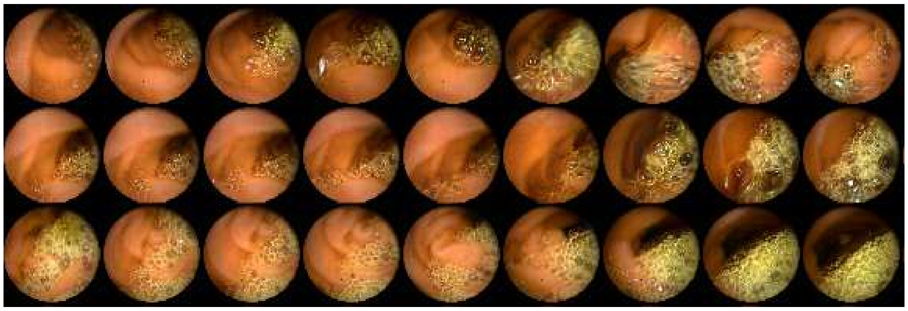


Figure 3.10: Three sequences of intestinal contractions with presence of turbid liquid, which hinders the correct visualization of the event.

3.4 A taxonomy of intestinal contractions in video capsule endoscopy

Following the previous analysis, we can state a taxonomy of intestinal contractions in video capsule endoscopy from two different perspectives: the *physiological* sources of variation, and the different patterns produced by the *camera motion* in the gut.

From a physiological point of view, intestinal contractions can be divided into the following three categories:

Occlusive contractions They correspond to sequences of phasic contractions, rendering an open-close-open lumen. The central frame of this kind of contractions shows the intestinal walls concentrically contracted, shutting the intestinal lumen completely.

Non occlusive contractions They correspond to the same pattern of occlusive contractions, but they show part of the lumen in the central frame. The origin of this kind of contractions is based on the physiological fact that the intestinal walls do not perform enough pressure during the contractile activity -for this reason, these non-occlusive contractions are hard to detect with techniques based on manometry-.

Sustained contractions They correspond to tonic contractions, where the closed lumen can be visualized during the whole sequence. The actual duration of tonic contractions can range from 5 to several seconds, i.e., from 10 frames on.

From the point of view of the different patterns strictly associated with the free movement of the camera throughout the gut, contractions can be divided into the following three categories:

Complete contractions The camera is continuously focusing the intestinal lumen, and the paradigmatic pattern of the contraction is clearly tracked throughout the whole sequence. This corresponds to a sequence showing the contraction from its beginning until its end in a complete way.

Incomplete contractions The camera moves in the very instant that the contraction takes place, pointing to the intestinal walls instead of the intestinal lumen. Thus, the lumen does not appear centered in the image, but laterally displaced, and as a consequence of this, part of the lumen can be found out of the scope of the visual field.

Out of plane contractions The camera suddenly moves and points to the intestinal walls at the moment when the contraction takes place. For this reason, half of the sequence contains images from the lumen, and half of the sequence does not. This kind of sequence can be viewed as a half complete contraction.

3.5 Two sources of capsule endoscopy studies: fasting and postprandial

The main source of video studies for motility analysis we were provided with is split in two separate clinical scenarios, namely: *fasting* and *postprandial*.

Fasting videos are provided by studies where the patients had been asked not to ingest anything within the 12 hours previous to the study. In *postprandial* studies the volunteers had been asked not to ingest anything within the 12 hours previous to the study in the same way as fasting, but during the capsule study, different kinds of nutrients were introduced in different places of the gut. As long as all the provided studies were accomplished in healthy volunteers with no apparent motility disfunctions, the clinical value of both the fasting and postprandial videos used in this research was circumscribed to the specific purpose of validation.

The fundamental difference between fasting and postprandial videos relies on the higher amount of turbid liquid in the postprandial cases. Due to the nutrients infusion into the gut, basically performed through a probe into the stomach, duodenum or jejunum, a higher presence of turbid liquid is undergone. On the other hand, the presence of bubbles as a specific shape of intestinal juices is shown to be more frequent in fasting videos.

Part II

AUTOMATIC DETECTION OF PHASIC CONTRACTIONS

Chapter 4

Introduction: Looking for classification alternatives for an imbalanced problem

A usual and widely extended way for referring the two different types of samples which can be found in a 2-class problem consists of assigning the label *positive* to one of the classes and the label *negative* to the other one. Although this kind of labelling may be randomly assigned, there exists a large number of situations in which one of the classes has got a central relevance. A typical example of this can be found in a clinical a diagnosis procedure performed over a population of patients showing a certain symptomatology. In this case, it is usual that the positive label is reserved for those patients which are likely to suffer the suspected pathology, while the negative label is assigned to those patients who do not undergo the pathology. In the problem of face detection in generic images, the positives a those regions of the image to which the system assigns a high likelihood of containing a face, while the negatives are the remaining regions.

If the *prevalence* of the positives is low (the number of negative samples far outnumbers the positive instances) we say that we are facing an *imbalanced classification problem* or the classification problem of an *imbalanced data set*. This occurs when the ratio between the number of ill subjects and not ill subjects in the population of study is low, the ratio between the number of regions containing faces and regions not containing faces for a given image is low, etc. In highly imbalanced problems where both classes are not separable, neglecting the positive class -classifying every sample as a negative- may yield to the lowest classification error. It is straightforward to create a classifier having a 99% accuracy (or 1% error rate) if the data set has a 99% majority class by simply labelling every case as belonging to the majority class, as a direct application of the principle of Occam's razor: The simplest solution among all other possible is the one selected. But in some real life imbalanced data set problems, we cannot afford the luxury of misclassifying samples of the minority class. One of the most graphical examples of this situation is cancer detection. No one can imaging a doctor that relies in this kind of "technique" for cancer diagnosis, although cancer prevalence -the rate of cancer in population- is low. It is absolutely crucial to detect each and every positive case in order to be able to have the patient a chance to save his life. If the doctor assessment is wrong, the *cost* of telling healthy patients that they may suffer cancer will not be as serious as rejecting a true cancer case. By the other hand, the false positives number, the number of patients not undergoing cancer who are told to be ill, should be as low as possible, to keep the diagnosis *predictive value*, and put confidence in doctor decisions. For other cases, the cost of a positive misclassification is not so high, and a tradeoff may be stated by the expert.

Finally, several consequences may be drawn from the fact that the minority class has *few* samples: 1) Positive samples may be so few in number that they may not be descriptive enough of the positive class -they lack generalization power over the positive class-. 2) As they are also few in comparison with the number of samples in the negative class, the performance of a classification system based on the classification error rate may be affected by means of the inclusion of a large number of *false positives*, which is strictly linked to the nature of the learning processes. The former consequence may not always be overcome. Sometimes, as usually happening in bio-molecular problems, the number of positive samples is strictly reduced by economic reasons. In some other contexts, this lack of discriminant power in the feature set may be tackled with the selection of a better set of descriptors, facing the context of feature extraction. The latter consequence is inherent to the skewed description of the problem -related to low prevalence of positives-, and must be tackled with specific strategies.

4.1 Some real life examples

Recent works have studied multiple strategies for the resolution of specific imbalanced problems in several real scenarios. For example, we highlight a few representative instances reported in the fieldwork of direct marketing solutions [57], speech recognition [58], spam detection [52], cellular calls fraud detection [31], detection of rooftops in overhead imagery [62] and oil spills in satellite radar images [53]. Basically, three different strategies can be enumerated, namely: the use of *stratified sampling*, which is based on re-sampling the original data sets in different ways: under-sampling the majority class or over-sampling the minority class, the use of *cost-sensitive* implementations, which are based on the assignation of different classification costs for the samples of each class, and finally, the design of *case-oriented* classifiers, modifying a specific part of the basic classification algorithm (SVM, AdaBoost, naive Bayes, decision trees,...) in order to suit it well to the problem of analysis.

Chawla et al. [20] investigated several issues concerning decision trees and imbalanced data sets: the impact of probabilistic estimates, pruning, and the effect of stratified sampling on the area under the ROC curve (AUC). They introduced a specific metric *m-estimate* as a smother of the probability assigned in each leaf, and two different over-sampling strategies (sample replication and synthetic minority over-sampling, SMOTE) and classical under-sampling. Akbani et al. [3] used SVM with imbalanced data sets. They combined over and under-sampling with the sensitivity control technique for SVM developed by [80]. Yang Liu et al. [58] used sampling, ensemble sampling, bagging and boosting strategies for detection of sentence boundaries and interruption points in speech recognition. Kolcz et al. [52] studied the impact of duplicate data in the quality of data mining performance for the example of spam detection, showing that the augment of duplicates reduces the AUC of classifiers such as naive Bayes and perceptron with margins. Fawcett & Provost [31] analyzed the imbalanced problem in the work field of cellular calls fraud detection. They used a rule-based system in order to categorize fraudulent calls for daily accounts. For dealing with skewed class distributions the authors used stratified sampling and empirical threshold adjustment with non uniform misclassification cost.

Maloof [62] examined the use of cost-sensitive learning algorithms and ROC curves analysis to cope with imbalanced data sets. Imbalanced data sets problems can be handled in a similar manner when misclassification costs are unequal and unknown. Over-sampling and under-sampling techniques generate the same ROC curves than varying the decision threshold or the cost matrix, which is shown to be linked also with prior class probabilities. Domingos [27] defined the metacost strategy by forming multiple bootstrap replicates of training set, learning a classifier on each and using the minimum Bayes conditional risk to relabel each training example with the estimated optimal class.

Ling [57] investigated in the fieldwork of direct marketing, which is defined as the process of identifying likely buyers of certain product and promoting the products accordingly. Potential buyers data are in a database, and the response rate to a promotion -final buyer- is low (typically about 1%). They used AdaBoost with naive Bayes and C4.5 with a *lift* index as a measure of performance in order to relate it with the net profit curve. AdaBoost was used by Viola et al. [85] for face detection, tackling the classification problem by means of a cascade of classifiers. Kubat and Matwin [54] introduced the g-mean as a measure of performance. They applies the technique known as one-side selection with nearest neighbor and and C4.5 decision trees.

4.2 A strategy for phasic contractions detection

The prevalence of phasic contractions in video frames is low (about 1:50-70), which states an imbalanced problem. We explored several alternatives for the implementation of a final system for the automatic detection of phasic contractions, which may be summarized in the following master lines: the use of a *cascade-oriented* implementation, the employ of *powerful classifiers*, such as support vector machines, and the application of *stratified sampling* techniques.

4.2.1 The cascade approach

Throughout the next chapter we develop our proposal for a final system for the automatic detection of phasic intestinal contractions in capsule endoscopy videos. In order to tackle the inherent complexity associated with the diverse visual patterns which an intestinal contraction may manifest, we focused our efforts on the design of a machine learning system for their automatic annotation. Our approach is constructed in a modular way following the model of a classification cascade, as shown in Figure 4.1. Each step of our cascade deals with a specific problem: the detection of dynamic patterns, the rejection of frames which are not valid for analysis and the final classification of the sequences into contractions and non-contractions. Each different step will process the input frames using a specific set of features specially chosen for that purpose. In the case of the visual pattern of the intestinal contractions, which corresponds to a lumen which is closed in the middle of a sequence of nine frames and which remains open at its beginning and end, it is reasonable to think that a good first approach to this paradigm should involve some numerical quantities conveying information about the probable presence of the intestinal lumen and about its size. In the case of turbid liquid, which appears as a hazing greenish fluid occluding the visual field of the camera, it makes sense to think that a color description may be useful for the detection of this sort of frames. In order to accomplish the former, we propose the use of two base image descriptors: the normalized intensity and the size of the lumen. The latter will be tackled by using color analysis techniques. The inclusion of textural descriptors, in addition to the former, conforms the feature set which is used for a final classification stage. This final stage consists of a support vector machine (SVM) classifier [79], which is trained by under-sampling the majority class. The following paragraphs provide an introductory description about these issues.

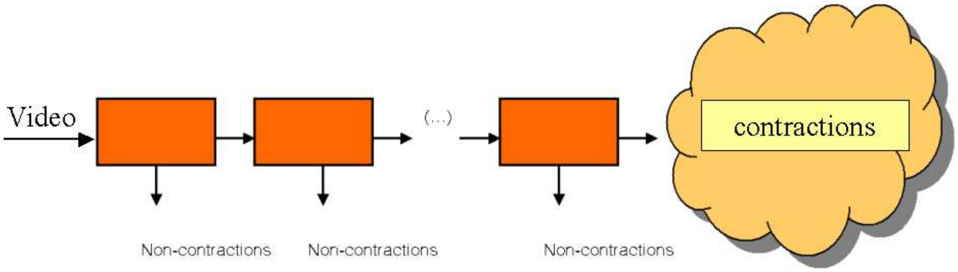


Figure 4.1: A general cascade strategy

4.2.2 The SVM classifier

Support vector machines were introduced by Vapnik [78] and they have been successfully applied in several areas such as image retrieval [76], text classification [77, 46], and handwriting recognition [22], just to cite a few -the interested reader can find a deep overview of applications where SVMs have been used in [16]-. SVMs look for the hyperplane which separates positive and negative samples, maximizing the distance from the hyperplane to the closest data points of each class (namely, the *margin*), as shown in Figure 4.2. This hyperplane is known as the *maximum-margin hyperplane* or the optimal hyperplane, and the vectors which are the closest to this hyperplane are called the *support vectors*. This approach is shown to minimize the risk function, while achieving a good generalization power by means of the optimization of the variance of the classifier with different training sets. Vapnik et al. [79] developed SVMs as non-linear classifiers by means of the use of a kernel-based strategy.

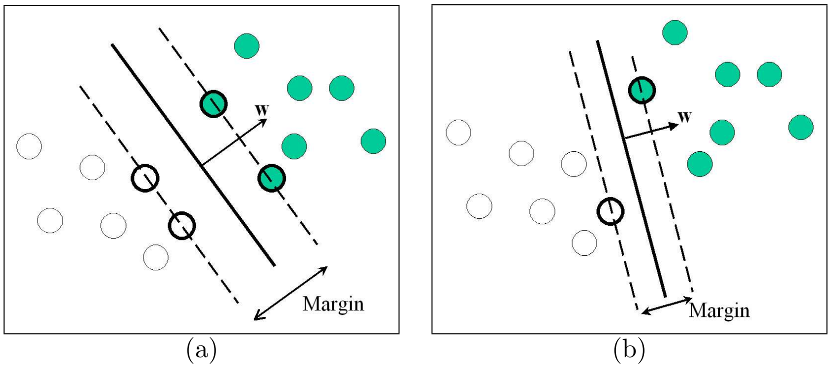


Figure 4.2: Two linear solutions for classification of a binary data set for (a) maximum margin and (b) a smaller margin solution.

The SVM scheme is endowed with a robust theoretical background which can be summarized in the following lines -for the interested reader in a deeper insight into the mathematical foundations of SVM, the introductory tutorials presented in [16] will result specially helpfull-. For a given a training set $\{\mathbf{x}_i, y_i\}, i = 1, \ldots, M$ with $y_i \in \{-1, 1\}$ and $\mathbf{x}_i \in \Re^d$, where M represents the number of samples and d represents the dimension of the feature space, the hyperplane $(\mathbf{w}, b)$ which solves the following optimization problem:

$$\text{Minimization of the functional:} \quad \tfrac{1}{2}\mathbf{w}^t\mathbf{w} \tag{4.1}$$

$$\text{constrained to:} \quad y_i(\mathbf{w}^t\mathbf{x}_i + b) \geq 1 \quad i = 1, \ldots, M \tag{4.2}$$

results in the hyperplane with the maximum margin, as pictured in the example of Figure 4.2 (a). In order to solve this optimization problem, we transform it into the equivalent Lagrangian dual problem [12]. First, we state the primal form of the Lagrangian:

$$L_P(\mathbf{w}, b, \alpha) \equiv \frac{1}{2}\mathbf{w}^t\mathbf{w} - \sum_{i=1}^{M} \alpha_i y_i(\mathbf{x}_i + b) + \sum_{i=1}^{M} \alpha_i \tag{4.3}$$

where α_i are the Lagrange multipliers. The minimization of L_P implies that the partial derivatives of L_P with respect to $\mathbf{w}$ and b vanish, and thus the following solutions are provided:

$$\begin{aligned} \mathbf{w} &= \sum_{i=1}^{M} \alpha_i y_i \mathbf{x}_i && (4.4) \\ 0 &= \sum_{i=1}^{M} \alpha_i y_i && (4.5) \end{aligned}$$

and since these solutions represent equality constraints in the dual formulation, we can substitute them in Equation (4.3) obtaining the dual Lagrangian:

$$L_D(\mathbf{w}, b, \alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{M} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^t \mathbf{x}_j \quad (4.6)$$

which yields to the equivalent optimization problem:

$$\begin{aligned} \text{Maximization of the functional:} \quad & \textstyle\sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{M} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^t \mathbf{x}_j && (4.7) \\ \text{constrained to:} \quad & \textstyle\sum_{i=1}^{M} y_i \alpha_i = 0, \quad \text{and} \quad \alpha_i \geq 0 && (4.8) \end{aligned}$$

It must be notice that this formulation represents the data $\mathbf{x}$ as dot products. This leads to the *kernel trick*, consisting of mapping the original data into a higher dimensional feature space, and calculate the dot product in that space. Since the only quantity required is the dot product, it is not necessary to calculate the mapping, we only need a formula for the dot product, which is provided by the kernel function. Popular kernels include polynomial kernels, radial basis functions kernels (RBF) and sigmoid kernels, among others. There is an extended opinion in the SVM community arguing that the choice of the kernel should not have a relevant impact in the performance of the classifier, although this assumption should be tested for each specific problem. Table 4.1 shows the formulae of the previously cited kernel functions. The kernel performance can be controlled by one or several parameters, such as the polynomial degree d, the σ value, or the sigmoid coefficients.

Table 4.1: Some popular kernels

Polynomial	RBF	Sigmoid
$K_{Pol}(\mathbf{x}, \mathbf{x}') = (\langle \mathbf{x} \cdot \mathbf{x}' \rangle + c)^d$	$K_{RBF}(\mathbf{x}, \mathbf{x}') = \exp(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{\sigma^2})$	$K_{Sig}(\mathbf{x}, \mathbf{x}') = \tanh(\kappa \mathbf{x} \cdot \mathbf{x}' + c)$

In addition to the former, it can be shown that the optimization problem related to SVM consists of a convex optimization problem, for which the Karush-Kuhn-Tucker (KKT) conditions are a necessary and sufficient. The KKT conditions state that:

$$\alpha_i(y_i(\mathbf{w}^t\mathbf{x}_i + b) - 1) = 0 \tag{4.9}$$

Although there exists one Lagrange multiplier α_i for each training point, the value of the Lagrange multiplier α_i is non-zero if and only if its corresponding training point $\mathbf{x}_i$ lies on the margin. The practical interpretation of this fact is that only a few samples are involved into the training procedure of the SVM, namely the support vectors, for which $\alpha_i > 0$. The rest of the samples can be reorganized or redistributed out of the margin region, and the resulting trained machine will be the same. This sparsity property is a desirable property if we are dealing with data sets containing a large number of samples.

The further extension of this scheme to the relaxation of the margin constraints, in order to deal with real cases in which the data sets are not separable, differing from the distribution shown in Figure 4.2, yields to the minimization problem:

$$\text{Minimization of the functional:} \quad \tfrac{1}{2}\mathbf{w}^t\mathbf{w} + C\textstyle\sum_{i=1}^{M}\xi_i \tag{4.10}$$

$$\text{constrained to:} \quad y_i(\mathbf{w}^t\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \quad i = 1, \ldots, M \tag{4.11}$$

where ξ_i are slack variables needed for the optimization procedure, and C controls the weight of a misclassification value of ξ_i, stating a trade-off between misclassification and maximal margin achievement. Several implementations have been presented for the resolution of this problem. A detailed explanation of the main approaches and the further extension of SVM to the multi-class scenario can be found in [23].

4.2.3 Sampling techniques on the training set

We performed a set of tests in order to analyze the optimal sampling methodology for the last SVM stage of the phasic contractions cascade. Thus, the classifier is trained with a balanced training set which can be obtained in two basic ways: a) selecting a random sample of the majority class, or b) artificially creating new data in the minority class. The former alternative is straight forward and can be stated in terms of widely used techniques, such as bootstrapping: random sampling with replacement of the data. The latter can show different and diverse implementations, such as data duplication or data interpolation. Synthetic minority over-sampling (SMOTE) [20] is a paradigmatic example of this kind of techniques. In SMOTE, the minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining all of the k minority class nearest neighbors. Depending upon the amount of over-sampling required, neighbors from the k nearest neighbors are randomly chosen. Synthetic samples are generated in the following way: we take the difference between the feature vector (sample) under consideration and its nearest neighbor, then we multiply this difference by a random number between 0 and 1, and add it to the feature vector under consideration. This causes the selection of a random point along the line segment between two specific features.

4.3 Other approaches for classification

4.3.1 Single classifiers and classifier ensembles

In order to obtain a global perspective of the performance of our approach, we tried diverse classification strategies, using some of the most widely used classifiers in the pattern recognition community. This list, which does not pretend to hold all the possible classification techniques but a sample of those which may be currently considered as a feasible basis for a gold-standard analysis, include: linear discriminant classifier (LDC), quadratic discriminant classifier (QDC), logistic classifier (LOGDC), Parzen classifier, decision trees (DT), and nearest neighbors classifier (k-NN). The LDC assumes that the data sets present Gaussian distributions with equal covariance matrix, estimates their parameters based on the sample scatter matrix and then calculates the classification error using the Bayes error. The QDC performs the same classification procedure, but there is no assumption that the covariance of each of the classes is identical. LOGDC performs the computation of the linear classifier for the data set by maximizing the likelihood criterion using the logistic (sigmoid) function. The Parzen classifier performs the estimation of the probability density functions underlying the data sets by means of the Parzen approach, and then calculates the classification error based on them. The DT classier which we refer to in this work corresponds to a binary decision algorithm, based on a certain splinting criterion, such as information gain, purity, etc, which distributes the training data into consecutive binary classifications, based on the minimization of the classification error under the selected criterion. A further pruning analysis can be performed in order to minimize the global error following different strategies: pessimistic pruning, test-set based pruning, early pruning, etc. The k-NN classifier assigns to each unseen sample the most frequent label among the k-nearest samples in the training set. These lines provide only a quick description of the main traits of each technique. A deeper analysis of the multiple mathematical formulations of each classification technique can be found in several handbooks of pattern recognition [28, 38].

In addition to these single classifiers, classifier ensembles constitute a useful tool for the improvement of the classification task. Classifier ensembles [55] are the result of the combination of multiple single classifiers in order to obtain a classification system showing a better performance than each of its single components. Moreover, each component of the ensemble may also be the resulting classifier of the combination of several single units. For instance, we can use a bagging ensemble, consisting of several versions of the same classifier trained by bootstrapping, as a single classifier which calculates its final output after and aggregation stage (performed by averaging all the individual outputs, for instance). The analysis of the theoretical background of the different classifier ensembles techniques is out of the scope of this work, and there currently exist several specific textbooks which the interested reader may consult in order to achieve a deeper insight -Kuncheva's book [55] might constitute one of the mandatory references in this field- .

4.3.2 The AdaBoost approach

During the last years, the pattern recognition community has devoted special attention to the study of boosting techniques, due to the optimal performance results obtained by particular implementations, such as AdaBoost, in some specific scenarios. For this reason, we found of special interest devoting a separate analysis to the study of the performance of AdaBoost in the domain of phasic intestinal contractions detection. AdaBoost is a boosting [35] classification technique which was first introduced by Freund and Schapire [36]. Original AdaBoost algorithm, and multiple modifications of it, have been tested in a wide range of different applications [37, 85]. AdaBoost consists of an iterative classification approach based on the use of weak classifiers which are trained by a weighted training set. The weights of the samples are changed on each iteration so that the wrong classified samples increase their weight. This is a boosting strategy used to force classifiers of late iterations to learn the difficult cases. The final classifier consists of the overall weighted sum of the weak classifiers.

The distinctive traits which differentiate *AdaBoost* from other boosting algorithms are coded in the Algorithm 4.3.2 [85].

Algorithm 4.3.2. *AdaBoost* **algorithm**:

```
 1: BEGIN
    Given example images $(x_1; y_1), ..., (x_n; y_n)$ where $y_i = 0, 1$ for negative and positive examples
    respectively.
 2: Initialize weights $w_{l,i} = 1/2m, 1/2l$ for $y_i = 0, 1$ respectively, where $m$ and $l$ are the number
    of negatives and positives respectively.
 3: for $t = 1$ to $T$ do
 4:    Normalize the weights,
                 $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$
       so that $w_t$ is a probability function.
 5:    For each feature, $j$, train a classifier $h_j$ restricted to a single feature and using weights $w_i$
       on the training data. The error is evaluated with respect to $w_t$, $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.
 6:    Choose classifier, $h_t$, with the smallest error $\epsilon_t$.
 7:    Update the weights: $w_{t+1,i} = w_{t,i}\beta_t^{1-e_i}$, where $e_i = 0$ if example $x_i$ is classified correctly,
       $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$
 8: end for
 9: The final classifier is:
                 $h(x) = \begin{cases} 1, & \sum_{t=1}^{T} \alpha_t h_t \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0, & \text{otherwise} \end{cases}$
10: END
```

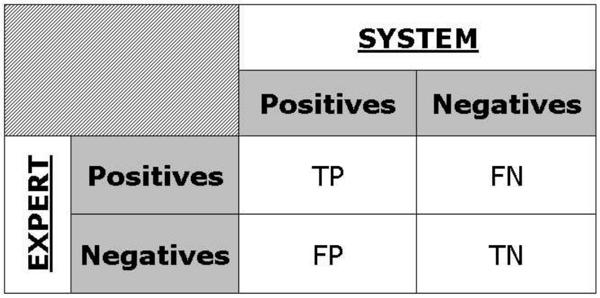
In *AdaBoost*, for each iteration t the weights of the samples are normalized according to step 4. Afterwards, the weak classifiers are trained as shown in step 5, and the weights are updated according to the error of the best weak classifier, as shown in step 7. We tested two modifications of the basic *AdaBoost* algorithm, which we refer to as $AdaBoost_{mod-1}$ and $AdaBoost_{mod-2}$ in Chapter 7, focusing on the sampling methodology of the training step 5, in order to investigate the impact of the sampling strategy on the global performance of the classifier.

4.4 Validation methodology: Performance assessment

The performance assessment consists of the assertion of the accuracy of an outcome in an objective way, allowing the comparison of the outcome obtained by one method with different outcomes obtained by other methods. This can be performed by means of objective metrics, which provide a quantitative assessment, or by means of graphical methods, which provide a qualitative assessment. Through this section, we describe the diverse assessment metrics and graphical methods which we propose for the validation of our system.

4.4.1 Numerical measures

In order to provide a performance result, several measures were used. These measures are based on the ability of the system in the detection of positives and the rejection of negatives. We expect the system to make mistakes during the classification process. Thus, some real contractions will be rejected when the system wrongly classifies them as non-contractions. On the contrary, there will be some real non-contractions which will be wrongly detected as contractions. This yields to a confusion matrix which can be stated as follows:



		SYSTEM	
		Positives	**Negatives**
EXPERT	**Positives**	TP	FN
	Negatives	FP	TN

Now, we define the *sensitivity*, *specificity*, *precision*, *false alarm rate (FAR)*, and *global error* as performance measures in terms of TP, TN, FP and FN. The formal definitions of these quantities are shown in Table 4.2. All these measures, except FAR, are bounded between 0 and 1. They convey different information regarding the accuracy of the system. The interpretation of sensitivity, specificity, precision, FAR, and global error in terms of the system performance is summarized in Table 4.3.

Table 4.2: Useful quantities for performance assessment

Positives:	Contraction sequences.
Negatives:	Non-contraction sequences
True Positives (TP):	Real sequences of contractions which are labelled by the system as contractions, i.e., the sequences at the output stage of the system.
True Negatives (TN):	Real sequences of non-contractions which are labelled by the system as non-contracttions, i.e., the sequences which are rejected by the system.
False Positives (FP):	Real non-contractions present at the output of the system.
False Negatives (FN):	Real contractions rejected by the system.

Sensitivity	$\frac{TP}{TP+FN}$	The proportion of the existing positives which our system successfully detects.
Specificity	$\frac{TN}{TN+FP}$	The proportion of the existing negatives which our system successfully rejects.
Precision	$\frac{TP}{TP+FP}$	The proportion of real positives at the output of our system.
FAR	$\frac{FP}{TP+FN}$	The ratio between the false positives and the number of existing positives (FAR may be greater than 1).
Global error	$\frac{FP+FN}{FP+FN+TP+TN}$	The ratio of errors.

Hypothesis testing

Hypothesis testing consists of the procedure of assessment of the *certainty* we have about a given conjecture or *hypothesis*, which must be *statistically proved* by means of a set of *experiments*. As a matter of fact, hypothesis tests are usually stated in the opposite way, providing the certainty value related to the fact that a given conjecture is not true. In this sense, hypothesis tests are useful for the assessment of the improvement or impact achieved by changing a usual procedure, and thus, we have to make up a set of experiments which yield to the comparison of diverse measures obtained with the old and the new procedure. The key point is that the base presumption is that nothing happens, and this yields to the *null hypothesis*, usually noted by H_0: both procedures perform in the same way. On the other hand,

Table 4.3: Interpretation of the different quantities for performance assessment

Sensitivity	1,	The system detects all existing contractions
	0,	The system detects no existing contractions
Specificity	1,	The system rejects all existing non-contractions
	0,	The system rejects no existing non-contractions
Precision	1,	The system provides only real contractions at the output
	0,	The system provides only real non-contractions at the output
FAR	1,	The system provides as many FP as the number of real contractions (FAR may be greater than 1)
	0,	The system provides no FP
Global error	1,	The system only provides real non-contractions and only rejects real contractions
	0,	The system only provides contractions and only rejects non-contractions

the alternative hypothesis H_1 consists of the actual claim to be tested: the new procedure performs in a different way than the old one. In order to *reject* the null hypothesis, some statistic must be calculated in order to obtain the probability of a significative divergence between the two procedures. Should a relevant divergence be observed, it is still possible that it has been obtained by chance, but the statistic provides us with the probability related to that event. If the divergence obtained occurs only one out of 98 times, we can assess with a 98% of *significance level* that the new procedure performs in a different way than the old one. In this work, we use two base hypothesis tests: the t-test and the Kolmogorov-Smirnov test [7].

The t-test is based on the *t-statistic*, which basically describes the probability of obtaining a certain value calculating the difference between the mean of a sample drawn from a given distribution and its real mean. Table 4.4 shows three different approaches for the t-statistic for the assessment of a mean value for a given normal distribution when the real mean μ and variance σ^2 are known, the assessment of the difference between the mean of two samples with equal variance, and the assessment of the difference between to samples when the variance are different.

Table 4.4: T-test statistics

One sample (mean & variance known)	Two samples (equal variance)	Two samples (different variance)
$t = \frac{\bar{x}-\mu_x}{s_x/\sqrt{n_x}}$	$t = \frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}}$	$t = \frac{(\bar{x}-\bar{y})-(\mu_x-\mu_y)}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}}$

where $\bar{x}$ and $\bar{y}$ represent the sample means, μ_x and μ_y are the population means, s_x and s_y represent the sample variances and n_x and n_y represent the number of elements in both samples. The values obtained with these statistics are compared with the analytical values, which are tabulated regarding the number of *degrees of freedom* associated with the number of elements in the samples [7], and the probability values, *p-values*, are obtained.

The Kolmogorov-Smirnov test (KS-test) compares the underlying density functions of two samples x and y. The KS-test calculates the proportion of x values less than a certain value v with the proportion of y values less than v, using the maximum difference over all v values as its test statistic. Mathematically, this can be written as $\max_v(|F_x(v) - F_y(v)|)$, where $F_x(v)$ is the proportion of x values less than or equal to v, and $F_y(v)$ is the proportion of y values less than or equal to v.

ROC curves

Receiver Operating Characteristic curves (*ROC curves*) [74] have their origins in signal detection theory. ROC curves were born due to the basic need of assessing the ability of a radar detector for reporting real targets, and at the same time, avoiding false alarms. This behavior is controlled by a parameter (or a set of parameters) that set up the radar in a characteristic operational point with a fixed sensitivity to real targets. In parallel to this, the number of false alarms are incremented as long as the number of desired real targets grows. The analysis of the ROC curve plots lets designers choose a deal between both successful detections and false alarms. This procedure has been widely used by the medical community framework, and only the semantics are different.

Parametric ROC analysis is based on the assumption that a two class population (positives and negatives) can be mapped into an one dimensional space, generating two class conditional probability density functions (the assumption of normality in this distributions cited in [63] is not mandatory, but it illustrates a paradigmatic situation in ROC curves). In this line, and if both classes are separable, positive samples will tend to fall on the right side and negatives on the left. These distributions eventually should have an intersection area if they are not absolutely separable. Now a threshold point is chosen, and all samples to the left of the threshold will be classified as *negatives* and all samples to the right as *positives.* Every positive sample falling on the left part will be a *false negative*, and viceversa with the *false positives.* ROC curves are built up as the result of simply moving the cutting point in order to seek for a compromise between TP and FP. The horizontal axis of a ROC curve correspond to the *FP-rate* (1-specificity), and the vertical axis correspond to the *TP-rate* (sensitivity). The resulting curve connects the point $(0, 0)$ with the point $(1, 1)$ in a monotonic way, and the area under the ROC curve is normalized to a maximum value of 1. Each pair (sensitivity, FP-rate) is obtained by displacing the threshold over the line in which the samples are distributed. The displacement of the threshold value through the probability density functions is shown to be equivalent to assigning different misclassification costs in the classifiers [15]. A big amount of literature is published in these issues, and most of the high relevance materials can be found in multiple references [90].

The area under the ROC curve (AUC) has been widely used in order to assess the performance of a classification system. However, it is known that this measure alone is not representative enough of performance when dealing with imbalanced data sets [62], [53], due to the high number of FP. Many approaches have been presented in the bibliography in order to evaluate the performance of a classifier instead of using the TP and FP as ROC analysis does. In this sense, the precision-recall curves appears as a helpful tool. Figure 4.3 shows an example of two ROC curves associated with different classifiers.

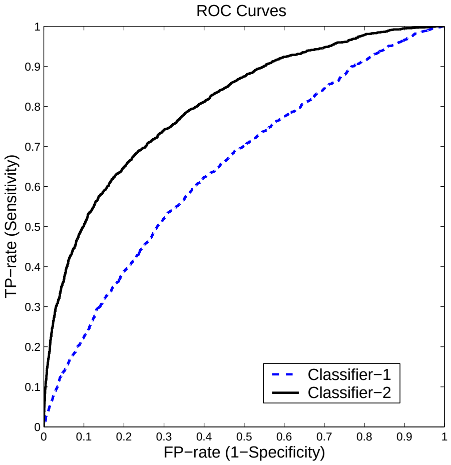


Figure 4.3: ROC curve plots the true positive rate (sensitivity) vs. false positive rate (1-specificity).

Sensitivity-FAR curves

Although ROC curves have been widely accepted as a useful tool for classifier performance comparison, they show an important pitfall which becomes quite significant in the case of imbalanced problems: the FP-rate defined by (1-specificity) does not convey information about the relevance of the false positives in the output of the system. This may be clarified by means of a simple example: Let us say that we have two different populations with the same number of positives $Pos_1 = Pos_2 = Pos$ and a different number of negatives Neg_1 and Neg_2, and one standard classification system. We apply the classification procedure to the different populations, and we obtain exactly the same ROC curve. Are we in a good position to say that the output of the system shows the same accuracy for both populations? The answer is no, because we do not have information about the specific imbalance rate in the different populations -the imbalance rate is here defined as the quotient of the number of negatives over the number of positives-. Let us get one operation point of our imaginary ROC curve, for instance, sensitivity = 80 and FP-rate = 20. Let us suppose that the imbalance rate of population 1 is $Ir_1 = 2$ and the imbalance rate of population 2 is $Ir_2 = 10$. In terms of positives detection, the system performance is the same for both cases: we detect 8 out of 10 existing positives, and as far as we stated that the number of positives is the same for both populations, we obtain the same number of positives, namely $Pos_{out} = 0.2 \times Pos$. But in terms of false positives detection, things are very different: the number of negatives for Neg_2 for population 2 is $Ir_2/Ir_1 = 5$ times larger than for population 1, and therefore the total number of false positives at the output of the system will be $0.2 \times Neg_1$ for population 1 and $5 \times 0.2 \times Neg_1$ for population 2. Thus, the precision for population 1 will be $0.8 \times Pos/(0.8 \times Pos + 0.2 \times Neg_1)$ and the precision for population 2 will be $0.8 \times Pos/(0.8 \times Pos + 5 \times 0.2 \times Neg_1)$, which undergoes a lower value. Conclusion: although the ROC curve plot is the same for both populations, the output of the system is much less accurate in terms of precision for the population with the greatest imbalance rate.

In order to tackle this pitfall, we propose the use of sensitivity-FAR curves as variation of ROC curves. Sensitivity FAR curves show a similar graph than ROC curves, but it is stretched or shrank depending on the ratio of false positives over the real positives. All the sensitivity-FAR curves start at (0,0), but different curves finish at different FAR levels, and FAR is not bounded to 1. In this sense, sensitivity-FAR curves convey insightful information for the case of imbalanced problems. The length of the curve can be used as an estimate of the performance of the system. Figure 4.4 shows an example of two sensitivity-FAR curves associated with two different classifiers.

PR curves

Precision-recall curves (*PR curves*) is a standard evaluation technique in the information retrieval community [11]. *Precision* is a measure in the interval [0,1] defined as the ratio of true positives detected over all the system output. It will be one when the system detects only positives, and zero when the system detects only negatives. In this way, a low precision value is associated with a high number of false positives, and it can be viewed as a measure of noise at the output stage. *Recall* -a synonym of sensitivity- is a measure in the interval [0,1] defined as the ratio of true positives detected over the total number of positives. It will be 1 when the system detects all the existing positives, and 0 when no positive is detected. For imbalanced problems, both measures together give more information than ROC curves. In ROC curves, precision is substituted by 1-specificity, or false positive ratio, and so the proportion

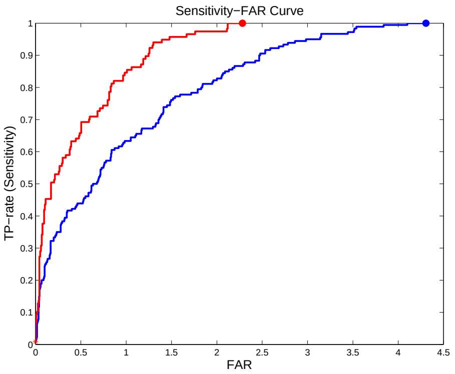


Figure 4.4: Sensitivity-FAR curve plots the true-positive rate (sensitivity) vs. the FAR value.

between positives and negatives is not taken into account. PR-curves are not necessarily monotonic, and they can start from a different coordinate point. The area under the PR curve has not the same straightforward analysis than in ROC curves, but it works as a good estimate for the classifiers performance. Figure 4.5 shows an example of two PR-curves for two imaginary classifiers in two different problems. The dotted lines corresponds to the threshold precision obtained by the trivial classifier which classifies every sample as a positive.

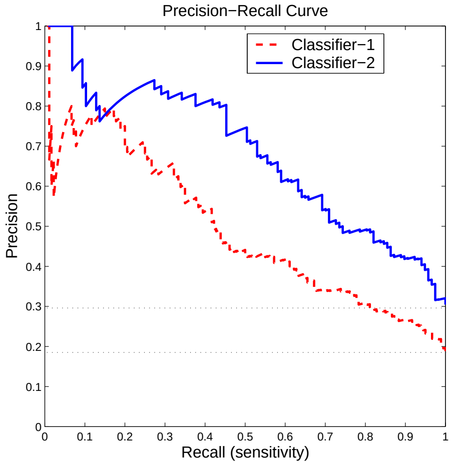


Figure 4.5: PR curve plots precision vs. recall (sensitivity).

4.4.2 Graphical methods

In addition to the numerical measures formerly described, which convey a quantitative assessment of the performance of the system from different points of view, we successfully applied diverse graphical methods for a qualitative visual assessment of performance. Among these methods, self organized maps (SOMs), framed-positives mosaics, and sequences sheets arise as the most useful techniques.

Self Organized Maps

Self Organizing Maps (SOMs) or Kohonen networks [51] are a specific type of neural networks which provide the possibility of reducing the dimensionality of complex multivariate data sets, allowing their visualization in a 2-dimensional representation. By producing easily comprehensible maps, SOMs offer a technique for visual understanding and interpretation of hidden structures and correlations in the input dataset. Several works have been published referring medical applications using SOMs in a wide range of fieldworks, including classification of craniofacial growth patterns [60], extraction of information from electromyographic signals [21], magnetic resonance image segmentation [40], cytodiagnosis of breast carcinoma [86], classification of renal diseases [11], and drug design [6], among others. In SOMs, the sample vectors are assigned to the most similar prototype vector, or best-matching unit (BMU), formally $c(x) = \text{argmin}_i\{\|x - m_i(t)\|\}$, where m_i are the prototype vectors, and x is the sample vector for which the BMU is determined. The learning process itself gradually adapts the model vectors to match the samples and to reflect their internal properties as faithfully as possible, which means that input vectors which are relatively close in input space should be mapped to units that are relatively close on the lattice. To achieve this, the training algorithm updates the model vectors iteratively during a number of training steps t, where a sample $x(t)$ is selected randomly, and then the BMU and its neighbors are updated as follows:

$$m_k(t+1) = m_k(t) + \alpha(t) h_{c(x)k}(t)[x(t) - m_k(t)] \tag{4.12}$$

where $\alpha(t)$ is the learning rate (which is decreasing monotonically over time) and $h_{c(x)k}$(t) is the neighborhood kernel. The neighborhood kernel determines the influence to the neighboring model vectors and its radius $s(t)$ is also decreasing with time. Thus, the learning process is gradually shifting from an initial rough learning phase with a big influence area and fast-changing prototype vectors to a fine-tuning phase with small neighborhood radius and prototype vectors that adapt slowly to the samples.

One of the scenarios where SOMs showed to provide a most useful help was related to the assessment of the characterization power of a plausible set of features. In this sense, we can use SOMs in order to validate whether certain descriptors are suitable for the characterization of a specific trait in frames or sequences. This technique takes profit of the ability of the SOM maps to create organized structures. If the descriptors are well defined, the provided maps are expected to present a coherent organization of the data, showing a gradient in the presence of the trait which the descriptors are set to characterize. This allows the definition of regions in the map where the presence and the intensity of the trait the descriptors address is more accentuated, and regions where the presence of the traits is not visible. In the case that the descriptors do not characterize properly the sought traits, the SOM map is expected to present a random organization. Figure 4.6 (a) shows a SOM created using a set of frames, with a set of useful descriptors for color characterization purposes. Each cell of the SOM represents a cluster of frames sharing similar descriptors, and a representative frame is visualized as the cell prototype. It can be seen that all the representative frames of each cell show a well defined gradient in terms of color. Figure 4.6 (b) shows the same set of frames, but using an inefficient set of descriptors for color. In this case, the organization of the map is apparently random, and the conclusion obtained is that the descriptors used in (b) are not suitable for color characterization. The bottom row shows a sample of frames belonging to the cluster for which the cell image in the SOM is the best matching unit.

In addition to the former, we used SOMs as a tool for the edition of training sets. As we stated before, if the descriptors are characterizing the frames in a proper way, the SOMs appearance is graded, well structured -not random-, and cells which are far away correspond to different patterns. Specifically speaking, if the elements of each cell are very similar, which numerically corresponds to a low *quantization error*, and the best matching units for a given vector are adjacent, which numerically corresponds to a low *topographic error* [79], we can select cells placed in opposite extremes of the SOM in order to define the positives and negatives examples of a training set. This training set can be used to train classifiers to automatically define the borders between the two classes.

Framed-positives mosaics

Framed-positives mosaics are a graphical tool consisting of a deployment of the frames from a video test in a sequential way, following the shape of a mosaic. Each frame which has been labelled by the system as a positive is surrounded within a square. Figure 4.7 renders an example of a framed-positives mosaic. This method lets the expert assess performance in a qualitative way at a quick glance just by examining the response of the system in terms of sensitivity (paying special attention to the positives which have not been labelled) and precision (paying special attention to the negatives which have been framed). This is necessary since for some applications, we cannot label all positive frames -for instance, the rejection of intestinal juices based on bubble detection-.

In addition to the former, a numerical analysis may be easily performed from framed-positives mosaics. The methodology used is straightforward and is explained as follows: 1) In the first step, the specialist checks the positives exclusively, marking all those frames where there is a discrepancy -i.e., the false positives-. 2) Then, the specialist checks the non-framed images marking every discrepancy -i.e., the false negatives. 3) Finally, the specialist counts the false positives and the false negatives

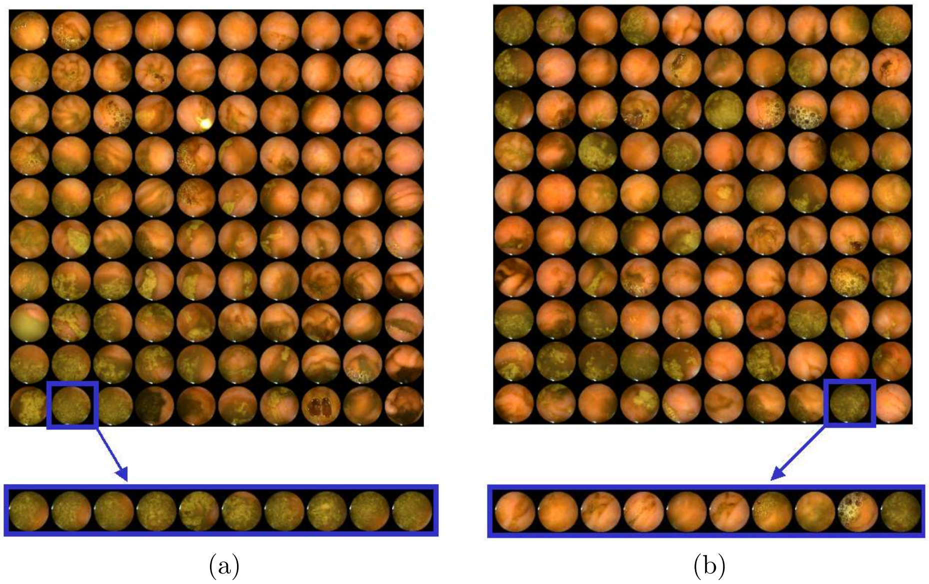


Figure 4.6: Two different SOMs. (a) This SOM was constructed with color meaningful feature. (b) This SOM was constructed using features which did not carry color information.

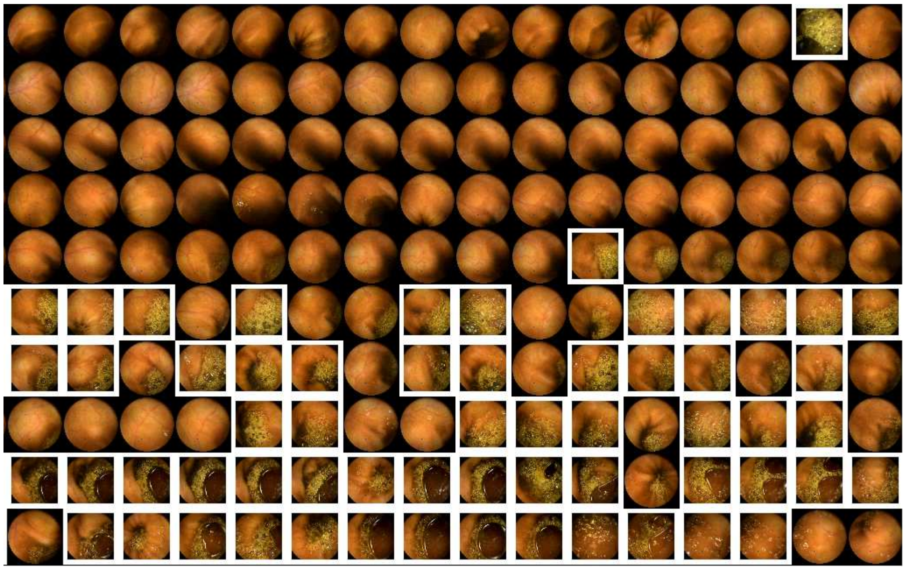


Figure 4.7: Framed mosaic. The positives are framed with a colored square, providing a tool for qualitative visual assessment. Quantitative assessment can by performed by manual checking and counting.

and uses them in order to calculate the performance measures, using the number of positives and negatives provided by the system. We used this strategy employing both paper impressions and image files. The former imply the use of pens for the image marking, the manual recount of the mistakes and the manual calculation of the performance measures. The latter are automatically implemented in a graphical tool.

Drag-and-drop color mosaics

Drag-and-drop color mosaics are a graphical tool which provides the specialist with a snap-shot of the distribution of the different events in one video. The video frames are deployed in a sequential way from left to right and top to bottom. Each frame is represented with a different color depending on its type. For instance, the central frame of a contraction finding can be visualized in black, the turbid liquid frames in beige, the wall frames in red, etc. Now, the specialist can have access to the visualization of each individual frame by clicking on it, or rendering a whole video sequence by dragging the mouse over the selected area. This provides a qualitative overview of each region, and may be used for the visual assessment of performance of the system when multiple events occur in a single frame -for instance, the contraction findings falling into a turbid sequence-. Figure 4.8 shows an example of a drag and drop color mosaic.

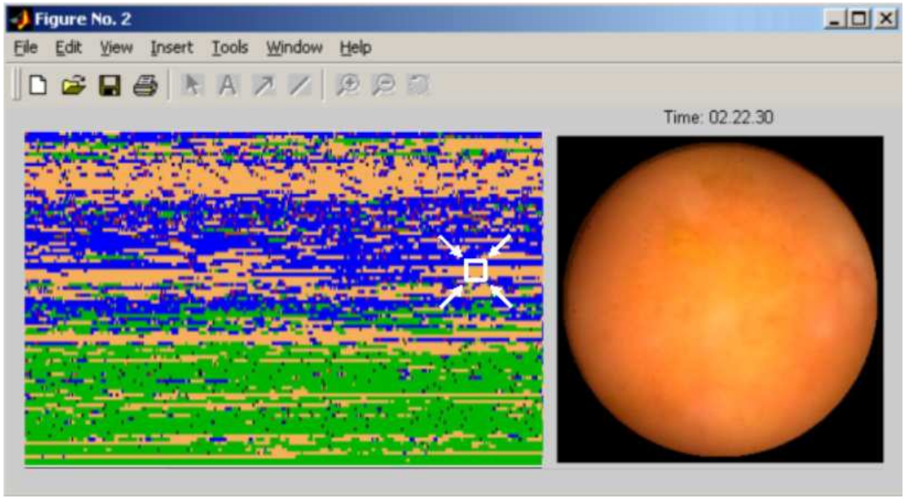


Figure 4.8: Drag-and-drop color mosaics allow the expert to select a region which the system identified as tunnel, wall, turbid of contraction frames for visual inspection -a wall frame (orange) is rendered-.

Sheets of positives sequences

The sheets of positives sequences allow the visualization of the dynamic behavior of a sequence of frames. They consist of 25 sequences of frames with the same length. The length of the sequences can vary depending on the application: for instance, the 9 frames sequences are suitable for the analysis of the contractions suggested by the system at the output stage. On the other hand, sequences of 23 frames can be used to assess the bound of 9 frames as the optimal bound for a contraction sequence characterization. Beside each sequence, the clinical time and frame index are shown, so the expert can easily access the video sequence in the video visualization tool and contrast its dynamic impression. The sheets of positives sequences can be used in order to measure the precision of the system when it is not possible to label positives previously. Figure 4.9 shows one sheet of positives sequences for 23 frames. The central frame corresponds to the center of the sequence, and therefore, the frame which is finally labelled as a positive -for this example, the central frame refers to findings of intestinal contractions-.

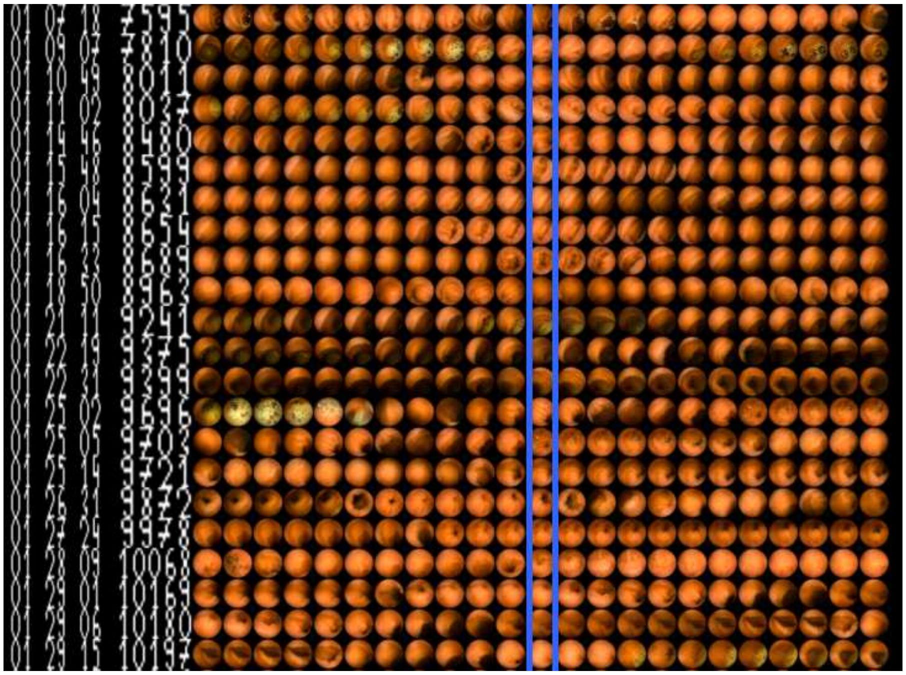


Figure 4.9: Sheets of positive sequences. The frame time and the frame index are referred to the central frame (in blue).

Chapter 5

A cascade system for phasic contractions detection

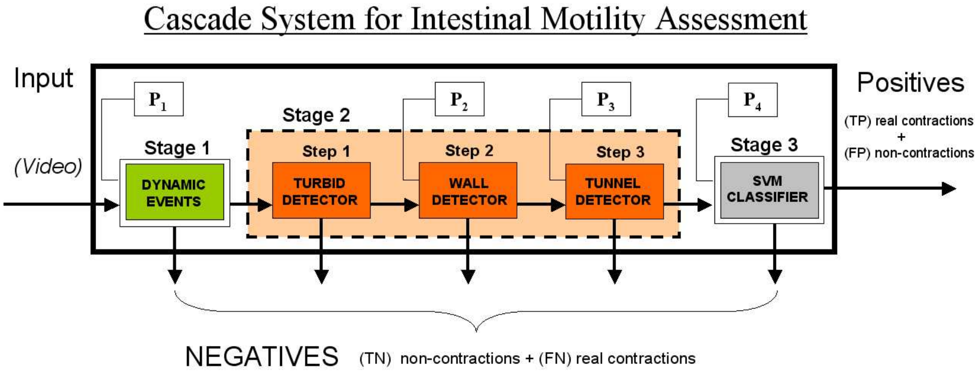


Figure 5.1: Cascade system for intestinal motility assessment. The input is the video study and the output are the intestinal contraction frames suggested by the system. Each stage rejects sequences of non-contractions. The global performance can be tuned by the set of parameters P.

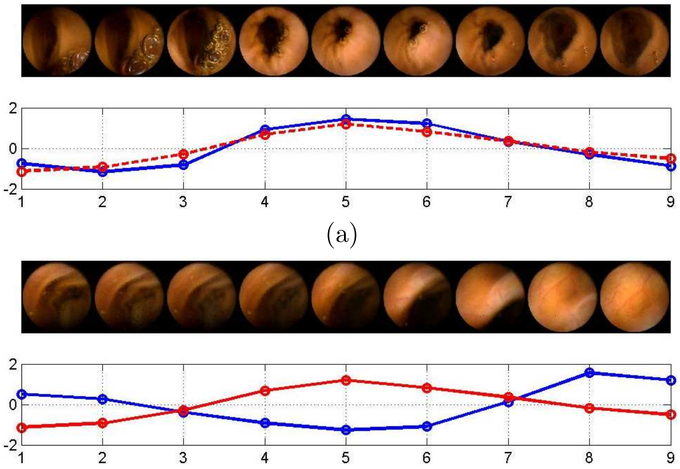
Our system is deployed in a sequentially modular way, namely, a cascade, as pictured out in Figure 5.1. Each part of the cascade receives as an input the output of the previous stage. The main input consists of the video frames, and the main output consists of the frames suggested as contractions. The rejected frames are distributed among three different stages: a first stage detecting dynamic patterns related to intestinal contractions, where most of the non-contractions frames are filtered; a second stage, removing non-valid frames due to occlusions or a wrong orientation of the camera, as described in section 3.3, and a final classification stage based on a support vector machine classifier (SVM) [79], where the final output defines the suggested contractions. The learning steps of each stage of the cascade involve a set of parameters P for tuning the classification performance. The turbid frames step and the final classification step consist of two support vector machine classifiers trained with a data set which has been labelled from previous studies. Through the following paragraphs, we provide a detailed description of each specific stage of the cascade.

5.1 Stage 1: Detecting dynamic patterns

The aim of the first stage is to pre-filter all the video frames according to the visual pattern of phasic contractions described in section 3.2. Taking into account that when the lumen is closed, all the light provided by the illumination leds carried by the camera is reflected by the intestinal walls, and that in a reciprocal way when the lumen is open the light is dissipated along the intestinal tube, a reasonable approach can be faced by using the image global illumination as an indicator of the presence of a closed or open lumen. This is implemented by means of the *normalized intensity* $f^1(n)$, defined in Equation 5.1.

$$f^1(n) = I_n - \frac{\sum_{i=-4}^{4} I_{n+i}}{9} \tag{5.1}$$

For each frame n, we take into account the 4 previous and the 4 following frames. For each one of these 9 frames, we calculate the overall intensity, $I_{n+i}, i = -4...4$, as the sum of the intensity values of its pixels. The final value of $f^1(n)$ represents a normalized intensity of the central frame within the 9 frames sequence. Should the central frame n be darker than its neighbors, the difference in $f^1(n)$ would tend to be negative, and viceversa. For the specific visual pattern of phasic contractions, the presence of an open lumen in the previous and following frames makes the central frame of a sequence of an intestinal contraction have a higher value of intensity than its neighbors. Thus, $f^1(n)$ is designed in order to present a high value when the central frame of a 9 frames sequence corresponds to an intestinal contraction, presenting a random pattern for non-contractions. A plot of $f^1(n)$ for (a) one contraction sequence and (b) one arbitrary sequence of 9 frames is pictured in Figure 5.2. Figure 5.3 shows a sequentially deployed mosaic of 135 frames from left to right and from top to down, spanning 1 minute and 7.5 seconds of a video sequence. The top plot corresponds to the value of f^1, showing one value for each frame in the mosaic; the contractions are bounded by a green frame, and their corresponding position in the plot is marked with a green line. The pictured contractions undergo a local maximum in f^1 corresponding to the closed lumen.



(a)

(b)

Figure 5.2: Pattern of f^1 for (a) one contraction and (b) one random sequence (solid blue line). The dashed red line corresponds to the averaged pattern of all the labelled intestinal contractions.

The discrimination of the frames into *positives*, which are to pass to the stage 2, and *negatives*, which are to be rejected by the system, is implemented by means of the application of a threshold P_1 over the normalized intensity $f^1(n)$, defined in Equation (5.1). Thus, a discriminant function g^1 may be formulated in the following way:

$$g^1(n) = \begin{cases} 1, & \text{if } f^1(n) > P_1 \\ -1, & \text{otherwise} \end{cases} \tag{5.2}$$

The dynamic patterns associated with intestinal contraction are those for which $g^1(n) = 1$, rejecting all the frames with $g^1(n) = -1$. This first stage has got one tuning parameter P_1 associated with it.

5.2 Stage 2: Rejection of wall, tunnel and turbid frames

The aim of stage 2 is to reject those frames which are to be not valid for analysis according to the specifications described in section 3.3, namely, the turbid frames and those frames where the camera is focusing on the intestinal wall (wall frames). In addition to this, those frames where the lumen appears static for a long sequence of time are rejected as well, as these frames do not carry motility information (tunnel frames). The wall and tunnel frames characterization is based on the intestinal lumen area descriptors, while the turbid liquid characterization is based both on texture and

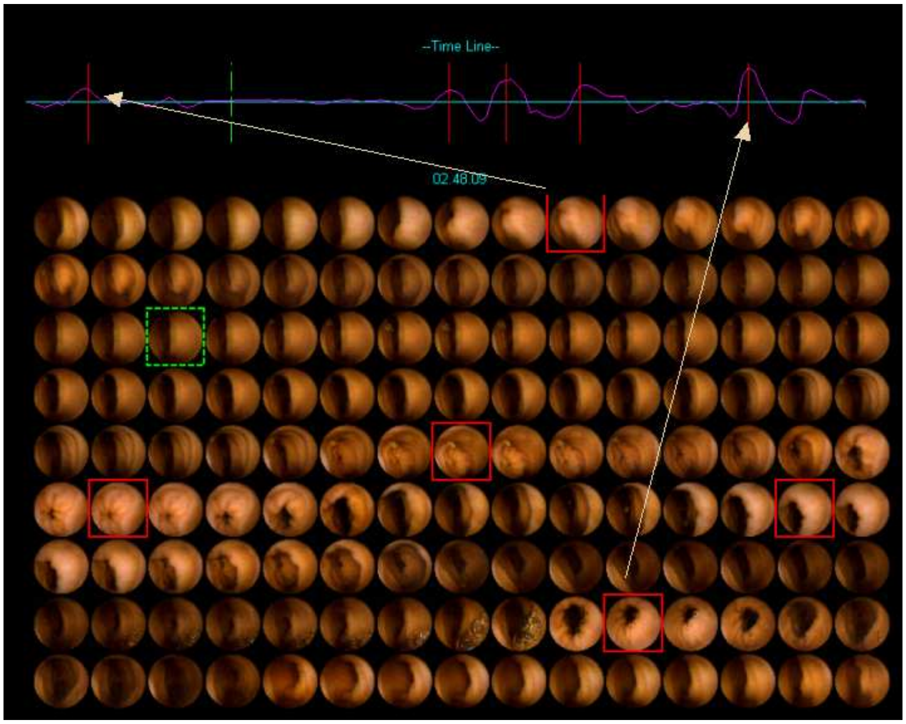
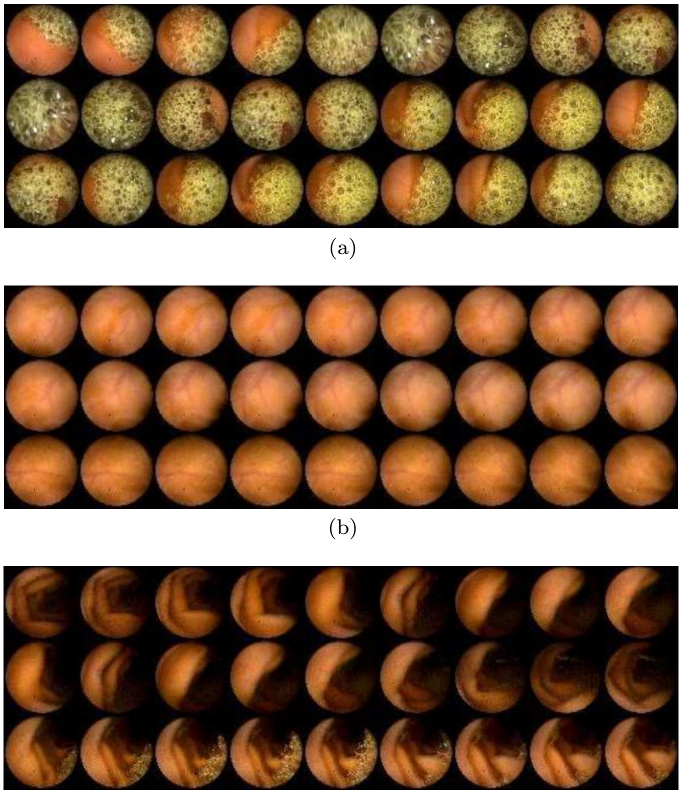


Figure 5.3: An example of capsule endoscopy video, which has been sequentially deployed for visualization purposes. Some intestinal contractions can be distinguished surrounded by a red square. The pattern described by f^1 shows a local maximum at the labelled contractions

color descriptors. Figure 5.4 renders a set of example sequences of (a) turbid, (b) wall and (c) tunnel frames.

5.2.1 Wall and tunnel frames

Both wall and tunnel frames are characterized by not carrying motility information, although they have different sources of origin. The former are due to the stable orientation of the camera towards the intestinal wall, keeping the intestinal lumen out of the field of view, while the latter corresponds to a stable orientation of the camera focusing the intestinal lumen for a span of time where no motility activity is present (in this sense, the resulting sequences show the intestinal lumen as a tunnel, during an undefined period of time). In these frames, the area of the intestinal lumen reveals itself as the most important source of information for characterization purposes.



(a)

(b)

(c)

Figure 5.4: Three example sequences of (a) turbid, (b) wall, and (c) tunnel frames. The system detects and rejects these sequences as system negatives in the second stage.

Lumen detection

The intestinal lumen appears as a dark hole surrounded by the clear intestinal walls in capsule endoscopy images. In order to estimate the lumen area, a Laplacian of Gaussian filter was applied (LoG)[71]. The LoG filter is a second derivative of Gaussian symmetric filter with a tuning parameter σ_{LoG} which plays the role of a scale parameter. The output of the LoG is high when a dark spot is found, providing a higher response the closer the diameter of the spot is, fitting the span of the Gaussian defined by σ_{LoG}, and the higher the contrast is between the dark spot and its bounds. Figure 5.5 renders the typical Mexican hat shape of a LoG filter; the σ_{LoG} parameter controls the width of the bell.

The value of σ_{LoG} was fixed to $\sigma_{lum} = 3$, the minimum size of the lumen in the central frame of a contraction sequence (this was straightforward to obtain after testing different values of several contraction sequences). The whole procedure for

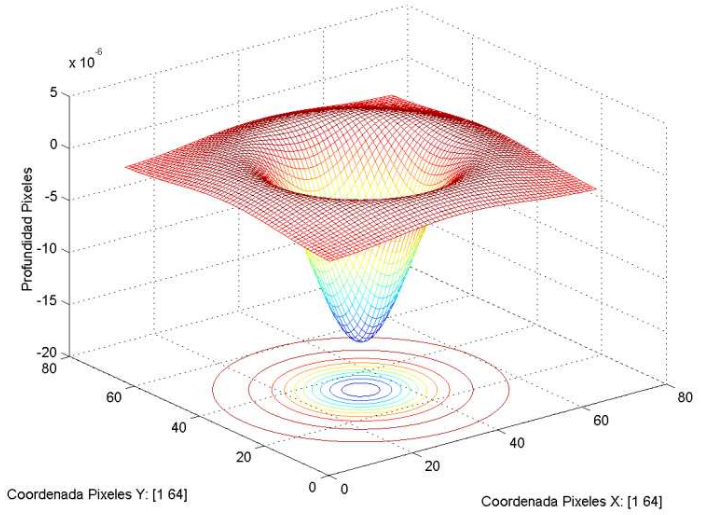


Figure 5.5: The Laplacian of Gaussian filter.

the lumen detection is represented in Figure 5.6 and graphically explained in Figure 5.7:

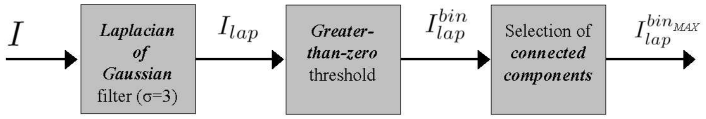

I
Laplacian of Gaussian filter (σ=3)
I_{lap}
Greater-than-zero threshold
I_{lap}^{bin}
Selection of connected components
$I_{lap}^{bin_{MAX}}$

Figure 5.6: Lumen detection procedure.

For each frame, the LoG filter is applied in order to obtain I_{lap}:

$$I_{lap}(n) = LoG_{\sigma_{lum}}(I_n) \tag{5.3}$$

Following, a binary image is created by means of a grater-than-zero threshold over I_{lap} as:

$$I_{lap}^{bin}(n) = F(x,y) = \begin{cases} 1, & I_{lap}(n)(x,y) > 0 \\ 0, & otherwise \end{cases} \tag{5.4}$$

Finally, we label all the connected components obtaining several blobs [43]. In case that only one blob is obtained, its area $f^2(n)$ is taken as the lumen area from:

$$f^2(n) = \sum_{pixels} I_{lap}^{bin}(n) \tag{5.5}$$

In case that several blobs are found, we obtain f^2 from $I_{lap}^{bin_{MAX}}(n)$, defined as the binary blob image containing the j-blob with the highest response of the filter (i.e., presumed the one with the highest contrast and best fitting in size) which is calculated based on the function:

$$f^3(n) = \max_i \sum_{pixels} \left(I_{lap}^{bin_i}(n) * I_{lap}(n) \right), i = 1...M \tag{5.6}$$

where $*$ represents the element-by-element product of the two image matrices, M is the number of connected components in I_{lap}^{bin}, and $I_{lap}^{bin_i}$ is the binary image containing the ith-connected component of I_{lap}^{bin}. The last row in Figure 5.7 shows an example of application of this technique in one video sequence: the first row shows the original sequence of frames, the second row shows the LoG filter response, the third row shows the binary blob image, and the final row shows the lumen segmentations obtained.

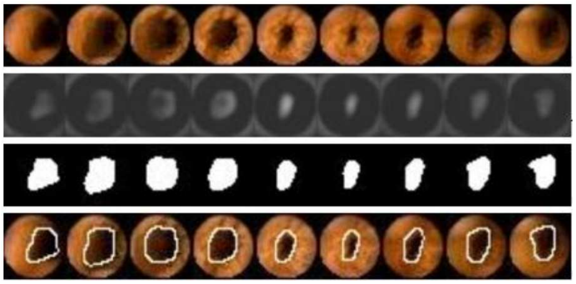


Figure 5.7: (First row) Original sequence, (second row) LoG filter response,(third row) binary blob and final lumen segmentation for the nine frames of an intestinal contraction sequence.

Tuning the σ_{lum} value

The LoG filter plays the role of a hole detector tuned to the minimum size which the lumen presents during an intestinal contraction. In this sense, it is important to notice that the response of the hole detector must distinguish between the image of a contracted lumen and the image of an intestinal wall, where no lumen is present. For the former, the filter should provide some value which indicates that a closed lumen is present, while for the latter the filter should provide a null value, indicating that no lumen is shown in the analyzed frame. With the aim of achieving the optimal value for σ, we developed multiple experiments over a pool of selected intestinal contractions, testing several values of σ_{lum} in an empirical intensive way.

Wall and tunnel frames discrimination

Both wall and tunnel frames are described by means of the sum of the area of the lumen throughout the sequence of 9 frames. The subsequent characterization of wall and tunnel frames is straightforward: the system classifies a frame as a wall frame if the sum of the lumen area throughout the 9 frames sequence is less than a certain threshold T_T, while the same frame is classified as a tunnel frame if the sum of the lumen area throughout the 9 frames sequence is greater than certain threshold T_W. This approach may be formulated through the discriminating functions g_W and g_T as:

$$g_W(n) = \begin{cases} 1, & \sum_{i=-4}^{i=4} f^2(n+i) < T_W \\ -1, & otherwise \end{cases} \tag{5.7}$$

$$g_T(n) = \begin{cases} 1, & \sum_{i=-4}^{i=4} f^2(n+i) > T_T \\ -1, & otherwise \end{cases} \tag{5.8}$$

Thus, the n-frame is characterized as a wall frame if $g_W(n) = 1$, and as a tunnel frame if $g_T(n) = 1$. Notice that both tunnel and wall are mutually exclusive.

5.2.2 Intestinal content

In those portions of the gut with a high content of intestinal juices, the visual field appears occluded by a hazing turbid liquid. If the visual area covered by the fluids is large enough, then no useful information about intestinal motility can be inferred from the images. This issue has a relevant importance, since both the normalized intensity f^1 and the lumen descriptors f^2 and f^3 show a random behavior in the presence of turbid liquid. For the characterization of turbid liquid we tested two different strategies, namely, color histogram matching and SVM classifiers. In some fasting studies, the intestinal content may be present in the fashion of bubbles. For this specific case, which may achieve a high relevance, we propose a textural characterization based on Gabor filters.

Color characterization: SVM classifier

A completely different approach was followed in order to try to avoid the implicit decision threshold associated with the histogram matching technique. In this sense, we relayed the decision of classifying a frame as a turbid frame on SVM classifier, which requires two main generalized parameters to be set, namely, the kernel type and the kernel parameter. Some of the most typical kernels used in the literature consist of linear kernels, polynomial kernels and radial basis functions kernels. We tested all these types of kernels to obtain the most suited to our specific problem. The SVM classifier is trained with samples of turbid and non-turbid frames. In order to establish the set of features to be used, we tested four different alternatives.

1. **RGB quantization**: We performed a color quantization grouping all the $256 \times 256 \times 256 = 16,777,216$ colors of the RGB space into a reduced space of $8 \times 8 \times 8 = 512$ values.

2. **RGB mean**: For each frame, the average value of the RGB components was calculated (3 featues).

3. **LAB quantization**: We followed the same procedure described for the RGB quantization, but using the LAB color opponent codification (512 features).

4. **AB mean** For each frame, the average value of the AB components of the LAB vector was calculated (2 features).

The RGB and LAB color codifications differ, mainly, in the way in which both methods distribute the intensity and chrominance information [88]: for the former, each channel carries all the information related to the color component (red, green and blue), and the final color of the pixel is the sum of the contribution of all its components. On the contrary, the LAB codification decouples the intensity and chrominance: the L component carries the intensity information, while the A and B components carry the chrominance information (explicitly, $A = magenta - green$ and $B = blue - yellow$).

Once the classifier is trained, the support vector machine classifies all the video frames into turbid and non-turbid. In order to incorporate the dynamic characteristics of the intestinal contractions as performed in the first stage, we adopted as a final criterion the rejection of those frames with more than 4 neighbors labelled as turbid frames within the 9 frames sequence (the number of 4 frames was strictly based on the experts' assessment), letting the remaining frames pass to the next step.

Texture characterization of bubble frames

Bubble frames are a specific case of intestinal juices, which are especially common in fasting videos, and which can provide useful information regarding physiological variables related to intestinal motility. The main trait that distinguishes bubble frames, as shown in Figure 5.8, is their specific aspect in terms of texture. In order to take profit of this issue, we tackled the texture characterization for bubble frames by means of Gabor filters [1, 56, 25], which have shown to provide good results in several applications involving texture analysis.

A Gabor filter can be viewed as a sinusoidal plane of particular frequency and orientation, modulated by a Gaussian envelope. These filters have been shown to possess good localization properties in both spatial and frequency domain and have been successfully applied in multiple tasks such as texture segmentation, edge detection, target detection, document analysis, retina identification, image coding, and image representation, among others [1, 81]. The filter response of a Gabor filter can be written as follows:

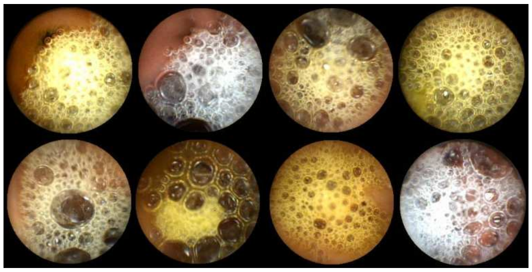


Figure 5.8: Different frames showing intestinal juices which occlude the visualization field

$$H(u,v) = \frac{1}{2\pi\sigma_u\sigma_v} e^{-\frac{1}{2}[\frac{(u-u_0)^2}{\sigma_u^2} + \frac{(v-v_0)^2}{\sigma_v^2}]} \tag{5.9}$$

where $\sigma_u = \frac{1}{2\pi\sigma_x}$ and $\sigma_v = \frac{1}{2\pi\sigma_y}$. In this sense, the Gabor function can be pictured as a Gaussian function shifted in frequency to *position* (u_0, v_0) -referred to as the Gabor filter spatial central frequency- and at an *orientation* of $\tan^{-1}\frac{u_0}{v_0}$. The σ_x and σ_y parameters are the standard deviation of the Gaussian envelope along X and Y directions and determine the filter bandwidth. Thus, the set of parameters $(u_0, v_0, \sigma_x, \sigma_y)$ completely defines a Gabor filter which is to provide a high response in all those regions of the image showing their energies concentrated near the spatial frequency point (u_0, v_0). By modifying these scale and orientation parameters, we obtain different Gabor filters, which make up a bank of filters. Gathering all the single filter responses so as to obtain the global response of the bank of filters, we defined an overall I_n^{bub} image as:

$$I_n^{bub} = \sum_{i=1}^{Num_{sc}} \sum_{j=1}^{Num_{or}} abs\{I_n^{i,j}\} \tag{5.10}$$

where $I_n^{i,j}$ represents the resulting response image of applying the filter with orientation i and scale j over the frame n, and Num_{sc} and Num_{or} are the number of selected scales and orientations in the bank of filter. As the response I_n^{bub} is expected to be high in those regions of the frame where the bubble pattern is present, the area of bubbles can be calculated by means of a threshold on the I_n^{bub} response.

5.3 Stage 3: The final classifier

The last stage of our approach consists of a SVM classifier with a radial basis functions kernel, represented by the formula:

$$K_{rbf}(x, x_i) = \exp\frac{-|x - x_i|^2}{2\sigma^2}, \gamma = 1/(2\sigma^2) \tag{5.11}$$

Stage three receives as an input the output of the second stage of the cascade, with an imbalance ratio which has been typically reduced from 1:50 to 1:5 frames. The output of the support vector classifier consists of frames suggested to the specialist as the candidates for intestinal contractions in the analyzed video. The choice of the SVM is underpinned by its robust mathematical background, being one of the most widely used classification techniques, with a remarkable success in multiple and diverse applications through the recent years [41]. An extended analysis about the suitability of the SVM classifier for this specific stage is developed in the discussion in Chapter 10. The γ parameter controls the operation point of the support vector classifier, and corresponds to the fourth tuning parameter of the system, P_4.

In order to characterize the intestinal contractions, a set of 33 features was computed. These features included: the 3 previously described functions f^1, f^2 and f^3; 6 textural features obtained from the co-occurrence matrix [75] of the original gray-level image after gray level normalization: energy, entropy, inverse differential moment, shade, inertia and promenance [65]; 18 features obtained form Rotation Invariant Uniform Local Binary Units operator (LBPriu2) applied in a circular symmetric neighborhood N of radius R, using $N = 16$ and $R = 2$ [66]; and finally, 6 statistical features: standard deviation, skewness and kurtosis calculated on the normalized gray level image and the local binary pattern vector. As performed in the previous stages, a feature vector was constructed taking into account the previous and following 4 frames, so that a final 33x9 = 297 dimensional feature vector was assigned to each frame. In order to address the high dimensionality of this feature space, a sequential forward feature selection method was applied [46].

Chapter 6

Experimental results

In this chapter, we present the experimental results of our research in the automatic detection of phasic contractions. We expose our outcome split into two parts. In the first part, we develop the study of each stage of the cascade system, providing their performance analysis and describing the different techniques used for their validation. We present the global performance of our system using different quantitative measures. In the second part, we assess the ability of our system to provide a reliable pattern of intestinal contractions, paying special attention on the assessment of its accuracy in terms of the intra- and inter-observer variability provided by the experts.

Our experimental tests were performed using 10 videos obtained from 10 different fasting volunteers (without eating or drinking in the previous 12 hours to the study), aged between 22 and 33, at the Digestive Diseases Dept. of the General Hospital de la Vall D'Hebron in Barcelona, Spain. The endoscopic capsules used were developed by Given Imaging, Ltd., Israel [39]. The capsules dimensions were 11x26 mm, contained 4 light emitting diodes, a lens, a color camera chip, two batteries with a mean life of about 6 hours, a radio frequency transmitter, and an antenna. The capsule acquisition rate was two frames per second with a resolution of 256x256x24-bit. For each study, one expert visualized the whole video and labelled all the frames showing intestinal contractions between the first post-duodenal and the first cecum images. These findings were used as the gold standard for testing our system, which was trained following a leave-one-out strategy. From now on, we associate *positives* with contraction sequences and *negatives* with non-contraction sequences. Our system provides an output in terms of positives, i.e., the output of the system consists of those video sequences which the system labels as "contractions", rejecting as negatives all the sequences which the system labels as "non-contractions". The following sections analyzes the performance of our system in terms of positives detection and negatives rejection. In this sense, each stage can be viewed as a black box with both an input and an output, and a set of configuration parameters P. In order to accomplish a detailed system performance analysis of our approach, we provide the study of each separate stage in the cascade, and finally, we provide the overall performance outcome.

6.1 First Stage: Detection of dynamic patterns

As we stated in section 5.1, the primary aim of stage one was to pre-filter as many frames as possible, reducing the imbalance rate without a significant loss in contractions, by means of a threshold P_1 on the feature f^1. This lets the system keep most of the dynamic patterns associated with contractions and reject a large amount of sequences. For each video j, we calculated P_1^j so that 99% of the labelled findings passed to stage two. We calculated the final value of P_1 as an overall mean over P_1^j, as:

$$P = avg_j(P_1^j), j = 1...10 \tag{6.1}$$

obtaining $P_1 = 0.20$ with $\sigma_{P_1} = 0.08$.

The validation of this stage consisted of a qualitative analysis of the false negatives, i.e., the contraction sequences which were rejected. Should the first stage be well defined, no clear pattern of intestinal contraction might be rejected, and we expected that only doubtful patterns should not pass to stage two. In order to accomplish this graphical analysis, we used the sheets of positives sequences to check the visual appearance of the false negatives.

6.1.1 Results

The performance results for this stage can be analyzed in Tables 6.1 and 6.2. For each stage, certain number of frames arrive at the input (column **Frames**), containing a number of intestinal contractions labelled by the expert (column **Findings**); the quotient *Non-contraction frames/Number of findings* represents the imbalance rate at the input of the stage (column **Imbalance Rate**). The output columns consist of the number and the percentage of frames and findings passing to the next stage, and the resulting imbalance rate. In addition to this, the rate of missing findings, i.e., findings which were wrongly filtered as non-contractions, and the rate of non-contractions frames, i.e., non-contractions which were wrongly detected as contractions, is provided. The following paragraphs deploy a detailed analysis of each stage, paying special attention to the reduction in the imbalance rate and the accuracy of the classification performed.

Table 6.2 shows that the overall number of frames at the output of stage one is about 11% the input, i.e., the system rejects 89% of the frames in this stage. But despite this high reduction in the number of frames, almost every finding was kept (97%), i.e., just about a 3% of the findings were wrongly rejected as non-contractions. At the output of stage one, the imbalance ratio was reduced about 10 times, from 61.3 to 6.9.

Table 6.1: Imbalance rate at the input of the first stage of the cascade

INPUT			
Study	Frames	Findings	Imb. Rate
Video 1	29444	747	39.4
Video 2	28803	529	54.4
Video 3	27816	575	48.4
Video 4	38885	733	53.0
Video 5	17619	356	49.5
Video 6	27360	476	57.5
Video 7	27176	918	29.6
Video 8	12620	150	84.1
Video 9	25994	206	126.2
Video 10	27967	397	70.4
Avg:			**61.3**

Table 6.2: Performance analysis for the first stage of the cascade

OUTPUT									
Study	Frames (%)		Findings (%)		Imb. Rate	Missed Findings (%)		Non-Cont. Frames (%)	
Video 1	3192	10.84	720	96.39	4.4	27	3.61	2472	77.44
Video 2	3027	10.51	502	94.90	6.0	27	5.10	2525	83.42
Video 3	3185	11.45	561	97.57	5.7	14	2.43	2624	82.39
Video 4	4025	10.35	717	97.82	5.6	16	2.18	3308	82.19
Video 5	1849	10.49	349	98.03	5.3	7	1.97	1500	81.12
Video 6	2943	10.76	459	96.43	6.4	17	3.57	2484	84.40
Video 7	2903	10.68	890	96.95	3.3	28	3.05	2013	69.34
Video 8	1366	10.82	143	95.33	9.6	7	4.67	1223	89.53
Video 9	2953	11.36	198	96.12	14.9	8	3.88	2755	93.29
Video 10	2948	10.54	385	96.98	7.7	12	3.02	2563	86.94
Avg:		10.78%		**96.65**	**6.9**		3.35%		83.01%

6.2 Second Stage: Detection of non-valid frames

6.2.1 Wall and tunnel frames

The wall and tunnel frames detection was performed by means of the parameters $T_W = P_2$ and $T_T = P_3$ of the discriminant functions g_W and g_T defined in Equations (5.7) and (5.8). A further analysis about alternative implementations of these discriminant functions and their impact on the obtained results is presented in the discussion in Chapter 10.

In order to set the P_2 and P_3 parameter values and assess the performance of the wall and tunnel detectors, we obtained the framed-positives mosaics of each video. We constrained the analysis of each video to a region of 30 minutes which was chosen after checking by visual inspection, using the video visualization tool, the presence of wall and tunnel frames. We applied the wall and tunnel detectors over the selected video regions, and repeated this procedure 10 times, modifying on each run the $P_{2,3}$

values in an increasing way. We tested the following interval values, which were split in 10 in a linear way: $P^2_{int} = [0, 100]$ and $P^3_{int} = [200, 1000]$. Based on the framed-positives mosaics analysis, we set $P_2 = 50$ and $P3 = 650$, and performed the wall and tunnel classification step over the 10 videos test pool.

6.2.2 Intestinal juices

Color characterization

The classification of the frames into turbid and not-turbid was performed by a SVM classifier. In this step, we focused our efforts on answering two main questions: 1) which the most suitable feature set for color characterization of turbid frames is, and 2) which the most suitable kernel for the SVM is.

Regarding the most suitable feature set for color characterization, we used SOMs in order to obtain a qualitative feedback about the distribution of the video frames in terms of the SOM map. For each video, we created the SOM associated with each feature space described in section 5.2.2, namely, RGB quantization, RGB mean, LAB quantization and AB mean, using all the frames from each video. Thus, for each video we obtained 4 different SOM maps. We selected the map which visually showed the highest degree of organization. This is a subjective selection which was performed by 4 different members of our group. The results was unanimous: for all the 10 videos, the mean AB SOM resulted the one with the highest degree of organization in terms of color, showing clear gradients and coherent contents for each cell -i.e., cells contained frames which were visually similar in terms of color-. The qualitative visual result was assessed by the SOM mean quantization error, which resulted smaller for the mean AB space in all the cases. Regarding the SVM kernel, we tested some of the most typical kernels used in the literature: linear, polynomial and radial bases functions. The best results in terms of classification error were obtained by the radial basis function kernel with $\gamma = 0.1$.

Texture characterization: bubble detection

In terms of color, the bubble frames share similar characteristics to the rest of turbid liquid. In this sense, the bubble detector does not detect new frames which have not been previously detected by the color characterization, but identifies those which share the textural characterization of bubbles, and which may carry clinical information for the specialists.

In order to characterize the textural information of the bubble filters we constructed a bank of Gabor filters. For this aim, we used 4 different directions oriented at 0^0, 45^0, 90^0 and 135^0, using 4 different σ values of (1, 2, 4, and 8), with an overall result of 16 filters in the bank -Figure 6.1 shows the power spectrum of this bank of filters-. The optimal number of directions and scales was obtained throughout an extensive empirical search and its analysis over framed-positives mosaics.

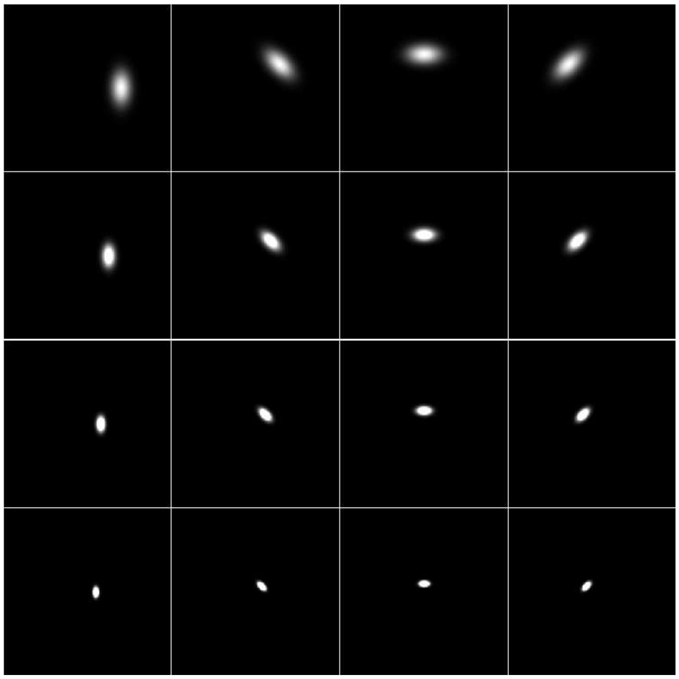


Figure 6.1: Filter response (power) for a bank of Gabor filters with 4 different orientations (columns) and 4 different scales (rows)

Figure 6.2 renders the original image, the I_n^{bub} filter response, and the segmented area of intestinal juices for 4 different sets of frames: (a) showing the wrinkles and folds of the intestinal wall, but low presence of bubbles; (b) showing intestinal juices covering about 50% of the visualization field; (c) intestinal juices covering a great part of the visualization field, and (d) sporadic single bubbles and rests of food. A high response can be observed in the area containing bubbles, independently of their size and distribution -even in the case of single bubbles (d)-. In contrast, those frames which do not show the bubble texture produce a low response image, even when some textural contain, such as wrinkles, or folds of the intestinal walls are present. Finally, some little pieces of food can be detected by the filter as small single bubbles (see the little spot on the top of the first frame of Figure 6.2 (d)). The response images were cropped in a 7% using a circular mask which was applied so as to eliminate the typical high response of the filter at the boundary of the field of view.

In order to characterize the filter output associated with the bubble pattern, by means of a threshold on I_n^{bub}, the specialists selected a pool of 100 frames, where the area containing bubbles was delimited by hand, and a random set of 300 frames where no bubbles were present. We applied an exhaustive search over a threshold on I_n to obtain a binary image defining two exclusively complementary areas in the frame: the region with intestinal juices and the region with no evidence of intestinal juices. The final threshold was searched in an iteratively convergent algorithm which

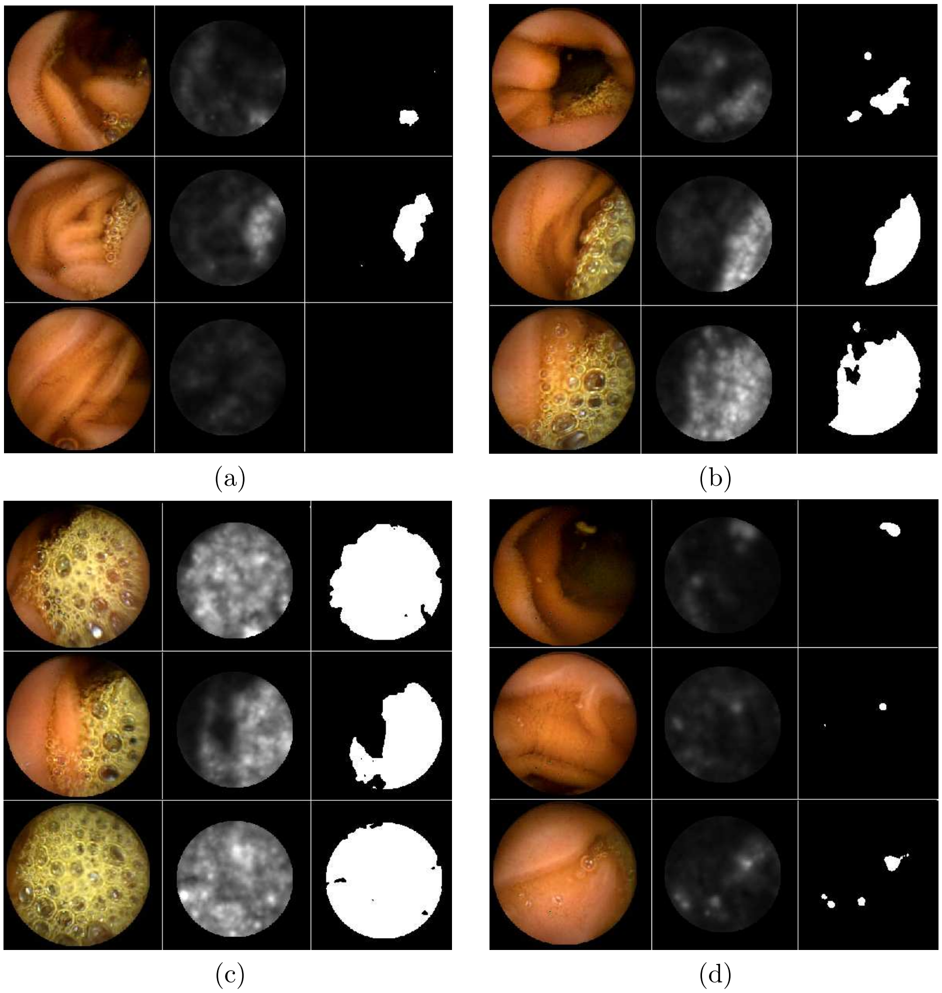


Figure 6.2: Original image, filter response and segmented area of intestinal juices for 4 different sets of frames (a, b, c and d)

compared the areas obtained by the given threshold value with those described by the specialists. We applied an error function under the criterion of the maximization of the overlap of the region of bubbles defined by the procedure with the region of bubbles defined by the specialists, and the minimization of the area of the wrongly detected of regions of bubbles.

Finally, we considered as not valid for analysis those frames where the detected region of bubbles was greater than 50% of the useful visualization area, following the specialists recommendations. We run the described method over the testing pool of 10 videos, obtaining the non useful sequences for all of them. For the system validation, we randomly chose 200 frames categorized as valid and 200 frames categorized as non valid, we shuffled them, and we provided them to the specialists to be labelled as bubble and non-bubble frames. Our system obtained an overall result of 98% of agreement in the case of the non valid frames -only 8 out of 200 frames labelled as non valid were wrongly detected by the system- and a 95% of agreement in the case of valid frames -10 out of 200 valid frames were wrongly labelled as containing intestinal juices in shape of bubbles-. The main source of error was detected in one single video where the intestinal walls presented an unusual colored spot-like texture, which underwent a high response in the bank of filters, resulting in a mislabelling as non valid frame. Figure 6.3 shows the positives framed mosaic for 5.25 minutes of a video. Notice the good precision of the detector and the low sensibility to the wrinkle patterns.

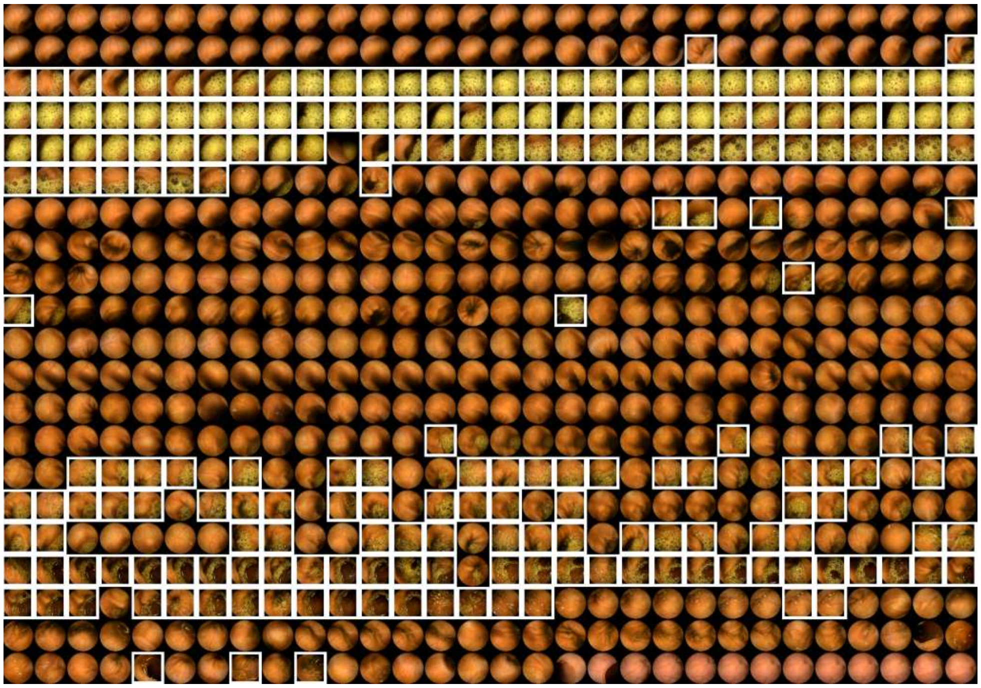


Figure 6.3: Positives framed mosaic for the intestinal content detection.

6.2.3 Results

For the analysis of this stage, we joined both the frames rejected by the wall and tunnel detector, and the turbid liquid detector. The performance results are shown in Tables 6.3 and 6.4. At the output of stage two, about 28% of the frames were rejected, keeping 96% of the findings provided by stage one. The imbalance rate was reduced to 4.6. In addition to this, the sum of the loss of findings, taking into account both stage one and stage two, set the rate of detected contractions at the output of stage two about 92.6%, as can be observed in the column **% Findings in video** in Table 6.4. As in the previous stage, the reduction of the imbalance rate is significant, while the loss in contractions appears to be reasonable -only 4.23% of loss for the second stage and an overall result of about 7.4% including stage one also-.

Table 6.3: Imbalance rate at the input of the second stage of the cascade

INPUT			
Study	Frames	Findings	Imbalance Rate
Video 1	3192	720	4.4
Video 2	3027	502	6.0
Video 3	3185	561	5.7
Video 4	4025	717	5.6
Video 5	1849	349	5.3
Video 6	2943	459	6.4
Video 7	2903	890	3.3
Video 8	1366	143	9.6
Video 9	2953	198	14.9
Video 10	2948	385	7.7
Avg:			6.9

Table 6.4: Performance analysis for the second stage of the cascade

OUTPUT										
Study	Frames (%)		Findings (%)		Imb. Rate	Missed Findings (%)		Non-Cont. Frames (%)		% Findings in Video
Video 1	2774	86.90	697	96.81	4.0	23	3.19	2077	74.87	93.31
Video 2	2346	77.50	474	94.42	4.9	28	5.58	1872	79.80	89.60
Video 3	2623	82.35	548	97.68	4.8	13	2.32	2075	79.11	95.30
Video 4	3170	78.76	673	93.86	4.7	44	6.14	2497	78.77	91.81
Video 5	1740	94.10	341	97.71	5.1	8	2.29	1399	80.40	95.79
Video 6	2288	77.74	453	98.69	5.1	6	1.31	1835	80.20	95.17
Video 7	2692	92.73	869	97.64	3.1	21	2.36	1823	67.72	94.66
Video 8	804	58.86	134	93.71	6.0	9	6.29	670	83.33	89.33
Video 9	678	22.96	184	92.93	3.7	14	7.07	494	72.86	89.32
Video 10	1538	52.17	363	94.29	4.2	22	5.71	1175	76.40	91.44
Avg:		72.41		95.77	**4.6**		4.23		77.35	**92.57**

6.3 Third Stage: Detection of intestinal contractions

Finally, the third stage, consisting of a SVM classifier with a radial basis function kernel with $\gamma = 0.01$, provides the output of the system. Thus, we can analyze the SVM from two perspectives: 1) the performance of the classifier, which takes into account only the number of the samples and remaining findings at the input of stage 3, shown in Table 6.5, and 2) the global system performance, which takes into account all video samples and findings which have been rejected along the previous stages.

6.3.1 Results

The output of stage three is at the same time the output of the system. Thus, we can analyze the output of stage three both in terms of stage performance and global performance. The stage performance is pictured in Table 6.6, while the global performance analysis is deployed in Table 6.7 and will be the object of analysis in the next paragraphs. Table 6.6 shows that the SVM classifier yields to a reduction about 71% in the number of frames at the output, keeping the 75% of the contractions provided by stage two. Moreover, the imbalance rate of the final data set is reduced to 0.7.

Table 6.5: Imbalance rate at the input of the third stage of the cascade

	INPUT		
Study	Frames	Findings	Imb. Rate
Video 1	2774	697	4.0
Video 2	2346	474	4.9
Video 3	2623	548	4.8
Video 4	3170	673	4.7
Video 5	1740	341	5.1
Video 6	2288	453	5.1
Video 7	2692	869	3.1
Video 8	804	134	6.0
Video 9	678	184	3.7
Video 10	1538	363	4.2
Avg:			4.6

6.4 Global performance

Finally, the global performance of the system, viewing all the steps in the cascade as a whole black box, can be faced in multiple ways. From a clinical point of view, the experts are interested in assessing: a) how many of the existing contractions our system is able to detect, namely, the system *sensitivity*, b) how many of the existing non-contractions our system is able to reject, namely, the system *specificity*, c) which

Table 6.6: Performance analysis for the third stage of the cascade

	OUTPUT										
Study	Frames (%)		Findings (%)		Imb. Rate	Missed Findings (%)		Non-Cont. Frames (%)		% Findings in Video
Video 1	904	32.59	595	85.37	0.5	102	14.63	309	34.18%	79.65
Video 2	607	25.87	343	72.36	0.8	131	27.64	264	43.49%	64.84
Video 3	646	24.63	405	73.91	0.6	143	26.09	241	37.31%	70.43
Video 4	981	30.95	547	81.28	0.8	126	18.72	434	44.24%	74.62
Video 5	433	24.89	266	78.01	0.6	75	21.99	167	38.57%	74.72
Video 6	768	33.57	339	74.83	1.3	114	25.17	429	55.86%	71.22
Video 7	835	31.02	603	69.39	0.4	266	30.61	232	27.78%	65.69
Video 8	189	23.51	111	82.84	0.7	23	17.16	78	41.27%	74.00
Video 9	228	33.63	130	70.65	0.9	54	29.35	98	42.11%	63.11
Video 10	363	23.60	248	68.32	0.5	115	31.68	115	31.68%	62.47
Avg:		28.42		75.70	**0.7**		24.30		39.65	**70.08**

the ratio between real contractions detected and the total number of sequences at the output of the system is, i.e., the system *precision*, and finally, d) a ratio between the false contractions at the output of the system and the existing contractions in the video, i.e., the system *FAR*.

Table 6.7 summarizes the performance results of the cascade system: Our approach achieves an overall sensitivity of 70.08%, picking 80% for the study referred as *Video 1*. The high overall specificity value of 99.59% is typical of imbalanced problems, and for this reason it is not generally useful for performance assessment tasks. However, FAR and precision carry insightful information about what the output is like. The resulting precision value of 60.26% tells us that 6 out of 10 frames in the output correspond to true findings. FAR is similar, but in terms of noise (the bigger the FAR, the larger the number of false positives), and normalized by the number of existing contractions. For different videos providing an output with a fixed precision, those with the highest number of findings in video will have lower FAR. In this sense, a FAR value of one tells us that we have obtained as many false positives as existing contractions in video. FAR and precision values are usually related, and Table 6.7 shows that the peeks of performance for both measures are found in the same two studies (*Video 6* and *Video 7*, outlined in bold type).

Table 6.7: Global system performance

Study	Sensitivity		Specificity		FAR		Precision	
Video 1	595/747	79.65	29135/29444	98.95	309/747	41.37	595/904	65.82
Video 2	343/529	64.84	28539/28803	99.01	264/529	49.90	343/607	56.51
Video 3	405/575	70.44	27575/27816	99.13	241/575	41.91	405/646	62.69
Video 4	547/733	74.65	38451/38885	98.88	434/733	59.21	547/981	55.76
Video 5	266/356	74.72	17452/17619	99.05	167/356	46.91	266/433	61.43
Video 6	339/476	71.22	26931/27360	98.43	429/476	90.13	339/768	44.14
Video 7	603/918	65.69	26944/27176	99.15	232/918	25.27	603/835	72.22
Video 8	111/150	74.00	12542/12620	99.38	78/150	52.00	111/189	58.73
Video 9	130/206	63.11	25888/25994	99.59	106/206	51.45	130/228	57.02
Video 10	248/397	62.46	27852/27967	99.59	115/397	28.96	248/363	68.32
Avg:	70.08(±5.81)		99.12(±0.35)		48.71(±17.91)		60.26(±7.84)	

Table 6.8 and Table 6.9 show the specific detection rate of the intestinal contractions corresponding to the occlusive and non-occlusive patterns for 8 videos. Table 6.10 shows the overall detection rate joining both types. The comparison of the detection rate for all the contractions in the specific case of occlusive + non-occlusive set is plotted in Figure 6.4, showing a parallel behavior, except for videos 7 and 8, where a slight decay in sensitivity is observed for occlusive contractions. An overall sensitivity around 85% was achieved for the occlusive+non-occlusive set, peaking at 93.2% for video 5.

Table 6.8: Individual detection rates for occlusive contractions

	Occlusive			
Study	Proportion		Sensitivity	
Video 1	125/747	16.7%	119/125	**95.2**%
Video 2	38/529	7.1%	36/38	94.7%
Video 3	59/575	10.2%	56/59	94.9%
Video 4	72/733	9.8%	66/72	91.7%
Video 5	32/356	8.9%	29/32	90.6%
Video 6	42/476	8.8%	37/42	88.1%
Video 7	178/918	19.3%	129/178	72.5%
Video 8	12/150	8.0%	8/12	66.7%
Avg:				**86.8(±11.0)%**

Table 6.9: Individual detection rates for non-occlusive contractions

	Non-occlusive			
Study	Proportion		Sensitivity	
Video 1	85/747	11.4%	76/85	89.4%
Video 2	34/529	6.4%	27/34	79.4%
Video 3	23/575	4.0%	17/23	73.9%
Video 4	176/733	24.0%	140/176	79.5%
Video 5	42/356	11.8%	40/42	**95.2**%
Video 6	44/476	9.2%	35/44	79.5%
Video 7	61/918	6.7%	49/61	80.3%
Video 8	12/150	8.0%	11/12	91.7%
Avg:				**83.6(±7.5)%**

Table 6.10: Detection rate for occlusive+non-occlusive contractions

	Oclusive + Non-occlusive			
Study	Proportion		Sensitivity	
Video 1	210/747	28.1%	195/210	92.8%
Video 2	72/529	13.6%	63/72	87.5%
Video 3	82/575	14.3%	73/82	89.0%
Video 4	248/733	33.8%	206/248	83.0%
Video 5	74/356	20.8%	69/74	93.2%
Video 6	86/476	18.1%	72/86	83.7%
Video 7	239/918	26.0%	178/239	74.5%
Video 8	24/150	16.0%	19/24	79.1%
Avg:				**85.3(±6.5)%**

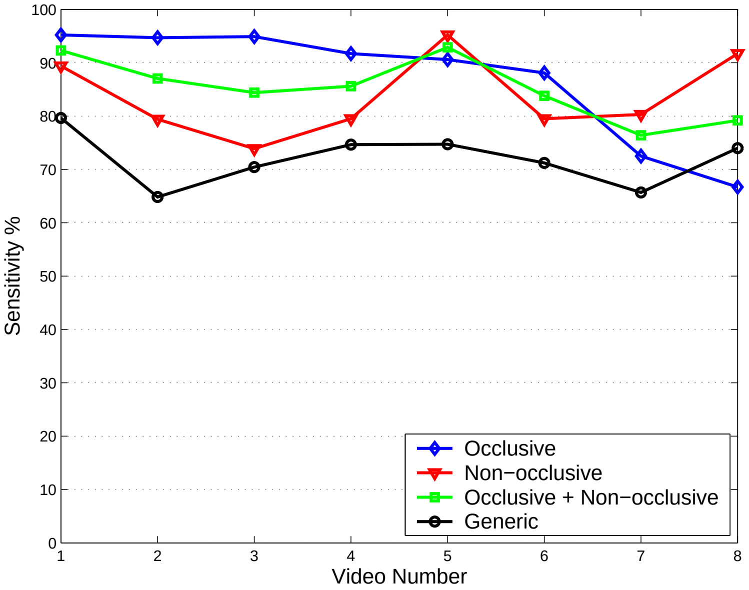


Figure 6.4: Detection rate for occlusive, non-occlusive and generic contractions

6.4.1 Assessment of the density of contractions

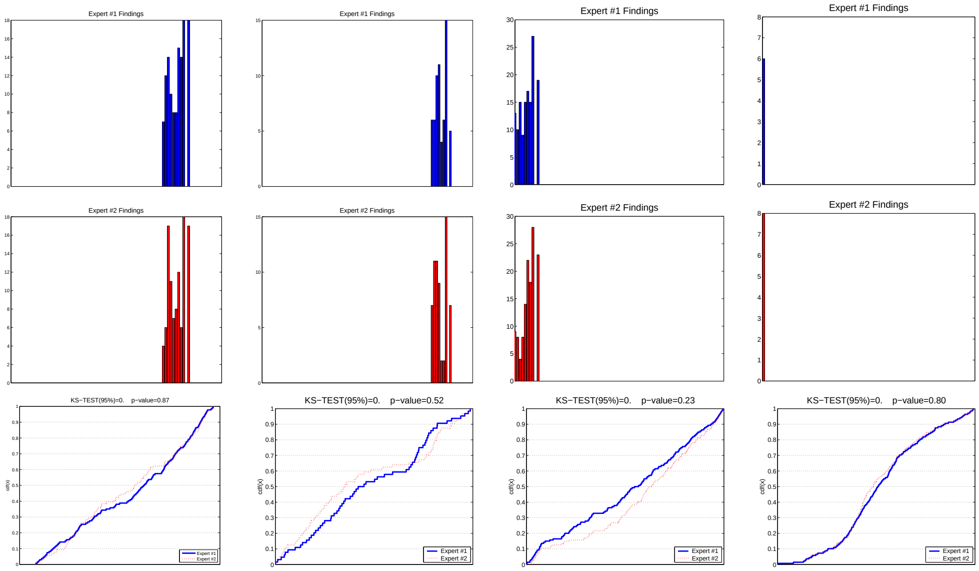
The main aim of this study is oriented to the assessment of the ability of our system to describe the pattern of the density of intestinal contractions. In order to assess this, we must answer two main questions: 1) which the divergence shown in the labelling of the same video by different specialists is, and 2) whether our system performs a labelling which is similar to that provided by the specialists. The former is linked to the inter-observer variability, while the latter is linked to the divergence of the labelling of our system in terms of inter-observer variability.

In order to perform this analysis, we estimated the inter-observer variability by means of two different specialists. Specialist 1 fixed the first post-doudenal and first cecal images of 8 videos, namely, the regions of analysis, and labelled their intestinal contractions. Then, the region of analysis of each video was divided into intervals of 30 minutes. For each video, Specialist 2 randomly selected one of its 30-minutes divisions, and labelled its intestinal contractions. All the contractions present in each video were gathered in a histogram of the intestinal contractions, grouping all the findings in bins spanning 3 minutes -360 video frames-. A Kolmogorov-Smirnov hypothesis test (KS-test) [18] was performed over this data. The null hypothesis for this test is that X_1 and X_2 have the same continuous distribution, where X_1 and X_2 are the labelling patterns provided by the specialists 1 and 2, respectively. The alternative hypothesis is that they have different continuous distributions. For each potential value x, the Kolmogorov-Smirnov test compares the proportion of X_1 values less than x with proportion of X_2 values less than x. The KS-test uses the maximum difference over all x values as its test statistic. For our test, we rejected the null hypothesis if the test was significant at the 99% level. The test resulted negative for all the 8 videos, providing a mean p-value of 0.80(±0.28). This result yields us to the conclusion that both specialists obtain similar patterns in terms of density of labels. Figures 6.5 and 6.6 plot the histograms of contractions provided by Specialist 1 (blue chart in the first row), the histograms of contractions provided by Specialist 2 (red chart in the second row), and the cumulative density functions of each distributions used in the the KS-test, together with its associated p-value (third row).

In order to answer the second question, we performed the same hypothesis test, but comparing now the Specialist 1 labelling, and the system output, for the whole zone of analysis of the video. The null hypothesis for this test is that X_1 and X_S have the same continuous distribution, where X_1 is the pattern provided by Specialist 1 and X_S is the pattern provided by the system. This test resulted negative for 7 out of 8 videos at the 99% level, showing a mean p-value of 0.25(±0.17). Figures 6.7 and 6.8 plot the histograms of contractions provided by Specialist 1 (red chart in the first row), the histograms of contractions provided by our system (blue chart in the second row), and the cumulative density functions of each distributions used in the the KS-test (third row), together with its associated p-value. It can be seen that the only video rejected by the test (*Video 5*), which showed a p-value $p = 0.0004$, presented a systematic over-detection of contractions, which was distributed in a similar amount along the video -only in the final part of the video the expert counts outnumbered the system counts-. This result yields to the conclusion that our approach and the specialists' obtain similar patterns in terms of density of labels, observing a unique case of divergence for *Video 5*, which presented a systematic increase in the labelling counts.

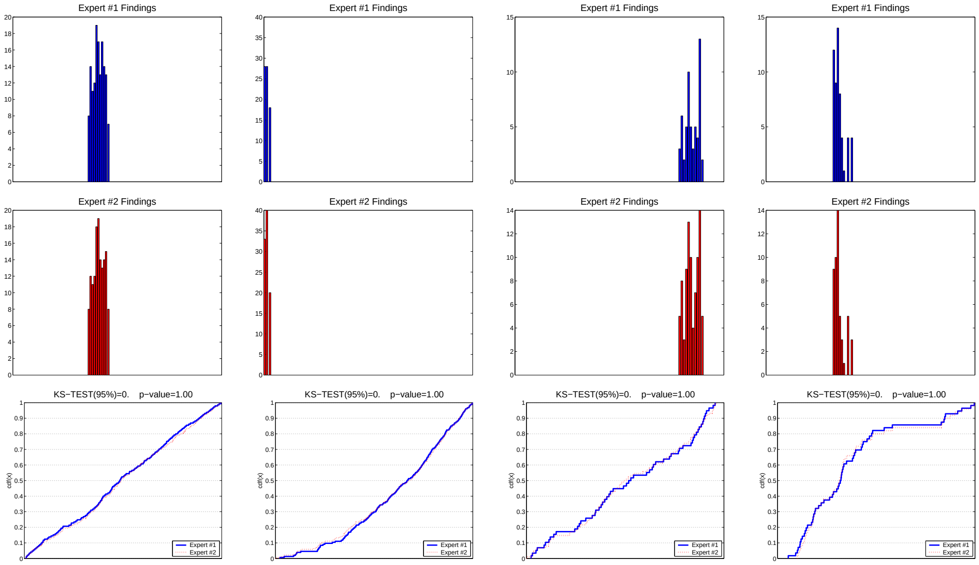
6.4.2 Qualitative analysis of the classification results

In addition to the former numerical analysis, we present a set of sequences provided by our system, which were carried out in order to obtain a qualitative impression of the classification results in terms of visual appearance. Figure 6.9 shows a selected set of intestinal contractions which had been labelled by the specialist, and which was successfully labelled by the system as a contraction, i.e., the true positives. Figure 6.10 shows missed contractions, which had been labelled by the specialists, but which the system rejected (false negatives). Finally, Figure 6.11 shows a set of sequences which had not been labelled by the experts, but which our system labelled as phasic contractions (false positives).



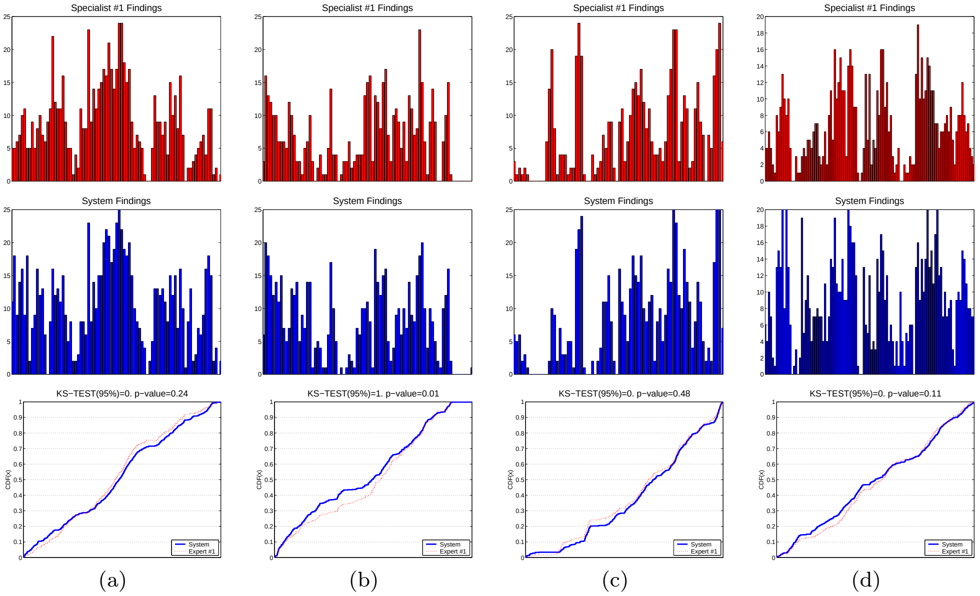
(a) (b) (c) (d)

Figure 6.5: Inter-observer variability for (a) *Video 1*, (b) *Video 2*, (c) *Video 3* and (d) *Video 4* findings.



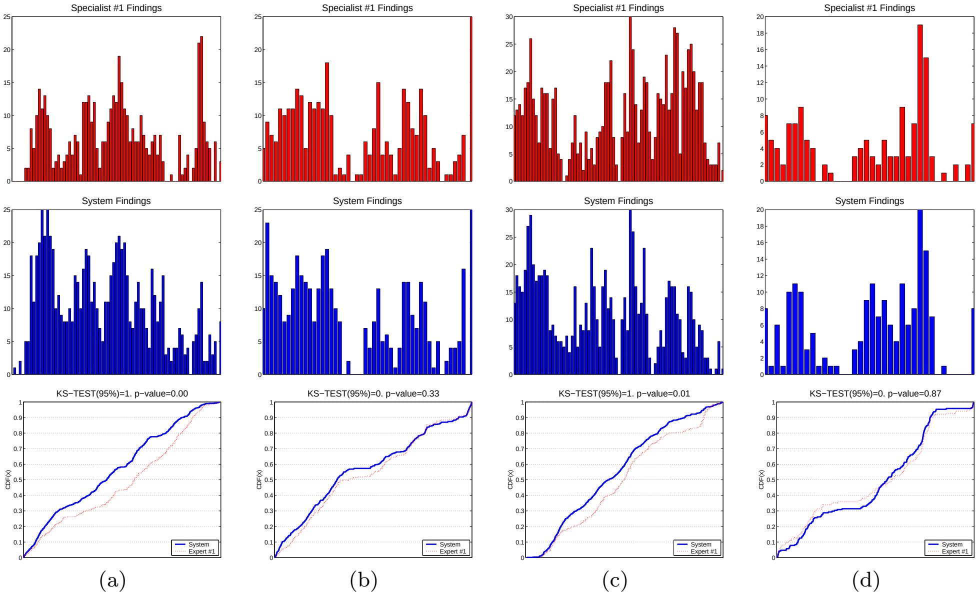
(a) (b) (c) (d)

Figure 6.6: Inter-observer variability for (a) *Video 5*, (b) *Video 6*, (c) *Video 7* and (d) *Video 8* findings.



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

Figure 6.7: Human (in red) vs. system (in blue) labelling histograms of phasic contractions for (a) *Video 1*, (b) *Video 2*, (c) *Video 3* and (d) *Video 4* findings.



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

Figure 6.8: Human (in red) vs. system (in blue) labelling histograms of phasic contractions for (a) *Video 5*, (b) *Video 6*, (c) *Video 7* and (d) *Video 8* findings.

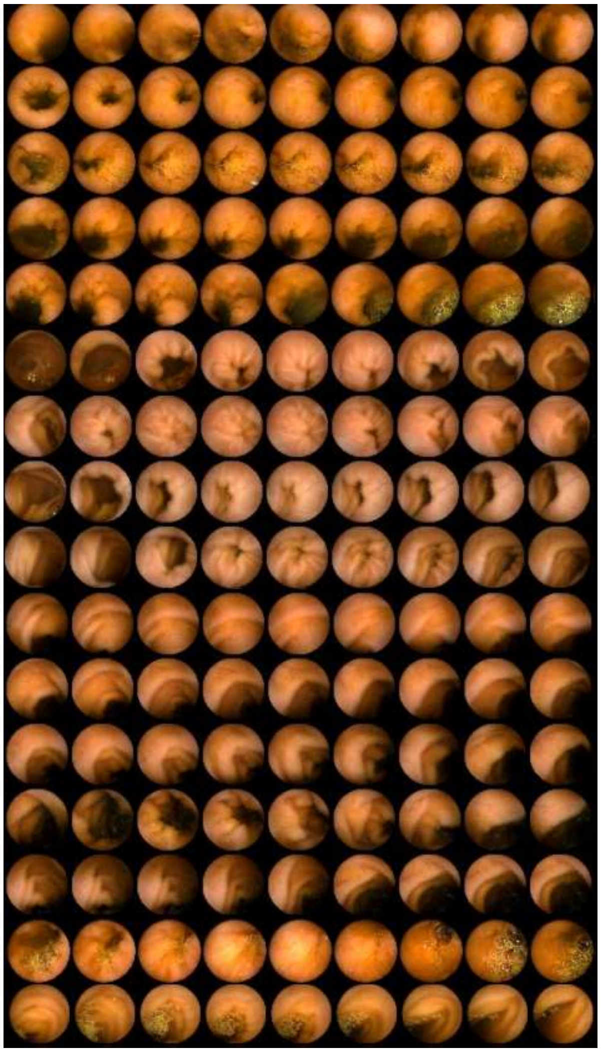


Figure 6.9: Correctly detected contractions by means of the automatic analysis (true positives).

The detected contractions basically correspond to the paradigm of phasic contractions described in section 3.2. In this sense, clear patterns of the intestinal lumen closing and opening are shown. It must be pointed out that the presence of turbid liquid in some frames does not result in a rejection of this sequence by the turbid detector, because only the clearest turbid sequences are rejected.

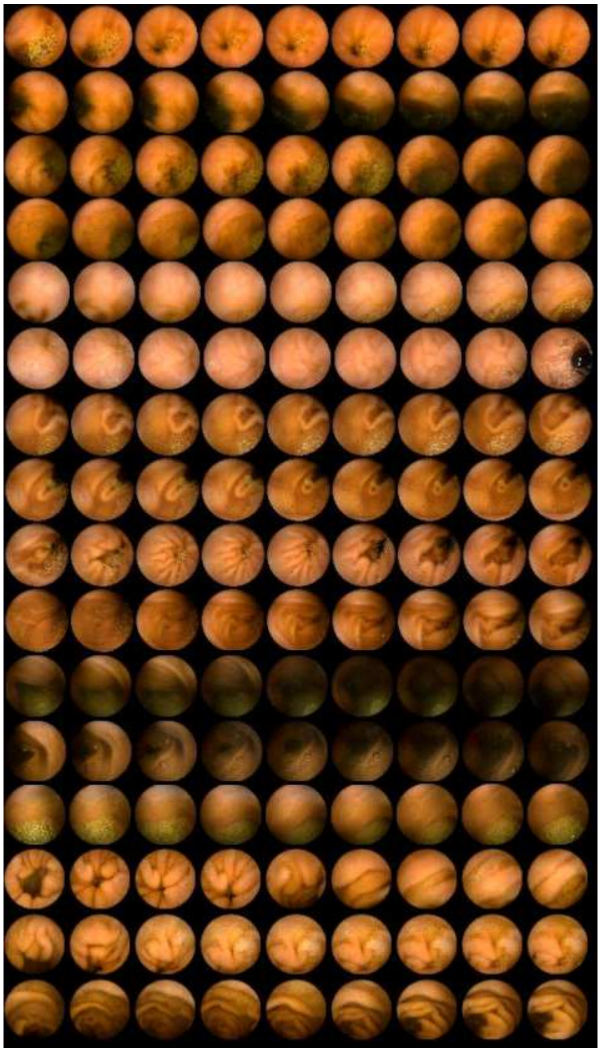


Figure 6.10: Missed contractions by means of the automatic analysis (false negatives).

The open lumen is not always present at the beginning and at the end of the sequence of most of the false negatives, as Figure 6.10 shows. This fact makes these sequences diverge from the expected pattern of phasic contractions which the system seeks for, and this outcome contributes to a decrease of the overall sensitivity. As a matter of fact, the random orientation of the camera, the duration over more than 9 frames of some contractions labelled by the experts, and the visual impression of motion during the video visualization contribute to a higher divergence in the patterns of contractions. An extended analysis of these issues can be consulted in the discussion Chapter 10.

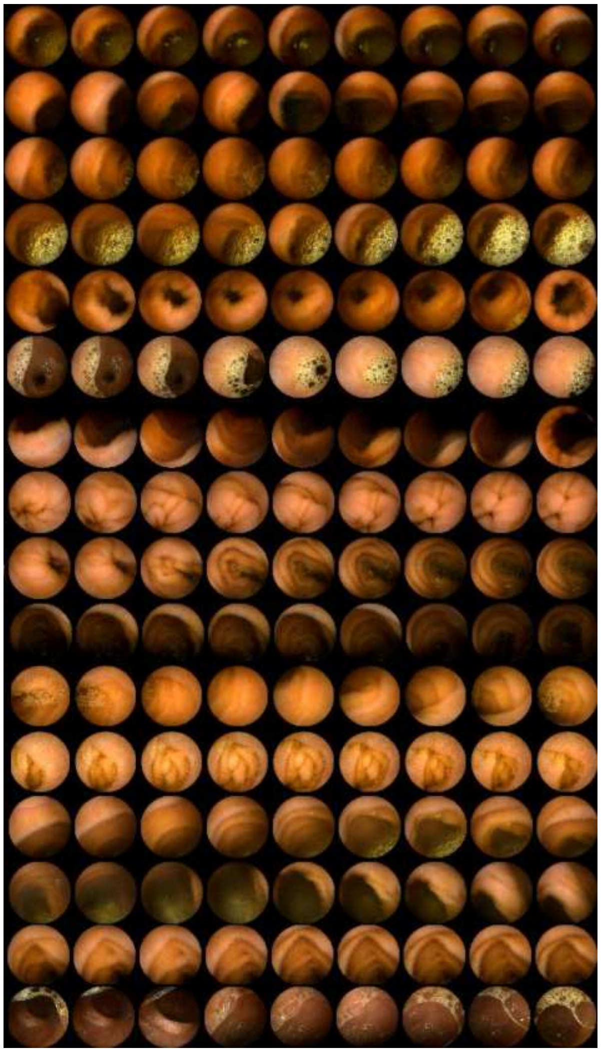


Figure 6.11: Sequences which have not been labelled by the experts, but detected as contractions by means of the automatic analysis (false positives).

The false positives analysis reveals that our system detected some clear cases of real phasic contractions which the experts did not label during the video visualization -an example of these sequences is rendered in the fifth row of Figure 6.11-. On the other hand, the sequences shown in Figure 6.11 display the inherent difficulty related to the high variability of patterns present at the output of the system: the lateral movement of the camera, the differences in illumination creating shadows which can be confused with the lumen, the multiple patterns of wrinkles which can provide a high response to the lumen detector, and the residual presence of patterns of turbid liquid, share the main responsibility in the false positives.

6.5 Validation of the system operation point

Providing that the set of parameters $P^0 = \{P_1^0, P_2^0, P_3^0, P_4^0\}$, was obtained in an exhaustive empirical search, we must assess that P^0 does correspond with an optimal operation point, in terms of system performance. In order to assess this issue, we proceeded with a forward-propagation algorithm for parameter selection, which is deployed in detail in algorithm 1. The procedure used essentially matches the following highlights: We reset all the parameters to P^0, and established a range of possible values for each of them: 16 values for P_1 within the interval [0:15], 11 values for P_2 within the interval [10:210], 11 values for P_3 within the interval [500:1000], and 8 heuristically selected values for P_4 [0.001,0.005,0.010,0.030,0.050,0.100,0.500,1.000]. The choice of these intervals was performed based on the minimum and maximum values for each stage. The interval of the last parameter γ was carefully selected based on the observation of a substantial variation of the classifier performance. After the initialization step, the system was evaluated for all the possible values of P_1 within the defined range, and the best operation point (P_1^{Best}) was selected according to the performance criteria defined in Algorithm 6.5. The value of P_1^0 was substituted by P_1^{Best}, repeating the same procedure for the rest of the parameters in a sequential way (P_2, P_3 and P_4). The whole procedure was repeated 5 runs and the final set P^{Best} was obtained by averaging $P^{Best_{i,i=1:5}}$.

Both the fast forward algorithm, and the performance criteria chosen are justified by the following reasons: On the one hand it must be taken into account that each single parameter modification has impact on the frames which are to be filtered by each specific stage, not only in the final assessment test, but also in the videos which are used for training in the leave-one-out strategy. In other words, when we vary one parameter, we must re-run all the system for each one of the 9 videos used for training, and we must apply a new leave-one-out strategy for each of them, training their classifiers using the 8 remaining videos. This clearly appears not to be computationally affordable using another parameter selection strategy which would imply a substantial increase in its computation time. On the other hand, a performance criterion function based on the global classification error does not appear to be a reliable metric in this context. In order to tackle the issue of performance assessment in imbalanced problems, several authors have proposed different solutions, including the use of the g-metric, the F-metric, and others [3]. Among the clinical community, the use of a trade-off between sensitivity and some other measure is widely extended. For our case, we demanded the experts to provide us with the reference of the performance threshold which should be used in the trade-off function, arriving at a final compromise around $sensitivity = 70\%$. Finally, we implemented this criterion within the criteria function defined in the fast-forward parameter selection algorithm:

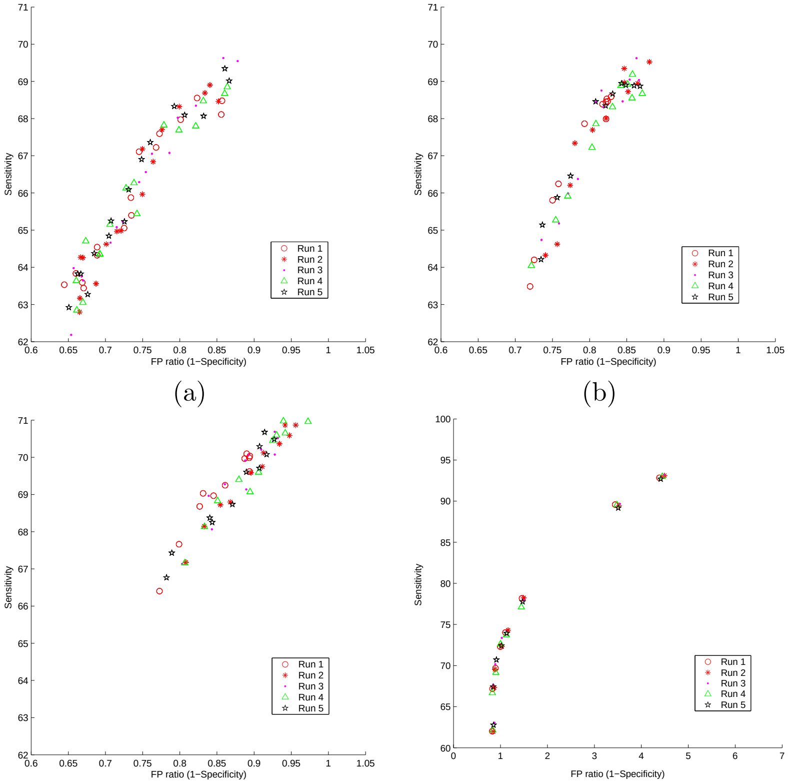
Fast-forward parameter selection. Algorithm 1

```
BEGIN
SET ranges for parameters:
$R_1 \rightarrow [0:15]$ {16 values}
$R_2 \rightarrow [10:210]$ {11 values}
$R_3 \rightarrow [500:1000]$ { 11 values}
$R_4 \rightarrow [0.001, 0.005, 0.010, 0.030, 0.050, 0.100, 0.500, 1.000]$
        {8 values}
for $i = 1$ to 5 do
  Set parameters: $P = P^0 = \{P_1^0 = 0, P_2^0 = 50, P_3^0 = 650, P_4^0 = 0.01\}$ {initialization}
  for $j = 1$ to 4 do
    Calculate the system performance substituting $P_j^0$ with each value of $R_j$.
    Apply the Performance Criterion to obtain $P_j^{Best}$.
    Substitute $P_j^0$ with $P_j^{Best}$ in $P$.
  end for
  $P^{Best_i} = \{P_1^{Best}, P_2^{Best}, P_3^{Best}, P_4^{Best}\}$
end for
$P^{Best} = avg(P^{Best_i})$
END

Performance Criterion:

For all the performance pairs (Sensitivity, FAR) obtained for each
parameter:
if For all the pairs, Sensitivity $\geq 70$ then
  We chose the parameter that achieves the higher sensitivity.
else
  We select the two parameters with a closest value to 70 (higher or lower)
  We choose the parameter which minimizes the error function:
  $sensitivity * (a * sensitivity^2 + b * FAR^2)$,   usign $a, b = 1$
end if
```

In order to accomplish a graphical analysis of this procedure, let us fix our attention on the ROC curves plots shown in Figure 6.12. In ROC curves, both sensitivity and specificity are plotted (properly speaking, the *FP-ratio* is plotted, defined as 1-specificity -notice the difference in the axes scale-) rendering the possible operation points of the system, and constituting a helpful tool for performance analysis. Figure 6.12 plots the points of the ROC curves segments corresponding to the different operation points provided by the different values of the parameter vector P after 5 runs. Each run is represented with a different symbol and color. Each graph (a), (b), (c) and (d) of Figure 6.12 corresponds to one parameter in P (P_1, P_2, P_3 and P_4). Figure 6.13 shows the points of the same ROC curves segments clustered by the same parameter. In these plots, each operation point is centered in the mean value of sensitivity and FP-ratio after the 5 runs, and the length of the ellipses axes is proportional to its standard deviation. ROC curves in Figure 6.12 show that our system appears to be robust, in the sense that the trade-off between sensitivity and specificity is kept for each run. The less FP-ratio, the less sensitivity is achieved. Furthermore, our system shows to be stable, in the sense that for several runs, the resulting operation point is confined in the ellipses drawn in the plots rendered in Figure 6.13, showing no hysterical responses. We can observe the monotonically growing curves for the differ-



(a)

(b)

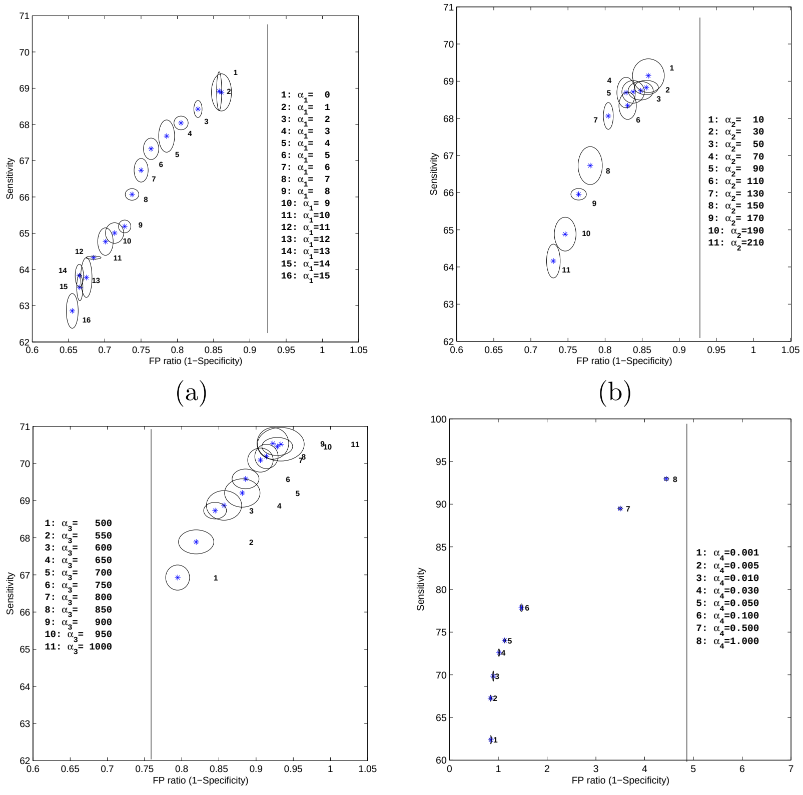
(c)

(d)

Figure 6.12: ROC curves segments for the forward parameter selection procedure for (a) P_1, (b) P_2, (c) P_3 and (d) P_4 grouped by runs.

ent parameter values, and the global displacement of the curve segment from 60% to 70% of sensitivity -(a) to (c)-. Parameter P_4 presents the widest range of variability, being consistent with the role of γ, which controls the margins which directly affect the support vectors used for classification.

The final performance of the system was calculated in two different ways: 1) averaging the performance point of the 5 runs of the validation procedure tuned with $P^{Best_{i,i=1:5}}$, and 2) averaging 5 runs of the system tuned in P^{Best}. Table 6.11 shows these results in comparison with the performance of the system tuned to P^0, exposed above. The final outcome confirms our hypothesis over P^0, since the confidence intervals of the performance values for the heuristically obtained parameters and those provided by the forward-propagation algorithm overlap both for sensitivity and FAR, assessing the equivalence of P^0 and P^{Best} in terms of performance.



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

Figure 6.13: ROC curves segments for the forward parameter selection procedure for (a) P_1, (b) P_2, (c) P_3 and (d) P_4 grouped by parameters.

Table 6.11: Performance operation point for the different parameters

Parameter	Sensitivity(std)		FAR(std)	
P^0	70.88	(0.51)	46.96	(0.79)
$P^{Best_{i,i=1:5}}$	71.35	(1.10)	47.96	(1.58)
P^{Best}	71.68	(0.44)	48.72	(0.54)

Chapter 7

Experiments with classifiers for phasic contractions detection

In this section, we provide the performance results of a set of classification experiments with contractions and non-contractions by using different advanced classification techniques, including SVM and Adaboost, among other powerful classifiers. We deploy a study about the relevance of different variables related to the classification process, such as the used classifier, the sampling methodology, and the use of combination strategies, including classifier ensembles techniques and cascades of classifiers. In the first section, we describe our first approach with the clearest patterns of occlusive contractions for the reduction of video inspection time. For this aim, we present the performance analysis of several ensembles of standard classifiers in a reduced set of features by means of the use of ROC curves and the AUC. In section 7.2, we test different variations of stratified sampling with SVM, which resulted the best classifier in our tests. We provide a comparative analysis by means of the use of PR-curves and the area under the PR-curve. In section 7.3, we investigate different alternative implementations of AdaBoost and compare them with the results obtained by the SVM classifier.

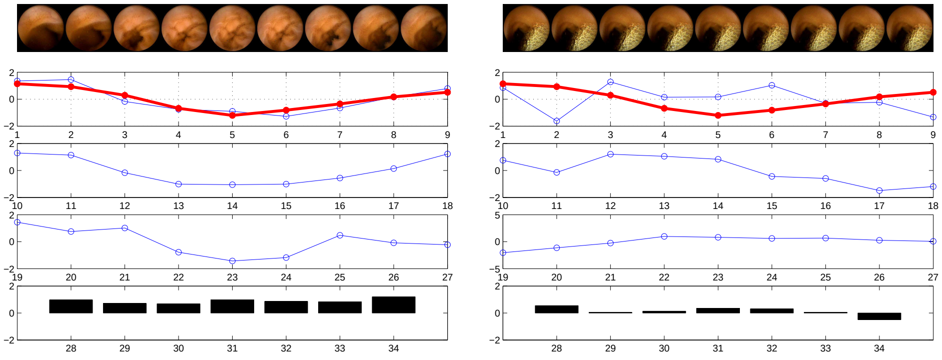
7.1 Classifier ensembles

7.1.1 Feature set for occlusive contractions

In order to characterize the pattern of occlusive contractions described in section 3.3, a 34-D feature vector $(x_1, \ldots, x_{34})$ was constructed as follows:

The first 9 features $x_1, \ldots, x_9$ correspond to the $-f^1$ value for the 9 frames sequence, taking into account the previous and following 4 neighbor frames. We used the same strategy with features $x_{10}, \ldots, x_{18}$ and $x_{19}, \ldots, x_{27}$, which correspond to the f^2 and $-f^3$ values, respectively. The patterns described by $x_1, \ldots, x_9$, $x_{10}, \ldots, x_{18}$

and $x_{19}, \ldots, x_{27}$ are normalized by taking out the mean and dividing by the standard deviation. The 7 remaining features are: x_{28} is the correlation between sequences $x_1, \ldots, x_9$ and $x_{10}, \ldots, x_{18}$; x_{29} is the correlation between sequences $x_1, \ldots, x_9$ and $x_{19}, \ldots, x_{27}$; and x_{30} is the correlation between sequences $x_{10}, \ldots, x_{18}$ and $x_{19}, \ldots, x_{27}$. Features x_{31}, x_{32}, x_{33} are the correlations between sequences $x_1, \ldots, x_9$, $x_{10}, \ldots, x_{18}$ and $x_{19}, \ldots, x_{27}$ on the one hand and the corresponding sequences averaged across the objects for the class "contractions". Feature x_{34} is the variance of intensity averaged across the 9 frames. This value is then normalized by taking out the mean across the whole video and dividing by the standard deviation. Features $x_1, \ldots, x_{34}$ are plotted in Figure 7.1 for examples of contractions and non-contractions. The sequences are joined by lines so as to see the shape patterns. The sequence $x_1, \ldots, x_9$ averaged across class contraction is overlaid in both subplots (the red line).



(a) (b)

Figure 7.1: Feature patterns for (a) contractions and (b) non-contractions. Each feature is identified by its number in the horizontal axis. The red line corresponds to the average value of the first descriptor for class contraction.

7.1.2 Single classifiers and ensembles

The prevalence of occlusive contractions in capsule endoscopy videos is very small: there are typically 30-80 contractions. After applying the dynamic events detector described in section 5.1, we found a typical ratio less than 1:100 in a video sequence of 5,000 frames, which implies an imbalanced problem. In order to test the behavior of several individual classifiers and the resulting ensemble of their combination, we carried out the following experiments:

We used 8 individual classifiers and two classifier ensemble methods. The individual classifiers were linear discriminant classifier (LDC), quadratic discriminant classifier (QDC), logistic classifier (LOGLC), nearest neighbor (k-NN) with $k \in \{1, 5, 10\}$, decision trees (DT), and Parzen classifier (Parzen) [28]. The two ensemble methods were: heterogeneous ensembles and bagging. The heterogeneous ensembles were built by taking a set of single classifiers of different types and aggregating their outputs. As we applied 8 classifiers, there are $2^8 - 1$(empty set) $-$ 8(single classifiers) $= 247$ possible heterogeneous ensembles. Bagging produces a classifier ensemble whereby each classifier is trained on a bootstrap sample. We constructed bagging ensembles of 25 decision trees. In this study, we used the average of the classifier outputs to be the ensemble output for both ensemble methods. This was done because we need a continuous-valued output as the ensemble decision. Our choice of bagging over AdaBoost (used for imbalanced problems in [85]) was based on the findings in the recent literature that bagging is the better of the two models for data sets with substantial amount of noise [9]. In order to rank the classifiers performance, we used the area under the ROC curve.

7.1.3 Results

Our experiments were built in the following way: the specialist analyzed 10 videos and manually labeled all contractions. A subset of 305 typical examples was then selected to be our class 'contraction' (positive). For the 'non-contraction' class (negative), 3050 examples were randomly chosen from all the videos, taking special care that the selected sequences did not belong to class contraction.

All 8 classifiers, the 247 heterogeneous ensembles and the bagging ensemble (25 decision trees) were trained and tested 100 times and the results were averaged. For each run we used the 305 contraction objects and a random bootstrap sample of size 305 from the class 'non-contraction'. This set of 610 objects was split randomly into 80/20 proportion for training and testing, respectively.

For all classifiers we used the Matlab toolbox PRTOOLS developed by R.P. Duin and his group at the Delft University of Technology [29]. We used the implementation of the single classifiers (LDC, QDC, LOGLC, 1-NN, 5-NN, 10-NN, decision trees and Parzen) and built our own ensembles and bagging routines. The continuous-valued outputs of all classifiers were used (these are available within PRTOOLS). For the calculus of the AUC, we used the trapezoidal rule, approximating the underlying function using linear interpolation.

The ROC curves for all classifiers were calculated on the testing set. The ensemble with the largest area under the curve appeared to be the one using just two classifiers: decision tree and Parzen, $AUC = 0.9603$. Figure 7.2 plots the ROC curve for this ensemble and also the ROC curves for the two component classifiers. The remaining single classifiers were very similar to one another and slightly worse than the Parzen classifier. The hybrid ensemble outperforms all single classifiers and follows the best behavior of its components in the different areas of the ROC curve. Table 1 displays the AUC for the individual classifiers as well as for the best 8 hybrid ensembles.

Table 7.1: AUC for single classifiers an the best 8 ensembles

Classifier	AUC	Ensemble	AUC
LDC	0.9040	PARZEN + DT	**0.9603**
QDC	0.8878	DT + 10-NN	0.9599
LOQLC	0.9033	DT + 1-NN + 10-NN	0.9598
PARZEN	0.9160	DT + 5-NN	0.9591
DT	**0.9463**	DT + 1-NN + 5-NN	0.9591
1-NN	0.8938	DT + 1-NN	0.9583
5-NN	0.9582	PARZEN + DT + 1-NN	0.9582
10-NN	0.9567	LOQC + DT + 1-NN	0.9567

Optimizing the classification by means of ROC curves

In imbalanced problems such as detection of contractions in endoscopy videos we are looking for operation points on the curve which will present the user with the best time-accuracy compromise. The overall performance of the classifier is of secondary importance. The physicians are interested in two different operation points on the ROC curve: *accuracy* of positive detection over a 98%, and *minimization of visualization time*, with a guaranteed positive detection over 80%. This brings to the fore two different areas of the ROC curve as marked in Figure 7.2. The shaded vertical stripe shows an example of a desirable time-optimization area. Its width denotes the maximum FP rate we are prepared to accept. In order to see how this is related to time-optimization, consider an example of an (unthresholded)video of 20,000 frames with 30 contractions in it. Assuming that all contractions were correctly labeled, the total number of frames which the system will leave to the expert to inspect is approximately $30 + 0.1 \times 19{,}970 = 2{,}027$. A lower acceptable FP rate, e.g., 0.05, will leave just over 1000 frames for inspection. Thus, in such a heavily imbalanced problem, the inspection time will depend exclusively upon the false positive rate. The interpretation of the shaded horizontal stripe is trivial: its height measures the amount of accuracy we are prepared to sacrifice. In the example in Figure 7.2, we accept at least 90% true positives, i.e., the maximum number of missed contractions in the example above should be 3 or less.

Classifiers that perform best in one area may not be the best in the other area. As the plot shows, the decision tree classifier is very good for the accuracy optimization, while the Parzen classifier is more successful for minimization of visualization time. The hybrid ensemble outperforms both, which re-confirms the widely accepted claim that ensembles are superior to single classifiers.

Bagging ensemble was constructed from 25 decision trees as base classifiers, with

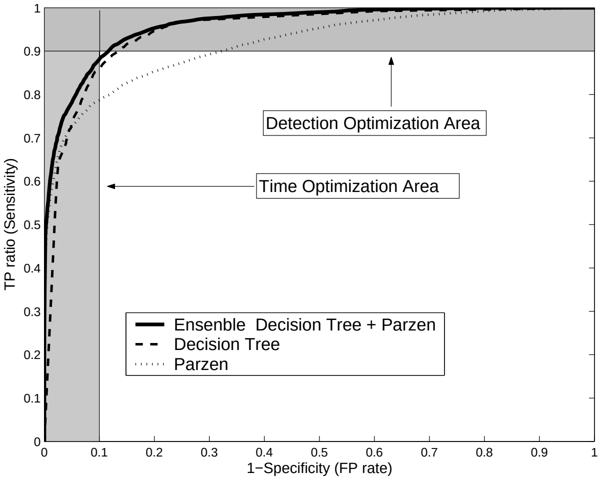
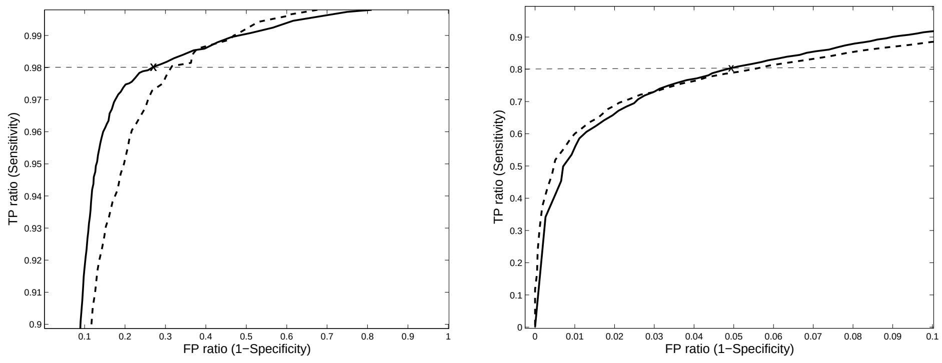


Figure 7.2: The ensemble with largest AUC (Parzen + decision tree) outperforms both single classifiers and follows the best behavior of its components in the ROC curve.

a resulting $AUC = 0.9647$. Figure 7.3 (a) shows the ROC curves for the bagging ensemble and the best heterogeneous ensemble at TP rate of 98% (accuracy zone). The best hybrid ensemble for this zone appeared to be the one consisting of a decision tree and 10-NN, with $AUC = 0.9599$. Bagging shows superior performance at the point of entering this zone. The heterogeneous ensemble outperforms bagging for FP rate over 50% which renders large inspection time. The same analysis was applied for TP rate of 80% and the result is plotted in Figure 7.3 (b). In this case, the best heterogeneous ensemble consists of a decision tree, 1-NN and 5-NN, with an $AUC = 0.9591$. The bagging ensemble only slightly outperforms this ensemble at the desired point. In contrast to the previous case, the heterogeneous ensemble is better for low accuracy rates, e.g., under 70%. Even by small differences, Figure 7.3 (b) favors the bagging ensemble as the best classifier for both operation points.

Since the main objective of this study is to look for a compromise between inspection time and accuracy, we suggest a variant of the ROC curve. On the horizontal axis we plot the inspection time required, and on the vertical axis the sensitivity of the classification. The inspection time is calculated in the following way: the output of a classifier is a set of frames with suspected contractions (true positive and false positive classifications). Each frame must be visualized as a middle frame in the sequence of 9 frames in order to create the dynamic impression. The typical visualization rate is 5 frames per second. That implies 1.8 seconds for each sequence (i.e., for each output frame), and a bound of 2 seconds can be used. The total visualization time for one video will be, therefore, the number of output frames multiplied by 2. The horizontal axis plays the role of a re-scaling, but not a mere re-scaling of the FP rate. Consider a thresholded video of 5,000 frames with 30 contractions in it. Take an (x, y) point from the standard ROC curve. To calculate the corresponding x' on our ROC variant,



(a) (b)

Figure 7.3: Bagging (solid) vs. best ensemble (dashed) for each zone of interest. (a) Bagging vs. hybrid ensemble (decision tree and 10-NN) in the sensitivity optimization area. (b) Bagging vs. hybrid ensemble (decision tree, 1-NN and 5-NN) in the time optimization area.

we use $x' = (5{,}000x + 30y) \times 2$.

We used the best ROC curve for any point, so different classifiers are responsible for different parts of the curve. This was done in the following way. Suppose that the ROC curves for all classifiers and ensembles are drawn on the same plot. For each value of FP we selected the curve with the maximum TP (the highest curve). In different parts of the ROC curve, different classifiers or ensembles might be the best. A system operating in a real environment should keep the collection of classifiers and ensembles which make up the overall "best" ROC curve. The operation point selected by the physician will translate into running the classifier or the ensemble responsible for this point.

Figure 7.4 shows the ROC variant (solid line). Two more ensemble ROC curves are shown for comparison, demonstrating that both ROC curves are inferior to the combined one. Table 7.2 shows a summary of the results for TP accuracy and visualization time. For the manual method, time is calculated assuming that the physician inspects the video at 5 frames per second. For the rest of cases, time is calculated as explained above. For 80% sensitivity, only 9 minutes and 10 seconds are needed, while 67 minutes are needed for manual labelling.

7.2 Sampling and SVM

Recently, several methods have been developed to improve the performance of SVM classifiers on imbalanced problems [3, 14]. Stratified sampling is based on re-sampling the original datasets in different ways: under-sampling the majority class or over-

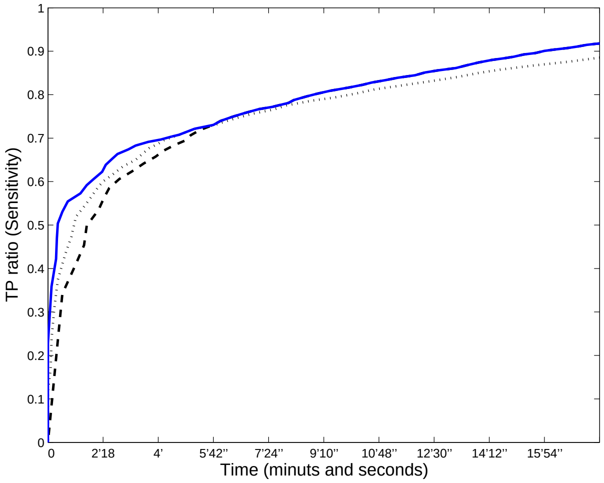


Figure 7.4: The time-accuracy variant of ROC curve (solid line). TP are plotted against visualization time. For comparison, two more ensemble curves are shown: (i) Bagging (dashed line) and (ii) decision tree, 1-NN and 5-NN (dotted line).

Table 7.2: Analysis of a 20,000-frames video with 30 contractions.

Method	**Positives**	**# Frames**	**Visualization Time**
MANUAL	30(100%)	20,000	1 hour 7 min
Dynamic events	29(99%)	5,000	2 hours 45 min
98% Detection	28(98%)	500	16 min 40 sec
80% Detection	23(80%)	275	9 min 10 sec

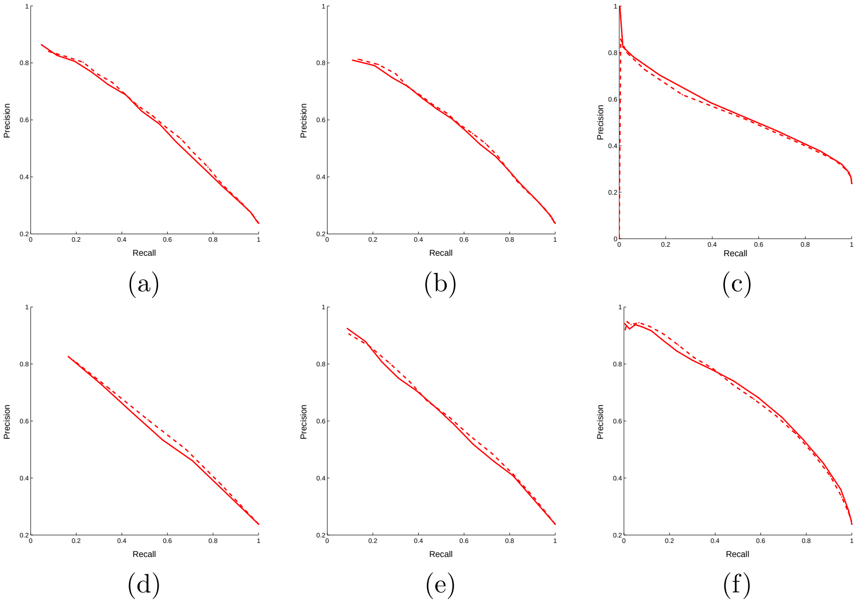
sampling the minority class. In these experiments, we tested under-sampling vs. SMOTE over-sampling technique [20]. We used both techniques with SVM. In order to obtain a comparative feedback, we also tested stratified sampling on a set of common single classifiers.

7.2.1 Results

Two sampling methods were tested: under-sampling and SMOTE. For under-sampling, we trained the classifiers choosing randomly from the training set a number of non-contractions equal to that of contractions, so the final training set for each leave-one-out run was built up with 500x7 contractions and the same number of non-contractions -around 7,000 training samples. For SMOTE, we replicated the positive samples up to the number of non-contractions using 5 nearest neighbors. Since around 2,500 frames typically pass the first stage, this yields to a training set of around 35,000 samples.

We carried out our experiments in order to answer two main questions: 1) in what measure the sampling technique used affects the performance of the classifiers tested, and 2) in what measure SVM outperforms the rest of classifiers for our dataset. With the aim of illustrating both questions, we used the precision-recall curves (*PR-curves*), which were introduced in section 4.4.1, and constitute a standard evaluation technique within the information retrieval community [11]. *Precision* is a measure in the interval [0,1] defined as the ratio of true positives detected over all the system output. It will be one when the system detects only positives, and zero when the system detects only negatives. In this way, a low precision value is associated with a high number of false positives, and it can be viewed as a measure of noise at the output stage. *Recall* -also known as *sensitivity*- is a measure in the interval [0,1] defined as the ratio of true positives detected over the total number of positives at the input of the classifier. It will be 1 when the system detects all the existing positives, and zero when no positive is detected. For imbalanced problems, both measures together give more information than ROC curves. In ROC curves [63, 73], precision is substituted by 1-specificity, or false positive ratio, and so the proportion between positives and negatives is not taken into account.

In order to give response to question 1, comparative plots are presented in Fig. 7.5. We can see that no substantial improvement is achieved by any of both sampling techniques, neither for single classifiers nor for the SVM, and the classifiers performance seems to be quite invariant to the sample method.

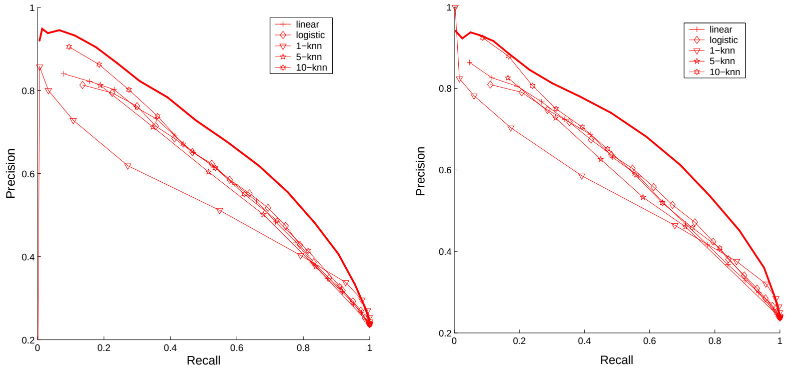


(a) (b) (c)

(d) (e) (f)

Figure 7.5: Under-sampling (dashed) vs. SMOTE over-sampling (solid) for detection of intestinal contractions with several classifiers: (a) linear, (b) logistic, (c) 1-NN, (d) 5-NN, (e) 10-NN, (f) SVM.

Regarding the performance of SVM with respect to the rest of the classifiers, both for under-sampling and SMOTE, SVM outperforms their competitors. Fig. 7.6 shows that only for low levels of recall -when almost no positive is detected- NN classifiers appear to be competitive. A quantitative approach can be analyzed by means of the area under the PR curve (AUPRC). The AUPRC gives us a metric for assessing the classifier performance. We calculated it using the polygonal rule, in the same way as it is done for the calculus of the area under a ROC curve [13]. Table 7.3 shows the AUPRC related to the graphs in Fig. 7.6. The statistical significance of the results was assessed by means of a t-test over the different values. An extended discussion about these results can be found in Chapter 10.



(a)

(b)

Figure 7.6: Single classifiers vs. SVM (solid line) for (a) under-sampling and (b) SMOTE.

Table 7.3: AUPRC for under-sampling and SMOTE experiments of Fig. 7.6.

Classifier	**AUPRC Under-samp.**	**AUPRC SMOTE**
Linear	0.6100	0.6021
Logistic	0.6067	0.6006
1-NN	0.5324	0.5485
5-NN	0.5946	0.5849
10-NN	0.6282	0.6216
SVM	**0.6934**	**0.6946**

7.3 Experiments with AdaBoost

We performed several experiments with AdaBoost. The main aim of this set of experiments was to verify the impact of different training strategies in the classifier performance. In order to do that, we tested three different approaches which are referred to as $AdaBoost$, $AdaBoost_{mod-1}$ and $AdaBoost_{mod-2}$. The following sections describe these different approaches.

7.3.1 $AdaBoost_{mod-1}$ implementation

$AdaBoost_{mod-1}$ splits the original training set $(X;Y)$, where X represents a set of N samples and Y its corresponding set of labels, into two separate sets: 1) the new training set $(X_{tr};Y_{tr})$ with N_{tr} samples, and 2) a remaining set $(X^{'}_{tr};Y^{'}_{tr})$ with $N-N_{tr}$ samples. All the classifiers are trained only with $(X_{tr};Y_{tr})$. The rest of the algorithm is identical to the original *AdaBoost* algorithm. With this modification, the error function is calculated using some samples which are not used in the training stage, so the classifier with the best performance is expected to be the one with the best generalization power for the whole set. The rest of the algorithm is identical to the original AdaBoost algorithm. The new edition of step 5 is:

$\mathbf{AdaBoost_{mod-1}}$:

5 : For each feature, j, train a classifier h_j with $(X_{tr};Y_{tr})$, which is restricted to using a single feature. The error is evaluated in (X; Y) with respect to w_t, $\epsilon_j = \sum_{x^n_{i=1}} w_i|h_j(x_i) - y_i|$.

7.3.2 $AdaBoost_{mod-2}$ implementation

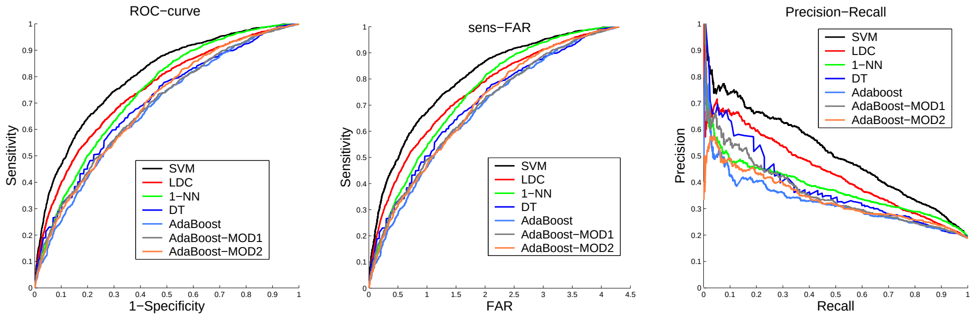
$AdaBoost_{mod-2}$ splits the original training set $(X;Y)$ into two separate sets: 1) the new training set $(X_{tr};Y_{tr})$ with N_{tr} samples, and 2) a remaining set $(X^{'}_{tr};Y^{'}_{tr})$ with $N-N_{tr}$ samples, exactly in the same way that $AdaBoost_{mod-1}$ does. The main difference with $AdaBoost_{mod-1}$ is that the training stage described in step 5 is performed by $(X^s_{tr};Y^s_{tr})$, consisting in N^s_t random samples from $(X_{tr};Y_{tr})$, which are re-sampled by bootstrap for each iteration t. Thus, all the classifiers in the same iteration are trained with the same training set, which is an under-sampling of $(X_{tr};Y_{tr})$, but for different iterations the training set changes. With this modification, we investigate the impact of the diversity introduced in the classification algorithm by means of the change of the basic training set using random under-sampling. The new edition of step 5 is thus:

$\mathbf{AdaBoost_{mod-2}}$:

5 : For each feature, j, train a classifier h_j with N^s_t samples randomly obtained from $(X_{tr};Y_{tr})$, which is restricted to using a single feature. The error is evaluated in (X; Y) with respect to w_t, $\epsilon_j = \sum_{x^n_{i=1}} w_i|h_j(x_i) - y_i|$.

7.3.3 Results

We show the results of the experiments performed with the previously explained strategies. We used decision stumps (single threshold over one single feature) as the weak classifiers for all the three AdaBoost implementations, and we under-sampled the majority class for the training data set. The split ratio between X_{tr} and $X_{tr}^{'}$ was 0.8. The performance curves were obtained by averaging 10 runs, and are shown in Figure 7.7 for one representative video. The classification results for other classifiers, such as SVM, linear discriminant (LDC), 1-nearest neighbor (1-NN), and decision trees (DT) are also shown. The LDC implementation was based on the Bayes classifier, assuming Gaussian distributions for both the classes. The DT implementation used information gain criterion and pessimistic pruning as defined by Quinlan [69].

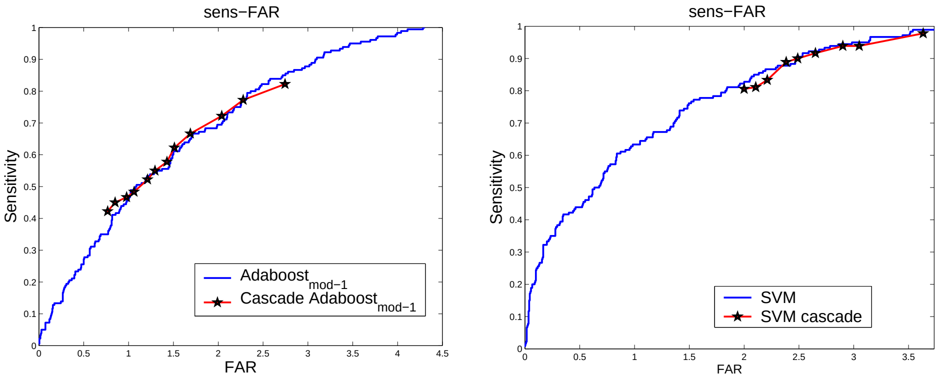


(a) (b) (c)

Figure 7.7: (a) ROC, (b) sensitivity-FAR, and (c) PR-curves for different classifiers.

The analysis of the curves clearly shows that SVM outperforms the rest of classifiers. LDC slightly outperforms the rest of classifiers, except SVM, in terms of AUC, which was assessed by means of a t-test, and also in terms of precision. The three Adaboost implementations showed a lower performance than the rest of classifiers. Both $Adaboost_{mod-1}$ and $Adaboost_{mod-2}$ appear to outperform *Adaboost* in precision for low recall values, although the three plots asymptotically converge for high values.

We finally performed a set of experiments oriented to test the model of cascade of classifiers with AdaBoost. In order to construct the cascade, we followed this procedure: We first split our data set into a training set with Num^{train} samples and a testing set with Num^{test} samples, using a ratio of 80% for training and 20% for testing. First, we trained one classifier with all the $Num_P = Num^{train}$ samples of positives, and $Num_N = Num_P$ (the same number) of negatives. The remaining negatives were kept as a repository of negatives for the next stages of the cascade. We used the output of the first classifier applied to the training data set so that to find the point in the ROC curve with 98% of sensitivity. In this way, a certain quantity of negatives $Num_{N_1}^{rej}$ were rejected by the system as true negatives, and Num_{N_1} were classified as false positives, while Num_{P_1} of the training positives were kept as true positives. In the next step, $Num_{P_1} - Num_{N_1}$ negative samples are obtained from the repository, so that the training set is balanced again. We repeat this process until the negative repository is empty or $Num_{N_i} > Num_{P_i}$ for the *ith*-stage of the cascade. The results of the application of this method in one video are shown in Figure 7.8 for (a) $Adaboost_{mod-1}$ and (b) SVM classifiers.



(a) (b)

Figure 7.8: Comparison of (a) $AdaBoost_{mod-1}$ vs. cascade and (b) SVM vs. cascade SVM.

Each point in the cascade plot corresponds to one stage. The plots of Figure 7.8 show that, for both the cases, the resulting ROC segment generated by the cascade approach appear to undergo similar values of performance than the single classifiers. Moreover, the AdaBoost cascade (a) present a lower generalization power, which is inferred by the starting point under 80% sensitivity when the cascade is applied to the test set.

Part III

AUTOMATIC DETECTION OF TONIC CONTRACTIONS

Chapter 8

Tonic contractions detection: Methodology

Tonic contractions present a completely different visual paradigm than phasic contractions, and therefore, the methods explained for phasic contractions detection are no longer valid. In this chapter, we deploy our proposal as a first approach to the detection of tonic contractions in video capsule endoscopy.

8.1 Sustained contractions as a visual pattern for tonic contractions

As we described in section 3.2.2, the paradigm of tonic contractions in capsule endoscopy corresponds to sequences of an undetermined number of frames [1], showing a completely closed lumen for the whole sequence. This can be viewed as a long phasic contraction which sustains the contraction event for several frames. Thus, the appearance of sustained contractions is mainly characterized by a star-wise pattern of wrinkles which is repeated for several frames. Figure 8.1 shows a sequence of a sustained contraction spanning 20 frames.

8.2 Wrinkle detection

Our hypothesis is that the characteristic wrinkles of sustained contractions are associated with the valley and ridge analysis of the intensity image. Thus, we orient our wrinkle detector to the detection of the valleys and ridges of the image. But before the application of a method for the valley and ridge detection, endoscopy images must

[1] In order to discriminate with phasic contractions, we fixed, following the physicians directives, the minimum length for a sustained contraction in 10 frames

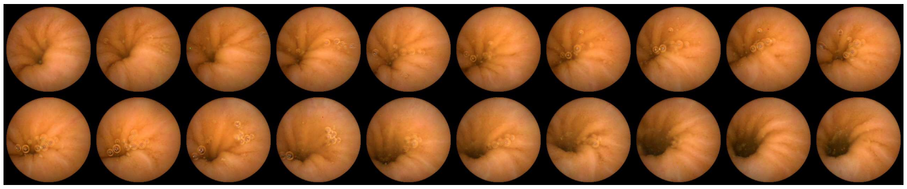


Figure 8.1: A tonic contraction. The characteristic visual pattern of the wrinkles is sustained for 20 frames

be pre-processed in order to smooth them. Thus, our proposal for wrinkle detection in endoluminal images of the gut is structured twofold: 1) smoothing of the original image, and 2) application of valley and ridge detection.

8.2.1 Smoothing of the original image

The pre-processing smoothing stage is performed by the application of a median filter, with a fixed rectangular window. The size of the median filter window is set to the mean width of wrinkles in sustained contractions, which is set to 6.5 pixels. The application of this filter is justified by the sharpness of the images in areas where a homogeneous view of the intestinal walls is rendered. This is mainly due to the physiological structure of the intestinal walls tissue, and the some amount of electronic noise. Figure 8.2 renders one example of the results of the median filter smoothing.

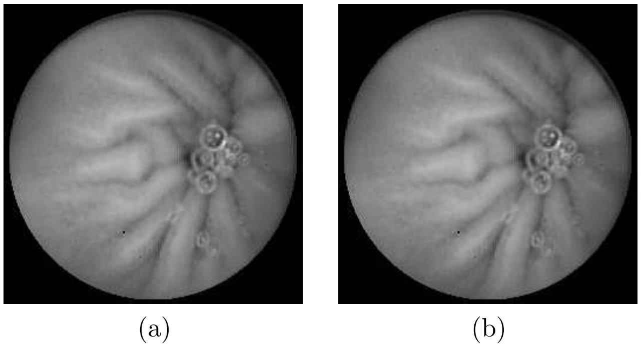


Figure 8.2: (a) Original image. (b) Diffused image by using a median filter.

8.2.2 Valley and ridge detection

The valley and ridge detection procedure is performed in the following way:

1. We create a filter mask by calculating the second derivative of an anisotropic Gaussian kernel [59]. The anisotropic kernel used had $\sigma_1 = 1$ and $\sigma_2 = 2$ for each direction.

2. We obtain 4 different filter responses $F^i(n)$ for an input image I_n as:

$$F^i(n) = I_n * \ kern_{\alpha_i}, \quad \alpha_i = \frac{i\pi}{4} \tag{8.1}$$

where, α_i represents 4 different orientations $0^0, 45^0, 90^0$ and 135^0, $kern_{\alpha_i}$ represents the anisotropic kernel rotated α_i radians, and $*$ represents the convolution operator.

3. The valley and ridges images F^{val} and F^{rid} are calculated as:

$$\begin{array}{rcl} F^{val}(n) & = & \max_{(x,y)}\{F^i(n)\} \\ F^{rid}(n) & = & \max_{(x,y)}\{(-1) * F^i(n)\} \end{array} \tag{8.2}$$

where $max_{(x,y)}$ represents the maximum value of the F^i functions for the (x, y) pixel. Figure 8.3 (a) shows the valley and ridges images F^{val} and F^{rid} for the example image of Figure 8.2.

4. We create a binary image by keeping the 75% percentile of $F^{val}(n)$ and $F^{rid}(n)$. Figure 8.3 (b) shows the binary images created by this procedure.

5. Finally, we apply a morphological skeletonisation [43] in order to obtain the lines with one pixel connectivity which describe the valleys and the ridges. Figure 8.3 (c) shows the skeletons created by this procedure.

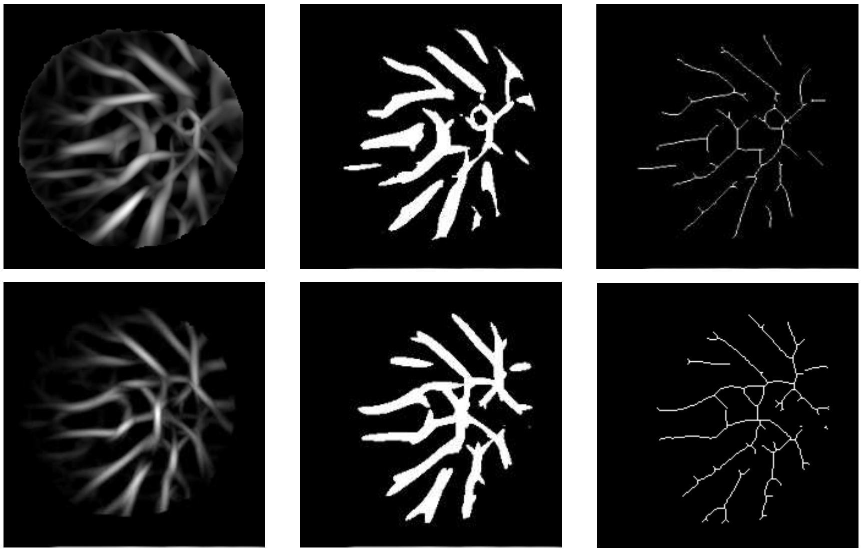

(a) (b) (c)

Figure 8.3: (a) Detection of valleys (top) and ridges (bottom). (b) Binary images. (c) Skeletonisation results

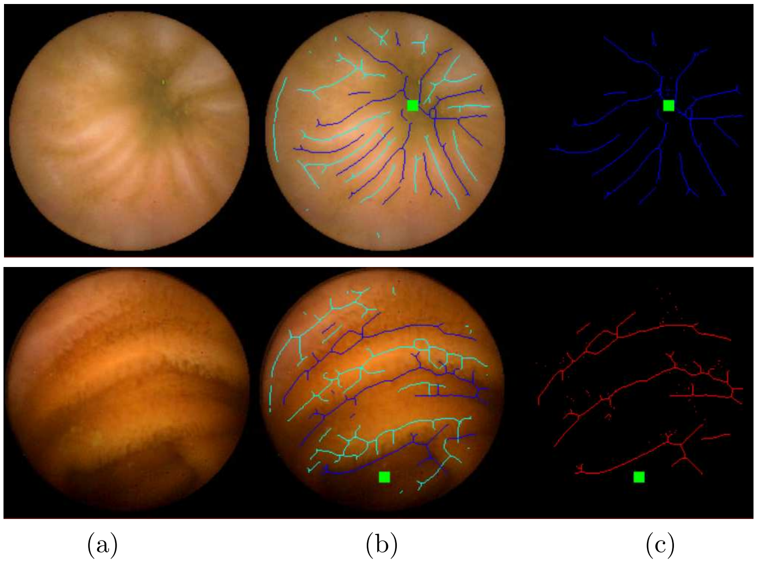


Figure 8.4: (a) Original image. (b) Wrinkles detected as valleys (blue) and ridges (cyan). (c) Valley wrinkles.

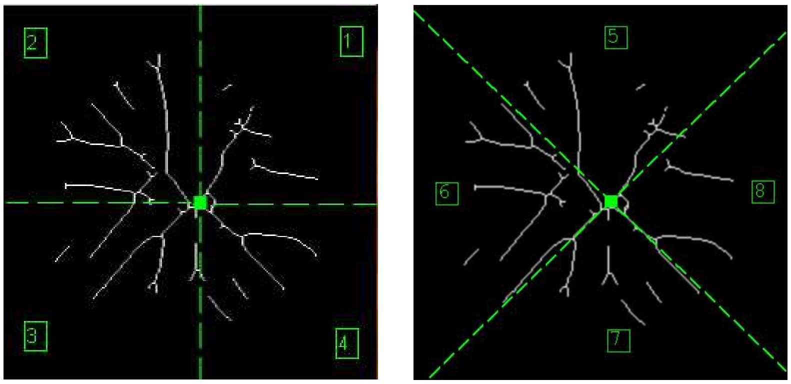
8.3 Wrinkle descriptors

The aim of this section is to explain the set of descriptors we used to characterize the radial patterns of wrinkles. Figure 8.4 shows the super-imposition of the valleys and wrinkles for two test frames: (top) a frame from a sustained contraction, and (bottom) a random frame. The green square corresponds to the centroid of the lumen estimated by means of the Laplacian of Gaussian detector. Notice that the centroid of the lumen appears in the middle of the radial wrinkle pattern for the contraction frame. On the contrary, for the random frame the position of the centroid of the lumen does not follow a fixed pattern.

We propose two different strategies to obtain directional descriptors for sustained contractions characterization. The first approach is based on 4 descriptors which code the multi-directional information of the wrinkles. The second approach simplifies the number of descriptors to 2, by means of a polar transform of the original data. Although the wrinkle pattern is defined both by valleys and ridges, we used only the valleys pattern as a source of wrinkle information. An extended analysis of this choice is deployed in the discussion in Chapter 10.

8.3.1 Multidirectional approach

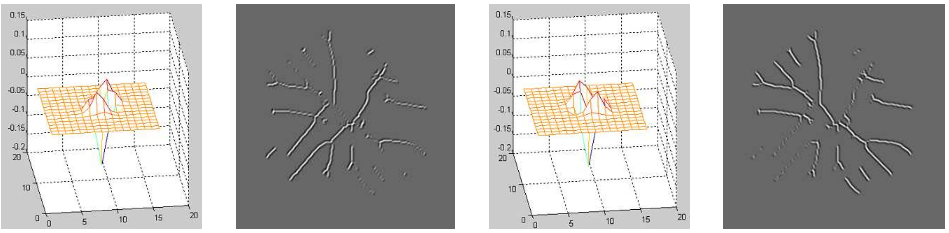
For a given wrinkle pattern, we define 4 different quadrants -see Figure 8.5 (a)-, which we denote by quadrant 1, 2, 3 and 4, using the centroid of the lumen, calculated as explained in section 5.2.1, as the quadrant division middle point.



(a) (b)

Figure 8.5: Quadrant division of a frame: (a) Diagonal and (b) vertical and horizontal directional analysis

The directional information is obtained by means of 2 second derivative of Gaussian steerable filters [34, 83, 82], oriented to 45^0 and 135^0 -the mathematical foundations of the design of steerable filters can be found in Appendix B-. For both filters, $\sigma = 1$, so they basically operate as line detectors in the direction towards which they are oriented. Figure 8.6 shows the 3D plot of the filters masks, and the response of both filters for an example of wrinkles pattern.



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

Figure 8.6: Filter masks and image responses for two orientations: 45^0 ((a) and (b)), and 135^0 ((c) and (d)).

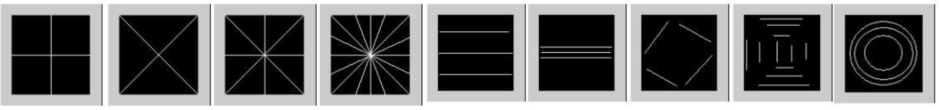
The former output is used to define two descriptors f^1 and f^2 as:

$$\begin{aligned} f^1(n) &= G^{45^0}_{1,3}(n) - G^{135^0}_{1,3}(n) \\ f^2(n) &= G^{135^0}_{2,4}(n) - G^{45^0}_{2,4}(n) \end{aligned} \tag{8.3}$$

where $G^{\theta}_{i,j}(n)$ represents the sum of the response of the filter with orientation θ over all the pixels of the image I_n in the quadrants i and j. Thus, f^1 and f^2 codify the global amount of directional information in the diagonal radial direction for each quadrant. This same analysis was repeated for a 45^0 rotated version of the quadrant distribution as shown in Figure 8.5 (b), defining the new quadrants labelled by 5, 6, 7 and 8. This new quadrant distribution provides two more descriptors f^3 and f^4 defined in Equation (8.4), which codify the global amount of directional information in the vertical and horizontal directions for each quadrant.

$$\begin{aligned} f^3(n) &= G^{0^0}_{6,8}(I_n) - G^{90^0}_{6,8}(I_n) \\ f^4(n) &= G^{90^0}_{5,7}(I_n) - G^{0^0}_{5,7}(I_n) \end{aligned} \tag{8.4}$$

In order to illustrate the behavior of this set of descriptors, we edited a pool of synthetic images and calculated f^1, f^2, f^3 and f^4. Figure 8.7 shows the feature values for several test images. Figures 8.8 and 8.9 show the feature values for a rotated segment oriented in radial direction and perpendicular to radial direction.



f^1:	00.00	10.00	09.30	26.19	00.00	00.00	-5.40	00.00	-19.65
f^2:	00.00	09.81	09.06	24.86	00.00	00.00	-8.64	00.00	-19.65
f^3:	10.04	-0.25	09.61	17.50	-14.03	-4.39	-0.95	-11.35	-12.29
f^4:	08.85	0.25	08.96	14.91	09.05	24.18	-0.12	-10.46	-10.39

Figure 8.7: Directional features for test images.

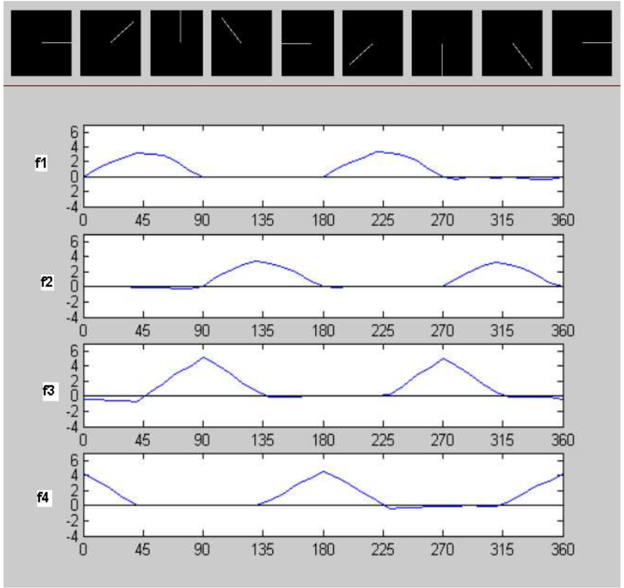


Figure 8.8: Features response for a rotating radial line.

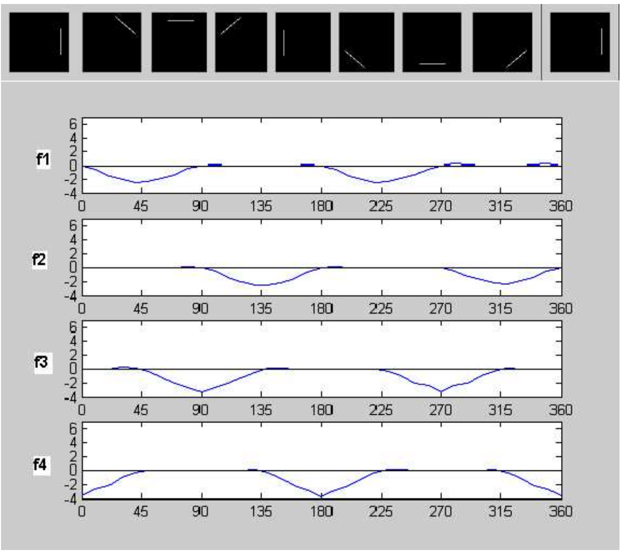


Figure 8.9: Features response for a transversal line which is rotated around the center of the image.

8.3.2 Polar transform approach

Polar transform [44] consists of a mapping from the original cartesian image, in which each pixel is referred to by the pair $(row, column)$, into a transformed image in which each pixel is referred to by a pair $(angle, dist)$. In order to perform a polar transform, we need to fix a center. For each pixel with cartesian coordinates $(row, column)$, the $dist$ value is its Euclidean distance to the center, while the angle value is the angle which the vector connecting the center and the pixel forms with the horizontal axis. Figure 8.10 shows a graphical interpretation of this transform. For a further mathematical background about the polar transform, the interested reader can consult Appendix A. In the polar image, the horizontal axis represents the *angle* parameter, ranging in the interval $[0^0, 360^0]$. The vertical axis represents the *dist* parameter, ranging in the interval $[0, max_{dist}]$, where max_{dist} corresponds to the maximal distance between two pixels within the camera field of view -which in capsule endoscopy frames corresponds to 240 pixels-. But, as for a given center, the most distant pixel distance p_{dist} is typically smaller than max_{dist}, the transformed image is zero padded from p_{dist} to max_{dist} for the angular orientation corresponding to the pixel p. The dist parameter is plotted from top to bottom, so the pixels which are closer to the center are rendered in the top part of the plot and the most distant points are plotted at the bottom part. After the polar transform, concentrically distributed lines appear as horizontal lines, while radially distributed lines appear as vertical lines.

We set the center of the polar transform to be the center of mass of the blob, relating the origin of the transform to the center of the intestinal lumen. Figure 8.11 shows the result of the polar transform on the wrinkles associated with valleys (blue) for a given frame.

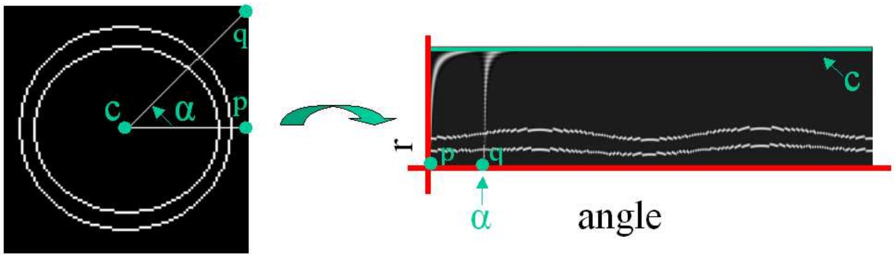


Figure 8.10: Graphical interpretation of a polar transform.

Finally, we calculate two descriptors $f^{1'}$ and $f^{2'}$ as follows:

$$\begin{aligned} f^{1'}(n) &= G^{0^0}(I_n^{polar}) \\ f^{2'}(n) &= G^{90^0}(I_n^{polar}) \end{aligned} \tag{8.5}$$

where $f^{1'}(n)$ and $f^{2'}(n)$ codify the global amount of directional information in the horizontal and vertical direction of the polar image I_n^{polar} of the n frame.

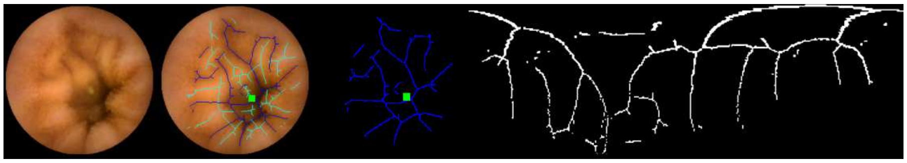


Figure 8.11: Wrinkles detected by valleys in polar coordinates.

8.3.3 Definition of the area of analysis

Tonic contractions wrinkles appear as radial lines in the original image, and as nearly vertical lines in the polar transform. However, it must be noticed that this pattern undergoes deformations which are more severe around the lumen center, and in distant points from the center -distant parts of the wrinkles usually occur to be curved and no longer respect the radial orientation-. Both regions correspond to the top and bottom areas in the polar plot. To minimize the influence of this phenomenon, we tested the exclusion of the area defined by the blob from the wrinkle analysis. In addition to this, we also excluded all the distant pixels. This exclusion was performed by a simple morphological procedure of dilation and substraction as defined in Figure 8.12. The region of analysis is defined by a ring-wise or donut mask, which is to be applied to the valleys wrinkle pattern previously to the feature extraction procedure. Figure 8.13 shows several examples of (a) frames from sustained contractions and (b) random frames, with their corresponding blobs, masks, wrinkles and polar transform of the valleys wrinkle pattern.

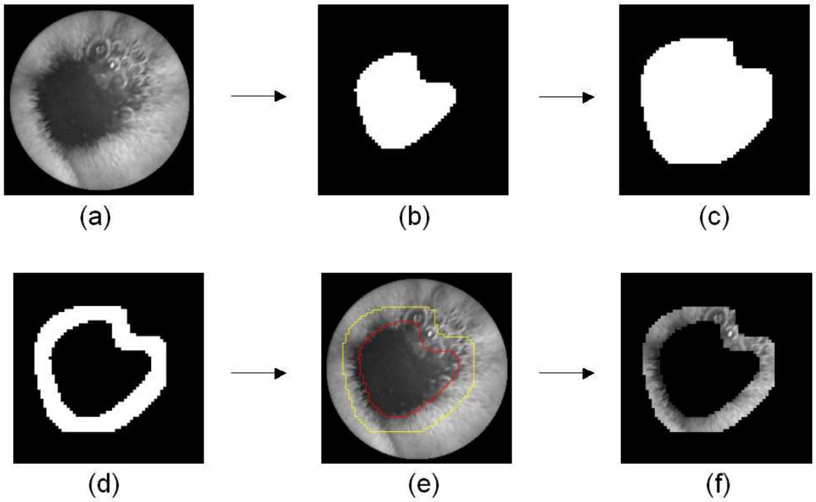
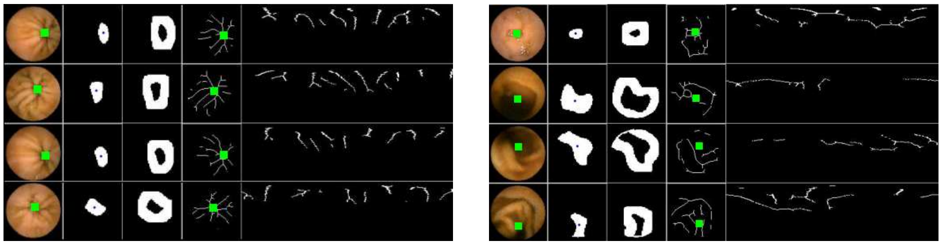


Figure 8.12: (a) Original image. (b) Blob. (c) Blob after dilation. (d) Morphological substraction of c-b. (e) Mask contours. (f) Segmented region of interest



(a) (b)

Figure 8.13: (a) Frames from tonic contractions. (b) Randomly selected frames.

8.3.4 A simple approach: individual analysis of each wrinkle

We implemented a simple approach by means of the characterization of each individual wrinkle by using the angle which it describes with the tangent to the closest point of the outer contour of the donut. The aim of this implementation is to serve as a reference for the results obtained by the alternatives previously explained, which perform a richer analysis regarding the directional information. The procedure applied is explained in the following lines and graphically depicted in Figure 8.14:

1. We find and delete all the t-junctions in the wrinkles image, by means of removing all the pixels connected to 4 or more pixels in a 3 by 3 neighborhood -Figure 8.14 (b)-.
2. The area of analysis is constrained to the donut region, as shown in Figure 8.14 (c).

3. We label all the connected components and remove all the segments containing less than 4 pixels -Figure 8.14 (d)-.

4. For each segment, the vector which connects its extremes is selected as an estimate of its director vector V_w.

5. For each segment, the tangent vector to the external contour of the donut, in the closest point to the exterior extreme of segment, is selected as the tangent vector V_t.

6. Finally, the descriptor f^w is calculated as:

$$f^w(n) = \sum_{i=1}^{N_n^{wrin}} (V_t^i(n) \cdot V_w^i(n))^2 \tag{8.6}$$

   where N_n^{wrin} represents the number of wrinkles, and $V_t^i(n) \cdot V_w^i(n)$ represents the scalar product between the director vector and the tangent vector to the outer contour of the donut, in the closest point to the i segment of the n frame. Figure 8.15 depicts a graphical representation of this scenario for the case of one single wrinkle.

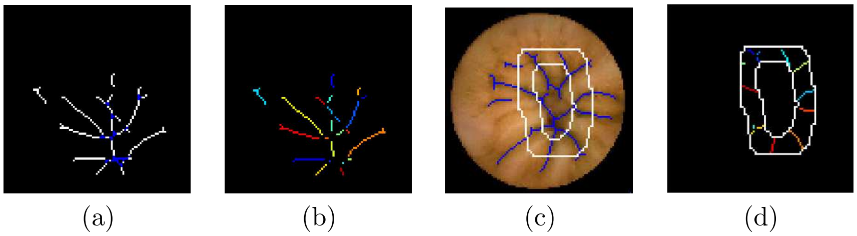

(a) (b) (c) (d)

Figure 8.14: (a) Original pattern. (b) T-junctions are removed. (c) A different frame showing the wrinkle pattern and the donut contours superimposed. (d) Connected components of the wrinkles.

In order to calculate the tangent vector, we used 6 neighbor pixels. With this implementation, the value of f^w tends to be minimized when all the segments in a pattern of wrinkles are radially distributed. With the aim of avoiding the inclusion of frames with low wrinkle information -few segments of wrinkles-, which would yield to a low value of f^w also, we rejected all the frames containing less than 3 segments.

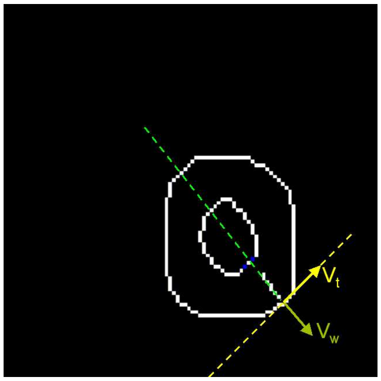


Figure 8.15: The descriptor f^w for each individual wrinkle is obtained from the scalar product of V_t and V_w.

8.4 Qualitative validation of the features sets

With the aim of obtaining a qualitative view of the description power of the proposed wrinkle features, we used SOMs. Figure 8.16 shows a SOM constructed using all the frames from a single video, and the features f^1, f^2, f^3 and f^4. It can be observed that most of the radial wrinkles patterns are situated on the top-left corner. The opposite corner -bottom-right- consists of frames showing concentric wrinkles from tunnel frames. This assesses our hypothesis about features f^1, f^2, f^3 and f^4, and it also assesses the results obtained with our synthetic experiments, which were shown above. Figure 8.17 shows a SOM constructed using all the frames from the same video, and the features $f^{1'}$ and $f^{2'}$ -the size of the SOM has been automatically set to optimize the topographic and topological errors in both cases-. It can be shown that the star-wise wrinkle pattern is placed on the left top corner. The scatter plot represents the value of $f^{1'}$ in the vertical axis, and $f^{2'}$ in the horizontal axis. As was expected, the star-wise patterns of sustained contractions correspond to high values for the vertical detector and low values for the horizontal detector.

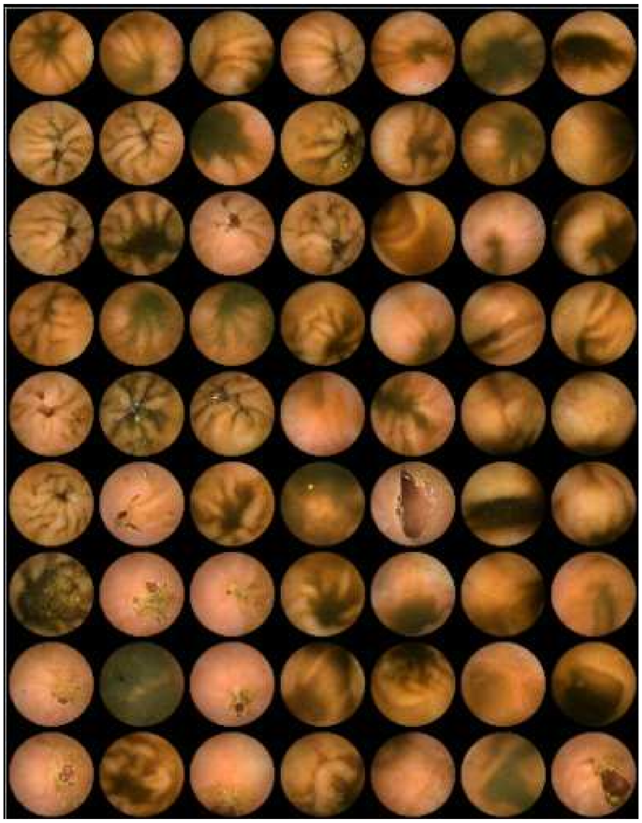


Figure 8.16: SOM constructed from all the video frames, using f^1, f^2, f^3 and f^4 as features

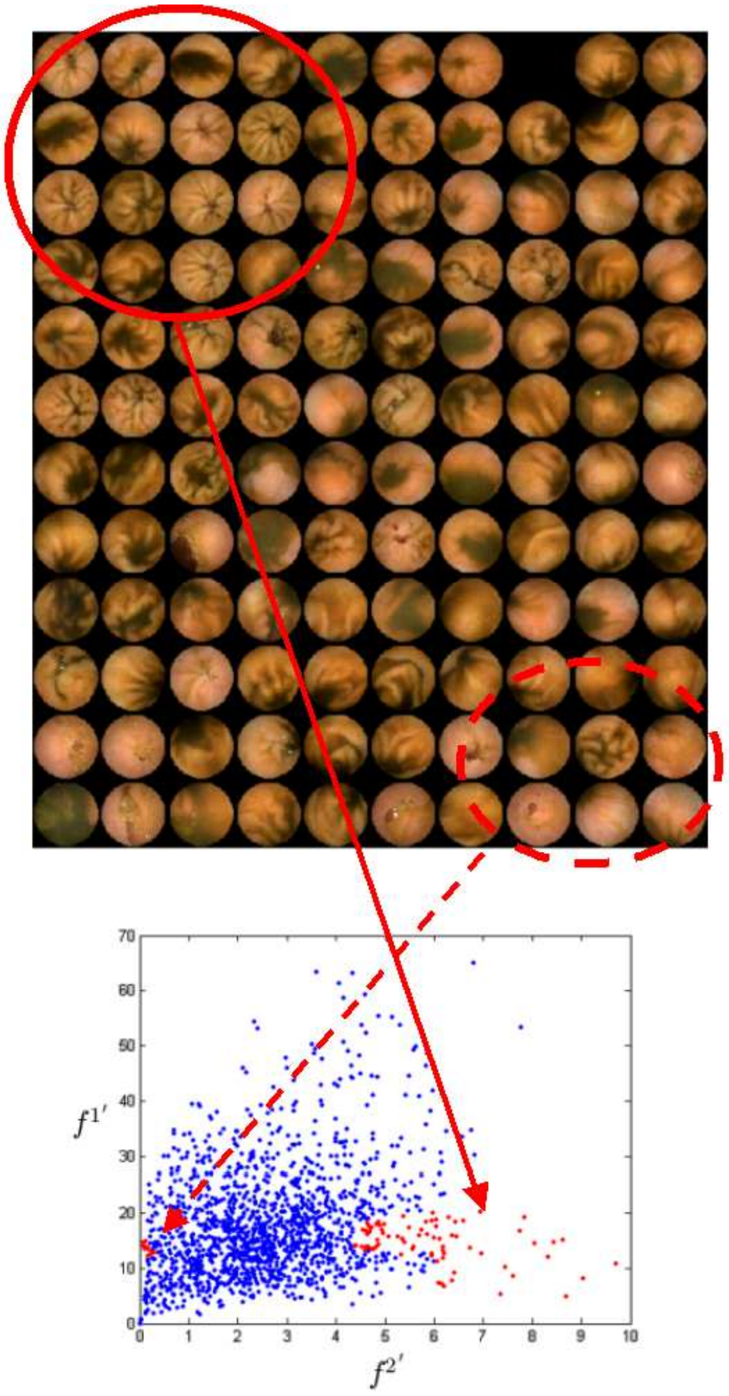


Figure 8.17: SOM constructed from all the video frames, using $f^{1'}$ and $f^{2'}$ as features. The corresponding features are plotted for two different regions of the SOM.

Chapter 9

Tonic contractions detection: Results

We tested the features explained in the previous section, defining 4 different sets:

Quadrant: Features f^1, f^2, f^3 and f^4.

Quadrant-donut: Features f^1, f^2, f^3 and f^4, restricted to the area defined in Figure 8.12.

Polar: Features $f^{1'}$ and $f^{2'}$.

Polar-donut: Features $f^{1'}$ and $f^{2'}$, restricted to the area defined in Figure 8.12.

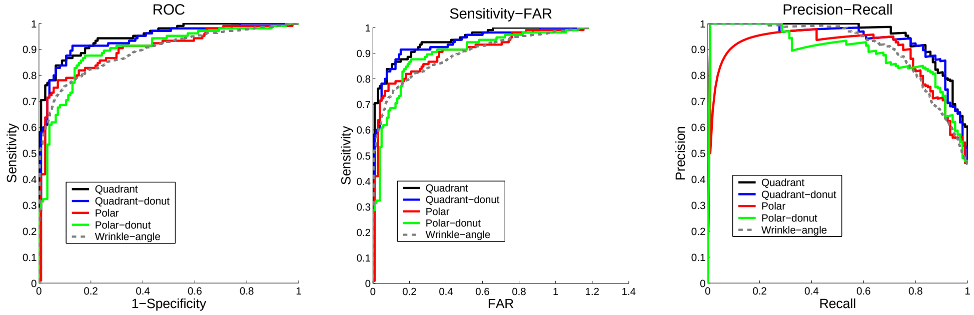
For the dilatation step of the donut generation, we used a 40 pixels squared structural element. For each feature set, we run 4 experimental tests in order to assess the performance of our approach in the detection of frames belonging to tonic contractions, and the final detection rate of tonic contractions.

1. Our first experiment was designed in order to test the ability of our system in the detection of frames presenting a clear wrinkle pattern.
2. Our second experiment consisted of testing the performance of our system with patterns of frames belonging to sustained contractions without any further restriction.
3. The third experiment tried to quantify the number of *frames* belonging to sustained contractions which were labelled as contraction frames -sensitivity over frames belonging to tonic contractions sequences-.
4. Finally, the fourth experiment provides the global performance, over one whole video, in terms of the number of tonic contraction *sequences* correctly detected

-sensitivity over tonic contractions sequences-, precision and FAR.

9.1 Performance on clear patterns of sustained contractions frames

For this experiment, the specialists selected a pool of 521 frames belonging to sustained contractions -from now on, we will refer to this set as *wrinkle frames*- and 619 random frames which did not belong to any sustained contraction -from now on, we will refer to this set as *non-wrinkle frames*-. For all our experiments we trained a SVM classifier with a radial basis function kernel and $\gamma = 0.05$. We used 80% of the samples for training and 20% for testing, performing 10 runs. Figure 9.1 shows the ROC, sensitivity-FAR and PR-curves of the classification experiments for wrinkles, wrinkles-donut, polar an polar-donut features:



(a) (b) (c)

Figure 9.1: (a) ROC, (b) sensitivity-FAR, and (c) PR curves of experiment one on sustained contractions.

Figure 9.1 shows that both *Quadrant* and *Quadrant-donut* feature sets outperformed the rest with equivalent performance, which was statistically assessed by a t-test over the AUC of the plots rendered in Figure 9.1 (a). All the proposed alternatives appeared to outperform, in terms of detection rate, the reference approach based on the assessment of the perpendicularity of the wrinkles with the outer donut contour, which we denote as *Wrinkle-angle*. In addition to this, from the results rendered in Figure 9.1 (c), two main conclusions can be inferred regarding the precision of the different methods: On the one hand, the *Quadrant* feature set obtained optimal results in precision for all the sensitivity values below 60%. On the other hand, the *Wrinkle-angle* feature set outperforms the polar approaches for clear patterns of wrinkle frames.

9.2 Performance on general patterns of sustained contractions frames

For this experiment, the specialists selected a pool of 2414 wrinkle frames and 2414 random non-wrinkle frames. An extended analysis about the used selection procedure is provided in the discussion in Chapter 10. Figure 9.2 shows the ROC, sensitivity-FAR and PR-curves of the classification experiments for the different features sets.

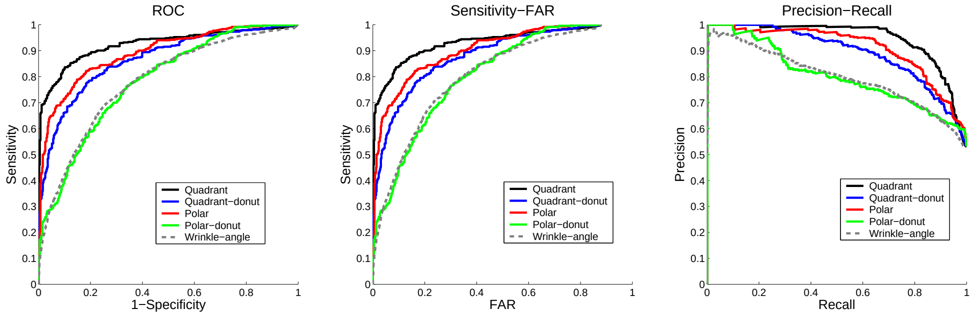


Figure 9.2: ROC, sensitivity-FAR, and PR curves of experiment two on sustained contractions.

Figure 9.2 shows that the *Quadrant* feature set outperformed the rest of methods both in detection rate and precision, which was assessed by a t-test of the AUC and AUPRC. It must be noticed that, for general patterns of wrinkle frames, the *Wrinkle-angle* and *Polar-donut* feature sets showed the worst results, performing in an equivalent way, while the *Polar* feature set clearly outperformed both of them.

9.3 Detection rate of frames belonging to sustained contractions

The set of features referred to as *Quadrant* resulted the one with the best performance results in the classification tests. We used the classifier trained in the previous experiment in order to label all the frames of the sequences of sustained contractions. The aim of this experiment is to assess the sensitivity of our system for frames of sustained contractions.

Table 9.1 summarizes the results provided by our system. We analyzed 42 sustained contractions with 737 frames. The mean length of the contractions was 14.2(±5.3) frames -there was one outlier of 123 frames, which was not included in the calculus of the mean value-. Our system detected 454 frames, which represents 61.5% of all the frames belonging to tonic contractions. In average, for each contraction sequence, our system labelled 60.9% of frames as frames showing the radial wrinkle pattern.

Table 9.1: Detection rate of frames belonging to sustained contractions

Number of sustained contractions:	42	(737 frames)
Number of annotated frames:	454	(61.5%)
Avg. frame annotation by sequence:	60.9%	(±0.3)

9.4 Detection rate of sustained contractions

The final step for the validation of our system consists of the automatic labelling of tonic contractions for one video. In order to define the criteria for the detection of a sustained contraction sequence, we followed the subsequent highlights proposed by the physicians: we consider that we detect a sustained contraction if we detect 5 or more radial wrinkle frames within a window of ±5 frames. In this sense, the detection pattern of sustained contractions diverges from that of phasic contractions, due to their different physiological nature. We defined a phasic contraction detection by providing the central frame of the contraction. For tonic contractions, we define a contraction detection by providing at least one frame holding the previous requirement. Summarizing, our criteria for the calculus of the system performance consisted of the following lines:

1. We automatically annotate all the wrinkle frames in the video.

2. We create all the sequences of sustained contractions following the criterion described above: all the frames belonging to the same sequence have, at least, 5 wrinkle frames within a ±5 frames neighborhood.

3. We consider that a sequence labelled by the experts is detected (a true positive) if there exists a sequence provided by the system which has, at least, one frame in common.

4. We consider that a sequence provided by the system is a false positive if none of its frames belongs to a sustained contraction labelled by the experts.

Following these criteria, we obtained the results shown in Table 9.2. Our system successfully detected 71.4% of the sustained contractions provided by the experts. In the final output 1 out of 3 suggested sequences are real tonic contractions. The impact of the detection of the remaining 2 sequences deserves a special analysis, which must be performed in a qualitative way by means of visual inspection.

Table 9.2: Detection rate of sequences of sustained contractions

Sustained contractions:	42	
System sequences:	106	
True positives:	30	(**71.4**% sensitivity)
		(**28.3**% precision)
False positives:	76	(**181.0**% FAR)

9.4.1 Qualitative analysis of tonic contractions detection

We deploy the qualitative analysis of tonic contractions detection by means of framed-positives mosaics. Figure 9.3 shows a set of representative sequences of sustained contractions detected by the system (TP). The frames detected as wrinkle frames are surrounded by a green square. Figure 9.4 shows a set of representative false negatives (missed findings). Finally, Figures 9.5 and 9.6 show two mosaics of the output of the system for the test video. The frames of the sequences detected by the system are surrounded by a green square, the experts' sequences are surrounded by a blue square, and the coincidences are in cyan. An extended analysis of the impact of these classification results is performed in the discussion in Chapter 10.

It must be noticed that, even with a precision about 30%, the output of our system supposes a valuable tool for the specialists, because they are directly driven to the suggested contractions, and they do not have to visualize the whole video. For the proposed example of our experiments, assuming the experts to take 10 seconds in the discrimination between a false positive and a true positive of tonic contractions, the total amount of analysis time is reduced to $106 \times 10 = 1060$ seconds, less than 20 minutes. The time consumed by the expert in the labelling of sustained contractions of the analyzed video which we show the mosaics from was reported to be more than 3 hours.

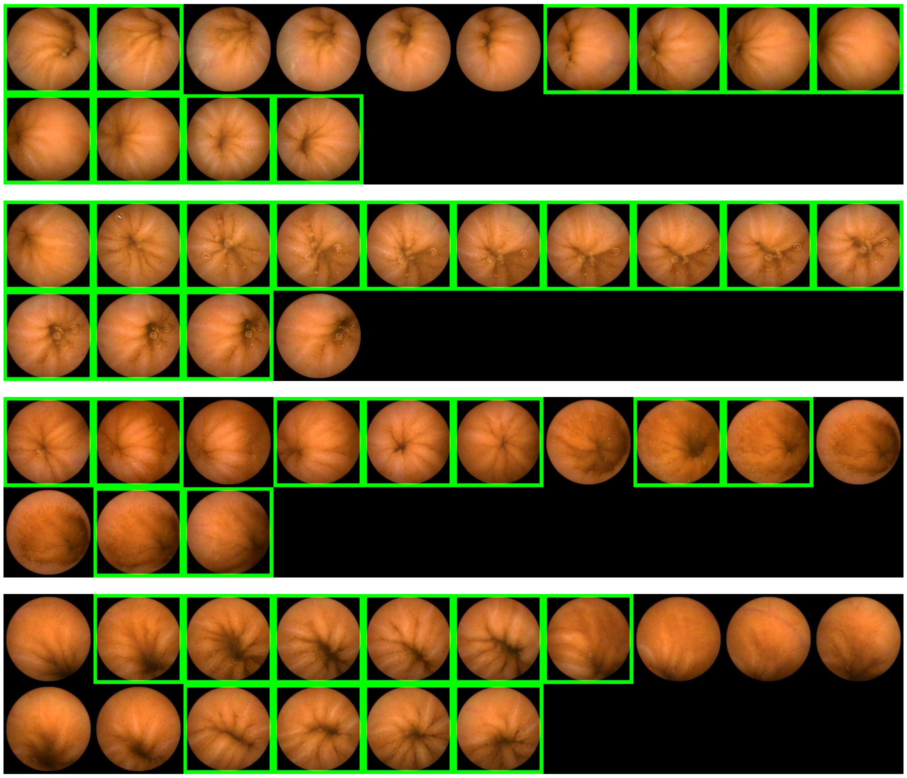


Figure 9.3: True positive sequences of sustained contractions

In the 4 true positive sequences rendered in Figure 9.3, the detected wrinkle frames (surrounded by a green square) show the sought radial pattern. Those frames which were not detected as wrinkle frames diverge from this pattern, mainly in the following ways: a) they show the center of the lumen displaced from the center of the image, and thus, part of the pattern remains out of the field of view if the camera, b) the wrinkle folds present orientations which no longer match the radial pattern with respect to the center of the lumen, and finally c) the wrinkles contrast decreases, especially in the case of the valleys.

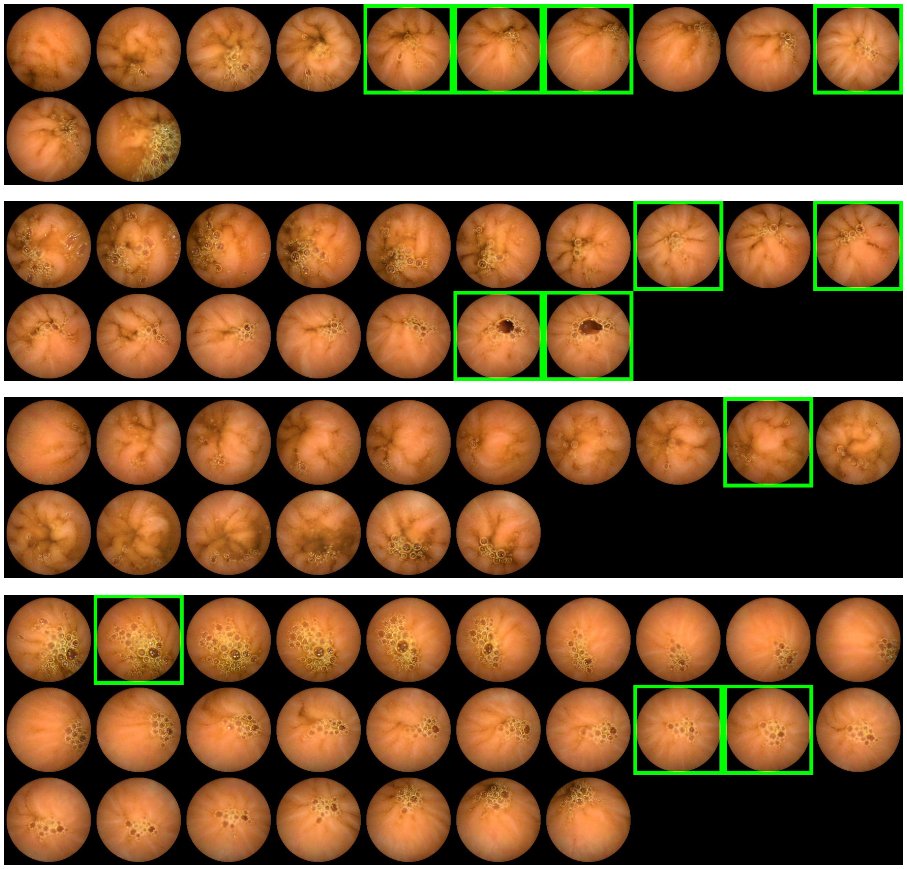


Figure 9.4: False negative sequences of sustained contractions

Regarding the false negatives analysis (Figure 9.4 renders 4 representative examples of them) the origin of the misclassifications is basically twofold: On the one hand, the same analysis formerly performed for the missed wrinkle frames of the true positive sequences may be applied here. On the other hand, the presence of bubbles in the center of the lumen hinders its correct detection.

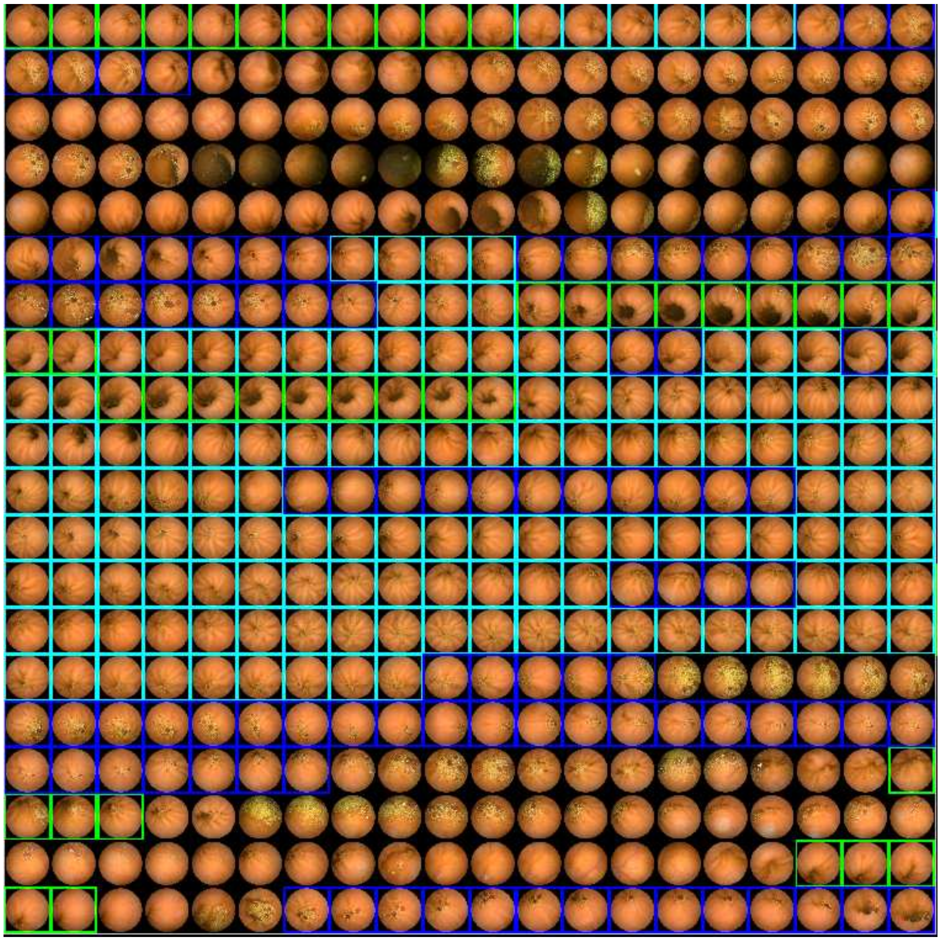


Figure 9.5: Mosaic. System output (green), experts' findings (blue), coincidences (cyan).

The mosaic rendered in Figure 9.5 shows a high degree in coincidence between the experts' characterization and the system output for sustained contractions (cyan frames). The sequences which were not detected by the system (blue frames) show images in which the wrinkle pattern is very subtle, or a considerable amount of bubbles is present.

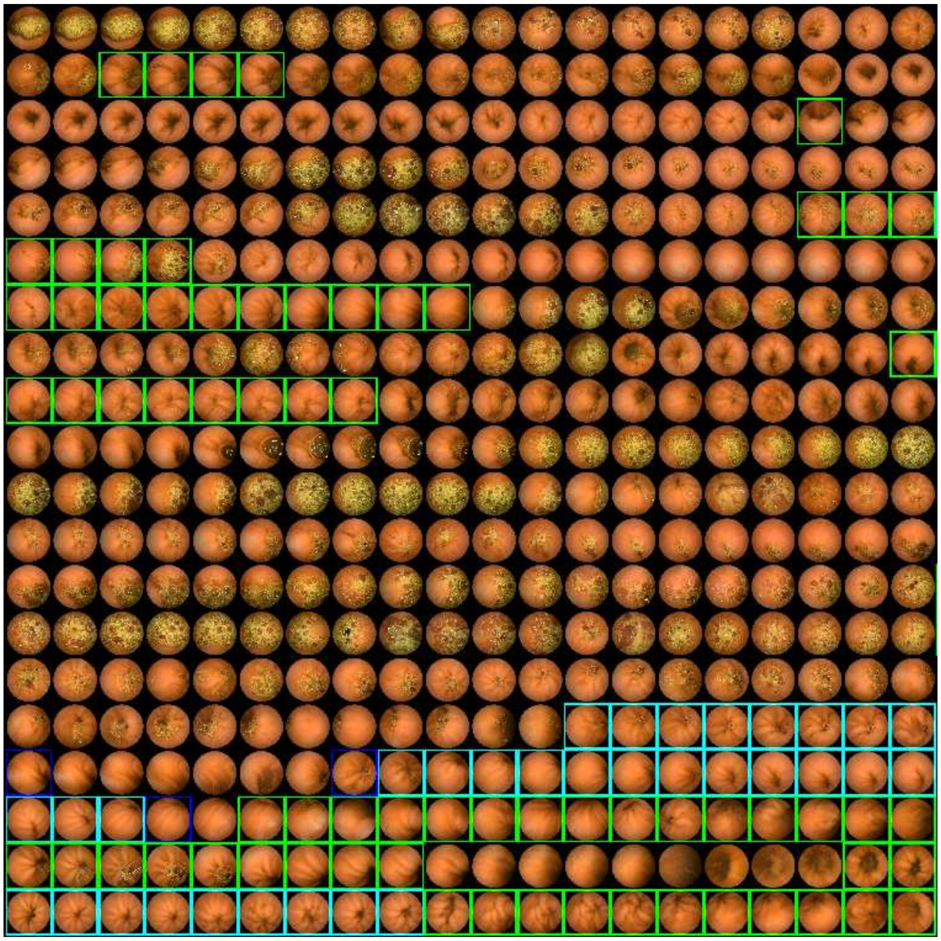


Figure 9.6: Another Mosaic. System output (green), experts' findings (blue), coincidences (cyan).

The mosaic pictured in Figure 9.6 shows several examples of sequences detected by our system as sustained contractions (sequences of consecutive green frames) but which were not labelled by the experts -i.e., the false positives-. It can be observed that some of these sequences actually match the paradigm of sustained contractions. In this sense, our approach showed its ability in the detection of real sustained contractions which the experts did not detect during the video visualization process.

Part IV

GENERAL DISCUSSION AND FINAL CONCLUSIONS

Chapter 10

Discussion

10.1 Phasic contractions detection

10.1.1 The cascade classification system

The choice of the cascade system is underpinned by the fact that each step is designed in order to reject an amount of frames which mainly include images which are not to be intestinal contractions -i.e., the system negatives-, letting pass through the sequential stages those frames related to intestinal contractions -i.e., the system positives-. This yields to an effective reduction of the imbalance ratio of the data set at the input of the last classification stage. Many authors have applied diverse strategies in order to tackle the impact the imbalance ratio has in the performance of classification, involving stratified sampling, cost-sensitive approaches, different implementations of decision trees and bagging, and the use of several metrics for performance measurement, mainly [20, 31, 53, 27, 64, 89]. In our strategy, each stage is tuned to prune as many non-contraction frames as possible, trying to minimize the loss of true positives, and achieving in this way an effective reduction in the imbalance ratio of the data. The last stage of the cascade, consisting of the support vector machine classifier trained by means of under-sampling, is to face a classification problem with an imbalance ratio about 1:5 -in contrast with the 1:50 at the input of the system-. This reduction in the imbalance ratio is shown to be an effective way of tackling the problem of classification in this kind of scenarios. In addition to this, one more important feature must be outlined: the modular shape of the system lets the expert identify new targets in the video analysis procedure, providing the chance to easily include them as new filter stages, and adding domain knowledge to the system in a natural and flexible way. It must be noticed that a previous study was performed in order to obtain a set of heuristical rules as a first approach for classification of phasic contractions. The interested reader will found a summarized description of this heuristical framework in Appendix C.

First stage

In order to set the final value of P_1, the parameter controlling the filtering of dynamic sequences at the first stage, we could have chosen to use in Equation 6.1 the P_i^j value which could guarantee the highest sensitivity for all the videos. But this alternative was rejected because it would imply an increase in the number of negatives passing to the second stage. Since, from the results obtained in our qualitative analysis, we observe that no pure pattern was rejected in stage one, we can state that this stage performs in the expected way, reducing the imbalance rate with no loss of sensitivity (the first stage showed a mean sensitivity value about 96.65%) and rejecting only those patterns which showed the highest divergence from the pure phasic pattern.

Second Stage

Tunnel characterization, performed by the discriminant function g_T in Equation (5.8), was chosen among different implementations which included subtle modifications. In this sense, the tunnel characterization could have been also performed if each frame of the 9 frames sequence fulfilled the condition of presenting a lumen area restricted to $A_{lower} < f^2(n)$, instead of using the overall sum along all the sequence. Our tests revealed that no relevant changes in terms of performance were achieved with this solution. The inclusion of an upper limit in the tunnel constraint, i.e., $A_{lower} < f^2(n) < A_{upper}$, did not provide better results either, since the maximum area of the lumen is implicitly restricted by the σ_{lum} parameter.

Regarding the validation stage of tunnel frames, the intrinsic difficulty of assessment must be noticed. The main point of the tunnel detection is the rejection of regions of frames sharing a lack of intestinal motility activity. In this sense, the tunnel detector labels each frame as a tunnel frame, although in order to perform this characterization the dynamic information of the neighbor frames is taken into account. In the validation process, we had to join these two aspects: 1) the ability of the detector to label a frame belonging to a tunnel sequence, and 2) the ability of the detector to label all the frames of the tunnel sequence. For this reason, the visual analysis cannot be performed exclusively in a frame-by-frame way, and the framed-positives mosaics appeared as an efficient solution for visual validation.

Color characterization of intestinal juices

The use of the SVM as the classifier for the intestinal juices color characterization is underpinned by two main reasons: On the one hand, the SVM with radial basis functions kernel showed the best qualitative results in the preliminary tests performed by using drag-and-drop color mosaics; this was assessed by the general comparative performance results in a general ranking competition with other classifiers. On the other hand, the use of one single classifier for the solution release which was delivered to the specialists simplifies the final application, since it is not necessary to include a new module for the turbid classification. The use of radial basis functions instead of other kernel functions is also based on slightly better performance results, but it must be noticed that the results achieved are quite invariant to the kernel choice. Different approaches for color characterization which we tested included color quantization for histogram matching -for a further insight about this technique, the reader may consult Appendix D-, although the best performance results were obtained by the SVM proposal that is presented in this work.

Bubble detection

In order to characterize the patterns of bubbles shown in Figure 5.8, we tested several alternatives, which focused on texture exclusively. Several works have tackled the problem of the characterization of different intestinal events in endoscopy using texture descriptors [48, 61], but as far as we know, no previous work has been oriented to the specific issue dealt with in this work. It must be noticed that the size of the bubbles may widely vary inter- and intra-frame, so different strategies based on Laplacian of Gaussian, wavelet analysis, and the use of derivatives of Gaussians and Gabor banks of filters were tested, which have been reported to perform well in multi-scale and directional characterization applications [1]. Our final decision on Gabor filters was based on the better empirical results they presented in all preliminary tests, performed over a selected set of frames containing intestinal juices and its validation through framed-positives mosaics.

The selection of the overall sum of the responses of the bank of Gabor filters as a descriptor of the textural information in bubble frames, in the way it was described in section 5.2.2, and the application of the threshold over I_n^{bub} in order to obtain the area of bubbles, deserve a comment about their performance. As was shown in the results in Chapter 6, this method performs well in terms of precision -almost no frames without bubbles are wrongly selected as bubble frames- and sensitivity -only very blurred cases of bubbles are missed-. It must be noticed that an increase in the number of filters, basically in terms of scale, yields to a corresponding increase in the detection of false positives, namely, different structures present in the endoscopic images which are not related to intestinal juices, such as the characteristic wrinkles of the intestinal contractions. On the other hand, a decrease in the number of filters yields to a loss in the sensitivity of the bubble detector, and some frames containing more subtle bubbles can be missed. Thus, the above mentioned set of parameters is the one that best matches the bubble pattern for the provided images. However, we must notice that this good behavior could be hindered if severe changes in sharpness were induced in the production line of the capsules. The provided results for the bubble detector are valid for all the videos analyzed so far by our team, but although the robustness of the procedure was tested in several experiments which included a sharpening pre-processing of the frames, the final outcome is not predictable for scenarios which widely differ from ours.

Third stage: the SVM for the final detection of intestinal contractions

One of the main considerations taken into account for the selection of the SVM classifier was its sensitivity to the skewed distribution of the data sets. It has been shown that the learning mechanisms of SVM make this classifier an attractive candidate for dealing with moderated imbalance ratios. The SVM takes into account samples which are close to the decision boundaries, namely, the support vectors, and it tends to be unaffected by samples lying far away. Additionally to the former, stratified sampling techniques (such as under-sampling the majority class, over-sampling the minority class, or artificially creating new samples) have been proved to be efficient in the improvement of performance of several classifiers, including support vector classifiers [3]. Our approach implements under-sampling in the learning strategy. Several methods of sampling were tested, and under-sampling resulted the one with the highest reliability (a detailed analysis and discussion about the design of these experiments can be found in [84]).

Inter-observer variability

Inter-observer variability was in the same order of intra-observer variability. The *30-minutes* division of the videos for the human vs. human inter-observer test was a constraint suggested by the specialists. The 3 minutes bins quantization conformed the specialists requirements for judging a result with statistical significance. The KS-test provides good results for our approach with a mean *p-value*= 0.20.

System computation time

Finally, regarding the computation time of the cascade system, our proposal presented the following behavior: The largest amount of time is devoted to feature extraction, which, depending on the length of the study, may span for about 2-3 hours. The diverse training stages of the system show a typical duration about half an hour. In this sense it must be said that although re-training the system after each new video study helps to increase its asymptotical classification performance, it seems reasonable to cut this feedback once an enough amount of characteristic training frames is gathered. This point might only be assessed after an exhaustive test in full operative conditions. Once the system is trained, the cascade classifier performs a classification result in less than one minute.

Qualitative analysis of the classification results

Regarding the true positives detection, the patterns shown in Figure 6.9 confirmed the ability of our approach in the detection of the paradigm of phasic contractions. This qualitative outcome is numerically corroborated by the high sensitivity values (over 90%) obtained for occlusive and non-occlusive patterns.

The missed contractions share some common features: On the one hand, the open lumen is not always present at the beginning and the end of the sequence, and this happens because the camera is not pointing towards the longitudinal direction of the gut, or because the selected contraction is spanning for more than the 9 frames -this could be likely linked to the blurring definition border between short tonic contractions and phasic contractions-. Moreover, the motion impression that the expert perceives during the video visualization is not present in the deployed sequence of frames. In this sense, we performed some tests which consisted of showing the experts a set of paradigmatic sequences containing doubtful contractions both by visualizing them in the video at a visualization ratio of 3 frames per second, and showing the same sequences deploying the 9 frames as in Figure 6.10. We found that the experts usually labelled a higher number of contractions during the video visualization, and a lower number looking at the deployed sequences. This fact drives us to think that the motility characterization should be performed in more subtle detail, so as to detect the apparently slight changes in some sequences shown in Figure 6.10, but which actually seem to be clear for the expert during the visualization process.

Finally, the false positives analysis gives very interesting results: On the one hand, our system shows its ability to detect real contractions which the experts did not get to label. This is a reasonable result, since one of the main drawbacks associated with motility assessment by manual labelling is the growing stress and fatigue which takes place during visualization, yielding to a loss of effectiveness in the final outcome. A rough study over the false positives of the ten analyzed videos showed that about $5-10\%$ of the false positives consisted of this kind of sequences. On the other hand, multiple factors are pointed out as sources of error for the false positives (namely, the lateral movement of the camera while focusing the lumen which can be confused with the pattern of its contraction, the differences in illumination creating shadows which can be confused with the lumen, the multiple patterns of wrinkles which can provide a high response to the lumen detector, and the residual presence of patterns of turbid liquid). We suggest that many of these problems may be tackled by a deeper study about the textural information provided by the lumen, both in the relaxed stage and the contraction activity. This issue and some other proposed approaches will be enumerated in Chapter 11, which presents our conclusions and highlights our proposed future lines of research.

10.1.2 Experiments with phasic contractions

Classifier ensembles

In this work, we show that ROC analysis shortens significantly the time required for a qualified professional to inspect videos of intestinal wireless capsule endoscopy with a minimal loss of performance. We tested 8 single classifiers, a hybrid ensemble model based on all combinations of these (247 ensembles) and a bagging ensemble of 25 decision trees. The best model according to the AUC criterion was the bagging ensemble. As we were interested in finding a compromise between accuracy and inspection time, a variant of a ROC curve was designed plotting the best achievable sensitivity versus inspection time. Operation points can be picked from this curve; this offers a significant reduction of the inspection time with reasonable sensitivity.

Experiments with sampling and SVM

The plots of section 7.2.1 show a comparative study between two different methods of sampling: under-sampling the majority class and over-sampling the minority class, both for single classifiers and SVM. The choice of these single classifiers is due to two important reasons: 1) On the one hand, linear and logistic classifiers are simple and fast; K-NN is a non-parametric technique that makes no assumption over the probability density function; we aimed at finding out if there was the same difference in behavior with these classifiers in order to infer conclusions for the SVM. 2) On the other hand, for the SMOTE implementation, the training set typically has around $35,000$ samples in a 54 feature space. In this situation, most of the software implementations -PRTools were used [29]- did not present the possibility of testing some common classifiers such as Parzen or decision trees, owing to this huge amount of data.

In this point, recent works have tested SMOTE with several classifiers in common public databases, such as UCI [20, 3], showing improvement with respect to under-sampling in some experiments. The results seem to show that for our problem, and from a statistical point of view based on the t-test analysis of the AUPRC, we cannot state that SMOTE outperforms under-sampling -maybe, slightly in the best of cases-. It must be pointed out that the aim of SMOTE is to provide a sufficient quantity of minority class samples to equilibrate the number of samples for the negative class. But this is not the only important issue: special attention must be paid to the fact that when we have very few samples for the minority class, they may not be enough to reconstruct the probability density function underlying them, which is basically the role of SMOTE. We suggest that this is the case of our data set, as well as of other examples presented in the papers mentioned above. Our efforts are focused on the research of this issue and the theoretical and empirical confirmation of this hypothesis.

The main outcome provided by this experiment is that both sampling techniques used yield to similar results for SVM. We cannot avoid deepening in the analysis of the consequences associated with our results. In fact, if the different sampling techniques used appear to have no effect in the final performance of the system, it is due to some relevant characteristics of our specific problem, closely related to the statistical properties underlying our datasets. Moreover, the sampling techniques we use, which seem to work for other studies -some of them already referenced above-, appear to give no performance improvement for our problem. This is a very challenging and interesting line of work, in which our group is completely involved.

We could re-formulate the previous paragraph into the next question: "Why is performance improvement problem-dependent using one sampling method in SVM with imbalanced datasets?". In this sense, the next experiment is to be of high interest and can help us to shed light in this scenario: Let us consider that we have two imbalanced datasets with samples N1 for the minority class -positives, contractions-, and N2 for the majority class -negatives, non-contractions- and, consequently N1$\ll$N2 (this is the typical scenario of an imbalanced problem). Now, keep N1 fixed and under-sample from N2 several times, to generate a smaller number of samples for training the classifier. Our purpose focuses on seeing the extent to which the SVM performance is affected as the skewness or imbalance rate gets worse. The same experiment can be repeated over-sampling N1. This analysis is extremely important to draw any firm conclusions about classification performance. Nevertheless, the inference that we can obtain from it is not so straightforward, and it is worthy of special and subtle attention. In this point, and in order to understand why a specific classification performance is achieved, two main questions should be stated: what the data distribution *around the decision border* is like for *both* datasets, and how *descriptive* the positive samples are for the probability density function (PDF) of their class. The first question is strictly related to the decision of the margins for the SVM, the second one is strictly related to the way in which over-sampling will perform. This analytical study is of high relevance, but it also holds intrinsic difficulties related to the statistical description and the geometrical analysis necessary to understand what the real distribution of the whole data in the n-feature space is like, and how well the distribution provided by the samples we use for training fits the real PDF of the minority class. Finally, an equally interesting question is open: to what extent the choice of the γ parameter in the SVM may affect the final performance of the classification in this scenario. We are preparing a more extended work with all these points of study, including the proposed experiment for a further piece of research.

Experiments with AdaBoost

The use of different sampling methodologies with AdaBoost could be useful to improve the classifier performance. The experiments performed in this line showed that although the sampling modification introduced in $AdaBoost_{mod-1}$ produced better results in terms of precision, this improvement is circumscribed to low values of recall. In this sense, we performed diverse preliminary tests using separate training, validation and test data sets, but the results were not improved. For the specific problem of the feature space defined for phasic contraction, AdaBoost alternatives show a low performance response, in general.

10.2 Tonic contractions detection

Sustained contractions represent a novel paradigm in motility assessment. As far as we know, there is not enough background knowledge about the typical duration of tonic contractions, their expected frequency in patients suffering from different diseases, or the impact of their analysis on clinical diagnosis regarding the above mentioned parameters. In addition to this, tonic contractions are more difficult to annotate than phasic contractions. It must be noticed that the physician must find the starting point and the final point of a tonic contraction for its labelling. It implies that when the expert comes across a motility event during the video visualization, the expert must go forward and backwards to be sure that the motility event is a tonic and not a phasic contraction, and that its duration exceeds 10 frames. Then, the expert must label the initial frame and the final frame. This protocol makes tonic contractions labelling very slow. For this reason, our experimental results were founded on the analysis of tonic contractions from two videos selected by the specialists. These two videos were a representative sample of videos for tonic contractions, and the experts performed an exhaustive labelling on them.

10.2.1 Reflections on the methodology of tonic contractions detection

The pre-processing smoothing stage is actually performed by a median filter, applied into a window with a fixed width to the typical width of a wrinkle. This parameter was statistically fixed and the tests accomplished showed a good behavior. We tried a more sophisticated smoothing method by means of anisotropic diffusion [44, 87], which respects the structural information of the image in order to apply the smoothing by keeping sharpness in edges. The final outcome in terms of the valleys and ridges was very similar, although anisotropic diffusion showed slightly better definitions of valleys for some difficult cases. However, the high computational cost of anisotropic diffusion in comparison to the median filter, makes its application not worthy, and the median filter was accepted as the final solution.

Valley and ridge detection

The choice of the anisotropic Gaussian kernel for the wrinkle detection is underpinned by the assumption that the wrinkles can be viewed, in a simplified way, as lines with a fixed width over a homogeneous background. In this sense, the choice of the anisotropic kernel instead of an isotropic one tries to boost the linear information of the wrinkles, in comparison to other types of dark structures such as shadows, bubbles, etc. The choice of the kernel size is strictly steered by the width of the wrinkles, in order to maximize its output.

Regarding the choice of the valleys as the unique source of analysis for wrinkles, several points must be highlighted: On the one hand, ridges correspond to the folds which the intestinal walls undergo towards the camera. Our experience revealed that frequently these folds are not so intense as valleys, they present more irregularities and near the intestinal lumen, they tend to get organized concentrically, not radially. On the other hand, valleys in sustained contractions usually show a higher degree of contrast, they show a regular radial organization, and keep a more constant orientation for each wrinkle. For these reasons, the inclusion of wrinkles from the ridge analysis does not help in the improvement of the results for sustained contractions detection.

Correcting the blob anisotropy

One of the main sources of deformation of the polar images comes from an assumption of circularity on the blob. As we assessed in section 8.3.3, the shape of the blobs can be very diverse, and though a typical blob tends to undergo a circular shape, the deviations of the blob profiles cause that concentric lines diverge from vertical lines in the polar image. In order to try to correct this anisotropy in the shape of the blob, we performed some trials focusing our attention on forcing a regression of the real blob into a circular shape by means of a warping [10]. The interested reader may consult our proposal of approach in Appendix E.

10.2.2 Qualitative analysis of tonic contractions detection

The qualitative analysis of the output of our system is an essential instrument to understand the real performance of the proposed approach. The visual validation on the false positives sequences shows the difficulty of the labelling of this kind of contractions by the specialists. The visual patterns of the sequences obtained as false positives rendered in Figures 9.5 and 9.6 match, in many cases, the paradigm of more than 10 frames of sustained contraction, although in some cases it is really difficult to separate the threshold between a phasic and a sustained for such a sort span of time. This does not happen for the case of the longest contractions, and both the specialist and the system provide sequences which intersect for all the cases without exception.

In true positives the wrinkle frames show the sought radial pattern. Regarding the false negatives, the origin of the misclassifications is twofold: some bubbles in the lumen center hinder the detection of the lumen. In other cases, the wrinkle pattern is so weak that no blob was in the binary Laplacian image. In this sense, the right detection of the lumen plays a fundamental role. In order to skip this dependency, we tested other approaches such as using as a centroid the center of mass of the wrinkle pattern, instead of the center of mass of the blob. However, the results provided by these alternatives were not satisfactory. Other tested alternatives for detecting the patterns of sustained contractions included the use of the relative position of the camera in the body, which can be calculated from a raw data source of position information which is provided by the capsule. A set of examples of preliminary tests using this information can be consulted in Appendix F.

Chapter 11

Conclusions

The basic conclusions of this thesis can be summarized in the following list:

1. Regarding the automatic detection of phasic contractions, we proposed a set of image descriptors based on image intensity, blob analysis and textural information, for the video frames characterization. We applied these descriptors to all the video frames and performed a classification based on a sequential system. This cascade approach showed its suitability for the reduction of the imbalance rate of the classification problem. In addition to the former, this type of design allows the experts to include domain knowledge in different modules, which are to be independent from each other. The final classification stage, consisting of a support vector machine trained by under-sampling provided the best performance results in our experiments. We tested and showed specific results about the impact of several modifications of powerful classification algorithms in the performance of our imbalanced problem.

2. In terms of individual contractions detection, our system showed good results of accuracy and precision. In terms of density of contractions (which represents the ability of our system to reproduce the temporal pattern of intestinal motility with phasic contractions), our statistical tests confirmed that our system provided equivalent patterns to those provided by the experts. The qualitative analysis of our results shows that our system misclassified only those patterns presenting a high deviation from the phasic pattern, and obtaining high values of sensitivity for the clear patterns. In addition to this, our system showed its ability to detect contractions which the experts omitted to annotate. This amount of rescued contractions typically corresponds to 5% of the total count of real findings.

3. Our proposal showed that the use of ROC curves provides an elegant and efficient way for the optimization of the classification process using multiple classifiers. This translates into a relevant reduction of inspection time for the case of occlusive contractions.

4. Regarding the automatic detection of tonic intestinal contractions, we proposed a set of image descriptors based on the detection of the radial wrinkles patterns which the intestinal walls show during the contraction event. We proposed different approaches for the characterization of these patterns, both in cartesian and polar coordinates, presenting a comparative study of their performance results. Our proposal showed to provide an important reduction in visualization time.

5. These results constitute a first approach to the study of intestinal motility assessment with video capsule endoscopy. The use of the software tools which we developed in our project provided the physicians with a helpful data source for this aim. The advances in intestinal motility assessment achieved by means of the analysis of the data provided by our tools, have shown their utility and relevance, which has been assessed by their publication in several pieces of research.

Future lines of research

Finally, we would like to point out our proposal for future lines of research in the scientific endeavor of enlarging the framework of the intestinal motility assessment by means of an accurate automatic detection of the dynamic events involved in it. For this aim, we would like to highlight some fundamental aspects which are mainly related to the definition of new visual paradigms of motility events, the identification of improved descriptors based on image and video analysis, and the achievement of more sophisticated classification techniques.

- We think that the fusion and combination of features from both phasic and tonic patterns constitutes a promising line of work. In this sense, the enrichment of the phasic categorization by means of the inclusion of the wrinkle information in the feature vector could help to provide a more accurate description in terms of textural content.

- The generalization of the dynamic models involved in intestinal motility constitutes an novel and open field of research. We are developing preliminary tests with different dynamic strategies, such as hidden Markov models, for the consequent representation of the video sequences by means of n-grams. This strategy precesses a rotation invariant approach for the video frames analysis which we already developed. In this context, the categorization of the variable length of the phasic and tonic contractions may play an important role.

- The automatic classification of diverse subtypes of intestinal contractions regarding other features, such as speed of the contraction process or the maximum and minimum size of the lumen during the contractile event, represents an interesting challenge from a physiological point of view. We think that the use of multi-class approaches for the classification framework could help to tackle this problem.

- We believe that the good categorization of static regions could help to achieve better results in classification. However, several drawbacks, mainly related to the presence of intestinal juices and the movement of the camera, drastically hinder the performance of the tested approaches. Our preliminary tests, based on color information and optical flow, showed promising results.

- We are currently investigating ways of taking profit from the positioning information which the camera provides. A registration process between the position, its speed and the visual data may be a useful source of information. The camera position data may be also useful for the assessment of the different patterns of motility for the different zones of the small intestine.

- Finally, we are in a continuous feedback with the experts in order to improve the current visual assessment methods, create optimal protocols and include faster and more efficient versions of our solutions for their use in a real clinical scenario in a close future.

Chapter 12

List of publications

- F. Vilariño, P. Spyridonos, J. Vitrià, F. de Iorio, F. Azpiroz and P. Radeva. Intestinal motility assessment with video capsule endoscopy: Automatic annotation of intestinal contractions. *IEEE Trans. on Medical Imaging* (under revision).

- F. de Iorio, C. Malagelada, F. Azpiroz, F. Vilariño and J. Vitrià, A. Accarino and J.R. Malagelada. In search for new parameters of intestinal motor activity in humans. *Gastroenterology*, (in press).

- F. Vilariño, L. Kuncheva and Petia Radeva. ROC curves and video analysis optimization in intestinal capsule endoscopy. *Pattern Recognition Letters*, 27(8):875-881, 2006.

- F. Vilariño, P. Spyridonos, J. Vitrià and P. Radeva. Experiments with SVM and Stratified Sampling with an Imbalanced Problem: Detection of Intestinal Contractions. *Lecture Notes in Computer Science*, 3680(2):783-791, 2005.

- P. Spyridonos, F. Vilariño, J. Vitrià and P. Radeva. Identification of intestinal motility events of capsule endoscopy video analysis. *Lecture Notes in Computer Science*, 3708(2):531-537, 2005.

- F. Vilariño, P. Spyridonos, J. Vitrià and P. Radeva. Self Organized Maps for intestinal contractions categorization with wireless capsule video endoscopy. In *Proceedings of the EMBEC*. 2005.

- F. Vilariño, P. Spyridonos, J. Vitrià and P. Radeva. Automatic detection of intestinal juices in wireless capsule Video endoscopy. In *Proceedings of the ICPR*. 2006 (Accepted).

- P. Spyridonos, F. Vilariño, J. Vitrià and P. Radeva. Anisotropic feature extraction from endoluminal images for detection of intestinal contractions. MICCAI 2006. *Lecture Notes in Computer Science*. (In press).

- F, Vilariño, P, Spyridonos, J. Vitrià, F. Azpiroz and P. Radeva. Cascade analysis for intestinal contraction detection. In *Proceedings of CARS.* 2006 (Accepted).

- F. Vilariño, P. Spyridonos, J. Vitrià and P. Radeva. Linear radial patterns characterization for automatic detection of tonic intestinal contractions. CIARP 2006 (Under revision).

- F. Vilariño, J. Salas, J. Vitrià and P. Radeva. Spatial tracking and video information registration for intestinal motility assessment in capsule endoscopy . CIARP 2006 (Under revision).

- F. Vilariño, M. Rosales and P. Radeva. Patch-Optimized Discriminant Active Contours for Medical Image Segmentation. In *Proceedings of IBERAMIA.* 2002.

- F. Vilariño and P. Radeva. Cardiac Segmentation With Discriminant Active Contours. In *Proceedings of the CCIA.* IOS Press, 2003.

- P. Radeva, J. Vitrià, J. Amores and F. Vilariño. Boosted Context Applied to Medical Imaging. In *APRMI International Workshop.* 2005.

Appendix A

Polar transform

The polar transform is a helpful tool for the representation of structures showing radial or concentric patterns. The polar transform for a center point $C = (c_x, c_y)$ maps an image $I(x, y)$ in the cartesian domain into the polar domain domain $I_{Pol}^{C}(r, \theta)$, where

$$\begin{aligned} r(x,y) &= \sqrt{(x-c_x)^2 + (y-c_y)^2} \\ \theta(x,y) &= \arctan(x-c_x, y-c_y) \end{aligned}$$

The inverse transform is given by the formulae:

$$\begin{aligned} x(r,\theta) &= c_x + r\cos\left(\frac{\theta\pi}{180}\right) \\ y(r,\theta) &= c_y + r\sin\left(\frac{\theta\pi}{180}\right) \end{aligned}$$

where $r \in [0, R_{max}]$ and $\theta \in [0, 360)$, and R_{max} is the maximum distance of a pixel to the center point C in the cartesian image.

Appendix B

Design of steerable filters

Steerable filters are designed in order to provide directional information for specific orientations. The foundation of this technique is based on a complete set of basis functions in the space of orientations for a given scale. This set of basis functions can be used to construct equivalent functions for arbitrary orientations. Derivative of Gaussian functions are commonly used for the design of steerable filters. The derivatives of Gaussian form a complete set with different dimension for each derivation order: two functions are needed for order one, three for order 2, etc. Figure B.1 shows the basis functions for the derivatives of Gaussian of different order at a fixed scale. The first row shows the Gaussian function. The following rows show the basis functions for the first, second and third order derivatives.

The mathematical development of the steerable filter approach is as follows. The Gaussian function in 2D has the formula (here, the σ parameter was set to 1 for convenience):

$$G(x,y) = e^{(-x^2+y^2)}$$

The first partial derivatives of the Gaussian function are:

$$\begin{aligned} G_1^0 &= \frac{\partial}{\partial x} G(x,y) = -2x\ e^{(-x^2+y^2)} \\ G_1^{\pi/2} &= \frac{\partial}{\partial y} G(x,y) = -2y\ e^{(-x^2+y^2)} \end{aligned}$$

These two functions form a basis set. One filter with an arbitrary angular orientation θ can be built up from this basis by means of interpolation in the following way:

$$G_1^\theta = \cos(\theta) G_1^0 + \sin(\theta) G_1^{\pi/2}$$

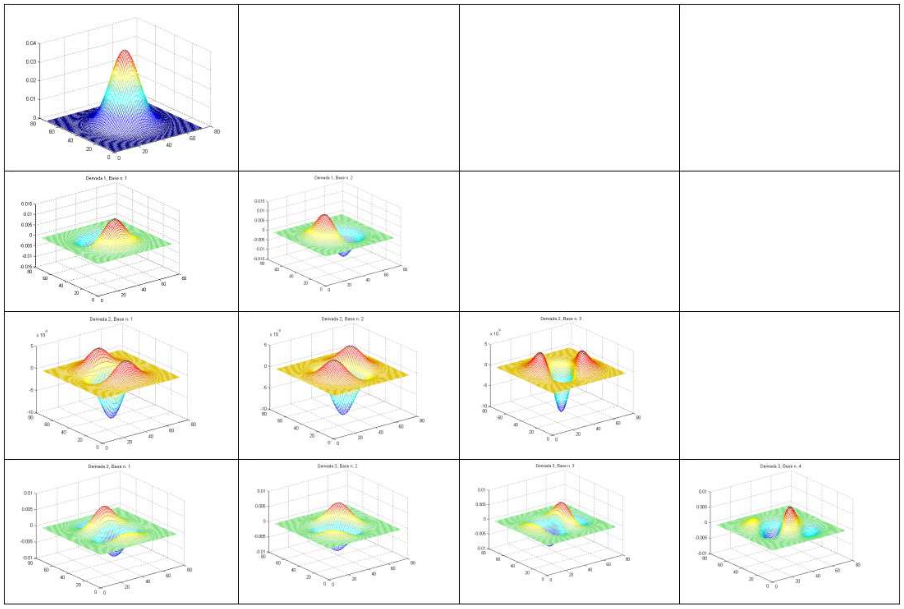


Figure B.1: Basis functions for a bank of filters of derivatives of Gaussian.

The general formula for the bases of higher order filters corresponds to:

$$G_n^{\theta_n}, n = 1, 2, 3 \ldots, \theta_n = 0, \ldots, k\pi/(n+1), k = 1, \ldots, n$$

And the general formula for an arbitrary oriented filter is:

$$G_n(\theta) = \sum_{i=1}^{n+1} G_n^{(i-1)\pi/(n+1)} k_{i,n}$$

For the specific case of second derivatives of gaussian, the three interpolants are as follows:

$$k_{i,2} = \frac{1}{3}\left[1 + 2\cos(2(\theta - (i-1)\pi/3))\right], i = 1, 2, 3$$

Appendix C

A simple set of heuristic rules for phasic contractions detection

These heuristics are exclusively based on intensity variation and lumen detection:

1. **Dynamic patterns detection**: We apply a zero-cross threshold over f^1.
2. **Rejection of the bubble and wall frames**:

   A The bubble detection is performed by means of a histogram matching. All the frames are quantized in color by using 256 bins. One frame is considered a bubble frame if more than 50% of its normalized color histogram coincides with, at least, another frame from a pool of 20 manually selected bubble frames.

   B One frame is labelled as a wall frame if the sum of the blob areas is lesser than 200 pixels.
3. **Rejection of lateral contractions**: We apply a circular mask with a radius 80% smaller than the radius of the circular visualization area of the frame. We reject all the sequences where more than 3 frames have more than 50% of the blob area out of this mask.
4. **Rejection of sequences with small variation in the blob size**: If the area of the blob which remains constant in the 9 frames is greater that 1500 pixels.
5. **Rejection of non-occlusive contractions**: If the area of the blob with the minimum value within the sequence is greater than 130 pixels.
6. **Rejection of false positives related to shadows wrongly detected as the lumen hole**: If the overall sum of f^3 for the pixels subscribed to the blob area is less than 10.

Appendix D

Color characterization of bubble frames using histogram matching

The presence of turbid liquid is characterized in terms of color, which is usually in a range from brown to yellow, mainly centered around green. The color characterization of the turbid frames may be performed by means of a color histogram quantization using, for example, the RGB space. This technique consist of two different steps: 1) color quantization, and 2) histogram generation for histogram matching. In the color quantization step, each of the possible 256 values of each RGB channel is associated with one of N bins linearly distributed between 0 and 255. In our case, we may apply a quantization value of $N = 8$. The resulting image has now $8 \times 8 \times 8 = 512$ possible colors, instead of $256 \times 256 \times 256 = 16,777,216$ colors of the original image. In the histogram generation step, we count the number of pixels for each of the 512 colors, and we construct a feature vector with these counts, normalized by the total number of pixels. Figure D.1 show several normalized histograms for different images. The horizontal axis corresponds to the bin number and the vertical axis corresponds to the numbers of counts for each bin.

We use the normalized histograms to compare the similarity of the video frames with a set of reference frames. For this aim, we manually selected a representative pool of 100 frames showing bubbles in the whole field of vision, taking special care in gathering the widest color range for the bubbles color description. Figure D.2 renders a subset of the frames used as reference colors for bubbles.

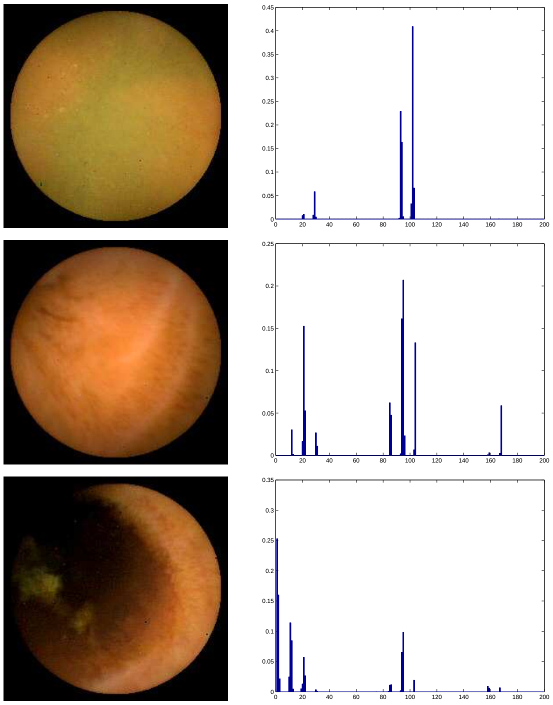


Figure D.1: Three different quantized color histograms for three different frames.

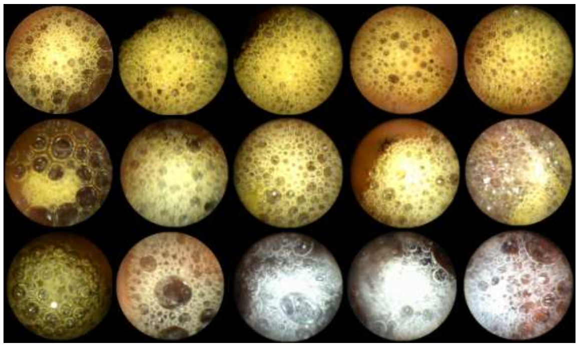


Figure D.2: Set of 15 bubble frames used in the reference set. The variety of colors range from yellow to green, including white

Our final goal consisted of obtaining a measure of the portion of the image which showed a color indicative of the presence of bubbles with a high likelihood. In order to tackle this aim, we applied a histogram matching [47]. We calculated the intersection of the normalized histogram of each video frame with each of the reference frames, obtaining the maximum intersection value in the following way:

$$Hist_{MI}(n) = \max_i \sum_{j=1}^{256} (Hist(n)^{(j)} \bigcap Hist(Ref_i)^{(j)}) \tag{D.1}$$

where $i = 1...100$ is the index of the i-reference frame, n is the index of the n-frame in the video, and j references the j-bin of the normalized histogram $Hist$. Thus, $Hist_{MI}(n)$ ranges in the interval $[0, 1]$, providing its minimum value when the n-video frame contains no color associated with bubbles, and providing its maximum value when all the pixels in the image share the same color distribution as one bubble reference frame.

Finally, we followed the same criterion described in the textural analysis, classifying a frame as a bubble frame if more than 50% of the image presented color similarity with one of the reference frames, i.e., $Hist_{MI}(n) > 0.5$.

Appendix E

Correcting the blob anisotropy

Warping consists of making an object change its original shape, by twisting or bending it, so that it adapts itself to a new shape. This technique requires a set of landmarks which are to guide the deformation of the whole image. In our scenario, we are interested in modifying the blob shape in order to adapt it to a perfectly concentric donut shape, as pictured in Figure E.1. Our conjecture is that the warping of the blob shape into a perfectly concentric shape may compensate the angular divergence of the wrinkles, leading to vertical patterns in the polar transform.

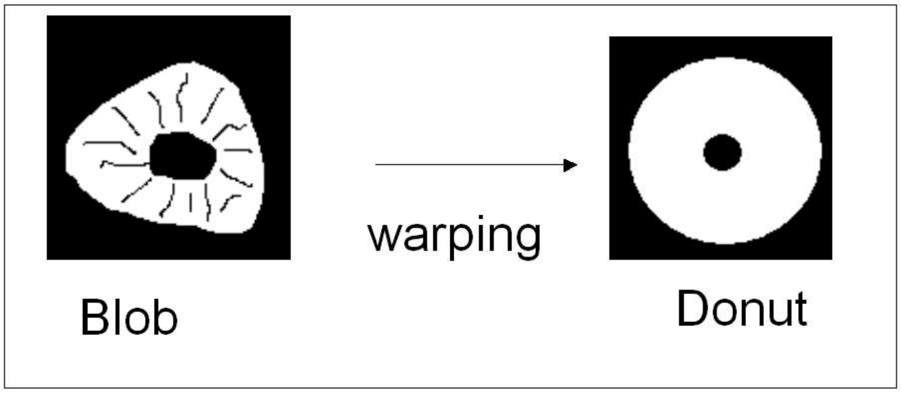


Figure E.1: Correction of the blob by means of warping.

Our proposal is summarized as follows:

1. Segment the original donut and distribute the landmarks along the internal and external donut contours in a homogenous way with a fixed distance of pixels.

2. Construct the target donut, set the landmarks number, and distribute the land-

marks along the internal and external donut contours in a homogenous way, using the same number of landmarks.

3. Associate each landmark of the inner contour of the original donut to one landmark of the inner contour of the target donut. Repeat the same procedure with the external contour -see Figure E.2-.

4. Perform the warping using the proposed associations of landmarks as evolution constraints for convergence.

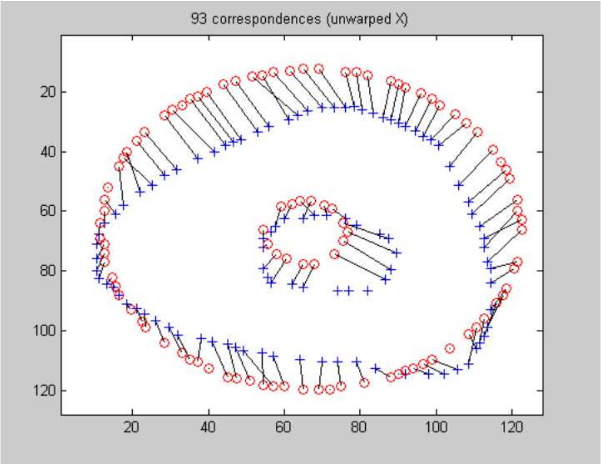
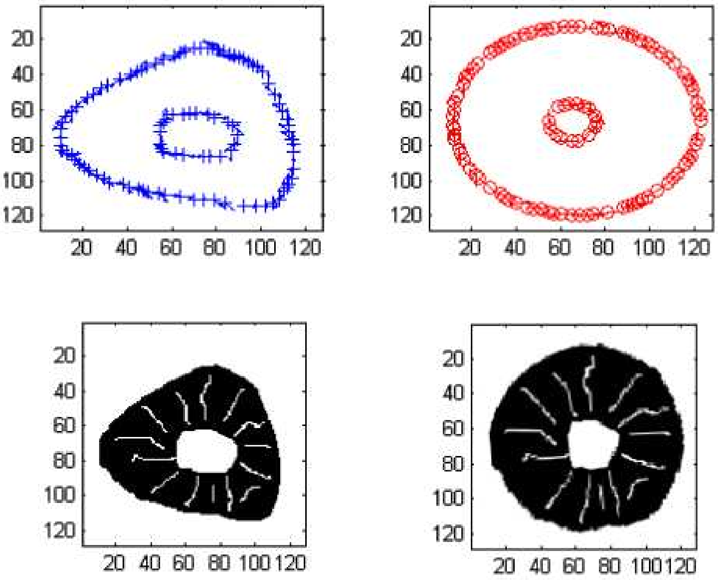


Figure E.2: Correspondences between the landmarks of the inner and outer contours of the original blob (crosses) and the synthetic donut (circles).

Figure E.3 shows on the top row the initial state for the original image (a) and the donut image (b), and on the bottom row the pattern of wrinkles (a) before and (b) after the transform. Finally, Figure E.4 shows the pattern of wrinkles for the original and warped images. The polar transform of the warped image shows lines with a higher vertical orientation.

We tested this method in tonic contractions, but we finally rejected its inclusion in the final procedure for the numerical problems associated with high anisotropic blobs, which led to a lack of convergence in the warping procedure. Furthermore, the improvement achieved showed to be relevant only for very clear samples, due to the large variability of the wrinkle patterns.



(a) (b)

Figure E.3: (a) Initial state for the original blob and (b) final result of the warping. The top row shows the control points and the bottom row the binary wrinkles images

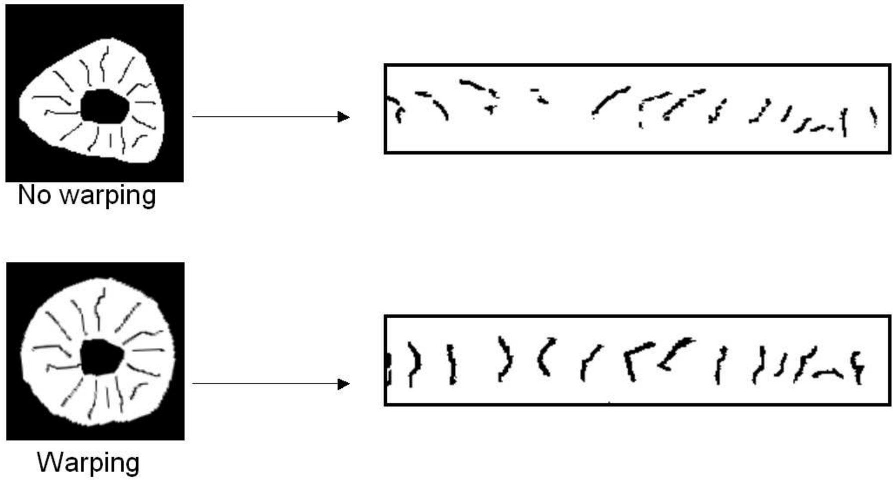


Figure E.4: Polar transform for the (top) original and (bottom) warped images.

Appendix F

Spatial tracking of the capsule in the body

The M2A© capsule stores, together with the video frames, the 2D spatial position of the capsule in the body. This information can be visualized as a path, and serves as a rough approximation of the real position of the camera. Thus, a pair of coordinates (p_x, p_y) is provided for each frame. These coordinates do not have a specific magnitude: they are obtained by power differentiation of the multiple aerials tapped to the patient, ant they refer to the pixel position in a visualization display provided by the commercial company. Figure F.1 pictures an example of path followed by the capsule for a clinical volunteer. The red and the green squares label the starting and the final point of the study.

The precision of the positioning is affected by several factors: directionality of the radiation pattern of the capsule, different absorbtion rates of the different soft and hard tissues, etc. Although and exhaustive calibration and validation procedure has not been performed for these values, we tried a first tentative approach to the use of the position information provided by the camera for the validation of a clinical scenario. The experts provided us with a set of studies in which a volunteer was administered with a drug which paralyzed its intestine for a determined span of time which could be controlled by the specialists. We plot the cumulative displacement from the positioning data provided by the camera. Figure F.2 shows some plots for different studies. The horizontal axis represents the time line and the vertical axis represents the cumulated displacement.

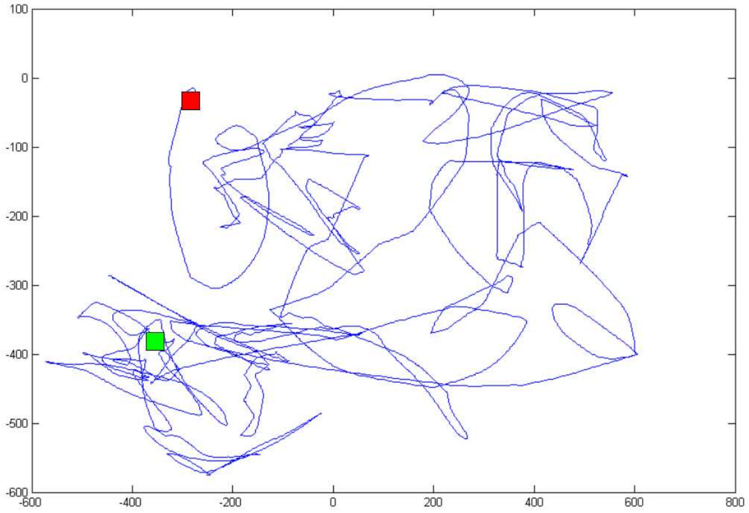


Figure F.1: Positioning information provided by the capsule.

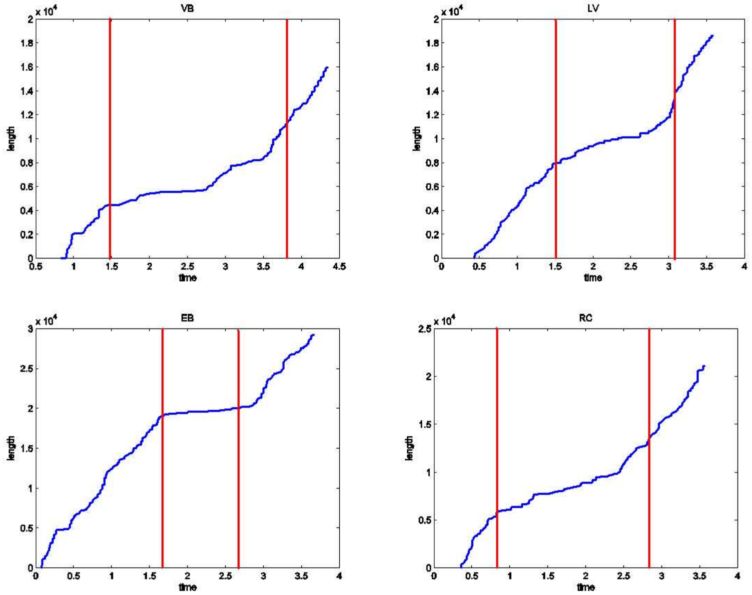


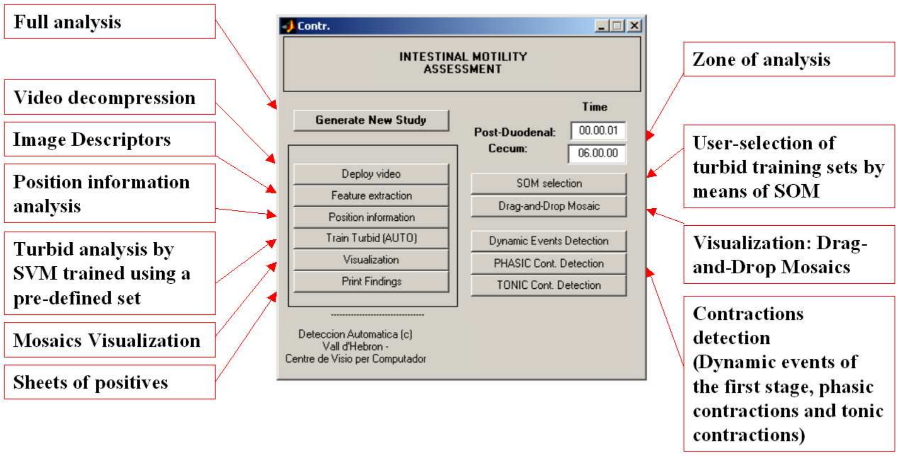
Figure F.2: Cumulated displacement for several studios. The red lines bound the expected effect of the paralyzing drug.

The red lines in Figure F.2 bound the estimated span of time in which the paralyzing drug is taking effect. For all the experiments, the evolution of the cumulated displacement tends to a static behavior within the expected area. These results show that, although the absolute position of the capsule is not a reliable measure, the relative position could be used as a helpful source of information for motility assessment.

Appendix G

Software applications

This section describes the software applications which we developed for the specialists, which include the automatic detection of phasic and tonic intestinal contractions, and the graphical tools for visual assessment. Figure G.1 pictures a snapshot of the main graphical user interface, in which all the functionalities are centralized.



Post-Duodenal: 00.00.01

Cecum: 06.00.00

Figure G.1: Main graphical user interface of the application.

The full analysis performs the feature extraction procedure over each video frame, generates the position data, and prints the findings provided by the experts into the sheets of positives. In order to adapt the turbid liquid classifier to the specificities of difficult cases, the specialists can manually create a training set of turbid and non-turbid frames by using the SOM tool. Figure G.2 shows a SOM in which the expert selected with the mouse a polygonal area containing turbid frames. This data set is used in the turbid detection step of the cascade.

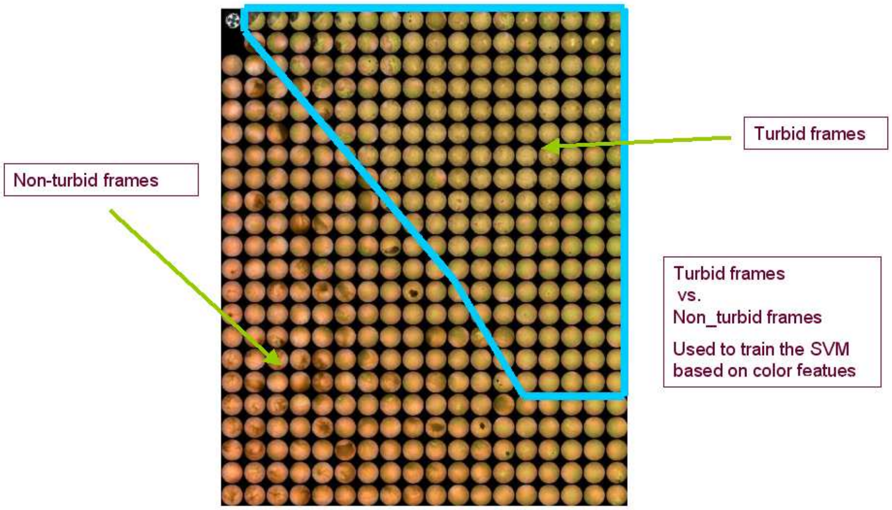
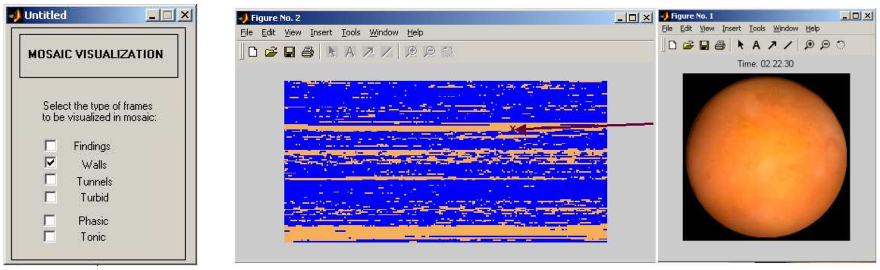


Figure G.2: SOM tool for the definition of a training set for turbid frames.

The visual assessment of the classification results can be performed by means of the drag-and-drop mosaics. A dialog frame -see Figure G.3 (a)- lets the specialists choose among the visualization of their own labels, tunnel, wall and turbid frames, and the contractions which were automatically detected by the system. Each type of frame is rendered with a different color. Figure G.3 (b) shows an example of a mosaic of wall frames, in which a specific region of the video was selected by the expert for its analysis. The video time is also provided during the video sequence visualization.


Untitled

MOSAIC VISUALIZATION

Select the type of frames
to be visualized in mosaic:

- □ Findings
- ☑ Walls
- □ Tunnels
- □ Turbid
- □ Phasic
- □ Tonic

Figure No. 2

Figure No. 1

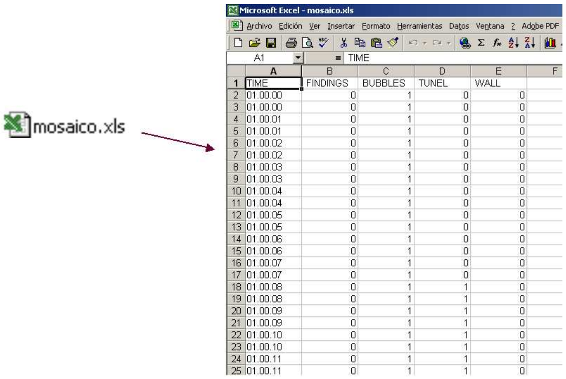
Time: 02.22.30

File Edit View Insert Tools Window Help
File Edit View Insert Tools Window Help

(a) (b)

Figure G.3: (a) GUI for the mosaic. (b) One mosaic of wall frames. The area selected by the expert is rendered for visual inspection.

In addition to the former, several useful files are provided to the experts. The file *mosaico.xls* consists of a spread sheet in which, for each frame, the expert can find out whether it has been classified as a turbid frame, a wall frame, a phasic contraction, etc. Figure G.4 shows a snapshot of this file.

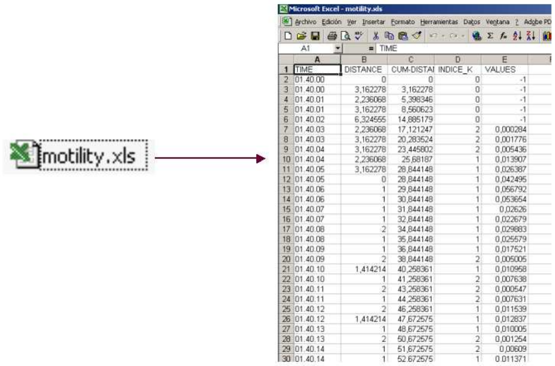


A1 = TIME

	A	B	C	D	E
1	TIME	FINDINGS	BUBBLES	TUNEL	WALL
2	01.00.00	0	1	0	0
3	01.00.00	0	1	0	0
4	01.00.01	0	1	0	0
5	01.00.01	0	1	0	0
6	01.00.02	0	1	0	0
7	01.00.02	0	1	0	0
8	01.00.03	0	1	0	0
9	01.00.03	0	1	0	0
10	01.00.04	0	1	0	0
11	01.00.04	0	1	0	0
12	01.00.05	0	1	0	0
13	01.00.05	0	1	0	0
14	01.00.06	0	1	0	0
15	01.00.06	0	1	0	0
16	01.00.07	0	1	0	0
17	01.00.07	0	1	0	0
18	01.00.08	0	1	1	0
19	01.00.08	0	1	1	0
20	01.00.09	0	1	1	0
21	01.00.09	0	1	1	0
22	01.00.10	0	1	1	0
23	01.00.10	0	1	1	0
24	01.00.11	0	1	1	0
25	01.00.11	0	1	1	0

Figure G.4: All the classification information is summarized in the file *mosaico.xls*.

The file *motility.xls* consists of another spread sheet which contains the position coordinates, together with the cumulative displacement and curvature values, as Figure G.5 shows.



A1 = TIME

	A	B	C	D	E
1	TIME	DISTANCE	CUM-DISTAI	INDICE_K	VALUES
2	01.40.00	0	0	0	-1
3	01.40.00	3,162278	3,162278	0	-1
4	01.40.01	2,236068	5,398346	0	-1
5	01.40.01	3,162278	8,560623	0	-1
6	01.40.02	6,324555	14,885179	0	-1
7	01.40.03	2,236068	17,121247	2	0,000284
8	01.40.03	3,162278	20,283524	2	0,001776
9	01.40.04	3,162278	23,445802	2	0,005436
10	01.40.04	2,236068	25,68187	1	0,013907
11	01.40.05	3,162278	28,844148	1	0,026387
12	01.40.05	0	28,844148	1	0,042495
13	01.40.06	1	29,844148	1	0,056792
14	01.40.06	1	30,844148	1	0,053654
15	01.40.07	1	31,844148	1	0,02626
16	01.40.07	1	32,844148	1	0,022679
17	01.40.08	2	34,844148	1	0,029883
18	01.40.08	1	35,844148	1	0,025579
19	01.40.09	1	36,844148	1	0,017521
20	01.40.09	2	38,844148	2	0,005005
21	01.40.10	1,414214	40,258361	1	0,010958
22	01.40.10	1	41,258361	2	0,007638
23	01.40.11	2	43,258361	2	0,000547
24	01.40.11	1	44,258361	2	0,007631
25	01.40.12	2	46,258361	1	0,011539
26	01.40.12	1,414214	47,672575	1	0,012837
27	01.40.13	1	48,672575	1	0,010005
28	01.40.13	2	50,672575	2	0,001254
29	01.40.14	1	51,672575	2	0,00609
30	01.40.14	1	52.672575	1	0.011371

Figure G.5: The position information is provided by one separate spread sheet.

The sheets of positives are printed in the bitmap files as the one rendered in Figure G.6. The video time, frame index and sequences of a length of 21 frames are provided.

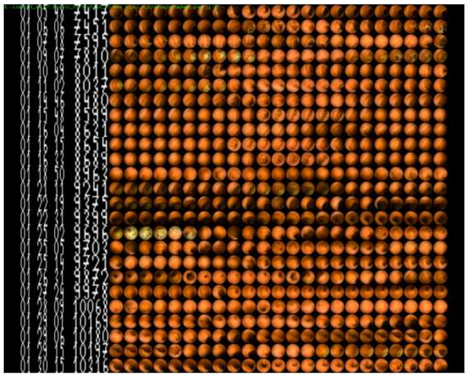
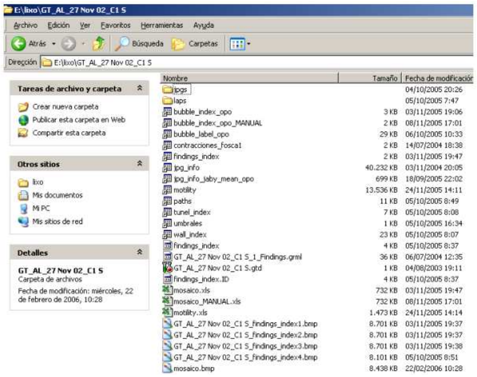


Figure G.6: The sheet of positives is saved as a bitmap file.

All the files are stored in the same folder -see Figure G.7-. This allows the specialist to have quick access to the whole information associated with each study.



GT_AL_27 Nov 02_C1 S
Carpeta de archivos
Fecha de modificación: miércoles, 22 de febrero de 2006, 10:28

Nombre	Tamaño	Fecha de modificación
jpgs		04/10/2005 20:26
laps		05/10/2005 7:47
bubble_index_opo	3 KB	03/11/2005 19:06
bubble_index_opo_MANUAL	2 KB	08/11/2005 17:01
bubble_label_opo	29 KB	06/10/2005 10:33
contracciones_fosca1	2 KB	14/07/2004 18:38
findings_index	2 KB	03/11/2005 19:47
jpg_info	40.232 KB	03/11/2004 20:05
jpg_info_jaby_mean_opo	699 KB	18/09/2005 22:02
motility	13.536 KB	24/11/2005 14:11
paths	11 KB	05/10/2005 8:49
tunel_index	7 KB	05/10/2005 8:08
umbrales	1 KB	05/10/2005 16:34
wall_index	23 KB	05/10/2005 8:07
findings_index	4 KB	05/10/2005 8:37
GT_AL_27 Nov 02_C1 S_1_Findings.grml	36 KB	06/07/2004 12:35
GT_AL_27 Nov 02_C1 S.gtd	1 KB	04/08/2003 19:11
findings_index.ID	4 KB	05/10/2005 8:37
mosaico.xls	732 KB	03/11/2005 19:47
mosaico_MANUAL.xls	732 KB	08/11/2005 17:01
motility.xls	1.473 KB	24/11/2005 14:14
GT_AL_27 Nov 02_C1 S_findings_index1.bmp	8.701 KB	03/11/2005 19:37
GT_AL_27 Nov 02_C1 S_findings_index2.bmp	8.701 KB	03/11/2005 19:37
GT_AL_27 Nov 02_C1 S_findings_index3.bmp	8.701 KB	03/11/2005 19:38
GT_AL_27 Nov 02_C1 S_findings_index4.bmp	8.101 KB	05/10/2005 8:51
mosaico.bmp	8.438 KB	22/02/2006 10:28

Figure G.7: All the files are stored in a common repository for each study.

These implementations, and the rest of the procedures presented in this work, were programmed in *Matlab 6.5*. We used the *LIBSVM* implementation [19] for the SVM classifier. We used the *CIS Laboratory SOM Toolbox 2.0* [4] for the SOM implementation. For the experiments with different classifiers we used the *PRTools* Toolbox [29]. For the AdaBoost and bagging experiments, we used our own implementations of the different methods.

All these applications are currently being used by the specialists for the experimental assessment of small intestinal motility with capsule endoscopy. The results obtained by the analysis provided using our approach can be consulted in the following pieces of research, presented at the American Gastroenterological Association congress, and their corresponding abstracts in the *Gastroenterology* journal.

- De Iorio F, Spyridonos P, Azpiroz F, Radeva P, Malagelada C, Malagelada J-R. *New insight into intestinal motor activity: correlation of endoluminal image analysis and displacement.* American Gastroenterological Association, Chicago, IL, USA, May 2005.
- De Iorio F, Malagelada C, Azpiroz F, Vilariño F, Vitrià J, Accarino A, Malagelada J.R. *In search for new parameters of intestinal motor activity in humans.* American Gastroenterological Association, Los Angeles, CA, USA, May 2006.
- Malagelada C, De Iorio F, Azpiroz F, Spyridonos P, Radeva P, Malagelada J-R. *Diagnosis of intestinal motor abnormalities by endoluminal image and displacement analysis: a novel approach.* American Gastroenterological Association, Los Angeles, CA, USA, May 2006.

Bibliography

[1] T. Aach, A. Kaup, and R. Mester. On texture analysis: Local energy transforms versus quadrature filters. *Signal Processing*, 45:173–181, 1995.

[2] D. G. Adler and C. J. Gostout. Wireless capsule endoscopy. *Hospital Physician*, pages 14–22, 2003.

[3] R. Akbani, S. Kwek, and N. Japkowicz. Applying support vector machines to imbalanced datasets. In *Proceedings of the ECML*, pages 39–50, 2004.

[4] E. Alhoniemi, J. Himberg, et al. *SOM Toolbox 2.0*, 2005. Software available at `http://www.cis.hut.fi/projects/somtoolbox/about.shtml`.

[5] A. Ali, J. M. Santisi, and J. Vargo. Video capsule endoscopy: A voyage beyond the end of the scope. *Cleveland Clinic Journal of Medicine*, 71(5):415–424, 2004.

[6] S. Anzali and S. Gasteiger. The use of self-organizing neural networks in drug design. *Perspectives in Drug Discovery and Design*, 11:273–299, 1998.

[7] P. Armitage and G. Berry. *Estadística para la investigación biomédica.* Ediciones Doyma, 1992.

[8] S. Bar-Meir and E. Bardan. Wireless capsule endoscopy - pros and cons. *Israel Medical Association Journal*, 4:726, 2002.

[9] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 36(1-2):105–139, 1999.

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

[11] W. Van Biesen and G. Sieben. Application of kohonen neural networks for the non-morphological distinction between glomerular and tubular renal disease. *Nepholiogy Dialisys Transplantation*, 13:59–66, 1998.

[12] S. Boyd and L. Vandenberghe. *Convex optimization.* Cambridge University Press, 2004.

[13] A. P. Bradley. The use of the area under the curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(6):1145–1159, 1997.

[14] J. Brank, M. Grobelnik, et al. Training text classifiers with SVM on very few positive examples. Technical report, MSR-TR-2003-34, 2003.

[15] L.J. Breiman, H. Friedman, et al. *Classification and Regression Trees.* Wadsworth & Brooks. Cole Advanced Books & Software, 1984.

[16] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):955–97, 1998.

[17] CCOHTA. Wireless capsule endoscopy. emerging issues in health technologies. Technical Report 53, Canadian Coordinating Office of Health Technology Assessment, 2003.

[18] I. Chakravarti, R. Laha, and J. Roy. *Handbook of Methods of Applied Statistics*, volume I, pages 392–394. John Wiley and Sons, 1967.

[19] C. Chang and C. Lin. *LIBSVM: a library for Support Vector Machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[20] N. V. Chawla, K. W. Bowyer, et al. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence and Research*, 16:321–357, 2002.

[21] C. I. Christodoulo and C.S. Pattichis. Unsupervised pattern recognition for the classification of emg signals. *IEEE Transactions on Biomedical Engineering*, 46(2):169–178, 1999.

[22] C. Cortes. *Prediction of generalization ability in learning machines. University of Rochester.* PhD thesis, 1994.

[23] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines.* Cambridge University Press, 2000.

[24] Blue Cross and Blue Shield Association. Wireless capsule endoscopy. Technical Report 17, TEC Assessments, 2003. 21.

[25] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Optical Society of America*, 2(7):1160–1169, 1985.

[26] R. De Franchis, E. Rondonotti, et al. Use of the Given video capsule system in small bowel transplanted patients [abstract]. *Gastrointestinal Endoscopy*, 55:129, 2002.

[27] P. Domingos. Metacost: A general method for making classifiers cost-sensitive. In *Proceedings of the ICKDDM*, pages 155–164, 1999.

[28] R. O. Duda, P. E. Hart, et al. *Pattern Classification.* Wiley-Interscience, 2n edition, 2001.

[29] R. Duin, P. Juszczak, et al. *PRTools 4. A Matlab toolbox for pattern recognition.* Delft University of Technology, 2004.

[30] R. Eliakim. Wireless capsule video endoscopy: Three years of experience. *World Journal of Gastroenterology*, 10(9):1238–1239, 2004.

[31] T. Fawcett and F Provost. Adaptive fraud detection. *Data Mining and Knowledge Discovery*, 1(3):291–316, 1997.

[32] Z. Fireman, A. Glukhovsky, et al. Wireless capsule endoscopy. *Israel Medical Association Journal*, 4:717–719, 2002.

[33] Z. Fireman, E. Mahanja, et al. Diagnosing small bowel crohns disease with wireless capsule endoscopy. *Gut*, 52:390–392, 2003.

[34] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.

[35] Y. Freund and R. E. Schapire. A decision-theoretic generalization of online learning and an application to boosting. In *Proceedings of the EUROCOLT*, pages 23–37. Springer-Verlag, 1995.

[36] Y. Freund and R. E. Schapire. Experiments with a new boosting algorithm. In *Proceedings of the ICML*, pages 148–156, 1996.

[37] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. Technical report, Dept. of Statistics, Stanford University, 1998.

[38] K. Fukunaga. *Introduction to statistical pattern recognition (2nd ed.).* Academic Press Professional, Inc., 1990.

[39] Given Imaging, Ltd. http://www.givenimaging.com/. 2005.

[40] J. O. Glass and W.E. Reddick. Hybrid artificial neural network segmentation and classification of dynamic contrast-enhanced MR imaging (DEMRI) of osteosarcoma. *Magnetic Resonance Imaging*, 16(9):1075–1083, 1998.

[41] I. Guyon, J. Weston, et al. Gene selection for cancer classification using Support Vector Machines. *Machine Learning*, 46(1-3):389–422, 2002.

[42] M. B. Hansen. Small intestinal manometry. *Physiological Research*, 51:541–556, 2002.

[43] R.M. Haralick and L.G. Shapiro. *Computer and Robot Vision*, volume I. Addison-Wesley, 1992.

[44] A. Hernandez, D. Gil, et al. Anisotropic processing of image structures for adventitia detection in IVUS images. In *Proceedings of Computers in Cardiology*, volume 31, pages 229–232, 2004.

[45] G. Iddan, G. Meron et al. Wireless capsule endoscopy. *Nature*, 405:4–7, 2000.

[46] T. Joachims. Text categorization with support vector machines: learning with many relevant features. In *Proceedings of ECML*, pages 137–142. Springer Verlag., 1998.

[47] F. Jou, K. Fan, and Y. Chang. Efficient matching of large-size histograms. *Pattern Recogn. Lett.*, 25(3):277–286, 2004.

[48] S. A. Karkanis, D. K. Iakovidis, et al. Computer aided tumor detection in endoscopic video using color wavelet features. *IEEE Transactions on Information Technology in Biomedicine*, 7:141–152, 2003.

[49] D.Z. Katz and B.S. Lewis. Capsule endoscopy in known or suspected crohns disease: the perspective of the referring physician and the patient [abstract]. *American Journal of Gastroenterology*, 97:S300, 2002.

[50] J. E. Kellow, M. Delvaux, et al. Principles of applied neurogastroenterology: physiology/motility-sensation. *Gut*, 45(2):1117–1124, 1999.

[51] T. Kohonen. *Self-Organized Maps.* Springer, Heidelberg Berlin, 1995.

[52] A. Kolcz, A. Chowdhury, et al. Data duplication: an imbalanced problem? In *Workshop on Learning from Imbalanced Datasets II, ICML*, 2003.

[53] M. Kubat, R. C. Holte, et al. Machine learning for the detection of oil spills in satellite radar images. *Machine Learning*, 30(2-3):195–215, 1998.

[54] M. Kubat and S. Matwin. Addressing the curse of imbalanced training sets: one-sided selection. In *Proceedings of the ICML*, pages 179–186, 1997.

[55] L. I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms.* Wiley-Interscience, 2004.

[56] Z. Li and G. Hu. Analysis of disparity gradient based cooperative stereo. *IEEE Transactions on Image Processing*, 5(11):1493–1506, 1996.

[57] C.X. Ling and C. Li. Data mining for direct marketing: Problems and solutions. In *Knowledge Discovery and Data Mining*, pages 73–79, 1998.

[58] Y. Liu, E. Shriberg, et al. Using machine learning to cope with imbalanced classes in natural speech: Evidence from sentence boundary and disfluency detection. In *Proceedings of the ICSLPl*, 2004.

[59] A. M. Lopez, F. Lumbreras, J. Serrat, and J. J. Villanueva. Evaluation of methods for ridge and valley detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(4):327–335, 1999.

[60] C. L. Lux and A. Atellzig. A neural network approach to the analysis and classification of human craniofacial growth. *Growth, Development, and Aging*, 62(3):95–106, 1998.

[61] G. Magoulas, V. Plagianakos, et al. Neural network-based colonoscopic diagnosis using online learning and differential evolution. *Applied Soft Computing*, 4:369–379, 2004.

[62] M. A. Maloof. Learning when data sets are imbalanced and when costs are unequeal and unknown. Workshop on Learning from Imbalanced Data Sets II. In *Proceedings of the ICML*, 2003.

[63] C. Metz. Basic principles of ROC analysis. *Seminars on Nuclear Medicine*, VII:283–298, 1978.

[64] M. C. Monard and G. Batista. Learning with Skewed Class Distribution. In *Advances in Logic, Artificial Intelligence and Robotics*, pages 173–180, 2002.

[65] P. Ohanian and R. Dubes. Performance evaluation for four classes of textural features. *Pattern Recognition*, 25(8):819–833, 1992.

[66] T. Ojala, M. Pietik, et al. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(7), 2002.

[67] E. M. Quigley. Gastric and small intestinal motility in health and disease. *Gastroenterology Clinics of North America*, 25:113–145, 1996.

[68] E. M. Quigley. Disturbances in small bowel motility. *Baillieres Best Practice and Research. Clinical Gastroenterology*, 13(3):385–395, 1999.

[69] J. Ross Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

[70] J. F. Rey, Gay G., et al. European Society of Gastroenterology. Guideline for video capsule endoscopy. *Endoscopy*, 36:656–658, 2004.

[71] J. C. Russ. *The Image Processing Handbook.* IEEE Press. 2nd ed, 1994.

[72] K. Schulmann, S. Hollerbach, et al. Feasibility and diagnostic utility of video capsule endoscopy for the detection of small bowel polyps in patients with hereditary polyposis syndromes. *The American Journal of Gastroenterology*, 100(1):27, 2005.

[73] J. A. Swets. Measuring the accuracy of diagnostic systems. *Science*, (240):1285–1293, 1988.

[74] J. A. Swets and R. Pickett. *Evaluation of Diagnostic Systems: Methods for Signal Detection Theory.* Academic Press, 1982.

[75] S. Theodoridis and K. Koutroumbas. *Pattern Recognition.* Elsevier, 2003.

[76] S. Tong and E. Chang. Support Vector Machine active learning for image retrieval. In *Proceedings of the ACM-ICM*, pages 107–118. ACM Press, 2001.

[77] S. Tong and D. Koller. Support Vector Machine active learning with applications to text classification. In *Proceedings of ICML*, pages 999–1006, 2000.

[78] V. Vapnik. *Estimation of dependences based on empirical data.* Nauka, Moscow. (English translation: Springer Verlag, New York, 1982), 1979.

[79] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag, 1995.

[80] K. Veropoulos, C. Campbell, and N. Cristianini. Controlling the sensitivity of support vector machines. In *Proceedings of IJCAI. Workshop ML3*, 1999.

[81] F. Vilariño and P. Radeva. Patch-optimized discriminant active contour for medical image segmentation. *Proceedings of IBERAMIA*, pages 100–105, 2002.

[82] F. Vilariño and P. Radeva. Cardiac segmentation with discriminant active contours. In *Proceedings of the CCAI. IOS Press*, 2003.

[83] F. Vilariño, M. Rosales, and P. Radeva. Patch-optimized discriminant active contours for medical image segmentation. In *Proceedings of IBERAMIA*, 2002.

[84] F. Vilariño, P. Spyridonos, et al. Experiments with SVM and stratified sampling with an imbalanced problem: Detection of intestinal contractions. In *Proceedings of the ICAPR (2)*, pages 783–791, 2005.

[85] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Porceedings of the IEEE Computer Society Conference on CVPR*, 2001.

[86] A.J. Walker and S.S. Cross. Visualisation of biomedical datasets by use of growing cell structure networks: A novel diagnostic classification technique. *Lancet*, 354:1518–1521, 1999.

[87] J. Weickert. A review of nonlinear diffusion filtering. In *Proceedings of the first International Conference on Scale-Space Theory in Computer Vision*, pages 3–28, 1997.

[88] G. Wyszecki and W. S. Stiles. *Color science: concepts and methods, quantitative data and formulae.* John Wiley & Sons, 2nd edition, 1982.

[89] B. Zadrozny and C. Elkan. Learning and making decisions when costs and probabilities are both unknown. In *Proceedings of the ICKDDM*, 2001.

[90] K. H Zou. ROC Analysis Literature Page. http://splweb.bwh.harvard.edu:8000/pages/ppl/zou. 2005.

Final Acknowledgment

This work was supported in part by a research grant from Given Imaging Ltd., Yoqneam Israel, Hospital Universitari Vall d'Hebron, Barcelona, Spain, as well as the projects FIS-G03/1085, FIS-PI031488, TIC2003-00654 and MI-1509/2005.