Farhan Riaz

Assisted Analysis of Gastroenterology Images Using Computer Vision Methodologies

Departamento de Ciências de Computadores
Faculdade de Ciências da Universidade do Porto
2012

Farhan Riaz

Assisted Analysis of Gastroenterology Images Using Computer Vision Methodologies

Tese submetida à Faculdade de Ciências da
Universidade do Porto para obtenção do grau de Doutor
em Ciência de Computadores

Departamento de Ciências de Computadores
Faculdade de Ciências da Universidade do Porto
2012

Dedicated to My parents

Acknowledgments

First of all, I would like to thank Prof. Miguel Tavares Coimbra and Prof. Mario Dinis Ribeiro from the University of Porto for supervising my work. I started my PhD without having background knowledge about the area of Biomedical Engineering. Through the professional guidance of my supervisors, I was able to adapt myself with this very new area in a seamless way. I have been getting very constant attention and feedback about my work and the criticisms from my supervisors have been very useful in forcing me to do high quality research in Portugal.

On a more personal level I would like to thank my parents for their support that has made it possible for me to achieve this task. Apart from my personal initiative, it was only because of the support and motivation from my parents that has made it possible for me to be able to do my PhD. The dearest thing for me in my life is my parent's wishes that I wanted to fulfil and that was my biggest motivation. I thank you now, and I will keep thanking you for the rest of my life for the support that I have always been having as I would probably never be able to even think about a PhD if you would not have desired. Whatever I am and whatever good that I do in my life is and will be primarily because of you.

One person who has been very supportive to me is my wife. We got married in the middle of my PhD and she joined me in Porto. This added much more diversity in my routine in contrast with the monotonous life that I was having before: Work, sleep, work, sleep ... She has stayed with me until the end and has been supportive to me during the times when I have been struggling in my research. The trips that we've had across Europe have been very refreshing and the most memorable times during my whole PhD. During my PhD, she also gave me the best gift that I could ever have: my daughter. I would like to thank my brother and sisters who have also been very supportive during my tenure in Portugal. Thanks for the skype chats continuously for 4 years that you've always had with me. It was the way of connecting with you and my parents which was only possible because of you.

I would like to thank Pedro-Pimentel Nunes and Prof. Mario Dinis Ribeiro from Portuguese Institute of Oncology Porto, Miguel Areia from Portuguese Institute of Oncology Coimbra and Francisco Silva from Karolinska Universitessjukhuse Sweden for providing data for carrying out this research. I would like to thank Andre Sousa for providing manual annotations of one of the datasets used in this work. I would thank the Fundação para a Ciência e a Tecnologia (scholarship reference: SFRH/BD/45066/2008) and Instituto de Telecomunicações (IT) for providing me financial assistance.

Resumo

Esta tese é motivada pelos potenciais ganhos que podem ser alcançados com o uso de sistemas de apoio à de decisão, baseados em computadores (CAD), no diagnóstico de cancro do trato gastrointestinal (GI) através de endoscopia. Dentro das vàrias linhas de investigação em sistemas CAD, este trabalho focaliza-se na àrea da visão computacional. O Estado da Arte sobre esta àrea aponta para o uso de um sistema de reconhecimento de padrões, focado principalmente nas técnicas de segmentação, extração de caractersticas, e classificação. Existindo sempre um estudo detalhado das caractersticas das imagens de endoscopia, para determinar quais as caractersticas mais relevantes para a elaboração do diagnóstico. Este estudo foi realizado recorrendo a gravações de vídeo realizadas durante a execução de exames endoscópicos de dois métodos de imagem complementares: Cromoendoscopia e imagens de banda estreita. Foi também desenvolvido um software capaz de extrair imagens clinicamente relevantes dos vídeos, imagens estas que foram manualmente anotadas por médicos especialistas.

Através do estudo das características relevantes da imagem, propusemos um método inovador de segmentação de imagem, através da integração de caractersticas de *creasness* com algoritmos de segmentação usando a técnica *normalized cuts*. As experiências realizadas demonstram a superioridade do uso de características de *creasness* para a segmentação de imagens gastroenterológicas. Adicionalmente uma combinação de caracterśticas de *creasness* e *edgemaps* é usada para transmitir informação complementar sobre as imagens. Tal como é verificado nas melhorias obtidas nos resultados.

Para a extração de características, e sobretudo devido às condições dinâmicas das imagens, a importância da invariância na rotação, escala e iluminação foi devidamente reconhecida. Este facto conduziu à apresentação de novos descritores de textura invariantes. Estes descritores foram integrados numa *texton framework*, usando uma abordagem *bag-of-words* para classificar imagens provenientes dos dois métodos mencionados no primeiro parágrafo. Onde os resultados se mostraram favoráveis, quando comparados com outros descritores de textura presentes no Estado da Arte.

Abstract

This thesis is motivated by the potential gains that can be achieved by the use of computer assisted decision (CAD) systems for diagnosis of cancer in the gastrointestinal (GI) tract using endoscopy. Although several lines of research are available for design of CAD systems, we have focused on the computer vision module in such systems. The state-of-the-art research hints at the use of a pattern recognition system consisting of segmentation, feature extraction and classification as its primary modules. A detailed study of the characteristics of endoscopic images is carried out to understand the imaging characteristics which are most relevant for making diagnosis. This was done by recording videos during live endoscopic examinations from two complementary imaging modalities: Chromoendoscopy and Narrow-band imaging. Softwares were developed to obtain clinically relevant images from the videos, that were manually annotated by expert physicians.

Based on our study on relevant image characteristics, we have proposed a novel image segmentation method by integrating *creasness* features with normalized cuts segmentation framework. Our experiments on the collected dataset show the superiority of the use of *creasness* features for segmentation of gastroenterology images. Later, we used a combination of *creasness* and *edgemaps* features, that transmit complementary information about the images. Experiments verify their complementary nature thus improving the segmentation results.

For feature extraction, owing to the dynamic imaging conditions and diverse anatomy of various organs in the gastrointestinal tract, the vitality of rotation, scale and illumination invariance was recognized. This led to the proposal of novel texture descriptors which are invariant to the said image transformations. The novel features were integrated into a *texton framework* using *bag-of-words* approach to classify images from the above mentioned imaging modalities. Classification results obtained from these novel texture features were compared with some other state-of-the-art texture features, showing the superiority of our proposed descriptor.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Cancer is the leading cause of deaths and accounted in 2008 for 7.6 million deaths (around 13 % of all deaths) worldwide (World Health Organization, 2008). The main types of cancer leading to overall cancer mortality reveal that Gastric (or stomach) cancer is the second most lethal cancer in the world (Fig. 1.1). Every year, approximately one million new cases of gastric cancer are diagnosed worldwide. For the whole digestive tract, the incidence rate for cancer is the highest. Gastrointestinal (GI) cancer occurs twice as often in men and it is common in people over the age of 50 years. Although it is a deadly form of cancer, changes in dietary habits and general awareness among the people has led to a steady decline in the incidence rate of cancer of the lower GI tract. The incidence of cancer of the upper GI tract has not however shown this trend. Research indicates that this is primarily as a result of the prevalence of obesity, and gastroesophageal reflux disease (American Cancer Society, 2008).

Gastroenterology imaging is a procedure, which is essential for the physicians to be able to diagnose any potential problems in the GI tract. In gastroenterology imaging, the GI tract is visualized using various equipments (depending on the imaging technology) to ascertain the health of the GI tract by analyzing the internal walls of various organs in the GI tract. Conventional diagnosis procedure for screening of GI tract suffers certain limitations. A computer assisted decision (CAD) system may be desirable, that can somehow alleviate some of the limitations thus assisting in screening of GI tract of the patients.

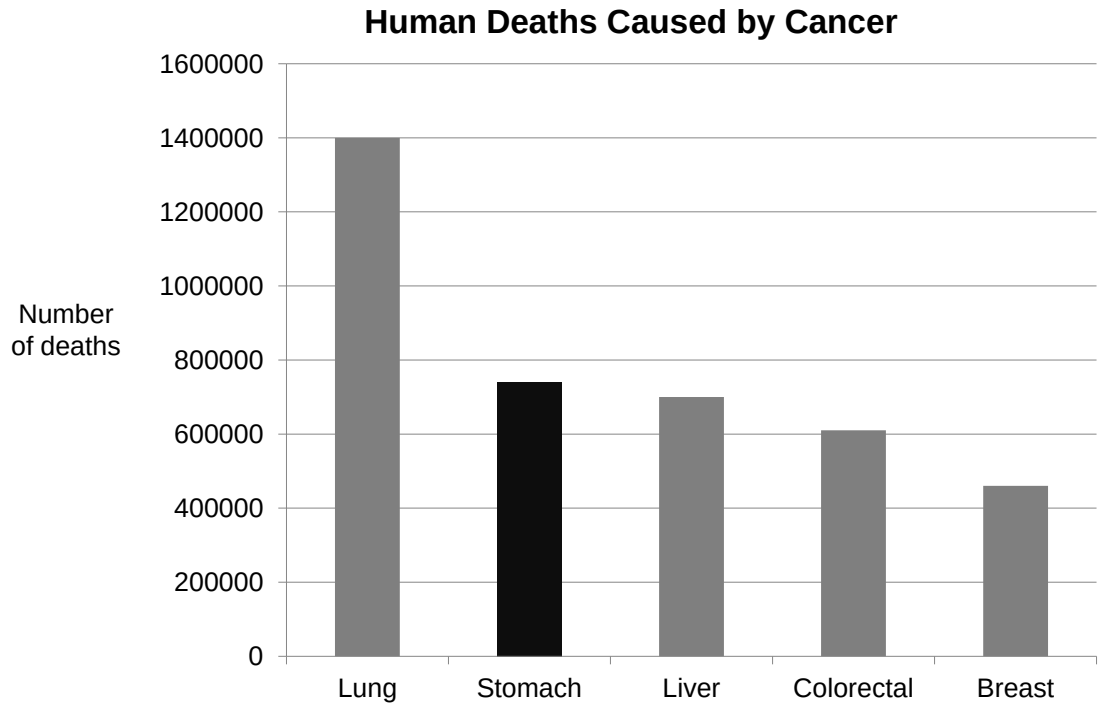


Figure 1.1: Mortality by cancer type (World Health Organization, 2008)

1.1 Conventional GI cancer diagnosis procedure

Conventional diagnosis of GI cancer is done primarily by using an endoscope. A clinician studies a patient file to learn his history. Given enough indicators, an endoscopy procedure is performed (Fig. 1.2). During this procedure, the physician traverses through the accessible parts of the GI tract to visualize any abnormalities such as the presence of ulcers, examination of internal walls etc. using an endoscope. It consists of a probe having a light source and a camera at its tip, that is inserted into the GI tract to visualize any abnormalities. The endoscope has a separate channel that is used for inserting instruments or manipulators in the GI tract. These instruments can be used to perform some operations such as collection of tissue sample (biopsies), cleaning the visualized tissue for removal of food traces etc. Biopsies of the tissue samples are also obtained for microscopic laboratory analysis of the tissue in clinically relevant images. The physicians usually record these images during live endoscopic examinations and make their decision about the diagnosis based on them. After the procedure, the physician makes appropriate annotations of his observations. He makes a diagnosis for a particular patient based on the visual examination of the tissue and usually waits for the biopsy report from the laboratory for a confirmed diagnosis.

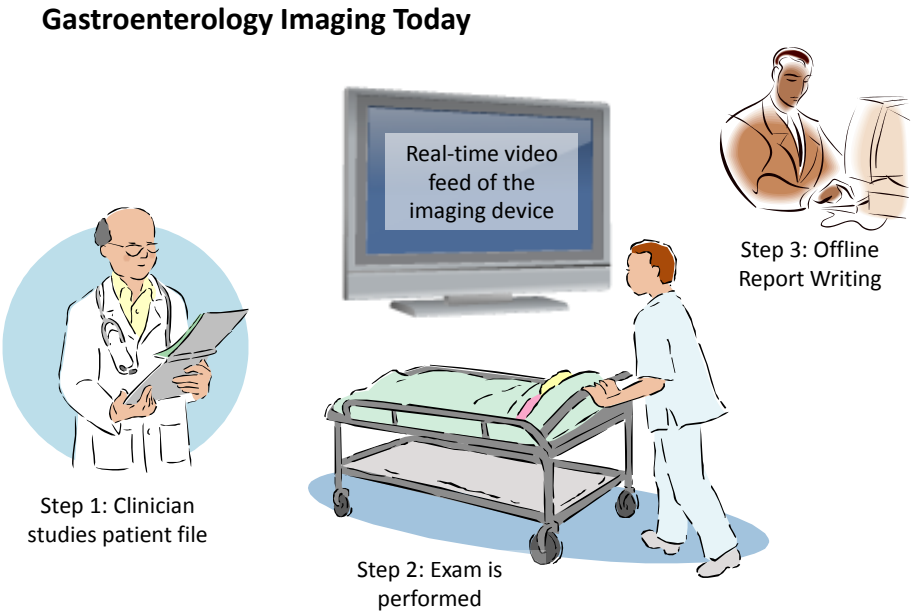


Figure 1.2: Current gastroenterology imaging procedure

1.2 Challenges in conventional diagnosis

There are several problems with the conventional GI cancer diagnosis procedure. One of them is that cancer screening is a time consuming procedure. Most of the pathologies related with the GI tract produce generic symptoms which make it hard to localize the organs which could potentially be responsible for them. This problem makes the diagnosis difficult since the physicians need to have a detailed visual analysis of the GI tract and require time to localize the problem during a typical gastroenterology imaging procedure. Due to this, a typical exam time for one endoscopy exam varies between 15 and 30 minutes. As the diagnosis is time consuming, a limited number of patients can be screened and therefore, the patients are prioritized according to their symptoms. Another important fact is that endoscopy is performed by an expert gastroenterologist as it is important to fully understand the anatomy and state of the tissues which constitute the walls of the GI tract. Therefore, an endoscopic procedure requires an extended amount of resources in terms of time, cost, clinical expertise and equipment. Also, the diagnosis varies with the expertise of the physician and there is lack of second opinion due to limited number of specialist gastroenterologists.

1.3 Future of GI cancer diagnosis: CAD system

Nowadays, the time-price factor for an experienced gastroenterologist is significant. Therefore, it is believed that a CAD system can be used to help educate young physicians, facilitate second opinion or help the physician to adapt to the changing technological trends in gastroenterology. A CAD system having the capability to raise alarms automatically, if installed, can be used to help expert gastroenterologist. In GI cancer diagnosis, even expert gastroenterologists can differ in their opinion on the state of the patient but this problem can be highly mitigated with the design of a CAD system as the diagnosis done by the automated system is expected to be robust. Alternatively, CAD systems can be employed for providing a reliable system, thus overcoming issues related with confidence level in some cases during manual diagnosis. They can be voluntarily used for making some quantitative measurements such as size of a polyp etc. Such systems can also be used for medical training. Inexperienced gastroenterologists can significantly benefit from the developments in CAD systems for gastroenterology to do effective training.

Recently there has been a phenomenal development in terms of the various imaging modalities, which can be effectively used for diagnosis of GI cancer. This is because, there is no single imaging modality which delivers all the relevant imaging characteristics, which are vital for visualizing the GI images. These recently established advancements have fast been a source of significant research in analysing the impact of various technologies on an early diagnosis of GI cancer which can potentially save millions of lives every year. The significantly important technological advancements and difficulties involved in cancer screening have made it very hard to cope up with the exponential growth experienced in the field of gastroenterology. With the design of a good CAD system, the interplay of most of the factors responsible for difficulties in screening of the patients and the explosive growth in the field of gastroenterology can be compensated, therefore emphasizing on the need to devise such systems, that can prove to be an asset for humanity.

1.3.1 Our vision on design of CAD system

We envisage a CAD system for diagnosis of gastrointestinal (GI) cancer, which consists of three main modules: information access, computer vision and human computer interaction (HCI). The information access system records the information of the patients for their effective follow up and has the ability to save exams. It consists of an atlas

of queries that can be made to retrieve/store the data, thus resulting in enhanced information access. The computer vision module is responsible for analysis of the images from the patient. This is mainly responsible for the processing part, and analyses the patterns available in the images to make an effective diagnosis. It can also be used to make some quantitative analysis such as polyp size etc. The HCI module is responsible for the interface. It can possibly use some voice commands, or recognize the gestures to take appropriate action or issue various commands to the system. Such a system will be able to provide improved diagnostic capabilities by reducing analysis time, helping in training of physicians, providing second opinion and increasing robustness of diagnosis.

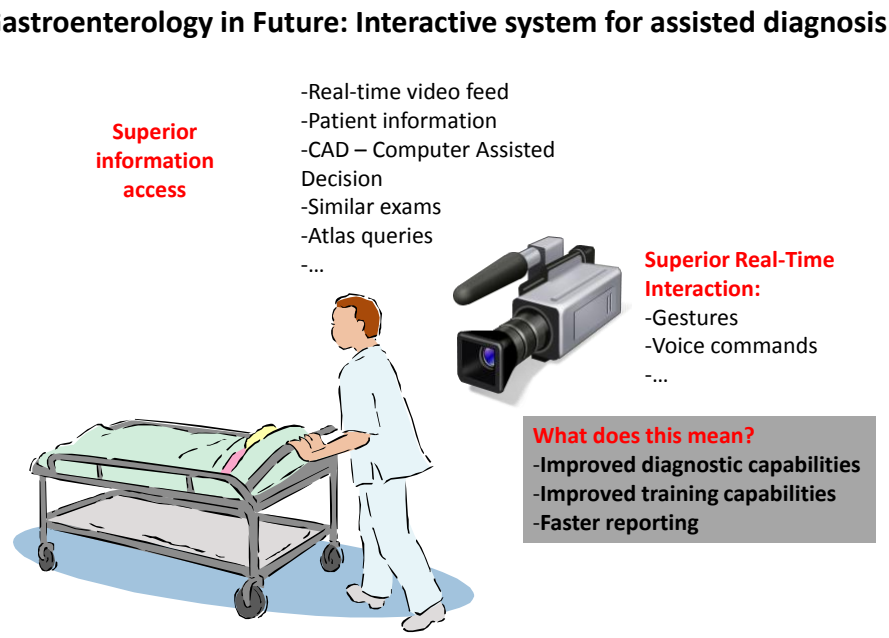


Figure 1.3: Future of Gastroenterology Imaging: Superior information access system provides patient history, assisted diagnosis is done using interactive commands for CAD system.

1.3.2 Our focus: Computer Vision

In this thesis, our objective is to identify various challenges related with the computer vision module in the design of CAD systems. These challenges are attributed to the dynamics related with both, the imaging modalities used for GI cancer diagnosis and the dynamic of the organs. There are various imaging modalities which are suitable for

addressing various challenges. For example we have conventional endoscopy which is used to analyse the internal walls of the organ superficially. For a detailed examination, enhanced methods have been developed allowing a much clearer visualization of the tissue, thus increasing the chances of detecting any potential abnormalities in the GI tract. These methods are suitable for addressing various complementary challenges as they provide complementary information about the health of the tissue.

Furthermore, the anatomy of the organs is responsible for providing dynamic imaging conditions. For example, esophagus is a tubular structure and therefore allows limited field of view for inspection of the internal walls. Stomach on the other hand is a sac-like structure that allows a wide range of views of a single tissue from various angles. Moreover, the visualization of the tissue is carried out by keeping the camera very close to the wall. Small changes in the distance between the camera and the wall can create a very different views of the same tissue. An automated system should have the ability to cope with these changes, resulting in a robust diagnosis. The small intestine and start of large intestine are not easily accessible using conventional endoscopy therefore capsule endoscopy (CE) is used for screening of small intestine. These are tube like structures and intestinal motility can cause very fast transitions of the capsule in the intestine, making it hard to do screening of various regions in the intestine, potentially overlooking some pathological condition. On the other hand, there is an illumination gradient in capsule images which arises due to a tube like structure of the intestine. In stomach and esophagus, the endoscopic probes are manually controlled and effort can be made to compensate for this gradient. In CE, the capsule movement occurs due to intestinal motility and therefore nothing can be done to improve the imaging other than finding methods that can normalize this gradient.

These are some of the few challenges from the imaging perspective. A generic computer vision module useful for a wide range of imaging modalities and various organs should be able to give a robust characterization of the tissues by providing rotation, scale and illumination changes in the images. On the other hand, they should also be able to calculate complementary image features such as color, shape, texture etc. to be generic enough to be used effectively for multiple imaging scenarios and organs.

1.4 Objectives of this thesis

In view of the requirements posed by the futuristic CAD systems for gastroenterology, we decided to pursue research specifically on the computer vision aspect to handle the

imaging dynamics of various imaging modalities or organs. This therefore leads us to the following objectives in this thesis:

To identify computer vision challenges in gastroenterology - In the recent past, several imaging modalities have been introduced which have distinct visual characteristics (e.g. narrow-band imaging exposes vasculature in images in contrast with white light etc). These characteristics demand an adaptation of the visual descriptors to incorporate the novel features from various imaging modalities. Other dimensions to the dynamics of imaging characteristics are enforced by the distinct organs (e.g. different anatomy of esophagus and stomach) or the clinician performing endoscopy (e.g. distance between the tip of the camera and the tissue) that can result in a different visualization of the same tissue. Given this, our objective in this thesis is to analyze the most generic set of requirements for the design of computer vision systems for gastroenterology.

To research on methods adapted to identified challenges - The identification of various challenges was followed by a study of conventional methods used in state-of-the-art research. The most vital objective of this thesis is that we intend to expand the existing methods to handle the imaging dynamics related with various imaging modalities and imaging sites. As such based on our research, we have focused on two important challenges in a traditional pattern recognition system for gastroenterology: segmentation and feature extraction.

To evaluate novel methods on gastroenterology scenarios - In this thesis, two experimental scenarios have been used to demonstrate the merit of the proposed computer vision methods. The choice of two scenarios is made due to the distinct visual characteristics of various imaging modalities. Given this, data acquisition for our experiments was done from two major European hospitals. A validation of the proposed methods should be done for both imaging modalities.

To identify future challenges in gastroenterology imaging diagnosis - In this thesis, our objective is to address the most generic set of challenges in gastroenterology imaging and as such, we expect that tackling these challenges will not give us the best performance for every imaging scenario. Handling more specific challenges for some specific imaging modality can improve the classification results further. Thus we make the final remarks about the strengths and shortcomings of our methods. This will help us highlight the challenges that need more research for a stronger impact of computer vision research on the design of CAD systems for gastroenterology in the future.

1.5 Contributions

The main contributions of this thesis are as follows:

- **Data acquisition** - Creation of two high quality datasets for computer vision research: one for vital stained magnification endoscopy for stomach and the another for narrow-band imaging endoscopy for esophagus. Annotation software was written for both these imaging modalities based on the requirement specifications provided by the physicians. All images in the dataset were manually classified, giving us a benchmark for performance evaluation of our methods.

- **Propose a novel image segmentation methodology** - This methodology integrates gradient based features in a global optimization framework of a popular segmentation method, normalized cuts, to give very good segmentation performance for images from both imaging modalities. The proposed methodology outperforms other popular state-of-the-art segmentation methods considered in this thesis.

- **Multiple image features for segmentation** - Usage of multiple complementary image features to further enhance the performance of segmentation methods on both imaging scenarios. We show that the novel methodology significantly outperforms other state-of-the-art methods and is generic enough to give consistent results on two complementary imaging scenarios.

- **Autocorrelation Gabor Features** (AGF) - Proposal of novel texture features based on Gabor filters, motivated by the need to extract rotation, scale and illumination changes in the images. A theoretical study of the properties of Gabor filters is done to propose novel invariant features. Empirical demonstration of the invariance of these features is provided.

- **Integrating AGF in a texton framework** - Use of the proposed texture features in a texton framework assuming redundancy in the filter banks. The proposed features are integrated in a bag-of-words approach to create histograms, that can be used directly for classification using feature based classifiers. These features have been used for classification of images from both imaging modalities giving superior classification performance as compared to the previously existing state-of-the-art methods.

- **Autocorrelation Homogeneous Texture** (AHT) - Proposal of another novel set of features based on the previously obtained texture features. An assumption

of homogeneousness of image regions done to obtain significant dimensionality reduction. Empirical demonstration of the invariance of these features is provided in the thesis. These features can be used directly for classification using feature based classifiers. Comparison of performance is done with its state-of-the-art counterpart showing its superior performance.

1.6 Thesis Organization

Chapter 1 describes the need for automated diagnosis systems for gastroenterology. In Chapter 2, we describe various imaging modalities that can be used for the screening of the GI tract. An identification of complementary challenges posed by these imaging modalities is also presented. In Chapter 3, we introduce a traditional pattern recognition system based on computer vision for design of CAD system for gastroenterology. We discuss the various mathematical methods used for the individual modules in the CAD system. In Chapter 4, we present the state-of-the-art in computer vision for gastroenterology and make critical remarks about existing computer vision methods for various gastroenterology imaging modalities. In Chapter 5, we discuss our methods of obtaining the datasets from two complementary imaging modalities from two different organs. In Chapter 6 and 7, we present our contributions and an experimental validation of the superiority of our proposed contributions is presented. These chapters introduce methods, which are proven to be complementary providing superior performance for two different imaging modalities. Chapter 8 concludes the thesis and identifies the various challenges and vital clues about the directions of future work for the design of CAD systems for gastroenterology.

Chapter 2

Gastroenterology Imaging

Digestion is the mechanical and chemical breakdown of food into smaller components, which are easily absorbed into the blood stream [128]. Digestive tract consists of a series of structures and organs through which food and liquids pass during their processing into decomposed forms absorbable into the bloodstream (Fig. 2.1). The digestive tract begins at the mouth and ends at the anus. It consists of the mouth or oral cavity, with its teeth for grinding the food and its tongue which serves to knead food and mix it with saliva; throat or pharynx; esophagus; stomach; small intestine, consisting of the duodenum, jejunum and ileum; and the large intestine consisting of the cecum, a closed-end sac connecting with the ileum, the ascending colon, the transverse colon, the descending colon, and the sigmoid colon which terminates in the rectum [128]. Glands contributing digestive juices include the salivary glands, gastric glands in the stomach lining, the pancreas, and the liver and its adjuncts - gallbladder and bile ducts. All of these organs and glands contribute to the physical and chemical breaking down of ingested food and to the eventual elimination of non-digestible wastes.

Sometimes a disruption in proper functionality of the gastrointestinal (GI) tract can cause several discomforts such as burning, vomiting, abdominal pain etc. Such symptoms can be related with abnormalities that need a proper evaluation by physicians [107]. For such cases, a procedure known as gastrointestinal endoscopy may be suggested. It is a diagnostic procedure which involves a visual examination of various organs in the GI tract that can help physicians in making the diagnosis. Over the past several years, research has led to significant technological advancements that have helped in evolution of the endoscopy examination. These have been motivated by the lack of a single procedure that can deliver all the required variables that can

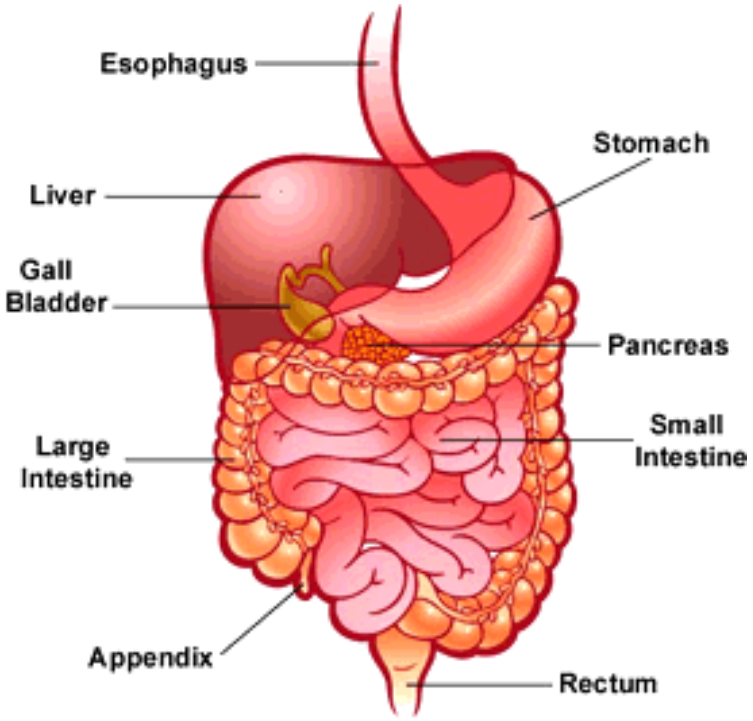


Figure 2.1: Gastrointestinal tract of a human (Reprinted from [33]).

assist the physicians in making their diagnosis. In this chapter, we will give a brief review of the conventional endoscopy examinations [63]. These examinations are very successful for making reliable diagnosis however they exhibit several shortcomings. These shortcomings led to a significant technological and procedural changes that include several other imaging modalities for conducting the exam, which are briefly discussed in this chapter. As a conclusion, we give some remarks about the imaging modalities discussed in this chapter.

2.1 Conventional Endoscopy

Conventional endoscopy uses white light for a visual examination of the GI tract [31]. The exam is performed using an endoscope.

2.1.1 The device: Endoscope

Endoscopy procedure is carried out using an endoscope (with the exception of capsule endoscopy). It is used for illumination of the internal part of the GI tract for

visualization. It consists of the following components [11]:

2.1.1.1 Endoscopic Probe

An endoscopic probe consists of a flexible tube having a light source, lens system and a separate channel for manipulation (Fig. 2.2). The flexible tube is inserted into the GI tract either through the mouth or the anus for visualization of the GI tract. It has a light source connected to its tip, that is used for illumination of the tissues in the internal body organs. The lens system is also connected to its tip, that helps in the transmission of the endoscopic video to the display device. There is a separate channel that allows the insertion of surgical instruments such as manipulators etc. These can be used to take some specimen of the tissue for laboratory examination (biopsy). There are separate air and water internal channels that are provided to clean the visualized tissue by removing the food traces that might block the view of the physician. The most important specifications of an endoscopic probe are [138]:

Figure 2.2: An endoscopic probe

- **Probe diameter** - It is one of the most important design parameters of an endoscope. Preferably, the diameter of the endoscopic tube should be small. This will allow easy navigation of the probe through the GI tract. The bigger the diameter of the probe, the higher the difficulty and discomfort in probe navigation.

- **Instrument channel diameter** - Another important specification is the diameter of the channel used for surgical instruments. The diameter of this channel should be small but the capability of inserting surgical instruments should not be compromised.

- **Angulation system** - All flexible digestive endoscopes have an angulation system guided by knobs on the control body. This system manipulates the bending section of the probe for insertion and viewing of the organ. The knobs are connected inside the control body housing a pulley-and-wire system or gear-and-chain system. These knobs control the direction of movement in the up-down and left-right direction providing enhanced manoeuvring capability for a wide angle of viewing.

- **Field of view** - The field of view is the area of inspection that is captured by the camera. The light emitted by the source of an endoscopic probe is expected to form a cone and therefore the field of view is usually represented as an angular size of the view cone.

2.1.1.2 Video processor

The endoscopic probe is connected to a video processor (Fig. 2.3) [11]. This video processor is responsible to direct the live endoscopic video to a display. This processor is also equipped with some controls that can help in enhancing visualization of the tissues e.g., changing the contrast, sharpness of images, electronic magnification etc. The processor has the video outputs in several formats such as HDTV, YCbCr etc. It is also connected to a video recorder that can record live videos on tapes during the endoscopic examinations. A pedal is also provided to the physician which can help in recording some clinically relevant images during the video examinations [53], that can be attached with the clinical report of the patient.

2.1.2 Procedure

Before undergoing the exam, the physician studies the patient file. It is one of the vital steps of the procedure because it is important to get acquainted with the patient history. Later, the exam is performed. During the exam, the expert physician inserts the endoscopic probe in the GI tract of the patient. The insertion can be done either through the mouth or the anus, depending on whether the physician wants to analyse

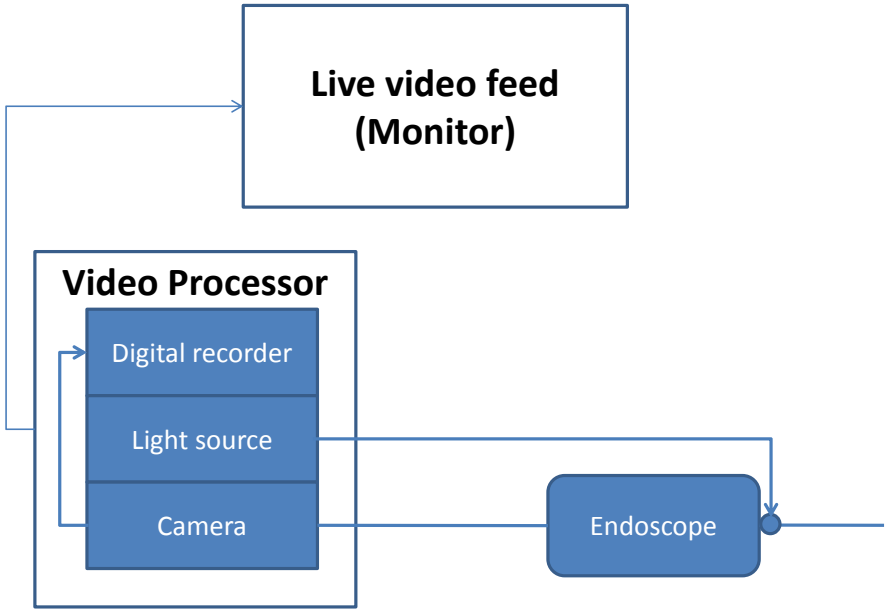


Figure 2.3: The video processor for a typical endoscope: It manages the light source, controls video acquisition, streaming to a device (monitor) and recording.

the upper or lower GI tract [9]. The tube is inserted fully and than withdrawn slowly while analysing the organ. A typical endoscopic image is shown in Fig. 2.4.

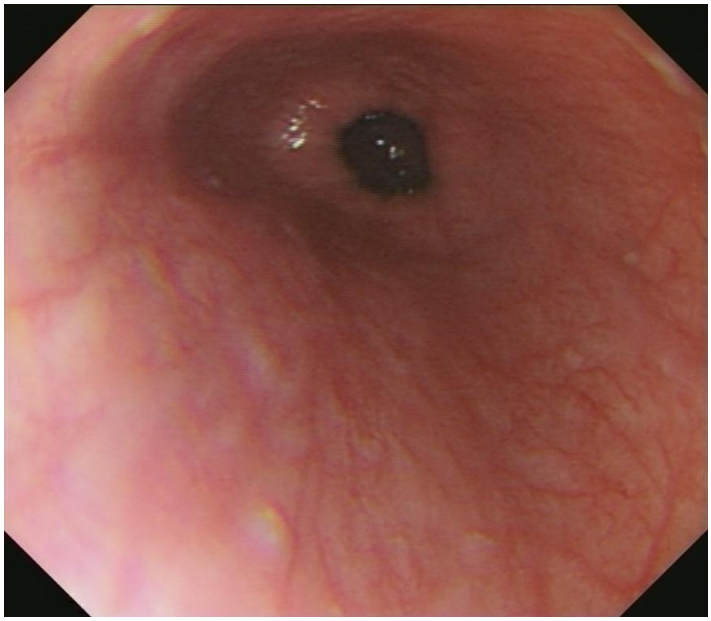


Figure 2.4: Visual illustration of a conventional endoscopy image.

Different instruments can be passed through the endoscope to directly treat many abnormalities causing little or no discomfort. For example, stretching a narrowed area, removing polyps (usually benign growths) or treat bleeding etc. Some tissue samples (biopsies) can also be removed for a detailed microscopic laboratory examination. A biopsy helps in distinguishing between benign and malignant (cancerous) tissues. Biopsies are taken for many reasons, and the physician might order one even if he does not suspect cancer.

2.1.3 Diagnosis

In conventional endoscopy, the diagnosis is done by the physician during the endoscopy procedure. During the exam, they record the images that were considered as exhibiting some clinical information, that was relevant for making their diagnosis. At the end of the procedure, the physician writes a report about the exam and if necessary, attaches the captured image for an effective follow-up of the patient.

2.1.4 Advantages of conventional endoscopy

Most of the advantages of conventional endoscopy are related with its accurate diagnostic capabilities. One of its most important advantages is that it allows detailed visualization of internal walls of the GI tract. Also, it is easy to take tissue samples from the GI tract. This results in an accurate diagnosis as the biopsies allow a microscopic analysis of the suspected tissue. Unlike some other imaging modalities such as capsule endoscopy [101] (discussed in detail later) in which the device is driven by the bowel/gastric movements, the navigation of the probe in the GI tract for conventional endoscopy is controlled manually. Therefore, the physician can focus on a particular site for as long as he wants. Also, the use of a separate water/air channel can be used to remove or disperse food traces that can lead to obstruction in the view of the physician. All these factors results in a clear inspection of the tissue thus making the diagnosis reliable.

2.1.5 Limitations of conventional endoscopy

One of the limitations of conventional endoscopy is that it is invasive. The insertion of probe in the GI tract can cause discomfort to the patient. Although it is a very safe

procedure, it carries a slight risk of some complications such as infection, perforation of organ, bleeding etc. which can require further treatment. Conventional endoscopic diagnosis using white light is based on morphological changes such as superficially elevated, flat, or depressed lesions, minimal changes in color etc. [46]. However, these findings are difficult to recognize, especially for inexperienced endoscopists. Often the issue is the detection of minute lesions that often do not stand out from surrounding tissue with standard available techniques (Fig. 2.5). As a result, the diagnosis may be inaccurate or a superficial cancer in the gastrointestinal (GI) tract may be overlooked. Furthermore, conventional endoscopy does not allow a detailed inspection of the mucosa on the inner lining of the organs. Research indicates that the inspection of structure of the mucosa can give indicators resulting in the early detection of various chronic diseases such as cancer [18]. Unfortunately, conventional endoscopy is not adequate for this purpose and therefore additional techniques must be used to improve their detection rate.

2.2 Magnification Endoscopy

There are two different ways to overcome the shortcoming of conventional endoscopy.

- High magnification endoscopy.
- High definition endoscopy.

High magnification endoscopy - Magnification endoscopes are characterized by their capacity to perform an optical zoom of the image by using a movable lens in the tip of the endoscope. A translucent cap on the tip of the endoscope may be used to stabilize the focal length between the lens and the target tissue to improve the image quality. Optical zoom obtains a closer visualization of the tissue while maintaining the image resolution. This gives a very detailed visualization of the tissue thus helping in an observation of the mucosal surface of the organs. The mucosa in contact with the lens is magnified without impairing the manoeuvrability of the scope. Degrees of magnification range from $\times 1.5$ to $\times 115$ and can be changed on the scope by turning a dial at the hand controls [111]. This examination has been reported to enhance the detection of abnormalities in the GI tract. Currently, the support for magnification endoscopy is usually being provided with conventional endoscopes. The controls with the probe have a knob which can be used for switching between normal and magnified view of the suspected tissue.

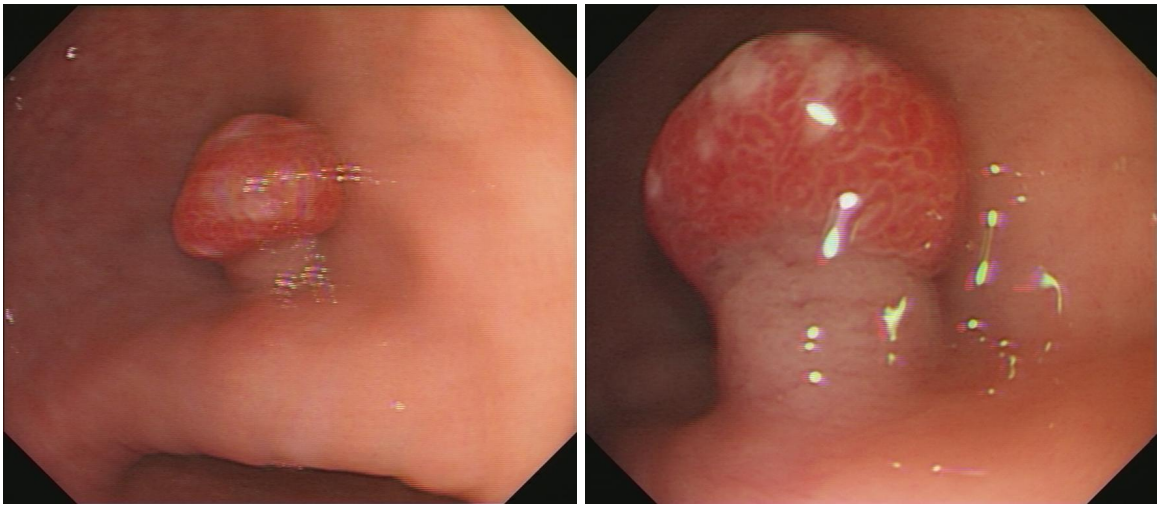


Figure 2.5: Difference between images from conventional (left) and magnification endoscopy (right). The Magnified image shows a better visualization of the texture of the polyp and thus is more useful for diagnostic purposes.

High resolution endoscopy - High resolution imaging improves the ability of the observer to discriminate the details in the image. High resolution endoscopes provide more detailed visualization of the tissue at the same magnification level: magnification is related with optical zooming and thus, image enlargement. High resolution is related with the capability of the video processor unit attached to the endoscopic probe. The technological advancement over the past few years has resulted in miniaturized video chips with much higher capabilities related with the number of pixels per unit area used to represent the images. This has laid the foundation of high resolution endoscopy which gives a very detailed view of the images. Modern conventional endoscopes typically provide both high magnification and high resolution endoscopy.

2.2.1 Advantages of Magnification Endoscopy

As discussed previously, conventional endoscopy is limited for visualizing only superficial morphological abnormalities in the organs. For an enhanced view of the suspected tissue, magnification endoscopy is used [18]. It gives a much more detailed analysis of the internal walls of the organs. This helps in an inspection of their mucosal layer. It is particularly useful for early detection of cancer. This is because the evolution of cancer starts by changes in the mucosal surface which are not easily visualizable using conventional endoscopy [61]. It later penetrates the walls of the organs thus going deeper into the muscular layers of the organs and starts to invade the

neighbouring organs. An early detection of cancer with high reliability is possible using high magnification endoscopy which is otherwise hard using conventional endoscopy.

The functionality of magnification endoscopy is usually provided with conventional endoscopes therefore the procedural disadvantages for both the imaging modalities related with the discomfort to the patient and the possible health risks are the same.

2.2.2 Limitations of magnification endoscopy

Although magnification endoscopy can be used for a detailed examination of the GI tract, it is not adequate for diagnosis. This is because the lesions must be detected first by conventional endoscopy before they can be examined under magnification [15]. Therefore, the role of magnification is to magnify the target area in which the physician detects some abnormality using conventional endoscopy.

2.3 Chromoendoscopy (CH)

Conventional endoscopic images prolong the clinical observation using white light. Human eye is particularly sensitive to color, shape and sizes of objects. It is therefore reasonable to emphasize on two important characteristics of gastrointestinal mucosa: colors and shape. This led to a significant interest in the application of dyes, which react with the internal walls of the GI tract and improve the contrast in mucosal surfaces [111]. This procedure that combines endoscopy with the application of dyes to enhance the contrast in mucosal surfaces is known as CH. Additionally, combining the functionality of magnification and high resolution endoscopy in conjunction with the application of dyes achieves enhancement of details that are not possible using purely optical methods. CH is used in many different settings, but the most common and important use is in the diagnosis of precancerous lesions or early cancer. This technique helps identify areas that can be targeted for biopsy or treatment. Different stains highlight areas of cellular change that can then be targeted for biopsy (Fig. 2.6). CH has been applied in a variety of clinical settings throughout the gastrointestinal tract.

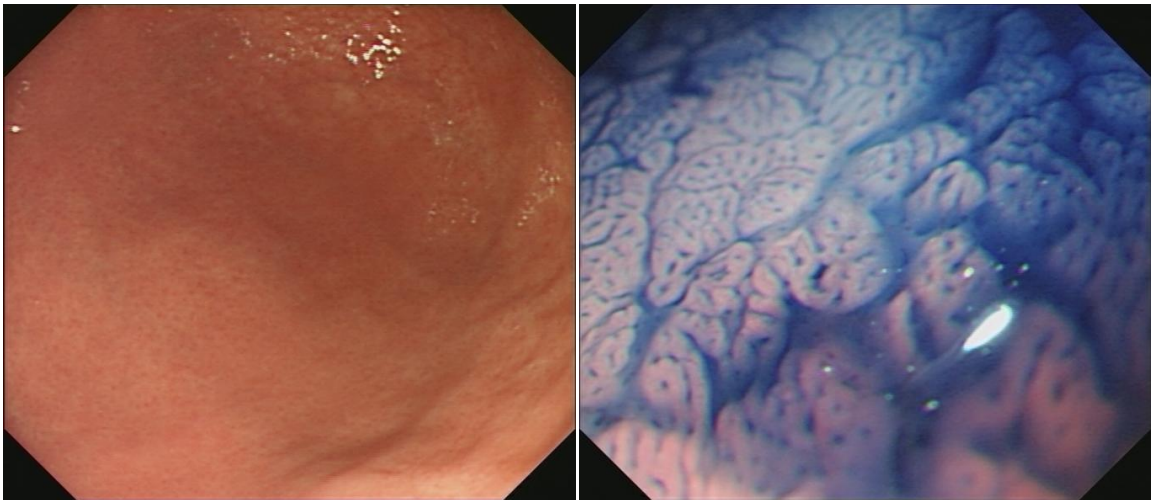


Figure 2.6: Difference between magnification endoscopy (Left) and chromoendoscopy (Right). Detailed visualization of mucosa is observed using chromoendoscopy

2.3.1 Procedure

During the endoscopic examinations, a dye is used to stain the tissue which is being visualized by the physician. The application of dye is done using the water channels available in conventional endoscopes. Dying agents used for CH are categorized according to their working principle. Vital stains like Lugol's solution and methylene blue are absorbed [23]. Contrast stains like indigo carmine are not absorbed but accumulate in pits and valleys between cells highlighting mucosal architecture [18]. Reactive stains like Congo red and phenol red react to changing conditions of acid secretion and carry a potential with regard to the early detection of gastric cancer and Helicobacter pylori infection [22]. The principle equipment required for conducting a CH exam is the same as a conventional endoscope with the additional need of a relevant dye that should be used during the procedure.

2.3.2 Advantages of chromoendoscopy

The most relevant advantage of CH over conventional endoscopy examination is an enhanced visualization of the GI tract. The enhancement is achieved by increasing the contrast of the mucosal surface, that is not an optical characteristic [111]. Several studies have indicated a higher sensitivity for detection of lesions using CH with magnification features [68]. The enhanced inspection is simple and assumes a very low cost. The procedure requires only an additional dye which is inexpensive and widely

available. CH exhibits the same disadvantages as that of conventional endoscopy.

2.3.3 Limitations of chromoendoscopy

Although chromoendoscopy enhances the mucosal patterns, it has several practical disadvantages. One of them is that the die which is sprayed can be easily displaced by lavage or diluted by intestinal secretions [137]. Also, chromoendoscopy is a cumbersome procedure and it requires more time than an endoscopic exam using careful inspection of the mucosa. In addition, spraying the dye solution occasionally causes agitation or heartburn and can also lead to a risk of allergic reaction [52].

2.4 Narrow-Band Imaging (NBI)

Different variant of white light endoscopy as discussed previously are widely used for a detailed visual inspection of the GI tract: they all share a shortcoming. They help in enhancement of the mucosal patterns, but the visual clues about the capillaries and vasculature is presented by none of the above mentioned imaging modalities. There are several inflammatory disorders that result in disorganization of the vascular patterns in the images. Visual detection of such disorders is usually not possible using white light endoscopy. It is therefore necessary to use some optical characteristics to enhance the contrast of vasculature in the images for their clear visualization.

NBI (Olympus Medical Systems, Tokyo, Japan) was developed primarily to emphasize the mucosal microvasculature and to identify vascular alterations indicative of various pathologic conditions. The technology consists of placing narrow bandpass filters in front of a conventional white-light source to obtain tissue illumination at selected, narrow wavelength bands (Fig. 2.7). These bands produce high contrast between vascular structures and the surrounding mucosa. The narrow band filters were selected on the basis of studies that determined a set of filters that achieved the preferred appearance for mucosal vascular patterns.

Using NBI, the tissue surface is illuminated with light of a narrow bandwidth with center wavelengths in the (415 nm) and green (540 nm) of the visible spectrum. Structures with a high haemoglobin content (i.e. blood vessels) strongly absorb these wavelengths making the vascular structure appear dark on a relatively brighter mucosal background thus creating high contrast images [24].

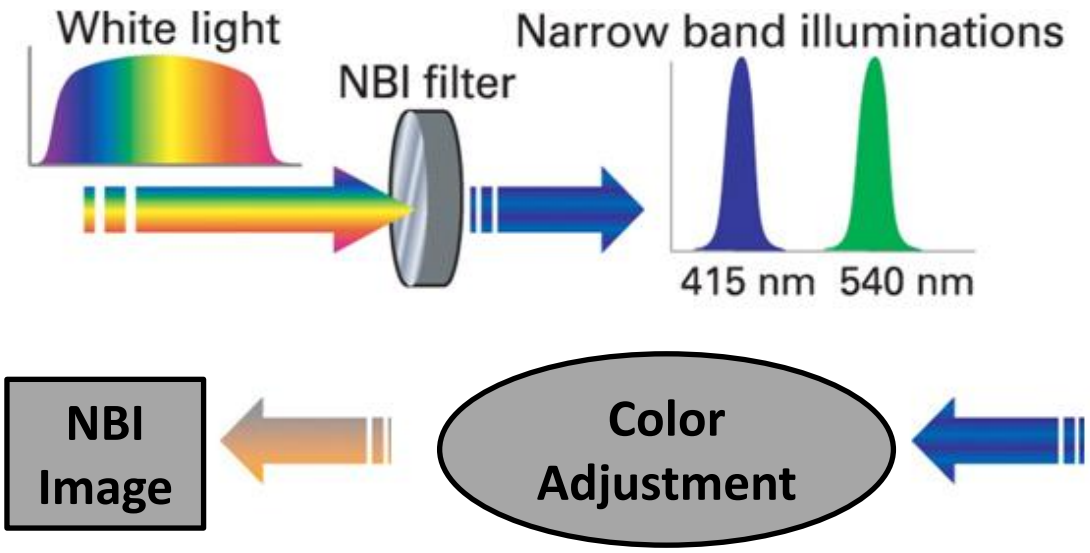


Figure 2.7: Generation of the NBI images, motivated by the penetration depth of shorter- and longer- wavelength light in the visible spectrum when incident on the tissue (adapted from British Medical Journal).

Capillaries in the superficial mucosal layer are emphasized by the 415-nm light and are displayed in brown, whereas deeper mucosal and submucosal vessels are made visible by the 540-nm light and are displayed in cyan [46]. Commercially available videoendoscope systems enable the user to alternate rapidly between white-light and NBI viewing modes by the touch of a button on the handle of the endoscope, front panel of the light source, or computer keyboard.

2.4.1 Applications

Classification of NBI mucosal patterns has been described for various conditions (e.g. Barretts esophagus and colon polyps). NBI generates a darker field of view than its white-light counterpart (Fig. 2.8). Consequently, to allow adequate inspection of the mucosal surface, the tip of the endoscope needs to be closer to the mucosa than is necessary with white-light imaging. Moreover, the presence of bile and blood (e.g., after biopsy) obscures the view under NBI because these fluids strongly absorb the narrow band light. NBI appears most useful when combined with zoom magnification (80x-150x) for targeted, detailed inspection of mucosal lesions [145]. For this purpose, a disposable plastic cap at the endoscopes tip is often used for image stabilization when the endoscope is apposed against the target tissue.

2.4.2 Advantages of NBI endoscopy

Compared with CH, the NBI observation has the advantage of convenient application without the necessity of dye spraying, thus the procedure can be shortened in time and an overlooked lesion with accumulation of dark-blue dye at the dependent portion of colon can also be avoided [145]. According to previous pathological studies, benign lesions have different features in terms of vascular architecture including vessel diameter and spatial distribution as compared to that of cancerous lesions in the colon [145]. Therefore, NBI endoscopy has a significant advantage in the diagnosis of colon cancer. In recent years, a number of researches have shown that the diagnostic accuracy of NBI endoscopy in distinguishing pre-cancerous and cancer lesions was higher than that of conventional colonoscopy and equivalent to CH.

2.4.3 Limitations of NBI endoscopy

Several studies have reported the disadvantages of NBI. One of the limitations is the non-standardization of NBI systems and prototypes used in research. In practice, the latest technological advancements in NBI shows its clear advantage over conventional endoscopy however due to the fast technological evolution in endoscopic imaging, there is an increasing need to train endoscopists in the basic principles and applications of these advanced technologies [41]. The interpretation of the contrast-enhanced images

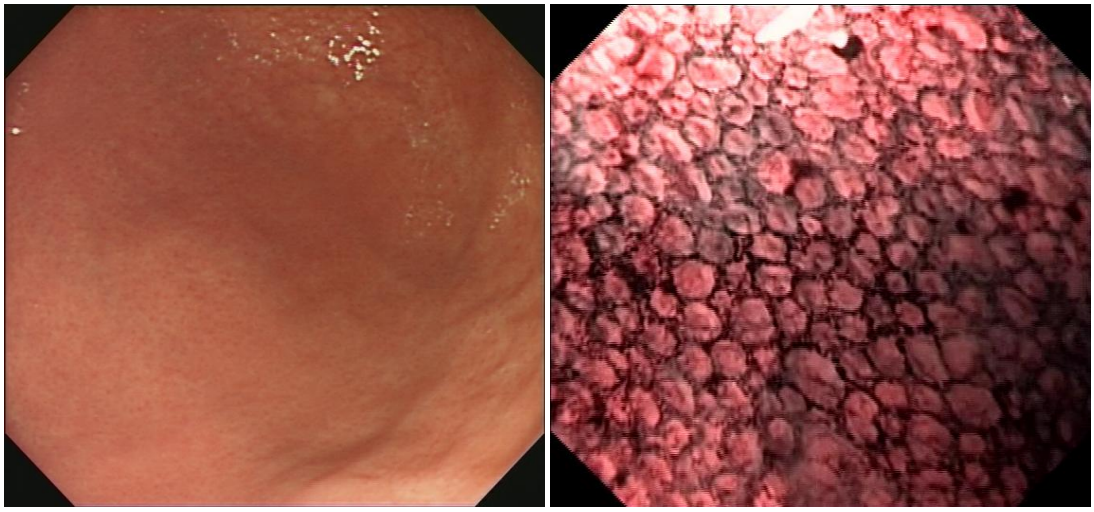


Figure 2.8: A visual demonstration of a white light endoscopy image (Left)vs. NBI image (Right). High contrast resulting in enhanced visibility of capillaries is observed for NBI image.

thus requires familiarity, may not always be straightforward, may increase procedure time, and is subject to observer variation.

2.5 Capsule Endoscopy (CE)

Since scope tests were first invented, clinicians have wished to be able to visualize the entire gut - all 9 metres. But, a direct view of the small intestine has remained elusive. Attempts have been made to develop longer endoscopic instruments. This technique called push enteroscopy has had only limited success. The longer instruments are difficult to control and manipulate and are hard to maintain. The accuracy of push enteroscopy is still limited since even in the best of hands the entire small intestine is not visualized. Modern endoscopic techniques have revolutionized the diagnosis and treatment of diseases of the upper gastrointestinal tract (esophagus, stomach, and duodenum) and the colon. The last remaining frontier has been the small intestine. The small intestine has been a difficult organ to make diagnoses and treat without performing surgery.

With wireless CE we can provide a simple, safe, non-invasive, reliable procedure, well accepted and tolerated by the patient, which has revolutionized the study of the small bowel. This technique evaluates endoscopically, the whole small bowel, avoiding any sedation, surgery or radiation exposure. Currently, CE is recommended as a third stage examination, after negative gastroscopy and colonoscopy in patients with obscure gastrointestinal bleeding. Also many studies have established, with a growing body of evidence, that this technique is cost-effective in other clinical situations, such as detection of small bowel lesions in Crohns disease in patients in which other methods have failed to provide a diagnosis, non steroidal anti-inflammatory drug enteropathies, celiac disease, small bowel polyposis syndromes and small bowel tumors etc [101].

2.5.1 The Device

The capsule endoscope is a disposable, small, swallowable, wireless, miniature camera which allows us to get a direct visualization of the gastrointestinal mucosa. The capsule which measures only 11 mm x 26 mm and weighs 3.7 g, holds a metal oxide semiconductor imaging-chip video camera, 6 white light-emitting diode illumination sources, 2 silver-oxide batteries and a radio telemetry transmitter. The image field is 140 degrees, magnification is x 8 and the depth of view is 1 to 30 mm [101]. The capsule

contains (Fig. 2.9):a camera, light emitting diodes, batteries and a radio transmitter. Before the capsule is swallowed, 8 skin antennas are taped to the patient's anterior abdominal wall and connected to the hard drive.

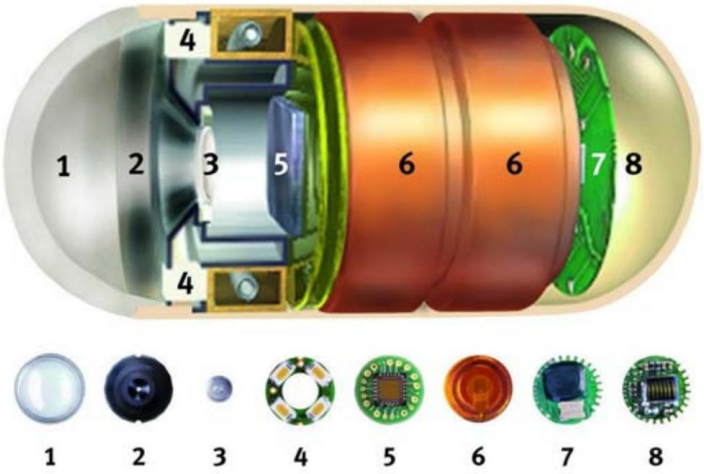


Figure 2.9: The wireless endoscopic capsule (1 - Optical dome; 2 - Lens holder; 3 - Lens; 4 - Illuminating LEDs; 5 - CMOS imager; 6 - Battery; 7 - ASIC transmitter; 8 - Antennas)

2.5.2 Procedure

After an overnight fast, the patient swallows the capsule with a few sips of water, then the capsule is passively moved along by peristalsis. Two hours after ingestion, the patient is allowed to drink, while eating is allowed after 4 h. During the procedure the patient may carry on with his daily activities. The camera is activated by removal of the capsule from its magnetic holder and takes 2 images per second and transmits these by means of radio frequency to a sensor array placed on the patients abdomen and from here to a recording device in a belt that the patient wears for the duration of the battery life (8 h) [101]. The use of the real time viewer may shorten procedures, as the patient can be disconnected once the cecum is visualized.

CE is usually performed as an outpatient procedure. The presence of intestinal contents or a motility disorder may cause the incomplete visualization of the intestinal mucosa. Several studies have examined the possibilities of improving bowel cleanliness and shortening transit time by means of different medications and different fasting periods. At present CE has some technical limitations: it cannot be used to obtain biopsy specimens or for endoscopic treatment and it cannot be controlled remotely.

CE has also some clinical limitations which are problems in sizing and locating small bowel lesions.

2.5.3 Offline analysis of CE images

After about 8 h, the sensor array and recorded data are removed and the recorded images are downloaded to the computer. It takes on average 40-60 min to read these images and since it is very time-consuming, one possible cost-effective strategy could be the use of expert nurse endoscopists to select images. Some studies have shown that highly motivated nurses can detect clinically significant lesions at a similar rate to physicians. Since its development, additional support systems have been added to the software to assist the reader, such as localization capability, suspected blood indicator, a multiviewing feature and quick view modality. The capsule is excreted with the feces, usually within 24 to 48 h [101]. Currently, there are several prototypes available for automated analysis of capsule endoscopy images. They have been used to do tasks such as topographic segmentation of images, detection of bleeding etc. Some sample CE images are shown in Fig. 2.10.

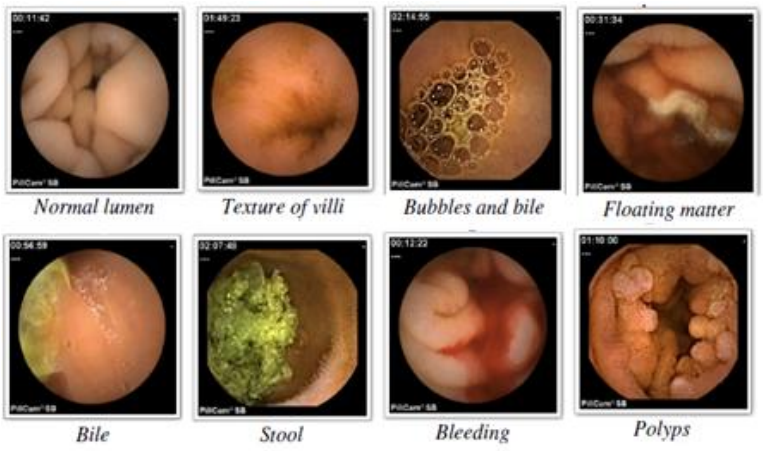


Figure 2.10: Typical capsule endoscopy images giving examples of abnormalities or obstructive visualization of the GI tissues (adapted from [10]).

2.5.4 Advantages of capsule endoscopy

Capsule endoscopy has some advantages over conventional endoscopic examinations. These mainly include the fact that it helps the physicians to visualize the small

intestine which is almost not possible using conventional procedures. It is non-invasive and very convenient as it allows the patients to continue their normal activities during the day. On the other hand, it enables the patients to forego the overhead of usually long waiting times for scheduling conventional endoscopic examinations.

2.5.5 Limitations of capsule endoscopy

Although CE has many advantages, it has numerous drawbacks. Capsule is a small device and therefore, it has limited source of power. In addition, it has limited storage and therefore high resolution images which may be required to analyze the gastric mucosa cannot be obtained using these. On the other hand, capsule cannot exclusively replace the other procedures available currently for analysis of small intestine. Since capsule has no therapeutic capabilities, any lesion discovered by capsule endoscopy must be investigated by other techniques. In addition, the use of capsule is contraindicated in patients with cardiac pacemakers, defibrillators or implanted electromechanical devices (due to the radio interference with the UHF signal) [100]. Last but not the least, the study of a capsule endoscopy video by physicians takes a long time which is stimulating significant amount of research being done on automated analysis of capsule endoscopy videos.

2.6 Virtual endoscopy

Endoscopy is uncomfortable for the patient and sedation and anesthesia may be necessary. Furthermore, it involves a degree of risk for the patient since it can cause perforation, infection and hemorrhage. Virtual endoscopy is a promising technique to explore hollow organs and anatomical cavities using 3D medical imaging and computer graphics (i.e., medical computer visualization techniques). The fields of application of virtual endoscopy are many such as:

- Non-invasive diagnostic endoscopy.
- Virtual endoscopy can be used for educational purposes such as endoscopist's training.
- The usage of virtual endoscopy for surgery planning.

2.6.1 A typical virtual endoscopy system

Virtual endoscopy simulates an endoscopic view by generating a virtual walk through of the organs and therefore help in visualization within the walls and beyond the walls of the organs. Virtual endoscopy usually involves a number of important steps [135]. First, the system is given input data which is a 3D image data set. The data is usually acquired by one of the medical imaging modalities (CT, MRI, etc). Several pre-processing techniques are used to enhance the data thus giving the enhanced volume data. Afterwards, segmentation of objects in 3D imaging dataset is being done. This is necessary to define the object in the whole volume data in which the user is most interested (e.g. colon, trachea).

Segmentation is followed by the navigation module which is concerned with the on-the-fly inspection of narrow tubular structures typical in endoscopy. It includes the interaction of the user to control camera movement, and deals with mapping the input device movements to camera parameter modifications. The user should get neither a lost-in-space feeling because of dealing with too many parameters nor a frustration feeling due to a heavily constrained navigation environment. A near real-time frame rate is also an important goal for a feasible navigation. The navigation module provides the data for the rendering module and also the camera parameters like camera position and orientation.

Once the camera position is determined and the data is prepared, rendering has to be performed. The results of rendering are later shown in an output device (Fig. 2.11). For user interaction with the virtual endoscopy systems, several input devices such as mouse, keyboard etc are being provided. These devices help the user to navigate through the 3D model of the object / organ created during virtual endoscopy.

2.6.2 Advantages of virtual endoscopy

One of the main advantages of virtual endoscopy is that it is non-invasive. Unlike conventional endoscopy which involves the insertion of a tube into the gastrointestinal tract which causes uneasy feeling for the patients, virtual endoscopy is a radiological method therefore it is much more convenient for the patient. Also, no sedation of required for performing the procedure. In this procedure, the entire organ can be visualized and the position of various lesions can be localized precisely (Fig. 2.12). Also, some parts of the human body which are impossible to access with a real endoscope (e.g., blood vessels, thoracic aorta) can be investigated with virtual endoscopy.

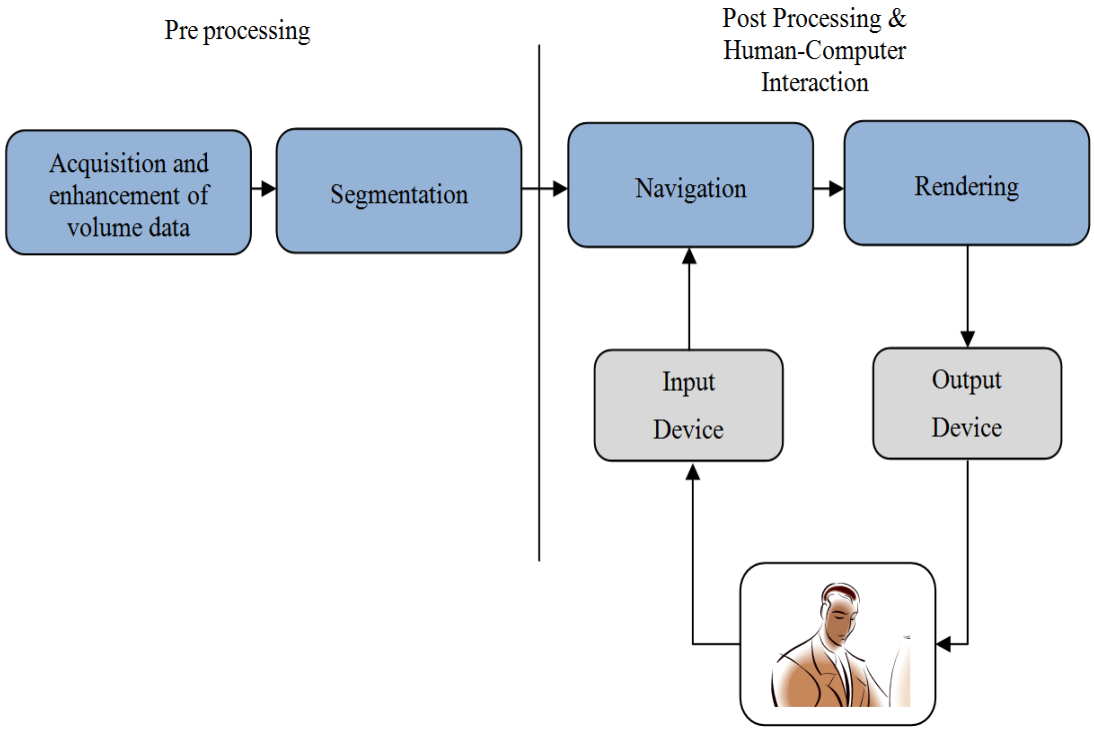
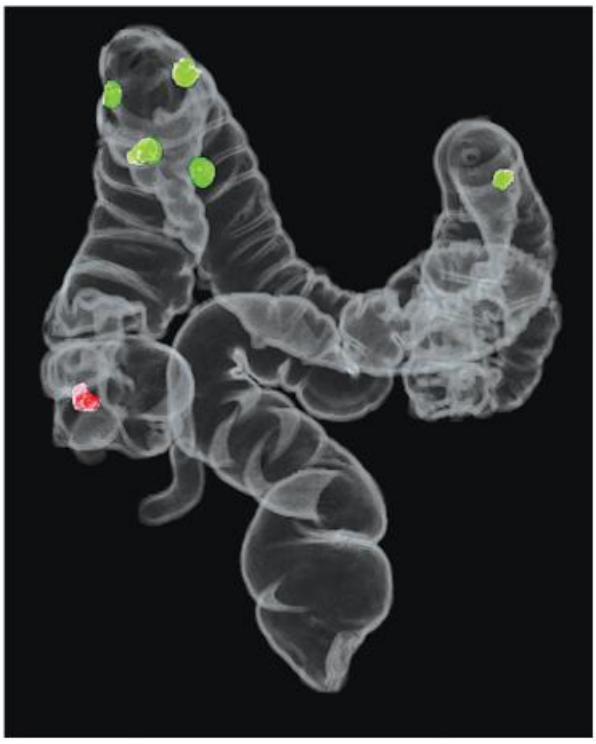


Figure 2.11: A typical virtual endoscopy system. Preprocessing generates a model of the organs while human-computer interaction and post processing are used for interactive visualization of the endoscopic site.

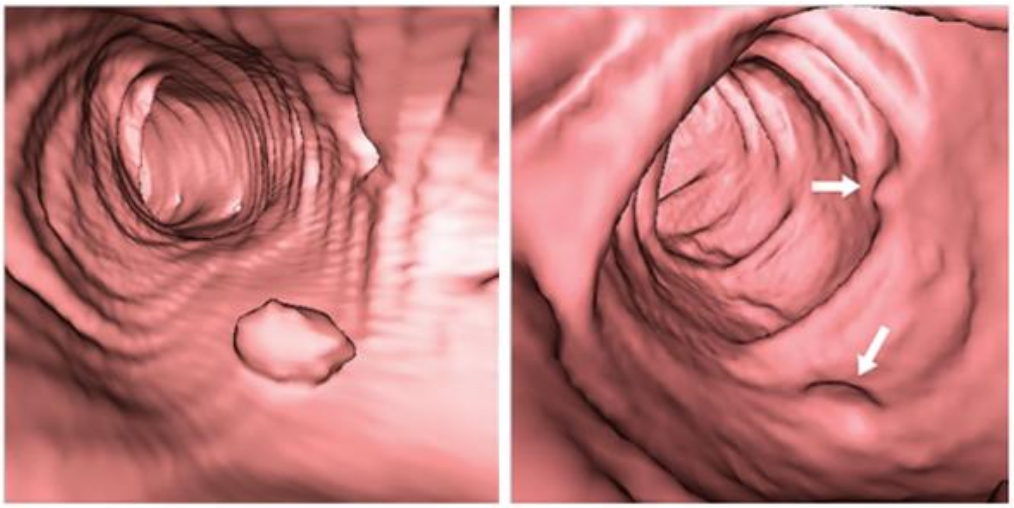
Endoscopes display only the inner surface of hollow organs and yield no information about the anatomy within or beyond the wall which can be visualized in virtual endoscopy. A virtual walk through can then be very targeted to analyse the locations of the lesions and therefore their size can be determined. Also, it is a fast method with sensitivity equal to that of that of colonoscopy. Last but not the least, it is technically less demanding. The conventional endoscopy procedures are usually performed by expert endoscopists. A clear understanding of the anatomy of the visualized tissues is necessary in such examinations. In virtual colonoscopy however, much of the work is computerized and therefore it is technically less demanding.

2.6.3 Disadvantages of virtual endoscopy

One of the disadvantage of virtual endoscopy is that it results in exposure of the patient to radiations. Therefore, it cannot be excessively performed and its use has to be limited for a particular patient. Another important fact is that it is non-



(a) Transparent 3D model of the colon generated using virtual endoscopy. Green and red spots represent the locations of polyps and masses respectively



(b) Virtual walkthrough of the colon with visualization of the polyps

Figure 2.12: Typical images generated as a result of virtual colonoscopy [36] (adapted from GE medical systems)

invasive and thus does not allow the physicians to take biopsy specimens. Usually the diagnosis is confirmed by biopsies therefore virtual colonoscopy cannot be exclusively

used for confirmed diagnosis. It also has some limitations on the minimum size of detectable polyps in the GI tract, which means that polyps smaller than that will be left undetected. Also, wrong illusions about polyps can be created by the remainder faeces. Last but not the least, virtual endoscopy cannot show texture and color details of the mucosa [139]. This is a serious shortcoming because the texture and structure of the mucosa are vital for making a diagnosis of various pathologies found in the GI tract. This is why, virtual endoscopy is most efficient and most effective to localize and detect polyps in the GI tract and not effective to visualize the texture of the internal walls of the GI tract.

2.7 Scope of this thesis

As discussed previously, there are several imaging modalities that can be used to diagnose the pathologies in the GI tract. Their optimal selection depends on the pathologies for which the physicians wish to screen the patient, or the site that has to be visualized by the physician. Also they present complementary information and therefore from a computer vision perspective, they present complementary challenges.

Conventional endoscopy and its variant are very reliable but it suffers lack of access to the small intestine. This leaves the physicians with not many options, but to screen the patients using endoscopic capsule for diagnosis. It however lacks a clear visual inspection of the organ. This is primarily due to the technological limitations for the miniaturized device that enforces capturing of low resolution video. In addition since it is navigated by the intestinal motility, it can result in some unexpected events such as very fast transition and the obstruction of the view of the tissue due to the traces of food. However, since there is no known method that can screen the small intestine, the physicians are limited to the option of an endoscopic capsule for patient screening.

Screening of the GI tract other than the small intestine is much more convenient using variants of conventional endoscopy. Their use is motivated by the nature of abnormalities for which the patient needs to be screened. Conventional endoscopy, is used to screen the patients for superficial morphological disorders such as elevations (polyps), depressions, blood detection etc. which are composed of the most basic visual primitives in computer vision. Magnification endoscopy enhances the view of the internal walls helping in a more clear visualization of the abnormalities and helping in their quantitative assessment such as size of very small polyps etc. which may not be visible using conventional endoscopy. Chromoendoscopy presents a different challenge

assuming that the human visual perception is sensitive to colors, shape and sizes of the objects. They correspondingly enhance the contrast in images by using dyes that stain the tissue with color, enhancing the mucosal structure exposing their shape, regularity and sizes. NBI presents a different challenge enhancing the capillary networks in the tissues.

It is vital to mention that there is one visual feature, which is common to the images from all these imaging modalities (with the exception of virtual endoscopy): *texture*. It is therefore expected that while designing the computer vision module in a CAD system for gastroenterology, image texture can be considered as one of the most generic image descriptors which can be useful for images from a wide range of imaging modalities. Following these guidelines, our aim in this thesis is to incorporate the most generic requirements of computer vision for CAD systems in gastroenterology in a pattern recognition framework, that can be useful in future for automated diagnosis for a wide range of imaging modalities.

Chapter 3

Background on Pattern Recognition

There are several ways in which computer assisted decision (CAD) systems can be useful such as simplify access to data, patient information systems, assisted decision making etc. In gastroenterology (GE), one of the most important challenges in the design of a CAD system is related with pattern recognition. Conventionally, a clinician observes the internal walls of the gastrointestinal (GI) organs to have a detailed look at the tissue. Specific observations related with color, texture, shape and regularity of the patterns are the main attributes which provide the clinicians with some vital information about the health of the tissue. Intuitively from a computer vision perspective, this manual diagnostic procedure involves pattern recognition (PR). Our design of CAD systems for gastroenterology (Fig. 3.1) envisions that there are certain steps involved in this whole process.

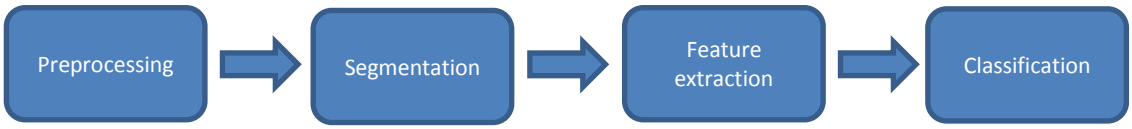


Figure 3.1: A traditional pattern recognition system

It starts with preprocessing since the data which comes from the sensor might contain noise which needs to be filtered. This task is accomplished by the preprocessing module in the traditional PR system. A single observation (i.e. an image) might consist of a combination of patterns, which might or might not be recognizable using

the trained classifiers. A mechanism for dividing an observation into smaller subsets (i.e. image segments) that represent individual patterns inside an image is thus highly desirable. The need for such a mechanism in GE imaging arises from the manual diagnosis procedures followed by the physician where they try to localize the abnormalities in the GI tract and visualize the suspected tissue for the presence of any lesions. The physicians are therefore looking for a specific region that is representing some clinical information in a particular image. The part of the image which is not clinically relevant is redundant and ignored by the physician. Typically, images having some clinical importance for diagnosis are captured and a sample of the suspected clinically relevant tissue in those images is taken for laboratory examinations. Intuitively for automated diagnosis, sub dividing an image into various image segments is inevitable. The segmentation layer addresses this objective by dividing an image into visually coherent regions, which hopefully map into different body tissues and thus communicate different clinical semantics. A good performance of this stage means that the following modules in the PR system get only clinically relevant information and redundant data is discarded. This ensures that information about a single pattern (and not a mixture of them) is provided to the later stages thus improving our chances of correct classifications.

The segmented parts from the input image have to be used now for feature extraction. Based on the objective of the classification task, the data from the sensors might not be useful directly for the final classification. This is because the original representation of the data may not be optimal representing the characteristics of the data and a transformation to a feature space is usually required. This feature space is designed to ensure that the most important characteristics of the data are well segregated in the feature space. In a traditional pattern recognition system for gastroenterology, feature extraction makes use of the most prominent set of visual features which are able to correctly describe the characteristics of the content of the images. If the system is able to obtain very good and highly discriminative features, the task of the subsequent stage in the traditional pattern recognition system i.e., classification is considerably simplified. The task of the classification is to assign the feature vector provided by the feature extractor to a class. The output of the classifier is typically a discrete selection of one of the predefined classes. All the preceding components of a pattern recognition system are designed and tuned for improving the performance of the classifier. The degree of difficulty of the classification depends on the similarity relations between the patterns of different classes. Therefore, its success is significantly affected by the feature extraction stage.

In this chapter, we will revisit the state-of-the-art methods used for segmentation, feature extraction and classification that will help us in motivating our choices for the design of a CAD system for gastroenterology.

3.1 Segmentation

Prior knowledge on image segmentation is vast. It can be viewed as a perceptual grouping problem where the images have to be divided into different homogeneous regions, which depending on the application can represent different features in images. Many different methods to perform image segmentation have been proposed in literature and it is very hard to find the most suitable method for a certain scenario. Conceptually, image segmentation methods can be mainly divided into two different categories according to Forsyth and Ponce [47]:

- **Segmentation by clustering** - Determine which components of a data naturally belong together according to some criterion. The components that belong together form a cluster.
- **Segmentation by fitting** - This view is similar to clustering; the main difference being that the model is explicit (e.g. 'look like' a line) and involves relations at a larger scale.

3.1.1 Segmentation by clustering

Clustering is a process whereby a dataset is replaced by clusters, which are collections of data points that belong together. It is natural to think of image segmentation as clustering; an image can be presented in terms of clusters of pixels that belong together. The specific criterion of clustering depends on the application. Pixels may belong together because they have the same color, same texture, they are nearby and so on. There are two natural approaches for clustering (Fig. 3.2):

- Divisive clustering - In divisive clustering, the entire data is regarded as a single cluster. This cluster is then recursively split up to yield a good clustering.
- Agglomerative clustering - In agglomerative clustering, each data item is initially regarded as an individual cluster. These clusters are than recursively merged to yield a good segmentation.

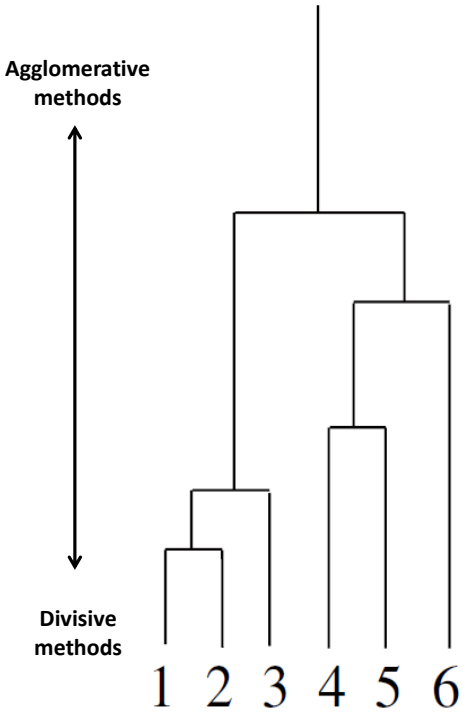


Figure 3.2: Dendogram representing types of segmentation methods ([47])

There are two main challenges when thinking about segmentation by clustering. The first major challenge is: How many clusters are there? This is an intrinsically difficult task if there is no model for the process that generated the clusters. Usually a hierarchy of the clusters is generated which is displayed to the user and an appropriate choice is made. The second major challenge is: What is good inter-cluster distance? Agglomerative algorithms use the intercluster distance to fuse the nearby neighbors. Divisive clustering uses it to split insufficiently coherent clusters. Even if a natural distance between data points is available, there is no canonical intercluster distance. Generally the one which is suitable for the dataset and feasible for the application is chosen.

Different clustering methods can be used for building image segmentation frameworks. In most of these methods, the outcome of segmentation depends on the measures of distance which have been used for inter-cluster distance. It can be based on color, texture, intensity based on the application. The main difficulty in using either agglomerative or divisive clustering methods directly is that there are a lot of pixels in an image. This mainly raises two important problems for image segmentation.

- When the data is huge, there is no reasonable prospect of examining the dendogram. This is because, a lot of data means that the dendogram is too big and effectively this means that the segmenter will decide when to stop splitting or merging by using a set of threshold tests. For example, an agglomerative segmenter may stop to merge the clusters when the distance between them is sufficiently small, may be less than a threshold or when the number of clusters reaches a value.

- Another difficulty is that it is not practical to look for the best split (divisive method) or the best merge (agglomerative method). Rather than working on a pixel level, the algorithms are usually modified to use some measures, which can summarize the characteristics of various clusters.

There are many methods based on segmentation by clustering, a few of them are discussed here.

3.1.1.1 Segmentation by K means - Pixel based method

K-means [129] is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. It is based on obtaining an objective function that assumes that there are k clusters (where k is known). Each cluster is assumed to have a center; we write the center of the i^{th} cluster as c_i. The j^{th} element to be clustered is described by a feature vector x_j. For example, if we were segmenting scattered points, then x would be the coordinates of the points; if we were segmenting an intensity image, x might be the intensity at a pixel. We now assume that elements are close to the center of their cluster, yielding the objective function

$$\sum_{i \epsilon clusters} \{ \sum_{j \epsilon i^{th} cluster} (x_j - c_i)^t (x_j - c_i) \} \tag{3.1}$$

Notice that if the allocation of points to clusters is known, it is easy to compute the best center for each cluster. However, there are far too many possible allocations of points to clusters to search this space for a minimum. Instead, we define an algorithm which iterates through two activities:

- Assume the cluster centers are known, and allocate each point to the closest cluster center.

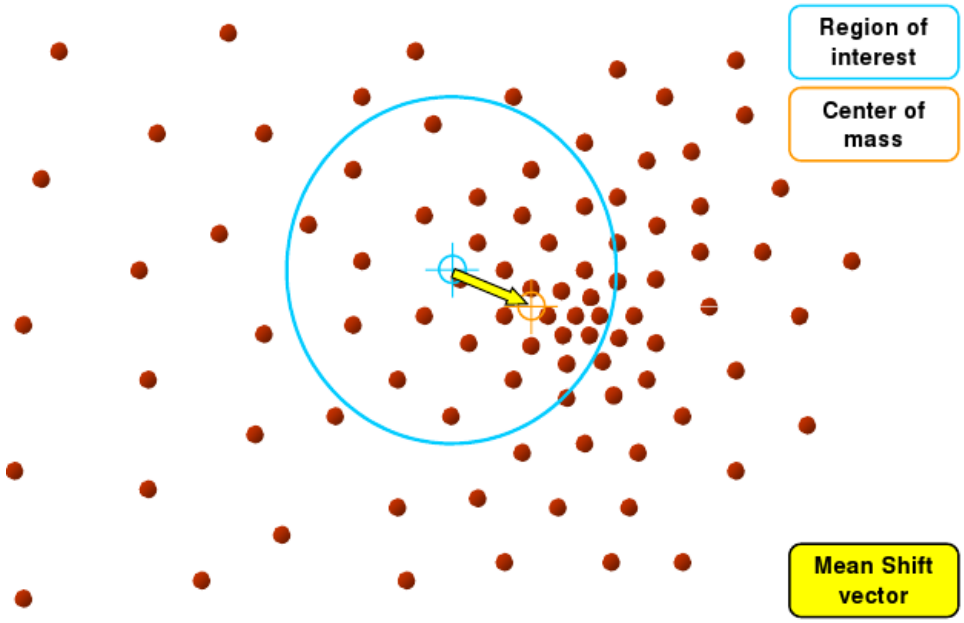


Figure 3.3: The mechanism of mean shift clustering.

- Assume the allocation is known, and choose a new set of cluster centers. Each center is the mean of the points allocated to that cluster.

We then choose a start point by randomly choosing cluster centers, and then iterate these stages alternately. This process will eventually converge to a local minimum of the objective function. This algorithm is usually referred to as k-means.

3.1.1.2 Mean shift - Kernel based method

The mean shift method was proposed in 1975 by Fukunaga and Hostetler [28] and it was mainly sidelined until Cheng's work [29] in 1998 in which he used it for mode seeking and clustering in a distribution. It is motivated by the iterative calculation of the gradient of the kernel density estimation to find the densest region in a distribution [34]. It is a technique to examine a complex multi-modal feature space and to find feature clusters. It is a kernel based method and the parameters of the kernel include the spatial bandwidth and the bandwidth of the feature space. Mean shift is composed of a number of steps. For each pixel in the image, following steps have to be performed ([1]Fig. 3.3):

[1]http://www.wisdom.weizmann.ac.il/~vision/index.html

- Compute the density estimate in a specific search window (kernel).
- Compute the direction and location of gradient of density estimation i.e. "mean shift vector".
- Move the search window by the magnitude of the "mean shift vector".
- Repeat the steps until the convergence of the "mean shift vector".

In practice, however, kernel bandwidth can have a strong impact on the algorithm's performance. The process of mode seeking is greatly affected by the selection of different bandwidths, effectively adjusting the sensitivity of the algorithm. An important detail of this method is that it takes into account both the spatial information (pixel location) and the range information (grey level, color etc). Although a visual data driven selection of the bandwidth is possible, some research has been done on calculating adapted kernel parameters for doing segmentation [57].

3.1.1.3 Watershed segmentation - Gradient based method

Watershed segmentation is a method which is related to the field of mathematical morphology [13], [136]. This is a technique in which the magnitude of an image gradient is interpreted as its altitude in a landscape. Therefore, an image is segmented into several catchment basins, which are regions of an image (interpreted as a high field) where the rain water will flow into the same lake ([2]Fig. 3.4). An efficient way to compute such regions is to start flooding the landscape at all of the local minima and to label ridges wherever different evolving components meet. In practice, this whole process can be simulated by the computation of the image gradient, where ridges/valleys in 2D can be identified as loci of minimum gradient magnitudes along the relief's level curves [90].

3.1.1.4 Graph cuts

Clustering can be seen as a problem of cutting graphs into pieces. In such a case, each data item (pixel) is associated with a vertex in a weighted graph, where the weights on the edges between the vertices summarizes the similarity between the pixels. These methods are basically divisive methods where an image is perceived as a

[2]Watershed - www.imagemet.com/WebHelp/spip.htm

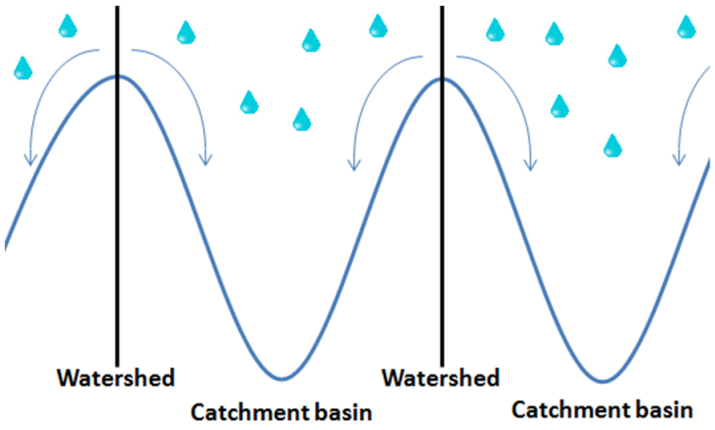


Figure 3.4: Mechanism of watershed image segmentation.

whole cluster and good clustering criterion are determined to recursively find the best split in the graph. The aim usually is to cut the graph into a number of connected components with low inter-component weights and large intra-component weights. This view of image segmentation has been widely used resulting in a large variety of segmentation methods. Graph cuts provide a clean and very flexible formulation of the problem of image segmentation. They provide a convenient way to use various visual features for expressing the similarity (affinity) between the pixels, and a set of powerful computational mechanisms to extract the global segmentation of these simple pair-wise pixel affinities. We discuss one of the graph based segmentation methods (Normalized cuts) in detail in Chapter 6.

3.1.2 Segmentation by fitting

One view of segmentation is to assert that the pixels in an image belong together because they conform to some model. This view is similar to clustering, the only difference being that it involves relation at a large scale than from pixel to pixel e.g., there is a program which tries to put tokens together into groups which "look like" a line. We can also think of this as clustering of tokens because together they form a familiar geometric model. This procedure can be regarded as "fitting". The most well known methods from this category are known as active contours. Active contour is an interface usually used to separate structures and background in an image. There are two principal approaches to build an active contour.

- Parametric deformable models - snakes.

- Geometric deformable models - level sets.

3.1.2.1 Parametric deformable models - snakes

The parametric deformable models represent curves and surfaces during deforming explicitly in parametric form. Firstly introduced in [80], the snakes method took its name due to similarity of evaluating of the contour with a moving snake. The parametric models can be described with the help of formulation of an energy minimization functional. It is based on searching of parametric curve that minimizes weighted sum of internal and potential energy of the curve. The internal energy specifies the tension of the contour whereas the potential energy is defined in the image domain. Therefore, the energy functional is defined in such a way that the curve had certain geometrical properties (such as smoothness) and would be attracted towards features of interest (such as the boundary of an object). Consequently, the most common approach is to assume that the object in the image to be expected has a boundary and a contour is initialized, that moves on the image gradient to settle on the object boundaries (Fig. 3.5). Snakes provides a global approach to shape segmentation problem than the ones based on gradient threshold. Some numeric methods are available which help in energy minimization such as gradient descend algorithm.

3.1.2.2 Geometric deformable models - Level sets

The parametric deformable models for segmentation have two main limitations. Firstly, it is hard to deal with topographical adaptations such as splitting or merging the model parts, which is useful for recovery of multiple objects. Secondly, the energy functional depends on the parametrization of the curve. Geometric deformable models provide the solution to this problem by doing the curve evolution as a signed difference function [104]. This gives an implicit representation of the evolving curve. Level sets are composed of terms representing the direction and speed of the evolving curves. The image data such as gradients representing object boundaries are incorporated in the term representing the speed of the evolution curve [25]. This helps in stopping curve evolution at a point of high gradients, representing object boundaries in the images. This formulation of curve evolution is more efficient at handling complex topographical surfaces (Fig3.6).

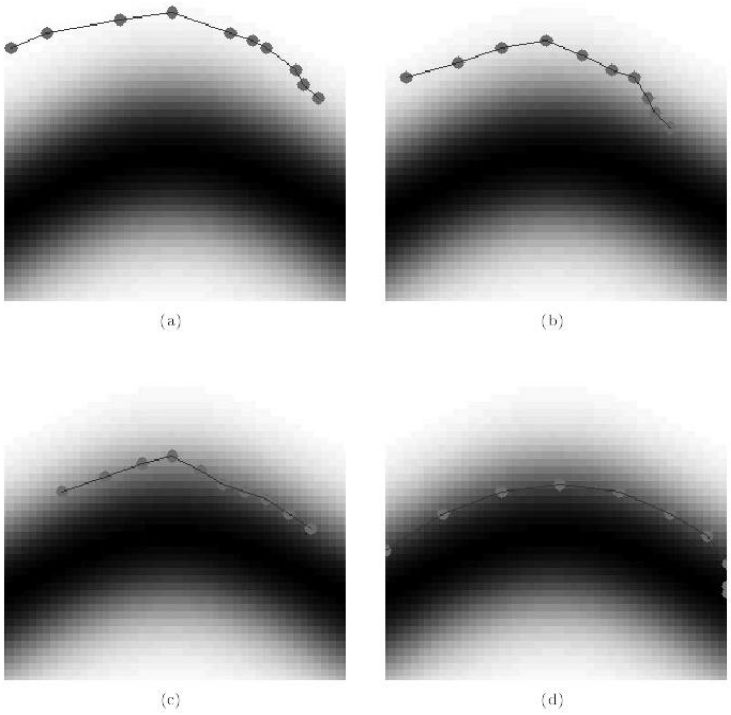


Figure 3.5: Visualization of the snake curve movement. The curve evolution starts in (a) and (b), (c), and (d) are the iterative steps in energy minimization[93]

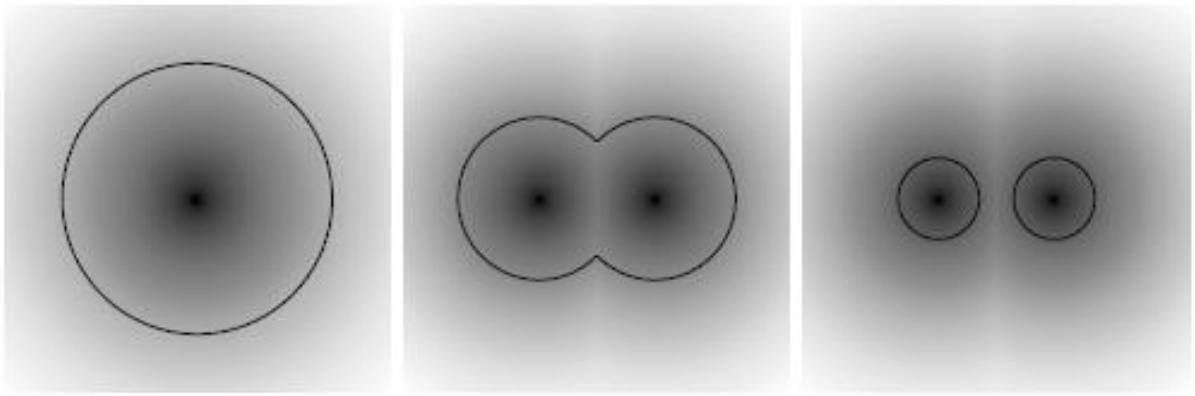


Figure 3.6: From left to right, the zero level set splits into two curves, handling topographical changes

3.2 Extraction of texture features

The visual data for an image may not be usable directly for various pattern recognition tasks. For this purpose, it is vital to obtain some important characteristics of the visual information. One of such characteristics is texture. In images, texture quantifies

local contrast (gray level differences) and local spatial structure. There is no precise definition of what texture is. This difficulty is demonstrated by the number of different texture definitions attempted by vision researchers, some (not all) of them are as follows:

- If the brightness is interpreted as elevation in a representation of the image as a surface, then the texture is a measure of the surface roughness [114].
- An image texture may be defined as a local arrangement of image irradiances projected from a surface patch of perceptually homogeneous irradiances [16].
- The basic pattern and repetition frequency of a texture sample could be perceptually invisible, although quantitatively present ... In a deterministic formulation texture is considered as a basic local pattern that is periodically or quasi-periodically repeated over some area [42].
- Texture regions give different interpretations at different instances and at different degrees of visual attention. At a standard distance with normal attention, it gives the notion of macro-regularity that is characteristic of the particular texture. When viewed closely and attentively, homogeneous regions and edges, sometimes constituting basic texture primitives also known as *texels* are noticeable [26].

The intensity variations which characterize the image texture are generally due to some underlying physical variation, which is very difficult to model mathematically. For this reason, there is no precise definition of texture which exists in the literature. However, some attempts have been made to identify some intuitive properties of texture, which are commonly used by the researchers.

- Texture in an image can be perceived at various levels of resolution [50].
- Texture is a property of areas; texture of a point is not identifiable [27].

A wide variety of techniques for describing image texture have been proposed. Within texture, the placement of texels can be periodic, quasi periodic or random. In natural scenes, they are usually random whereas in artificial scenes they are often deterministic or periodic. These differences motivate different approaches in analyzing texture within the image processing community. There are a lot of methods which can be used to quantify the image texture. Following Tuceryan et al. [131], we can list them into four categories.

- **Statistical methods** - Dealing with statistical properties and spatial distribution of the gray level values in an image.
- **Model based methods** - Construction of models that can describe textures (more widely used in texture synthesis).
- **Structural methods** - Assume that textures are composed of some basic primitives known as 'texels'.
- **Signal processing methods** - Based on filtering the images in spatial or frequency domain.

3.2.1 Statistical methods

Statistical approaches are suited for random textures. They yield characterizations such as smoothness, coarseness, graininess etc. They represent the texture indirectly according to the non-deterministic properties that manage the distributions and relationships between the gray levels of an image. This technique is one of the first methods in machine vision. By computing local features at each point in the image and deriving a set of statistics from the distributions of the local features, statistical methods can be used to analyze the spatial distribution of gray values. Some common statistical texture feature measures are as under.

3.2.1.1 First order statistics

The simplest form of statistical methods work on individual pixels. Each pixel in a gray scale image is represented by an unsigned integer. If 'b' is the number of bits used to represent a pixel value, a particular pixel can take up to $2^b - 1$ distinct values. For most images, 8 bits are used for such a representation giving us gray level values ranging from 0 to 255. The spatial distribution of gray-level variations can be described by a probability distribution of pixel intensity. It is this histogram which is used to generate a class of texture features. One direct way to characterize the qualities of textures is to use the shape of an image histogram e.g., a histogram with a narrow gray-level distribution indicates a low-contrast image and a multimodal histogram suggests regions of different brightness. A group of statistical measures which describe the histograms can be calculated from the gray level values of individual pixels in an image, including mean gray level of pixels, variance and their standard deviation etc.

A further characterization of the histogram includes skewness and kurtosis, effectively measuring the symmetry and peakiness of the distribution.

3.2.1.2 Second order statistics

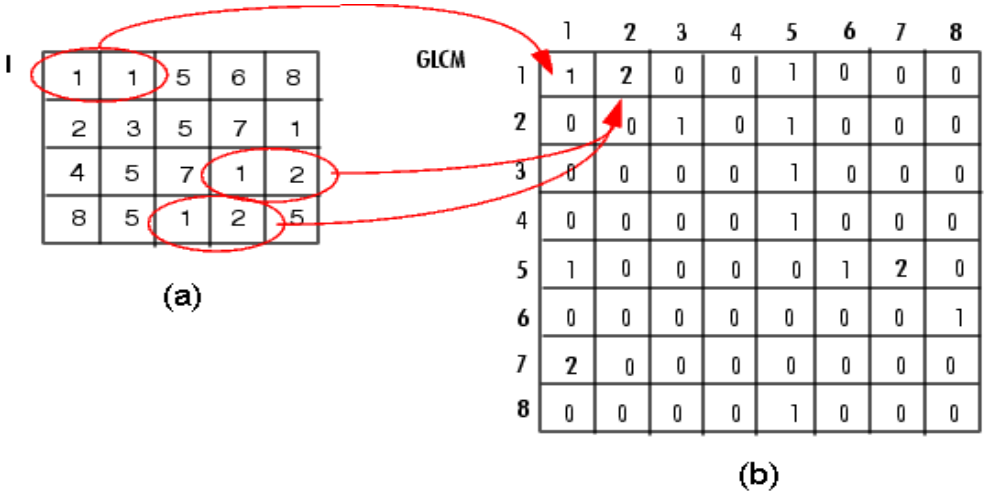
Second-order texture measures are mainly based on the joint gray-level histogram of pairs of geometrically related image points. The most widely used second-order statistics methods are based on Gray Level Co-occurrence Matrices (GLCM), gray level differences, autocorrelation and Local Binary Patterns (LBP). GLCMs (Fig. 3.7) proposed by Haralick [65] have become one of the most well known and widely used texture measures. Haralick proposed two steps for texture feature extraction: the first is computing the co-occurrence matrix and the second step is calculating texture features based on the co-occurrence matrix. Some of the widely used features are energy, entropy, contrast and correlation etc. This technique is useful in wide range of image analysis applications ranging from biomedical to remote sensing techniques.

The most widely used method currently based on second order statistics is LBPs. The original LBP operator was introduced by Ojala [103]. This operator typically works with the eight neighbours of a pixel, using the value of this center pixel as a threshold. If a neighbour pixel has a higher grey value than the center pixel (or the same grey value) then a one is assigned to that pixel, else it gets a zero (Fig. 3.7). The LBP code for the center pixel is then produced by concatenating the eight ones or zeros as a binary code. A region is then represented by a histogram of these codes.

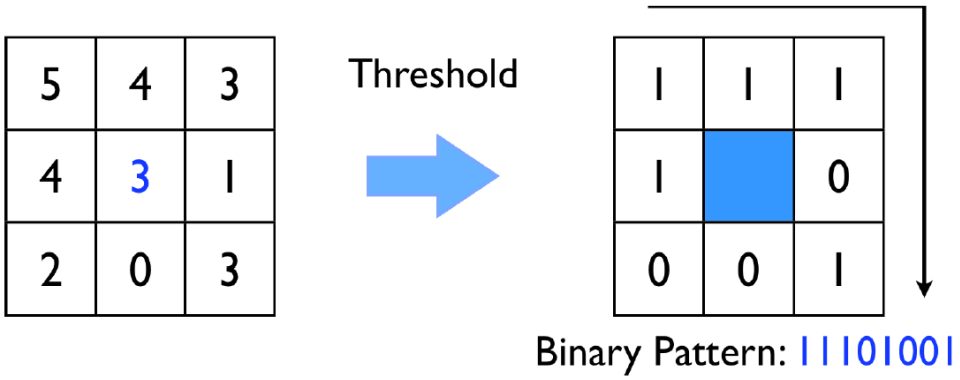
There are several other methods in the literature and their usage depends on the underlying application. Although use of higher order statistics has also been investigated, the computational complexity for feature extraction increases exponentially with the increase in the order of the features.

3.2.2 Model based methods

Model-based methods were originally developed in the texture synthesis field. They are based on the construction of an image model that can be used not only to describe texture, but also to synthesize it. These models assume that the intensity at each pixel in the texture image depends on the intensities of the neighboring pixels. The model parameters capture the qualities of texture [27]. They try to capture the process, which resulted in the generation of a specific texture. With such methods, some model is



(a) Gray level co-occurrence matrix (Reprinted from [4])



(b) Local binary patterns [122]

Figure 3.7: Statistical methods based on second order statistics

assumed and the parameters of that model are estimated based on sub-images (i.e., in neighborhoods) and those model attributes can be used as features. There are three fundamental model based methods found in the literature [102]: Markov Random Fields introduced by Dubes and Jain [39], fractals proposed by Pentland [106] and Multiresolution Autoregressive features proposed by Mao and Jain [99]. Model based methods have been applied to applications of such as texture synthesis.

3.2.3 Structural methods

Pure structural based approaches on texture analysis are based on the view that texture are composed of texels (texture primitives) which appear in near-regular repetitive spatial arrangements. These are obtained using some placement rules and

thus a texture is recognizable using the knowledge of texels and the placement rules. Therefore, structure and spatial organization of texture elements are the key components of the resulting models. Examples of texture elements include edges, shapes etc. The placement rules in structural methods correspond to global properties of textures. The advantage of the structural approaches is that they provide a good symbolic description of the image; however, this feature is more useful for synthesis than analysis tasks. These kinds of methods are limited unless regular or semi-regular textures are examined.

Structural approaches employ a variety of spatial analytical techniques for detecting the periodicity and analyzing the regularity of textures in order to recover the geometric structure and placement rules of texture elements. Many applications use texture primitives for extraction of regional properties such as the use of morphological features such as erosion and dilation as structural elements [64].

3.2.4 Signal processing methods

Psychophysical research has given evidence that the human brain does a frequency analysis of an image [19]. The primary visual cortex of the human beings can be perceived as a combination of frequency selective, orientation sensitive filters. Additionally, most of the former methods are suited for regular, semi regular texture or micro textures. Such limitation can be overcome by signal processing based methods which have the diversity of providing both micro- and macro- texture analysis. There are several types of filtering based methods.

3.2.4.1 Spatial domain filtering

Spatial domain filters are the most direct and most widely used methods to capture the texture characteristics of an image. It is one of the most basic low level signal processing methods which are used for a variety of tasks such as noise reduction, image sharpening, edge detection etc. In these methods, convolution of an image with a mask is usually done.

$$f(t) * h(t) = \int_{-\infty}^{\infty} f(\tau)h(t-\tau) \tag{3.2}$$

Where $f(t)$ is the signal, $h(t)$ is the filter, '*' is the symbol for convolution and

τ represents shifts in time. An example of the use of spatial filtering for feature extraction is edge detection. According to Gonzales and Woods [54] 'an edge is the boundary between two regions with relatively distinct grey-level properties'. Based on this definition, most of the edge detection methods work on the assumption that an edge occurs where there is a discontinuity in the intensity function in the images. Using this assumption, if we take the derivative of the intensity values across the image and find points where the derivative is a maximum, we will mark our edges. Two of the most common types of edge detector masks i.e., Sobel and Prewitt edge detectors are shown in Fig. 3.8.

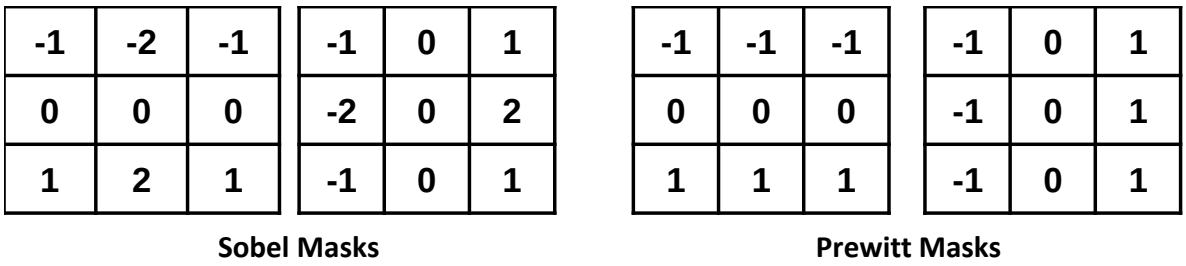


Figure 3.8: Edge detection using spatial filters

The most basic texture features using these methods are using edge densities. Fine textures have higher edges per unit area than coarse ones. Edgeness is usually measured over an image by computing a magnitude from the responses of the image filtered using masks.

3.2.4.2 Frequency domain filtering

In frequency domain filtering methods, images are usually filtered in the Fourier domain. All the masks used in the spatial domain have their frequency domain counterparts which can directly be used for filtering in the frequency domain. There are certain advantages of filtering in the frequency domain. In spatial filtering methods, convolution is a computationally complex task. Filtering in spatial domain corresponds to multiplication in frequency domain therefore much of the computational overhead is overcome in the frequency domain. Apart from the advantages related to the reduction of computational complexity, the frequency domain presents an alternative representation of images which is very useful for rotation invariant feature extraction. This is because a Fourier transform of a signal has two components: magnitude, phase. If rotation of an image takes place, this rotation is reflected in the phase component of the transform and the magnitude part remains unaffected. This follows from the shift

property of the Fourier transform [20]. This allows for research on rotation invariant descriptors while working in the frequency domain.

3.2.4.3 Spatial-frequency domain filtering

Although the frequency representation of an image has several advantages, it also has an important limitation: conventional frequency representation of an image ignores its spatial information. This problem is usually overcome to some extent by the use of wavelets. The wavelet transform or wavelet analysis is the most widely used solution to overcome the shortcomings of the Fourier transform.

$$F_\psi(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} \psi(\frac{x-b}{a}) f(x) dx \tag{3.3}$$

Where $f(x)$ is the signal which is decomposed and ψ represents the mother wavelet, which is scaled by a and shifted by b to obtain various self similar wavelets representing various filters in the frequency domain, obtaining a decomposed signal $F_\psi(a,b)$.

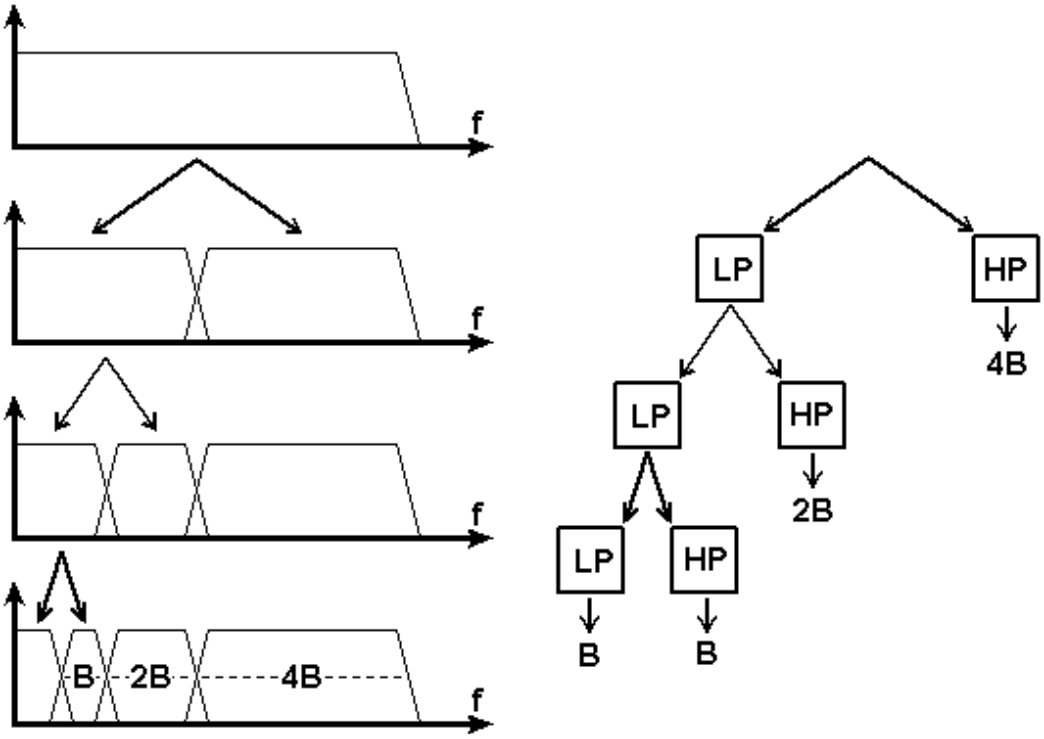


Figure 3.9: Wavelets filter bank (HP - High pass; LP - Low pass)

Wavelet transform has the ability to give the representation of a signal in both spatial

and frequency domain. The scaling parameter controls the spatial confinement of the decomposed signal, which is inversely related to the frequency confinement. When a higher localization of a signal is achieved in the spatial domain, the frequency localization is lost, and vice versa. This is because there is an upper limit to the uncertainty of both spatial and frequency information that can be resolved at the same time between time and frequency domains (discussed in detail in Chapter 7). Various self similar wavelets are used for decomposing the signal in wavelet transform.

Figure 3.10: Wavelet decomposition of an image

Each wavelet is represented as filters in the frequency domain, whose center frequency is scaled according to its bandwidths. Because of this collection of representations wavelets give a multiresolution analysis of the signals. The signal is decomposed using high pass and low pass filters and the low pass filtered signal is recursively filtered until a desired decomposition level (Fig. 3.9). In this decomposition, we obtain certain signals containing various frequency contents of the signal depending on the frequency and bandwidth of the bandpass filters. Effectively, the filters with higher center frequencies represent the high frequency contents of the image (edges). Repeating the filtering process recursively gives us high frequency contents of the signal on various levels of spatial resolutions of the signal. An example of multiresolution decomposition of an image can be seen in Fig. 3.10. After achieving this, statistics from high frequency contents (edges) are used to extract some features, which are usually used for pattern recognition.

In the case of wavelets we normally do not speak about time-frequency representations but about time-scale representations, since the term frequency is reserved for the

Fourier transform. Due to the multiresolution nature of wavelet transforms, they have been widely used for various texture recognition tasks in the literature [84],[109], [8]. One of the spatial-frequency filtering methods for feature extraction is known as Gabor filters. They have recently been used successfully for various applications involving texture feature extraction. The most important characteristic of Gabor filters is its optimal joint spatial-frequency resolution capability (discussed in more detail in Chapter 7).

3.3 Extraction of color features

Colors are one of the most important visual primitives, which facilitate various visual tasks in the real life. The use of color in computer vision is motivated by two important factors. First, color is a powerful descriptor that often simplifies object identification from an image. Second, humans can distinguish between thousands of color shades and intensities, compared to about only a few dozen shades of gray level. This helps us distinguish between many different objects, which otherwise are very difficult to identify when using simple gray level values to represent the images (Fig. 3.11). In 1666, Sir Isacc Newton discovered that when a beam of sunlight passes through a glass prism the emerging beam of light from it is not white, but instead it consists of a continuous spectrum of colors (Fig. 3.12). The visible color spectrum is divided into several regions spanning a narrow band of frequencies in the electromagnetic spectrum. Most of the colors that we see in nature do not correspond to a single wavelength but rather a mixture of wavelengths from the spectrum. The amount of energy at each wavelength is represented by a spectral energy distribution.

The colors that we perceive in an object depend on the nature of light reflected from that object. For example, an object appears green because it reflects light primarily in the 500 to 570 nm range, while absorbing most of the energy of the other wavelengths, thus effectively creating the sensation of green color in the visual scenes. Something that reflects light that is balanced in all visible wavelengths appears white to the observer. Color is therefore only a perceptual result of light in the visible region of the spectrum - spectral power distributions among different wavelengths exist physically, whereas their visual result appears in the form of colors in the eye and brain.

Given these facts, there is one important question: How can different colors be represented? In color theory, color models mathematically describe how colors may be represented. A color space is one where the components of the color model are precisely

(a) Gray scale image (b) Color image

Figure 3.11: Visual example of discriminative power of color. Players from two teams cannot be easily distinguished without color (Reprinted from [102]).

defined. Using these definitions in color models, the colors in the real world can be quantified thus helping in various vision tasks. There are many color models which have been proposed in vision research, however we are interested in the following:

- HSV color space.
- CIELUV color space.

3.3.1 HSV color space

The HSV color space is one of the most common cylindrical-coordinate representations of colors. This space rearranges the colors to be a more intuitive and perceptually relevant than the conventional RGB (Red, Green, Blue) cartesian representation. This is because, the human eye is incapable of dividing color into its spectral components [102] as done in the RGB space. Therefore they cannot think in terms of red, green and blue colors. The perceptual relationships between colors can be represented by

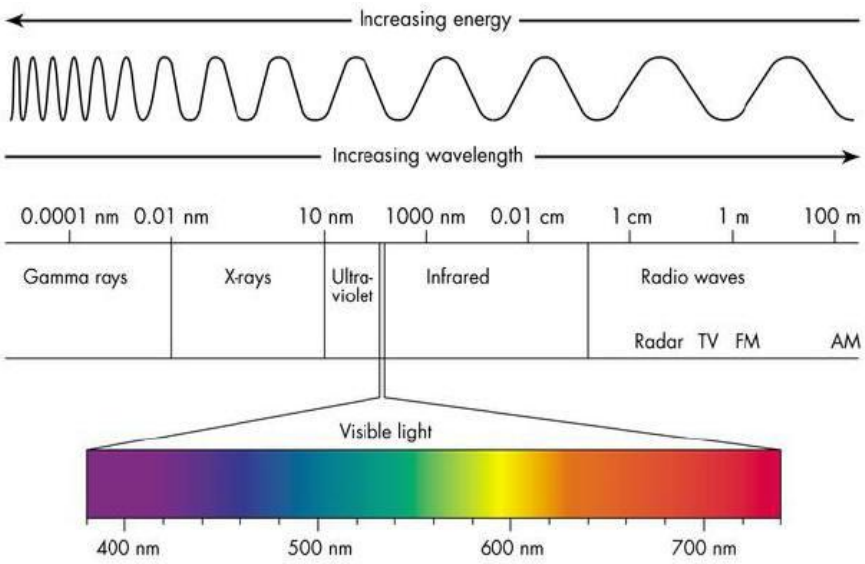


Figure 3.12: Visible wavelengths in the electromagnetic spectrum (Reprinted from [2]).

some other color models, one of them is HSV (Hue, Saturation, Value) [5]. This is because human sensations are best described by these three components (Fig. 3.13):

- Hue describes the color type (tone of the color). It identifies the dominant wavelength of the light. It is perceived as ranging from red through yellow, green and blue.
- Saturation describes the strength of the dominant wavelength i.e., the amount of energy at color's dominant wavelength relative to the amount of white light.
- Value is the brightness of the color. It ranges from 0-100%. When the value is 0, the color will be black.

HSV color model describes colors as points in a cone (black point at the bottom and fully saturated colors around a circle in the top). One of the well known HSV color descriptors is Scalable Color descriptor, which has been used as color descriptor in the MPEG7 standard [96].

3.3.2 CIELUV color space

Although HSV color space closely approximates the human visual perception of colors according to shades of color and their lightness, they are not perceptually uniform. A

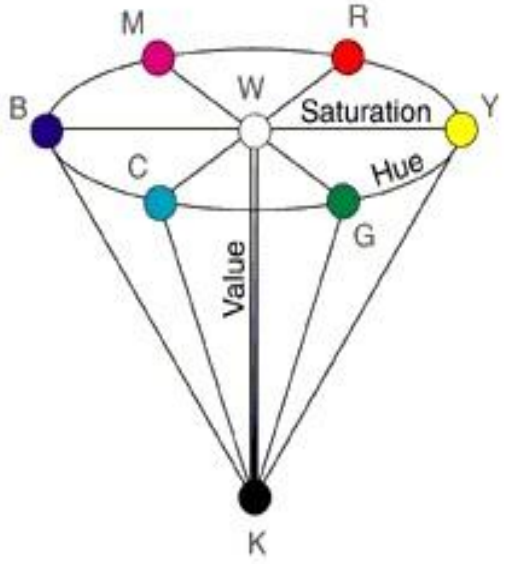
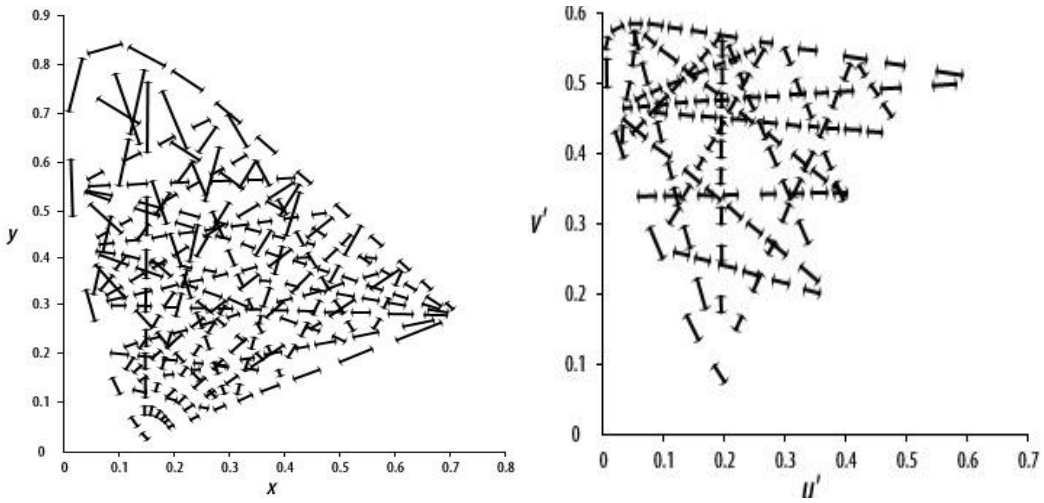


Figure 3.13: A cone representing colors in an HSV space (Reprinted from [3])

perceptually uniform color model implies that a unit distance in the color space corresponds to a uniform perceived distance in color. The color spaces such as RGB, HSV are far from exhibiting perceptual uniformity. The CIE (*Commission Internationale de l'Eclairage*) in 1931 defined three standard imaginary primaries (X, Y and Z) to replace red, blue and green because all visible colors could not be specified with positive values of these three components. Therefore all color in the visible spectrum cannot be represented by RGB space. This problem was overcome with this newly created CIEXYZ space using which, all visible colors could be specified with only positive values of the individual components. However, the CIEXYZ space is not perceptually uniform. In 1976, Commission Internationale de l'Eclairage (CIE) has defined a nearly uniform chromatic space i.e. CIELUV color space. It is composed of three components, the $L*$ component corresponds to lightness and $u*$ and $v*$ correspond to the chromaticity components.

A visual illustration of color difference is shown in Fig. 3.14. Each line here represents color difference of equal proportion. The distance between the end points of lines in each line segment are perceptually the same according to the observer. However the line lengths in (Fig. 3.14(a)) vary, sometimes greatly depending on their location in the space. To remove these perceptual differences, CIELUV was proposed (Fig. 3.14(b)) which shows a much more uniform spacing of all line segments, therefore hinting at a better uniformity of color differences in color space. Given that CIELUV color space preserves the visual perception of color differences, it has been widely used



(a) CIE x,y chromaticity diagram (b) CIE u,v chromaticity diagram

Figure 3.14: CIE diagrams for representation of perceptual color differences in (x,y) and (u',v') spaces (Reprinted from [1]).

for applications such as segmentation where color differences between two coherent regions having different colors has to be measured.

For implementation purposes, conversion between color spaces needs to be done. For conversion of images from RGB to HSV and CIELUV color space, please refer to Appendix B for the relevant formulas.

3.4 Pattern classification

Pattern classification is the organization of observations into groups sharing the same properties, i.e. patterns. The characteristics of these properties may include criteria such as structure, intent or applicability. A classification schema organizes patterns according to their properties which can be used for various applications such as retrieval, searching, recognition etc. There are many pattern classification methods which exist in the literature and they can be mainly divided into two categories.

- **Supervised methods -** Inferring a pattern from labeled data used for training.
- **UnSupervised methods -** Find structure in unlabeled data.

3.4.1 Supervised methods

Applications in which the training data comprises examples of the input vectors along with their corresponding labels are known as supervised learning methods. Such methods analyze the training data and infer a function from that data which is called classifier. For the testing data, this classifier is used to predict the output label. This predicted output can be compared with the actual label of the testing data to evaluate the performance of the classifier. There are several supervised learning methods, we will discuss only about *Support Vector Machines* (SVM) given their application on pattern classification using novel feature extraction methodologies in this thesis.

SVMs were introduced by Vapnik [132] and they have been widely used in various machine learning tasks intended for image retrieval and classification, just to name a few. SVMs rely on preprocessing the data to present a pattern in high dimension, typically much higher than the original feature space. With an appropriate nonlinear mapping function $\rho(.)$ to a higher dimensional space, data from two different categories can always be separated by a hyperplane. If $\mathbf{x}$ is an input pattern, it is transformed to $\mathbf{y}$ such that $\mathbf{y} = \rho(\mathbf{x})$. For each of the n patterns, let $z = \pm 1$ according to whether pattern is in class w_1 or w_2. A linear discriminant in an augmented $\mathbf{y}$ space is

$$g(\mathbf{y}) = \mathbf{a}^T\mathbf{y} \tag{3.4}$$

where both the weight vector and transformed pattern vector are augmented. The separating hyperplane ensures

$$zg(\mathbf{y}) \geq 1 \tag{3.5}$$

The goal in SVM is to find the hyperplane with the largest margin of separation; we expect that the larger the margin, the better generalization of the classifier (Fig. 3.15). The distance from any hyperplane to a transformed pattern $\mathbf{y}$ is $|g(\mathbf{y})|/\|\mathbf{a}\|$ and assuming that a positive margin b exists, equation 3.5 implies

$$\frac{zg(\mathbf{y})}{\|\mathbf{a}\|} \geq b \tag{3.6}$$

The goal is to find the weight vector a that maximizes b. The solution vector should be scaled arbitrarily and still preserve that hyperplane and thus to insure uniqueness the following constraint is imposed: $b\|\mathbf{a}\| = 1$; i.e., we demand that the solution should

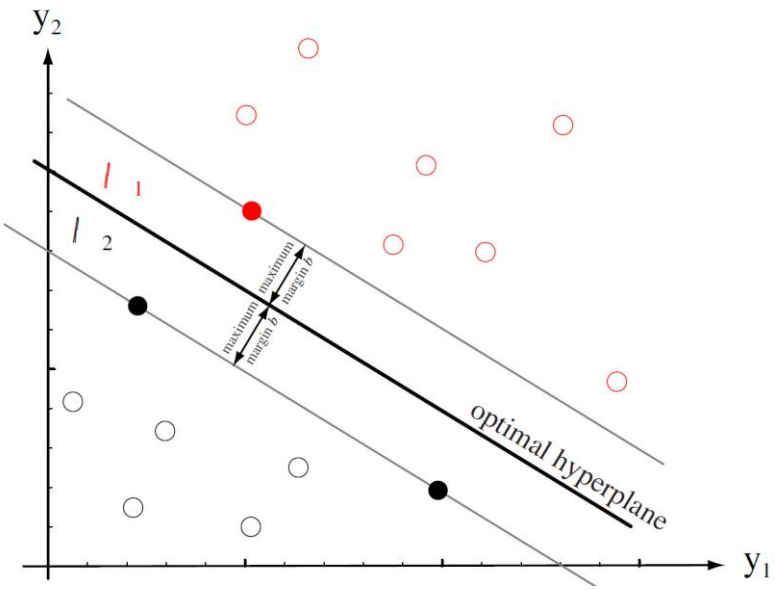


Figure 3.15: Training a SVM consists of finding an optimal hyperplane, i.e., the one with maximum distance from the nearest training patterns. The support vectors are those nearest patterns at distance b from the hyperplane [40].

also minimize $\|\mathbf{a}\|^2$. Support vectors are the training patterns for which equation 3.6 should represent equality i.e., the support vectors are equally close to the hyperplane. The support vectors are the training samples which are the most difficult to classify, i.e., they are most informative for classification task.

3.4.2 Unsupervised methods

In supervised methods, it is assumed that training samples are used to design a classifier. In unsupervised methods, unlabeled samples are used for classification. It is closely related to the problem of density estimation in statistics. However, unsupervised learning also encompasses many other techniques that seek to summarize and explain key features of the data. There are many methods of unsupervised learning which include k-means clustering, self-organizing maps etc [40].

3.5 Classifier performance evaluation

In a typical pattern recognition system, the feature extraction task is followed by the classification stage. At this stage, one of the most important choices pertain to the

selection of an appropriate method for training of the classifier to define the decision boundaries. Typically various strategies to segregate the whole data are done for training the classifier and establishing these boundaries. Later on, the novel data is classified using the decision boundaries and the resulting classifications have to be evaluated using certain metrics.

3.5.1 Model selection

For almost all real pattern recognition systems, one of the most vital problems is the selection of an appropriate method for training a classifier. For a system in which we have an unlimited number of samples this question has a straightforward answer: choose the model which yields the highest classification accuracy. However in most of the real applications, we have access to only a finite number of samples. So, the straightforward approach is not adequate for real scenarios. One approach is to use all the available data for training the classifier and estimate the error rate. This approach is however not good as it is overly optimistic and it is not uncommon to have a 100% correct classification. A much better approach is to split the data into disjoint sets. There are several such approaches:

3.5.1.1 Hold out methods

In this approach, the available data is split into two sets: a training set and a testing set. The training set is used to train the classifier and establish the decision boundaries. The testing set is used to evaluate the performance of the classifier and estimate the error rate. However, in this method, the error rate of the classifier might be misleading if we happen to get an unfortunate split where the training data is not good and does not completely define the overall characteristics of the data.

3.5.1.2 Random subsampling

This is another approach of model selection which is considered much better and yields much better estimates of the true performance of the classifier. In random subsampling, several splits of the dataset are done. Each split selects a fixed number of examples. For each data split, we retain the model with the training examples and estimate the error rate of the classifier with the test examples. The true error estimate is calculated as the average of separate estimates from every training example. There

are several approaches which can be used for subsampling the data. One such method is known as K-fold cross validation. In this method, K partitions (folds) of the whole dataset are created. For assessing the performance of the classifier, K experiments are performed. For each of the K experiments, $(K-1)$ folds are used for training the classifier and the remaining one fold is used for testing it. The advantage of the K-fold cross validation is that all the examples in the dataset are eventually used for both training and testing of the classifier. The true error is estimated as the average error rate calculated on all the K folds of the classifier. If E represents the error estimates, the true error estimate is given as:

$$E = \frac{1}{K}\sum_{i=1}^{K} E_i \tag{3.7}$$

In cases where we have a lot of data for training and testing, K-fold cross validation gives a very good estimate of the classifier performance. If however we have a very small amount of data, Leave-one-out cross validation (LOOCV) can be used. In this method if we have N samples, the experiment is repeated N times: for each experiment $(N-1)$ samples are used for training and the remaining one sample is used for testing the classifier. The error rate is estimated in the same way as that of K-fold cross validation. LOOCV is computationally expensive because it requires many repetitions of training.

Random subsampling is not good in scenarios where there is a lot of repetition of data. One such example is the use of consecutive frames in a video to classify its contents. In such a case, similar structures/contents are expected to appear in consecutive frames from the video. If these images fall in both training and testing set, the estimate of the overall classifier performance will be optimistic. Typically, folds are created manually such that images belonging to the training set do not falling in the testing set.

3.5.2 Evaluation metrics

Typically the output predictive classifications after model selection have to be compared with the available ground truth. This comparison is vital to evaluate the performance of a particular classifier. Classifier evaluation therefore helps us to assess the performance of the classifier in an objective way. To date, there are several metrics used for this purpose. In various fields such as bioinformatics, there is an uneven distribution of data between different classes and the use of simple metrics such as

error rate or average accuracy is typically not enough. We discuss certain additional metrics, that can be used to correctly determine the performance of the classifier.

3.5.2.1 Numerical measures

Certain numerical measures can describe classifier performance. These metrics combine the ability of the system to detect the positives for every class while rejecting the negative ones. They are typically composed of the following components:

- True Positive (TP): Real positives which were indicated as positives by the classifier.
- True Negatives (TN): Real negatives which were indicated as negatives by the classifier.
- False Positive (FP): Real negatives which were indicated as positive by the classifier.
- False Negative (FN): Real positives which was indicated as negative by the classifier.

Specificity, sensitivity and accuracy are used as performance measures derived from these four quantities. The former two (specificity and sensitivity) are used for quantifying the results of individual classes while the latter (accuracy) is used to assess the overall performance of the classifier. These measures are usually bounded between zero and one. Their interpretation is summarized in Table 3.1.

Specificity	$\frac{TN}{TN+FP}$	The proportion of negatives which the system successfully rejects
Sensitivity	$\frac{TP}{TP+FN}$	The proportion of positives which the system successfully detects
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	The percentage of predictions made by the system which are correct

Table 3.1: Performance evaluation metrics

3.5.2.2 Confusion matrix

Confusion matrix is a visualization tool typically used for classifier evaluation particularly for cases where there is an uneven distribution of data between various classes. Each column of the matrix represents the instances in the predictive class whereas each row indicates the instances of the real class. The advantage of the confusion matrix is that it enables an intuitive understanding of the performance of the system.

		Outcome prediction	
		Positive	Negative
True Class	Positive	TP	FN
	Negative	FP	TN

Figure 3.16: A typical confusion matrix in a two-class system

3.5.2.3 Receiver Operating Characteristics (ROC) curves

The ROC curve is a fundamental tool for diagnostic test evaluation. In an ROC curve, the true positive rate (sensitivity) is plotted as a function of the false positive rate (1-specificity). For different cut off points of the latter, the values of sensitivity are calculated (Fig. 3.17) and plotted as a graph (Fig. 3.18). This graph defines the operating points of the classifier. The area under the ROC curve (AUC) is a measure of how well a parameter can distinguish between two groups.

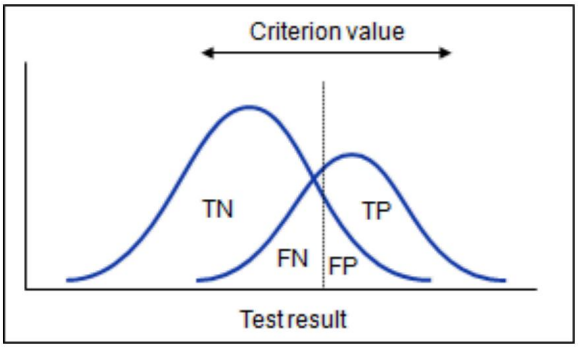


Figure 3.17: Calculation of operating points on an ROC curve

Each point on the ROC curve represents a sensitivity/specificity pair corresponding

to a particular decision threshold. Usually, certain discrete points are used as thresholds to calculate the operating points and the classifier's performance is interpolated between those points. A classifier having perfect discrimination capability i.e., having no overlap between the distributions in Fig. 3.17 has an ROC curve passing through the top left corner of the graph. For a classifier which is not perfect, the closer the curve is to the corner, the better the classifier is. On the other hand, an ROC curve consisting of a line passing through the diagonal (dotted line in Fig. 3.18) represents a random classifier.

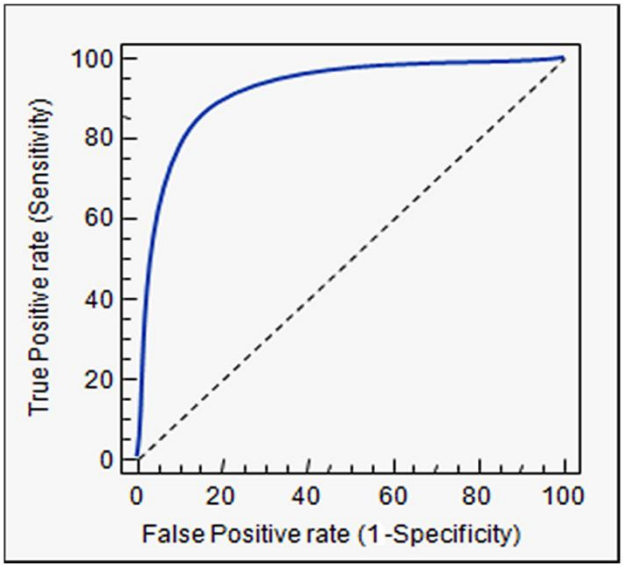


Figure 3.18: ROC curve

Chapter 4

Computer Vision for Gastroenterology

Recent years have seen an increased interest in research on Computer Assisted Decision (CAD) systems for gastroenterology (GE). The post 2000 era has been revolutionary given the introduction of new imaging modalities creating much more room for research on endoscopic imaging diagnosis both from clinical and CAD perspective. Significant research in the past has been done on assisted analysis of endoscopy images. In this chapter, we will discuss the latest state-of-the-art in computer vision (CV) for diagnosis of abnormalities in the gastrointestinal (GI) tract. We have modularized this chapter according to the detection of the following abnormalities (an explicit division into CV based objectives does not apply to most contributions): Ulcer and bleeding detection, Polyp detection, abnormality detection and Diagnosis of cancer. This chapter will lead us to the possible future directions and challenges for designing more adequate feature extraction methodologies for GE imaging (details in Section 4.5).

4.1 Ulcer and bleeding Detection

Hwang at al. [70] proposed an algorithm that they claimed to have used for detection of bleeding in wireless capsule endoscopy images. The authors did a manual selection of blood and non-blood pixels to train their respective probability models. For this method, it is important to know that there is a certain color range distinguishing blood pixels from non-blood pixels. However, most dark pixels representing blood can overlap with the dark pixels in non-blood pixels having no chrominance. The

Figure 4.1: Different types of ulcers. *Left*: Bleeding ulcer, *Middle*: Narrowing ulcer, *Right*: perforated ulcer (adapted from [78]).

EM clustering algorithm is used to train the probability models for blood and non-blood pixels. The theoretical foundation of the paper is interesting and good results are obtained. However, the proposed methodology is not invariant to illumination changes in the images. Also, the model parameters are learned using the training dataset. The results obtained are therefore highly dependent on the training data. Another shortcoming is that the testing pool used in these experiments is quite small and an extended validation of the method needs to be carried out.

M. Mackiewicz et al. [92] chose a histogram based approach for the detection of bleeding. They claim that this is necessary for fast model adaptation which is easier using non-parametric methods contrary to Hwang's [70] method (where the model parameters are learned using training data). Their approach uses adaptive color histograms to track the moving background and bleeding color distributions over time in capsule endoscopy (CE) videos. It therefore addresses the problems of drastic changes in blood color distributions due to GI fluids and thus helps in detection of red lesions when the difference between their color distribution and background is large enough. In the second stage of the algorithm, they analyzed all candidate blood frames using Hue-Saturation-Intensity (HSI) and Local Binary Pattern (LBP) features. Results show that their method compares favourably with Suspected Blood Indicator (SBI) [120] by Given Imaging. Since this method does not produce illumination invariant color features, it is not expected to discriminate between dark bleeding tissues and those that appear dark and are normal.

Li and Meng [87], [86] proposed the idea of using a combination of chrominance moments and Local Binary Patterns (LBP) as the image features to discriminate between normal and abnormal regions. HSI color space was adopted for the desired experiments and hue and saturation components were used to achieve illumination

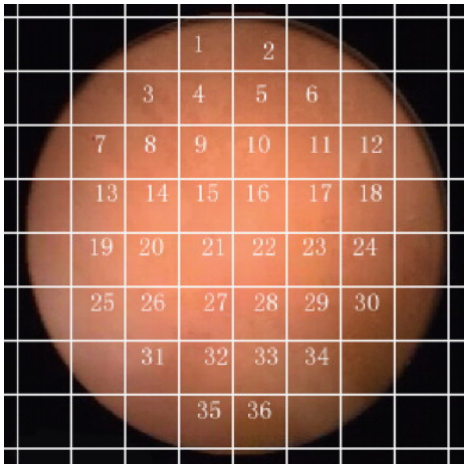


Figure 4.2: The most 'relevant' image segments chosen from one WCE image.

invariance. The wireless capsule endoscopy (WCE) images were divided into a number of block based regions (Fig. 4.2) that were chosen for feature extraction. A Neural network classifier was used for the classification task. Although good results were obtained, this study needs a stronger validation as the training and test sets were allowed to be composed of image blocks from the same patient sequence due to the lack of adequate data for relevant statistical evaluation of the algorithm.

Li and Meng [88] proposed a novel approach for detection of ulcers in capsule endoscopy images. They used uniform LBP of curvelet transformed images as texture features to distinguish between normal regions in the images and ulcers. The proposed features capture multi-directional features and show robustness to illumination changes. Multilayer perceptrons and support vector machines were used as classifiers. Although the results obtained were very good, features were extracted from image blocks that were obtained by a grid based segmentation of capsule images (Fig. 4.2). The rationale that the authors have adopted for this strategy is that for feature extraction from full images, some detailed micro image characteristics might be disregarded. The proposed method has been used for ulcer detection on large amount of data, however it was collected from only 5 patients. For an extensive testing of the proposed method, a dataset from more patients should be included in the study.

4.2 Polyp detection

Karkanis et al. [79] proposed a computer aided tumor detection method in endoscopic images using color wavelet features. These are based on covariance of second order

textural feature over wavelet frames. The methodology has been applied to a rich dataset of color colonoscopy videos with very good overall performance achieved for the detection of polyps.

S. Hwang and M. Celebi [69] proposed a novel method for the detection of polyps. They used watershed segmentation using a novel initial marker selection method using Gabor filters and K-means clustering. After a preliminary segmentation of the image, polyp candidates are identified assuming that they should be circular or elliptical. Curvature centres are calculated using the boundary of the polyp candidates where most of them should be located inside the polyp region. Good results for identification of polyps were obtained with a very high sensitivity but slightly lower specificity. The authors pointed out that the low specificity was due to the richness of texture exhibited by some non-polyp regions, which were classified by the algorithm as polyps.

A. Karargyris and N. Bourbakis [78] proposed a method to detect polyps in the bowel. They used Log-Gabor filters followed by SUSAN edge detector to do segmentation of capsule images. The algorithm runs around the boundary of the segmented part calculating the curvature at each point. For points having a certain curvature, the centres of curvature are detected. A cloud of curvature points inside the segmented image regions is found. A simple two-threshold sequential clustering is later used to cluster the curvature points. Starting from those cluster centres, active contours are used to expand the cluster on the gray level image to obtain the desired polyps. As

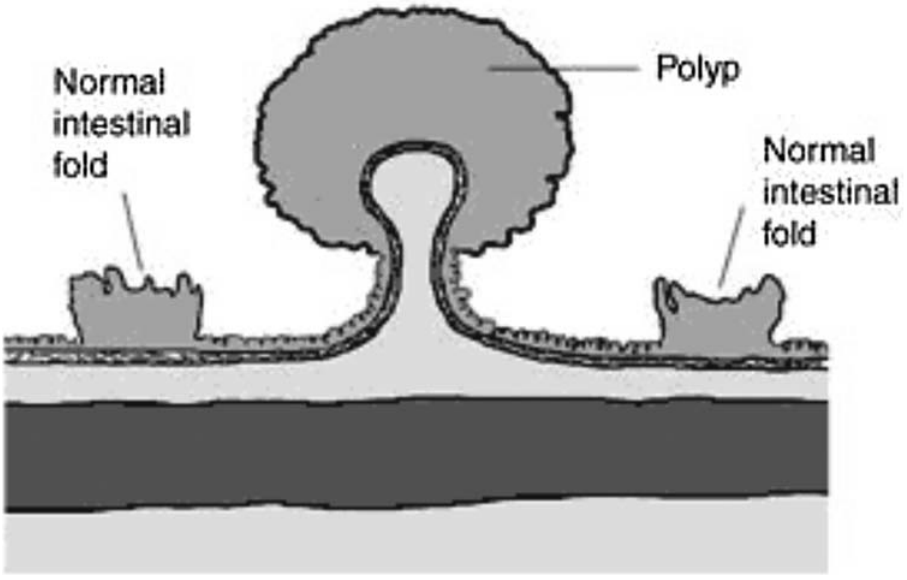


Figure 4.3: Graphic representation of a polyp in human colon. National Institute of Diabetes and Kidney Diseases, National Institutes of Health.

in the previous method, the specificity of the algorithm is low and it is attributed to the regions having crisp texture while being a non-polyp. Rules based decision making in such cases to identify polyp candidates as done for SUSAN edge detection and curvature calculation could be manipulated to strongly effect specificity of the algorithm. These decisions are therefore expected to be data driven. Also, it is a synergistic approach and the use of both SUSAN edge detector and active contours significantly increases the complexity of the proposed methodology.

B. Li et al. [85] propose a texture descriptor which is built upon a combination of wavelets and Local Binary Patterns (LBP). Wavelets have been known for their great success in acquiring multiresolution texture features whereas LBP are known to give discriminative texture features while providing illumination invariance. The strength of both these methods is combined by taking LBPs of the wavelet sub-bands resulting from decomposition yielding powerful texture features. Good results for tumor detection are obtained using this methodology however it would be interesting if the authors had used rotation invariant LBPs to complement the wavelet based features having poor directionality with rotation invariant image features.

M. Hafner et al. [59] performed pit pattern classification to support the assessment of colonic polyps using zoom endoscopy. They have studied the application of texture analysis methods in the wavelet domain (Dual-Tree Complex Wavelet Transform - DTCWT) for feature extraction from the images. The authors justify the use of DTCWT due to its approximate shift invariance property that can better handle the imaging dynamics in colonoscopy and better directional selectivity as compared to Discrete Wavelet Transform (DWT). Authors have proposed to use the wavelet transform coefficients from all the color channels for an increased detection of abnormalities in the colon. Classification is done using 1-nearest neighbor (NN) classifier. While the feature extraction appears solid, we suspect that usage of a good feature space classifier could improve the performance of the proposed methodology. It is important to note that the authors have emphasized the need for shift and rotation invariant texture features due to varying imaging conditions in endoscopy exams.

J. Bernal et al. [12] have devised a methodology for segmentation of polyps in colonoscopy images using a model of polyp appearance. Their aim was to divide an image into a minimum number of informative regions. Their method consists of applying an image segmentation algorithm on an image to obtain over-segmentation (such as watershed) and then applying a region merging algorithm that takes into account the model of polyp appearance. They have compared the performance of their method with normalized cuts and show that their algorithm shows superior

segmentation results. However, it is important to note that since the region merging algorithm incorporates a model of a polyp, it is not expected to give good segmentation results for a different GE objective such as lesion detection etc.

4.3 Abnormality Detection

P. Hiremath et al. [67] proposed a method for detecting possible presence of abnormality in endoscopic image from lower esophagus. Color segmentation of images was done based on parameters learned from the training images. The image is segmented using 3σ intervals around the mean RGB values. Edge detection of the obtained image segment is done followed by contour curvature computation. Thresholding was later used to analyze the curvature change along each edge contour and consequently, abnormal regions in the image were identified. Although the authors achieved good performance, this method is parametric where hard thresholds are used and therefore it is not expected to be robust. It is however important to note that the authors did not mention, what kind of abnormalities they have detected using this methodology.

M. T. Coimbra and J. P. Cunha [32] performed a statistical analysis of various MPEG-7 visual descriptors for classification of capsule endoscopy images. Their objective was the assessment of usefulness of various descriptors in event detection such as polyp, bleeding etc. Experiments showed that the Scalable Color (SC) and Homogeneous Texture (HT) descriptors are most adequate for event detection. The good performance of SC compared with other color descriptors is due to its higher 'resolution'. On the other hand, the good performance of HT when compared with Local Edge Histogram indicated that too much attention on small variations in the image is not desired for adequate abstraction of the desired abnormalities.

S. Bejakovic et al. [10] tried to detect Crohn's disease by the use of a combination of color and texture features. They used three MPEG-7 visual descriptors i.e. the Dominant Color Descriptor (DCD), Homogeneous Texture Descriptor (HTD) and Edge Histogram Descriptor (EHD) and Haralick texture features to perform the detection. The classification was done using support vector machines. A large dataset was used for validation of the performance of the methods and experiments showed that the DCD descriptor outperformed the other descriptors in overall classification of the images, giving the most discriminative image features. This is because the color information is distinct in case of blood, ulcers etc from other image characteristics including villi, bile, bubble and other extraneous matter. Also, they advocate that the

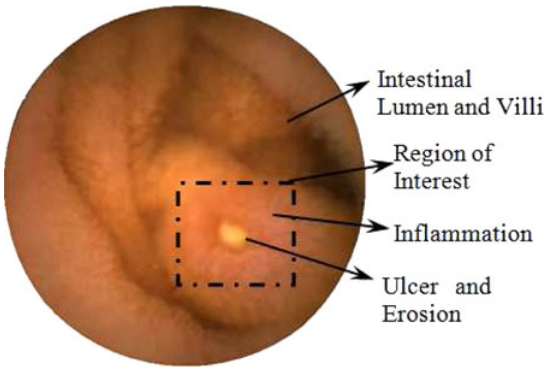


Figure 4.4: Annotation and export of capsule endoscopy images including a Region of Interest (ROI) and a severity classification (Adapted from [82]).

color information is least affected due to noise and compression while distorting other visual features such as texture.

R. Kumar et al. [82] have performed an assessment of Crohn's disease lesions in capsule endoscopy images. They have made use of the most popular image descriptors for color (DCD), texture (HT) and edge (EHD) in the MPEG-7 in all their possible permutations to assess their performance in classifying the lesions. Features were extracted from regions of interest that were obtained using manual annotations (Fig. 4.4). These were classified as normal, mild, moderate or severe based on the size and severity of the lesions and any surrounding inflammation in the lesions. Support vector machine were used for classification of the lesions. It is important to note that the best performance was obtained when a combination of all the image descriptors was used for classifying the lesions.

I. Figueiredo et al. [45] have used the variational level sets framework for segmentation of Aberrant Crypt Foci (ACF) in colonoscopy images (Fig. 4.6). They integrated the use of physiologically important image features used for detection of ACF in the traditional level sets methods, which instead use more generic image features such as edges in the images. The resulting integration gives very good segmentation results. This paper however uses only a limited number of test cases. An extensive validation of the proposed method on a bigger dataset is required for more robust conclusions. It is important to note that the presented method is not expected to give good performance for detection of some abnormalities other than the detection of ACF as the image features used in this method are not generic.

V. Charisis et al. [30] proposed a novel method for detection of abnormal patterns in wireless capsule endoscopy images. They used Bidimensional Empirical Mode

Decomposition (BEEMD) on RGB color images of small bowel to extract the structure of the bowel on various scales. Afterwards, lacunarity analysis was employed as a method to quantify and extract the texture patterns of the image regions. The method is novel and shows very good performance however this paper lacks a comparison of the proposed technique with more traditional methods for texture feature extraction. It lacks a proper motivation of the use of the proposed methodology and also does not include an analysis on how it compares with the other methods.

4.4 Diagnosis of Cancer

C. Demir and Y. Demer [38] presents a systematic survey of various methods that have been used for the design of computer assisted decision system for cancer diagnosis. Based on this survey, the authors have identified three important steps, which are usually involved in all such systems i.e., pre-processing, feature extraction and classification. They discuss various methods that have been used by different researchers for these tasks. This paper highlights an important issue in the area of automated diagnosis of cancer which is the lack of availability of data. Several research groups are working on this line of research however it is not possible to make a fair comparison of their performance achieved due to the lack of an open gold standard database. It is therefore essential to form a benchmark of datasets with the classification results backed by their biopsy reports.

M. Hafner et al. [58] make use of ring filters for feature extraction followed by classification using support vector machines (SVM). The ring filters are a combination

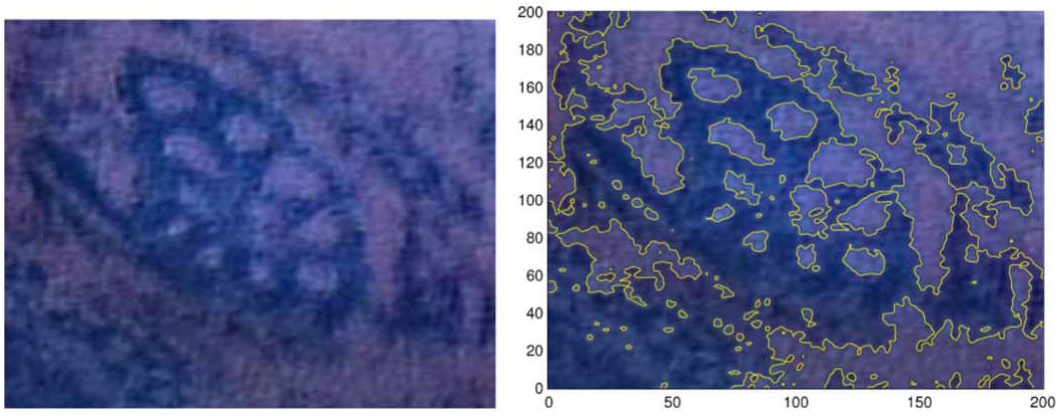


Figure 4.5: An image representing Aberrant Crypt Foci in colonic images (Adapted from [45]).

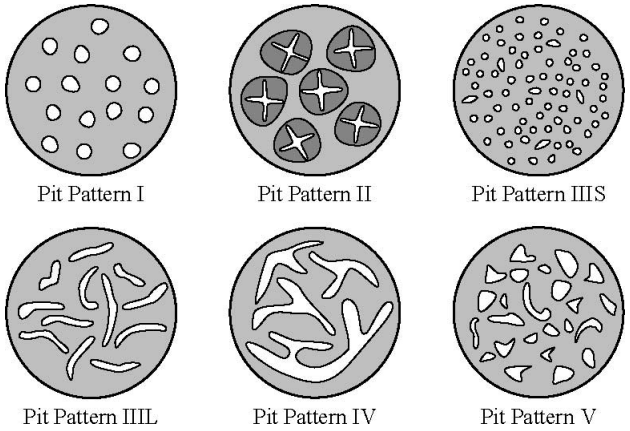


Figure 4.6: Pit pattern classification (Adapted from [58]).

of isotropic band pass filters. The authors have pointed out that their fourier based features are among the best performing however they are slightly inferior as compared to the Dual-Tree Complex Wavelet Transform (DTCWT) and Local Binary Pattern (LBP) features. We suspect that this happens because contrary to the DTCWT, the ring filters are not capable of directional selectivity thus limiting their use in extracting certain texture characteristics. The authors also hinted that it would not be sensible to make a direct comparison between the performance reported by various authors in their papers as compared to their methods since the performance should be measured on the same datasets for being directly comparable.

A. Sousa et al. [123] made use of the adapted color histograms for classification of cancer in gastroenterology (GE) images. They did a survey on these images to conclude that the colors in GE images do not span the full spectrum of the color spaces. The richness of colors therefore lies in some particular regions in the color space. They introduced adapted color spaces to enhance color resolution in the space where the occurrence frequency of the colors is high and compromised on resolution where occurrence frequency was low. The proposed descriptor was combined with Local Binary Patterns to classify endoscopic images into three different classes: normal, pre-cancer and cancer. They obtained very good classification results. It is important to note that these are very specific histograms for which the validation was done on vital stained images (using mythelene). Other staining agents can lead to a different distribution of colors in the color space thus the histogram bins need to be recalculated by vector quantization of the color space.

We [110] proposed the use of Gabor filters for feature extraction from the images. Images from these filters were integrated in a bag-of-words approach (Gabor textons -

Chapter 7) for classification of gastroenterology images from two imaging modalities: chromoendoscopy and narrow-band images. Results showed that the proposed framework performs well for identifying cancer in images from these imaging modalities. As a conclusion from this work, although the color features are very useful for a faithful description of gastroenterology images, their use for classifying images using narrow bands of light is rather limited due to lack of rich colors. Texture features on the other hand are useful for classifying images from most of the imaging modalities.

4.5 Discussion

In this chapter, we have reviewed the state-of-the-art on computer vision for gastroenterology. This review has led us to reach a number of conclusions related to various feature extraction approaches for assisted diagnosis.

Dynamic imaging conditions - As pointed out in [59], gastrointestinal endoscopy poses a challenging computer vision problem to the researchers. This is because of the lack of full control of the camera, leading to arbitrary views of a tissue on various scales and various orientations. The feature extraction methodology is expected to be able to cope up with these changes, presenting robust methods which are invariant to rotation, scale and illumination changes in the images.

Importance of multiresolution (MR) features - From our review, it is clear that many researchers have resorted to the use of wavelet based MR image characteristics. This is because of the presence of coarse or fine texture in the images due to the gastric mucosa, capillary network etc. which will only be distinguishable using MR techniques. Typically, these methods can also yield directional sensitive features. This creates the possibility of producing rotation invariant image features thus helping in covering most of the aspects originating from dynamic imaging characteristics.

Segmentation of images - Some authors have resorted to "perfect" segmentation of the images using manual clinician's annotations. Some have resorted to grid-based solutions which are better as compared to analysis using full images. Segmentation of images from various imaging modalities is very subjective (depending on pathologies and imaging sites) and therefore it poses a very challenging task in gastroenterology. It is therefore very hard to present generic segmentation methods that will work for most of the gastroenterology imaging scenarios.

Quantized color spaces - The use of full color spaces for classification of gastroen-

terology images is usually not the optimal. It is more useful to vector quantize the color spaces since the images manifest reduced color spaces. The use of adapted color spaces gives a finer resolution color representation of the images, which helps in determining small changes in color characteristics such as the detection of inflammation, which might otherwise not be detectable using full color spaces.

Based on our findings from the review presented in this chapter, we will proceed in the rest of the thesis with our novel contributions on segmentation and feature extraction of gastroenterology images. These contributions will be followed by a detailed performance comparison of our proposed methodologies with some other relevant methods.

Chapter 5

Materials

Data collection and annotation is one of the most vital steps for the development of computer assisted decision (CAD) systems. Chapter 1 lists the objectives of this thesis, one of them being the research on generic computer vision tools for CAD systems in gastroenterology (GE). In this context, the traditional pattern recognition (PR) discussed in Chapter 3 highlights various important modules that are necessary for the design of computer vision module. The experimental validation of the two important stages i.e., segmentation and classification in the proposed PR system is necessary for a final evaluation of their performance. It is therefore imperative to collect the data and obtain its manual annotations from the physicians. These annotations involve the information about identifying clinically relevant regions and their final classification in the images. Given all this, we have acquired images from two complementary GI organs using complementary imaging modalities:

- **Chromoendoscopy (CH) -** Obtained using staining of gastric tissues with mythelene blue. Lesions are characterized using color and texture of the gastric mucosa.
- **Narrow-band imaging (NBI) -** Obtained using two narrow bands of light in the visible spectrum to enhance the vascular structure in the images from esophagus.

In this chapter we will discuss in detail the procedures adopted for collection of the acquired data and discuss the specifications of the data that was finally used for testing of our algorithms.

5.1 Chromoendoscopy - CH

Our CH dataset was collected at the Portuguese Institute of Oncology (IPO), Porto. Videos were captured during scheduled endoscopic examinations at the hospitals in the past years using Olympus Q240Z endoscope. Optical characteristics of this endoscope include 140° field of view and four way angulation (210° up, 90° down and 100° right/left). The endoscopic videos were recorded on tapes during real endoscopic examinations in the DV format. After acquisition of the videos, relevant frames were selected from them and were annotated by the physicians.

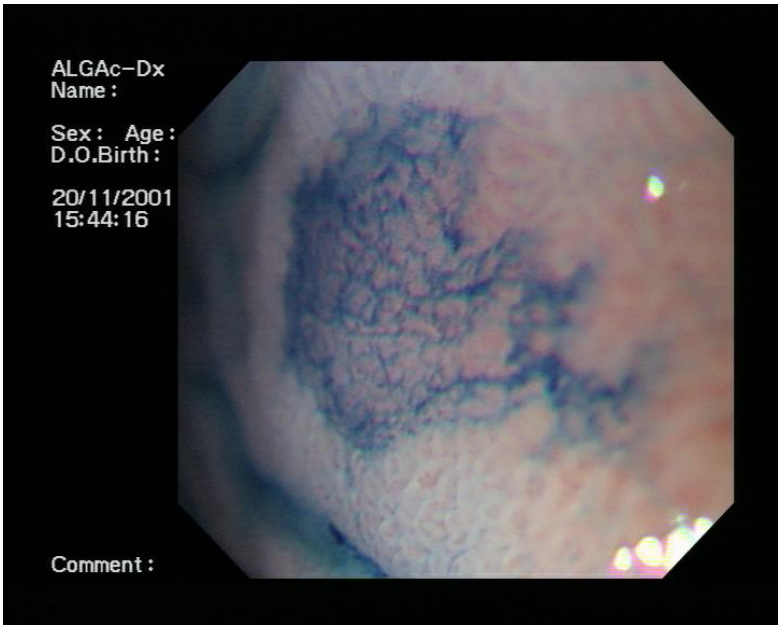


Figure 5.1: Example of an acquired CH image.

5.1.1 Selection of frames from videos

In a real exam, the movement of the probe around the clinically relevant tissues is typically slow therefore, we have many redundant images all of which belong to the same *event*. We will define an *event* as an image of a tissue, which encapsulates some clinical information which is relevant for making diagnosis. In reality, the physician usually selects the most clear shot related with a particular event and integrates it in the final report. If the redundant images are not discarded and all the frames are used for classifier evaluation, a lot of images representing the same *event* may have been used for both training and testing of the classifier when doing cross validation. This will give optimistic results thus yielding an unrealistic estimate of classifier performance

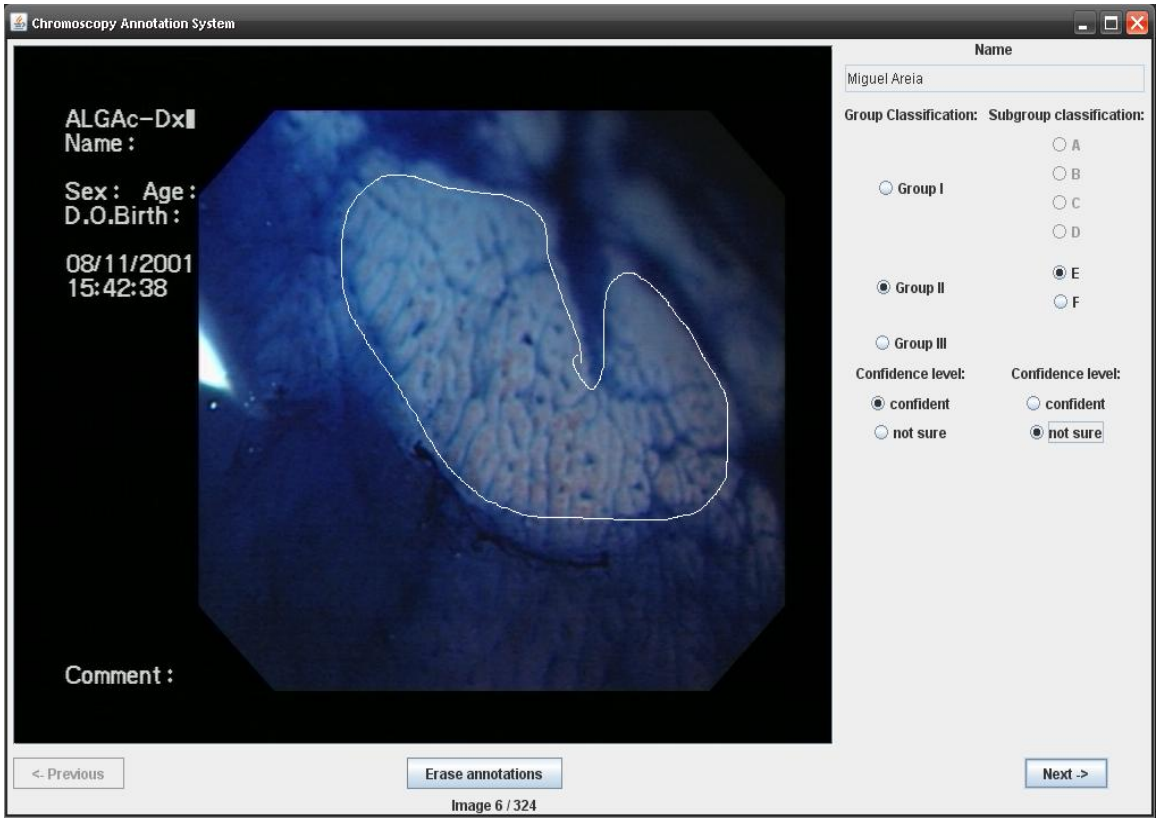


Figure 5.2: A snapshot of annotation software for CH images.

which is highly undesirable. To avoid this, we only use those clear frames that were integrated in the final report by the physicians discarding all others of the same event. With the help of expert clinicians, videos were analysed and a total of 176 images from 28 patients were selected from CH videos given their clinical relevance. A typical image captured during the examinations is shown in Fig. 5.1.

5.1.2 Annotation of frames

Selection of clinically relevant frames from the videos is followed by their manual annotation and classification. Therefore, it is important to develop an annotation software for this purpose. In principle, the software should have high usability for the physicians. Its requirement specifications were discussed with two clinicians and its functionality was based on the Dinis-Ribeiro classification proposal for stomach cancer [111]. The software (Fig. 5.2) was installed in the computers at IPO, Porto for annotation of the acquired images. According to the recommendations of the clinicans, the following functionalities were added to it.

- Ability to mark and erase manual annotations, depicting clinically interesting regions.
- Classify the images into groups and subgroups following Dinis-Ribeiro proposal.
- Specify the confidence level of physician while making the annotation.

The usability of the software is the most important non-functional requirement and therefore a constant interaction and progressive recommendations were taken from its potential users at the IPO, Porto.

5.1.3 Clinical taxonomies for CH images

For the study presented here, we have used the clinical taxonomies introduced by Dinis-Ribeiro classification proposal [111] that was externally validated in [7]. The authors showed that it was possible to effectively diagnose certain gastric lesions and adequately predict neoplasia occurrence using this proposal for CH. According to this proposal, CH images are classified into their respective classes based on color, shape and regularity of pit patterns (Fig. 5.3)

- Group I was defined when the mucosa showed a regular mucosal patterns and no change in color after staining with methylene blue was observed.
- Group II was defined when the mucosa showed regular pattern and stained in blue.
- Group III was defined if neither a clear pattern was noticeable nor a change in color was observed (heterogeneous staining).

Clinical studies show that Group I images are considered normal, Group II cases are considered metaplasia lesions and could lead to cancer lesions. So in this case, the patient should be regularly followed up. Group III are considered metaplasia lesions and the patient should be immediately treated to prevent late and reserved prognostic [111].

5.1.4 Analysing manual annotations

Two expert gastroenterologists annotated and classified the whole set of images in the CH database:

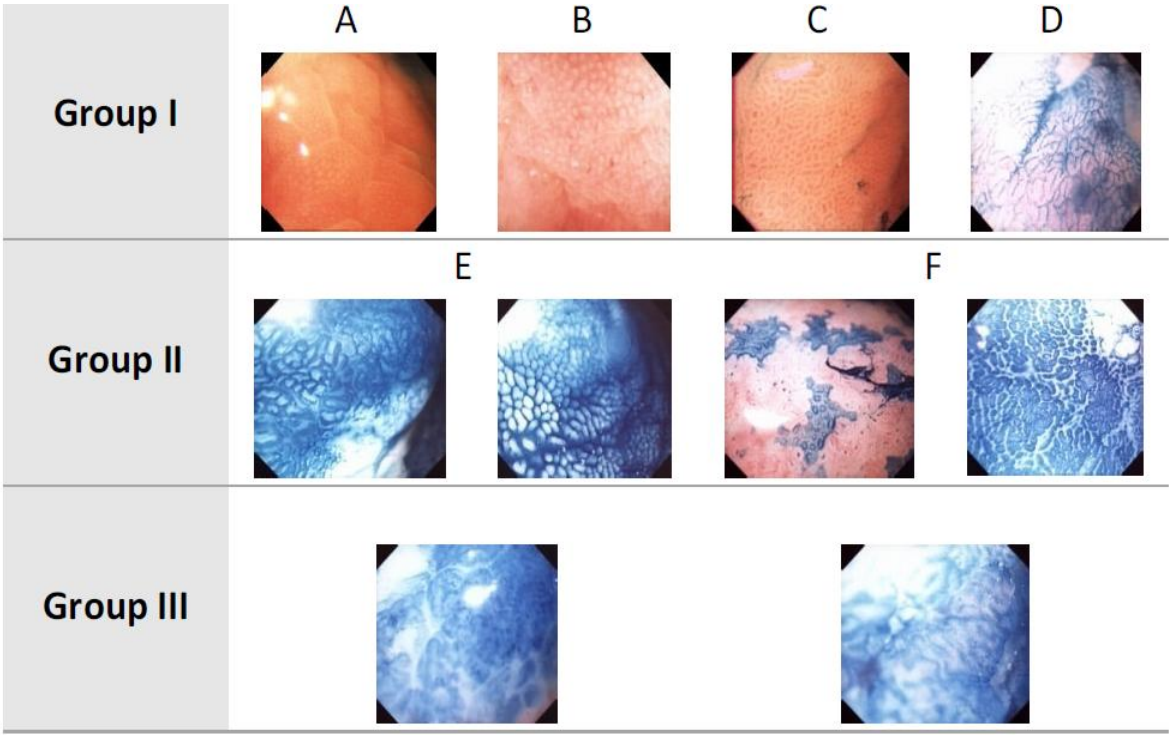


Figure 5.3: Dinis-Ribeiro classification proposal (adapted from [111])

- Mario Dinis Ribeiro, MD, PhD is working with IPO, Porto at Gastroenterology Department and the Faculty of Medicine at the University of Porto.
- Miguel Areia, MD is working with IPO, Coimbra at the Gastroenterology Department.

With the manual classifications made by these two physicians having vast experience in the area of gastroenterology, a high confidence clinical dataset was acquired to be used for our experiments. Afterwards, the manual classifications obtained from both clinician's were compared and a statistical analysis of their annotations was done (Fig. 5.4), which is composed of the following components:

- Confident means that both clinicians were confident of their classification.
- Insecure means that at least one clinician was not confident of his classification.
- Agreement means that both specialists classified an image into the same group.
- Disagreement means that specialists classified the images into different groups.

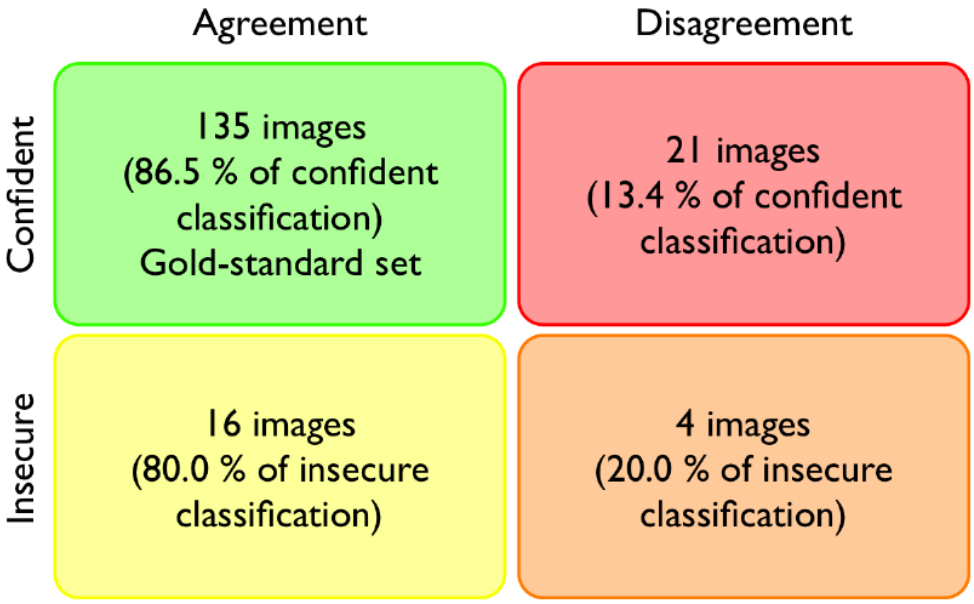


Figure 5.4: Comparison of classification using two physicians

A careful analysis of those scenarios where the physicians disagreed was made by discussing them with both the physicians. Observations showed that in those cases where the physicians disagreed, the physicians marked different clinically relevant regions for classification (Fig. 5.5). To confirm this, the clinically interesting regions of these cases marked by one physician were provided to the other physician and vice versa for classification. In these cases, the classifications of both the physicians for the respective image regions coincided.

For a final computer assisted diagnosis the system should identify the worst possible diagnostic scenario and mark the specific region relevant for such classification. However from a computer vision perspective this restriction is not needed: both regions from the same image can be used for training and testing of classifiers as the physicians are confident of their classification for the particular regions and their respective classifications are same. Also, it is important to maximize the usage of the data to obtain statistically relevant results. Therefore for 20 of those 21 images where the physicians disagreed (Fig. 5.4), two different regions with their respective classifications were generated. For one remaining image, one of the physicians agreed that he made a mistake making the annotation of the other physician valid. This gave us $20 \times 2 + 1 = 41$ images that were added to the gold standard of 135 images to give us a dataset of 176 CH images.

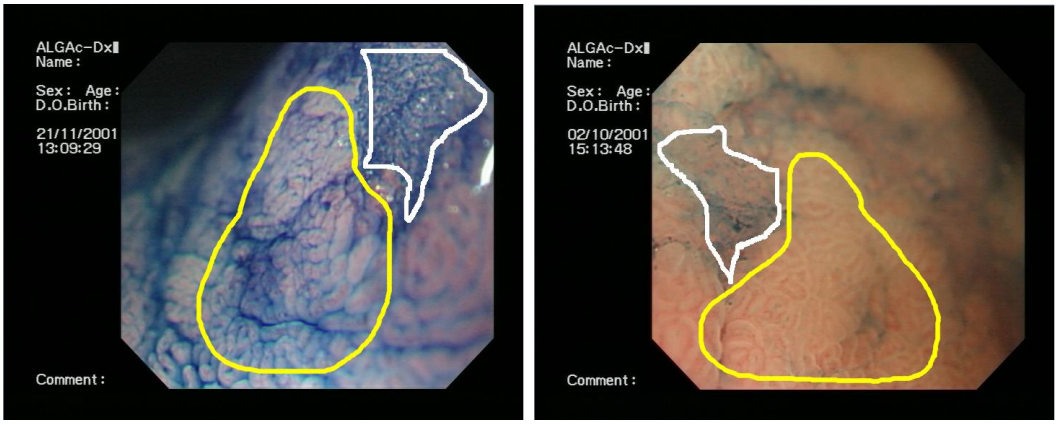


Figure 5.5: Examples of cases where the diagnosis made was confident but there was disagreement between the physicians

A careful visual analysis of the data showed several images from a video sequence having very similar annotated regions from the physicians. These were the images that were probably replicated by the doctors while saving the clinically relevant images. These repetitive images were removed to give us a dataset of 142 images (Table 5.1).

Chromoendoscopy Dataset		
Group I	Group II	Group III
44 (30.98%)	79 (55.63%)	19 (13.40%)

Table 5.1: Composition of CH dataset

5.2 Narrowband imaging (NBI) images

Our NBI dataset was collected at the Karolinska Universitessjukhuse, Sweden during routine clinical work. Videos were captured during scheduled endoscopic examinations at the hospitals in the past years using Olympus GIF-Q160Z endoscope. Optical characteristics of this endoscope include 140° field of view and four way angulation (210° up, 90° down and 100° right/left). The endoscopic videos were recorded on tapes during real endoscopic examinations in the DV format. After acquisition of the videos, the following steps were involved in annotation of the dataset.

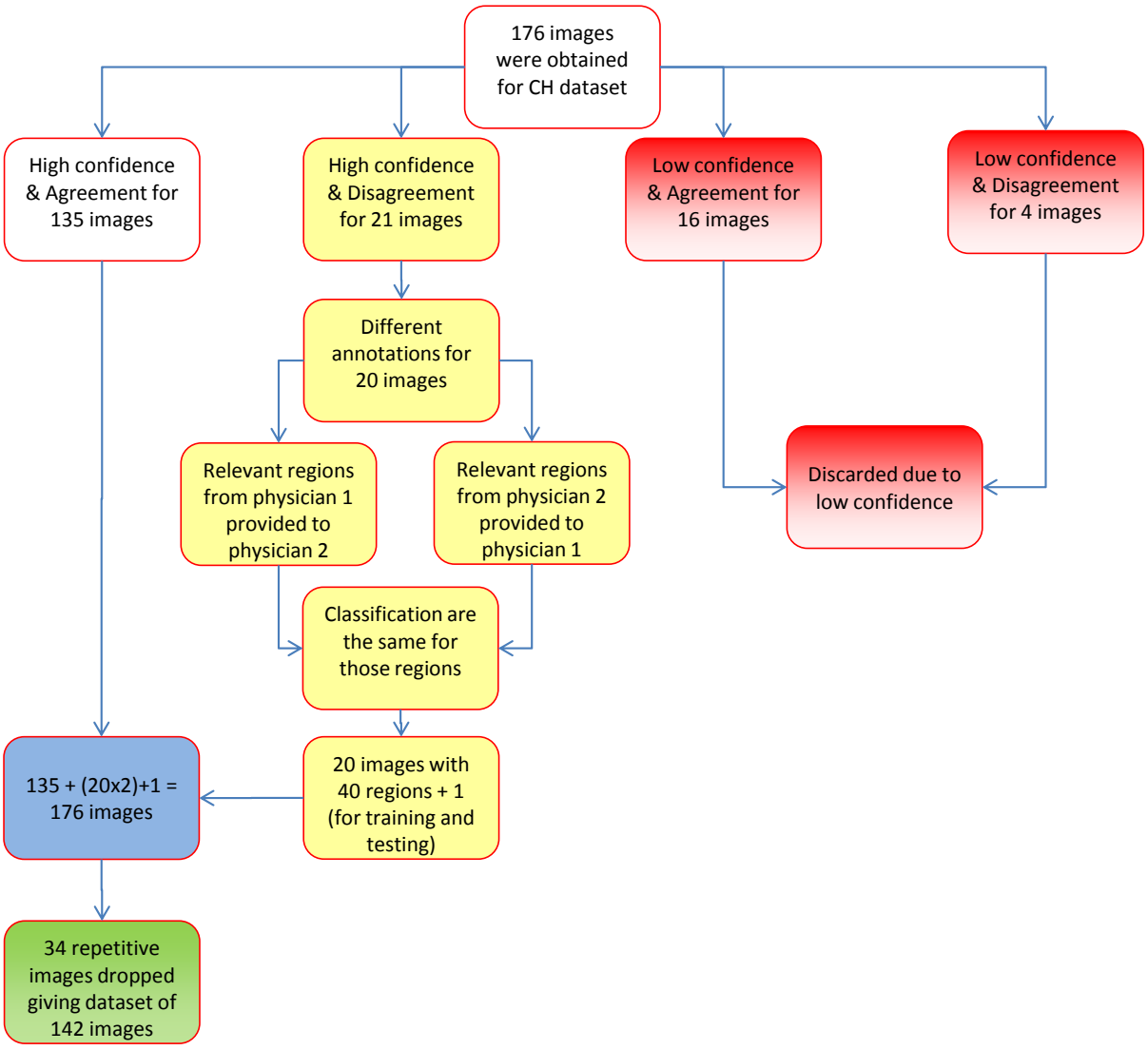


Figure 5.6: Procedure of annotation of CH dataset

5.2.1 Selection of frames from videos

As in CH images, the diagnosis in NBI images is also done based on some clinically relevant frames in the videos. Consequently, acquiring individual clinically relevant images from the videos is one of the first steps towards data collection for NBI. With the help of expert clinicians, videos from 84 patients were analyzed and a total of 250 images, each representing an *event* or tissues surrounding an *event* were selected from

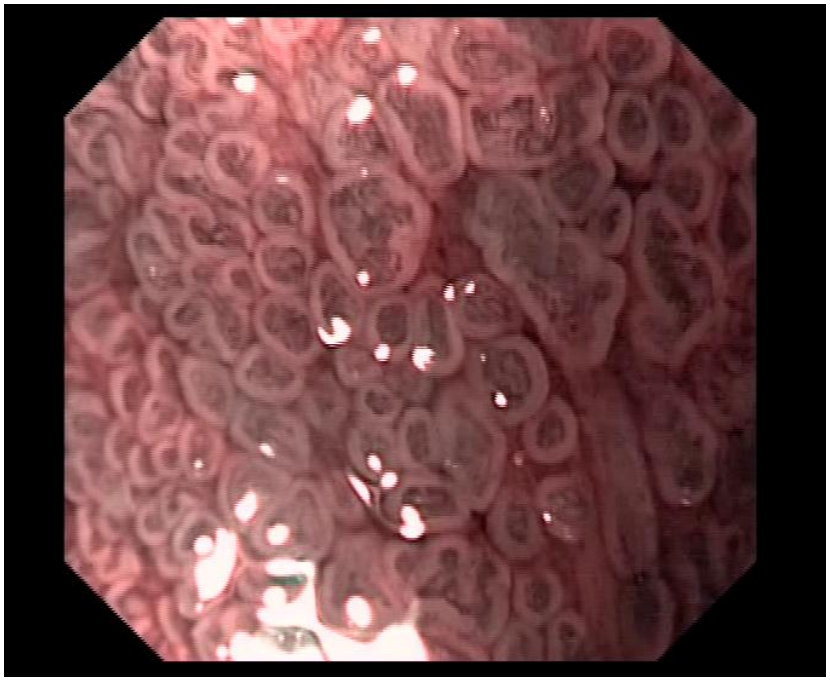


Figure 5.7: An NBI image acquired during endoscopic exam

NBI videos given their clinical relevance.

5.2.2 Annotation of frames

Selection of clinically relevant frames from the videos is followed by their annotation and manual classification. For this purpose, a software was developed to be used by the physicians. The software requirement specification were discussed with the physicians and its functionality was based on Singh's classification proposal for Barrett's esophagus [121] based on physician's recommendations. Progressive interaction with physicians while designing the software was done to ensure full functionality while providing high usability for the physicians. According to the recommendations of the physicians, the following functionalities were added to the annotation software (Fig. 5.8):

- Selection of clinically interesting region from an image.
- Classification according to pit patterns.
- Classification according to vasculature.
- Overall classification of image.

- Confidence level of the physician.

During the design of the software, frequent coordination with the physicians was done to ensure its high usability with minimum effort by the physicians.

5.2.3 Clinical taxonomies for NBI images

Several clinical taxonomies for diagnosing Barrett's esophagus have been proposed in the literature [6][118], these are considered tedious and difficult to apply in routine practice [121] during endoscopy examinations. We have used the clinical taxonomies proposed by R. Singh et al. [121] for diagnosing Barrett's esophagus. According to Singh's proposal, NBI images are classified into their respective classes based on regularity of shape and regularity of pit patterns and vasculature (Fig. 5.9):

- Group I was defined when the round pit patterns and regular microvasculature are observed.

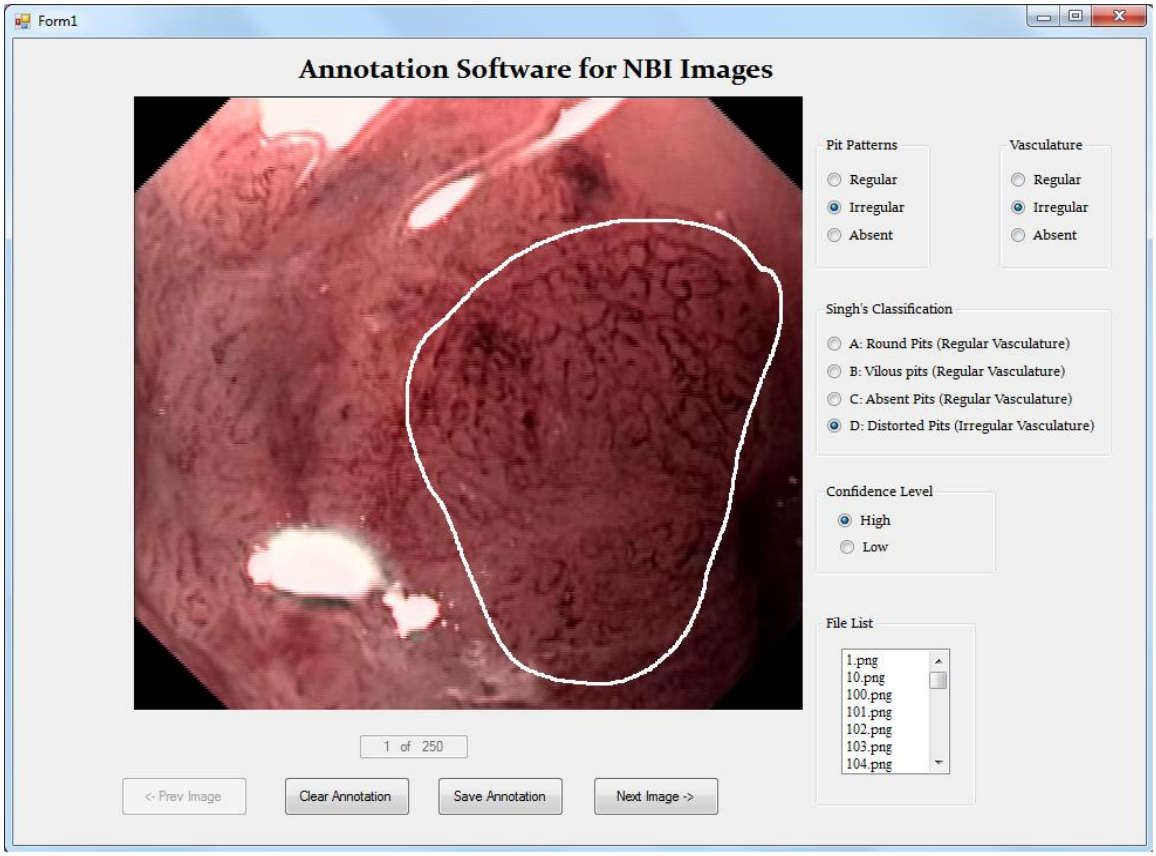
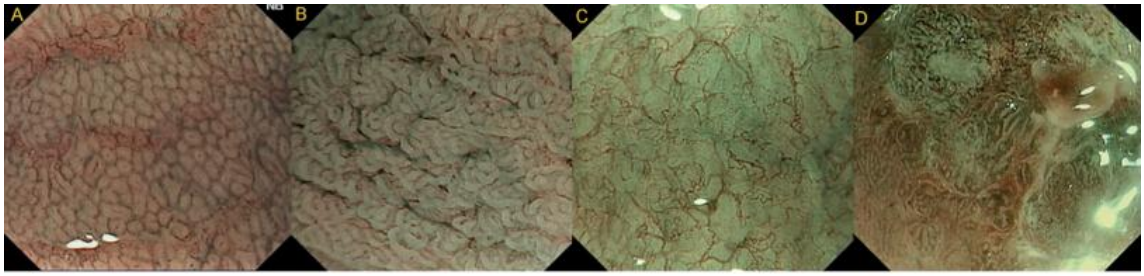


Figure 5.8: Snapshot of NBI annotation software



Type	
I	Round pits with regular microvasculature
II	Villous/ridge pits with regular microvasculature
III	Absent pits with regular microvasculature
IV	Distorted pits with irregular microvasculature

Figure 5.9: Singh's classification proposal (adapted from [121]).

- Group II was defined when villous pit patterns and regular microvasculature are observed.
- Group III was defined when there are no pit patterns but microvasculature is regular.
- Group IV was defined when the pit patterns are distorted and the microvasculature is irregular.

5.2.4 Analysing manual annotations

The NBI database was annotated and classified by Dr. Miguel Areia, who is working with IPO, Coimbra in the gastroenterology department. A clinical dataset was acquired (Fig. 5.10), which was characterized by the following parameters:

- Confident: clinician is confident of his manual annotations and classification.
- Insecure: clinician is not confident of his manual annotations and classification.

Since there was only one physician which annotated the NBI dataset, its inter-observer variability could not be analysed. The NBI images for which the physician exhibited low confidence for classification were discarded. We performed a detailed visual

analysis of the remaining images to ensure that there is no visual similarity between two distinct images in the dataset. The final NBI dataset therefore consists of 224 images (Table 5.2).

NBI dataset			
Group I	Group II	Group III	Group IV
57(25.44%)	73(34.60%)	16(7.14%)	78(34.80%)

Table 5.2: Composition of NBI dataset

5.3 Conclusions

The digestive tract is composed of various organs having different anatomy and structure. For instance, small intestine or esophagus are tube like structures which allow limited fields of view for inspection of the internal walls due to limited number of positions for placing the camera for observing the tissues. Stomach on the other hand is a sac-like structure and it allows a much wider range of fields of view leading to unpredictable camera rotation and illumination changes. Both of these organs therefore pose complementary challenges from a computer vision perspective. An image descriptor which is specifically designed for feature extraction from the esophagus may prove to be comparatively limited for feature extraction from the stomach. The dynamic imaging conditions are therefore caused by two factors:

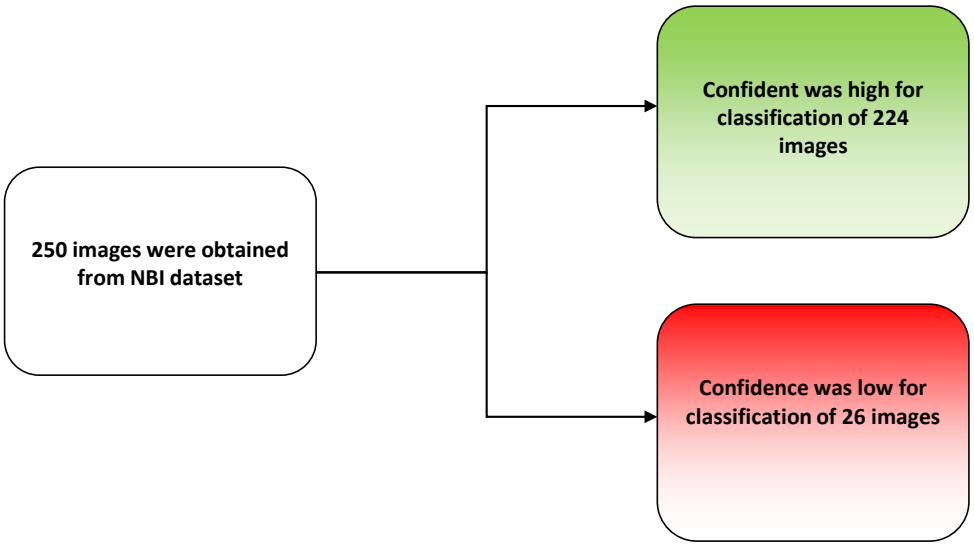


Figure 5.10: Confidence of annotated and classified NBI dataset

- Distinct anatomy of the organ.
- Visual features of images from various imaging modalities.

As discussed in Chapter 1, one of the main objectives of this thesis is to research generic computer vision tools that can be used for various imaging modalities. Given all this, we have acquired images from two complementary GI organs using complementary imaging modalities:

- Vital stained magnification endoscopy images from stomach.
- Magnification narrow-band images from esophagus.

The images obtained from these modalities pose complementary challenges: the former uses staining to obtain enhancement of color and shape of the mucosa whereas the latter uses two narrow bands of the white light exposing the structure of the vessels. The choice of these imaging modalities from a clinical perspective was based on a high degree of accuracy that these modalities have for the detection of cancer in the respective sites in the gastrointestinal tract. Based on the recommendations of clinicians, annotation softwares were developed for both imaging modalities based on their respective proposals for classification of the images. In addition to their use for classification of images, additional functionalities such as representation of confidence level of physicians and subgroup classification etc were also added. Frequent coordination with the physicians was done to ensure easy usability of the software.

From a computer vision perspective, we believe that the complementary nature of the images from the chosen imaging modalities and organs can help us in determining the generic nature of our proposed algorithms in this thesis. In the following chapters we will discuss various novel contributions in computer vision, applied to the specific scenario of gastroenterology.

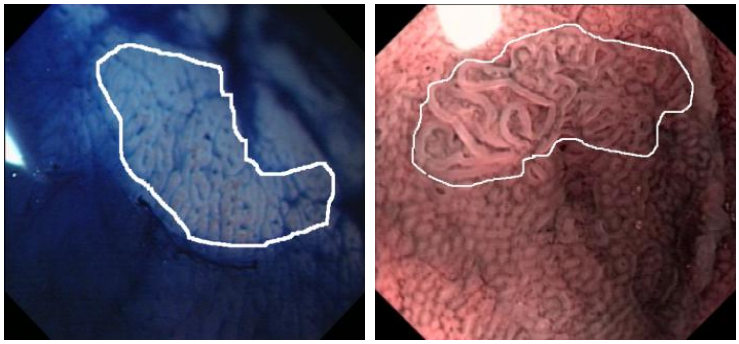
Chapter 6

Segmentation for Gastroenterology

Computer Assisted Decision (CAD) systems in the future can be useful in improving our ability to screen the gastrointestinal tract by providing the physician a second opinion, but for that happen they must overcome one of the most fundamental challenges in computer vision: segmentation. Its need is motivated by the manual diagnosis procedure where the physician is interested in inspecting a specific clinically relevant region in an image. Such a region is expected to encapsulate information that can be relevant for diagnosis (Fig. 6.1). On the other hand, segmentation of endoscopic images is considered a hard task given the very specific nature of gastroenterology (GE) images. This is mainly introduced by the camera conditions and the imaging site itself. Lack of full control over the camera, organ position and their motion are the main reasons for imperfections such as bad illumination, poor focus, severe lens distortion etc. Even under ideal imaging conditions, the specific visual characteristics of GE images such as reduced color spaces or the absence of geometric structures make their segmentation challenging.

Although many segmentation methods exist in the literature and some of them have been used for segmentation of GE images but to the best of our knowledge, there are no articles in the literature which report generic image segmentation methods which are suitable for segmentation of images from multiple imaging modalities. This is because different imaging modalities focus on different challenges and their objectives vary diversely. They expose various distinct features of the images and it is hard for a single segmentation method to give satisfactory results for subjective segmentation of GE images from multiple imaging modalities.

In this chapter, we discuss the use of Normalized Cuts (NCut) [119] for the segmen-



(a) CH image (b) NBI image

Figure 6.1: Typical examples of CH and NBI images with annotations of clinically relevant parts made by the clinicians

tation of GE images. Our choice is motivated by the robustness of NCut to highly textured images and its ability to encapsulate various visual features for image content description (discussed in more detail later). Our contributions are as follows:

- We model the boundaries of clinically relevant regions in GE images as creases in a topographic surface using *multilocal creaseness* features [90]. These are used to construct undirected, weighted graphs of the images. NCut is applied to these graphs for doing image segmentation. In this setting, *intervening contours* are used to measure dissimilarity between the pixels in an image [94].
- The proposed methodology is applied to the scenario of segmentation of GE images from two complementary imaging modalities, Chromoendoscopy (CH) and Narrow-band Imaging (NBI) from two different organs i.e. stomach and esophagus respectively.
- We propose the use of multiple image features including edgemaps, creaseness and color features to further enhance the segmentation performance thus giving high quality image segmentations for the above mentioned imaging modalities.
- We compare the results obtained by NCut with two popular state-of-the-art segmentation methods, i.e. mean shift (MS) [35] and level sets (LS) [89], conceptually using different strategies for image segmentation.

The outline of this chapter is as follows: We will motivate the choice of NCut (Section 6.1) and discuss briefly about NCut image segmentation strategy (Section 6.2), followed by our affinity calculation for GE images (Section6.3). We will later explain our experimental results (Section 6.4) and discuss them (Section 6.5).

6.1 Segmentation for gastroenterology

There is a large variety of segmentation methods available in the literature. Their use depends on the specific objectives and the nature of the image content. GE images are characterized by their diverse texture ranging from coarse to fine texture for normal and cancerous images respectively. One of the considerations for selecting good segmentation methods for GE is therefore avoiding over-segmentation of the images. As discussed in Chapter 3, segmentation according to Forsyth and Ponce [47] can be done using two approaches: segmentation by fitting and segmentation by clustering. Methods based on the former assert that the pixels in an image belong together because together they conform to a model (discussed in detail in Chapter 3). These methods typically rely on gradient information in the images around clinically relevant regions for good segmentation [80][93]. Weak boundaries around these regions result in a compromise on the quality of segmentation for such methods due to lack of convergence of the algorithm. Segmentation by clustering methods can be characterized as either bottom-up (agglomerative) or top-down (divisive) (details in Chapter 3). Agglomerative methods start with individual entities (pixels in an image) and combine them to form large clusters and are most widely used clustering methods. Given that GE images are highly textured, focusing too much on the details can lead to over-segmentation of images. These methods are also expected to be more sensitive to noise (Fig. 6.2). Methods using spatial kernels such as mean shift will perform well as they work on a spatial neighborhood however these require a dynamic adjustment of kernel parameters based on image contents which is not trivial.

Divisive methods, according to [81] offer advantage over agglomerative clustering methods because most users are interested in the main structure of the data, while ignoring its very local characteristics. In this context, several approaches exist which perform feature based clustering [72][74][116] using global characteristics of the image through calculation of image features such as color, texture etc. They however may not be suitable for segmentation as they lose the spatial relationships among pixels in the images [126]. Graph based clustering approaches have the ability to fuse the spatial and feature information in the images. In such approaches, a visual group can therefore be represented as a function of feature similarity, proximity and continuation [126].

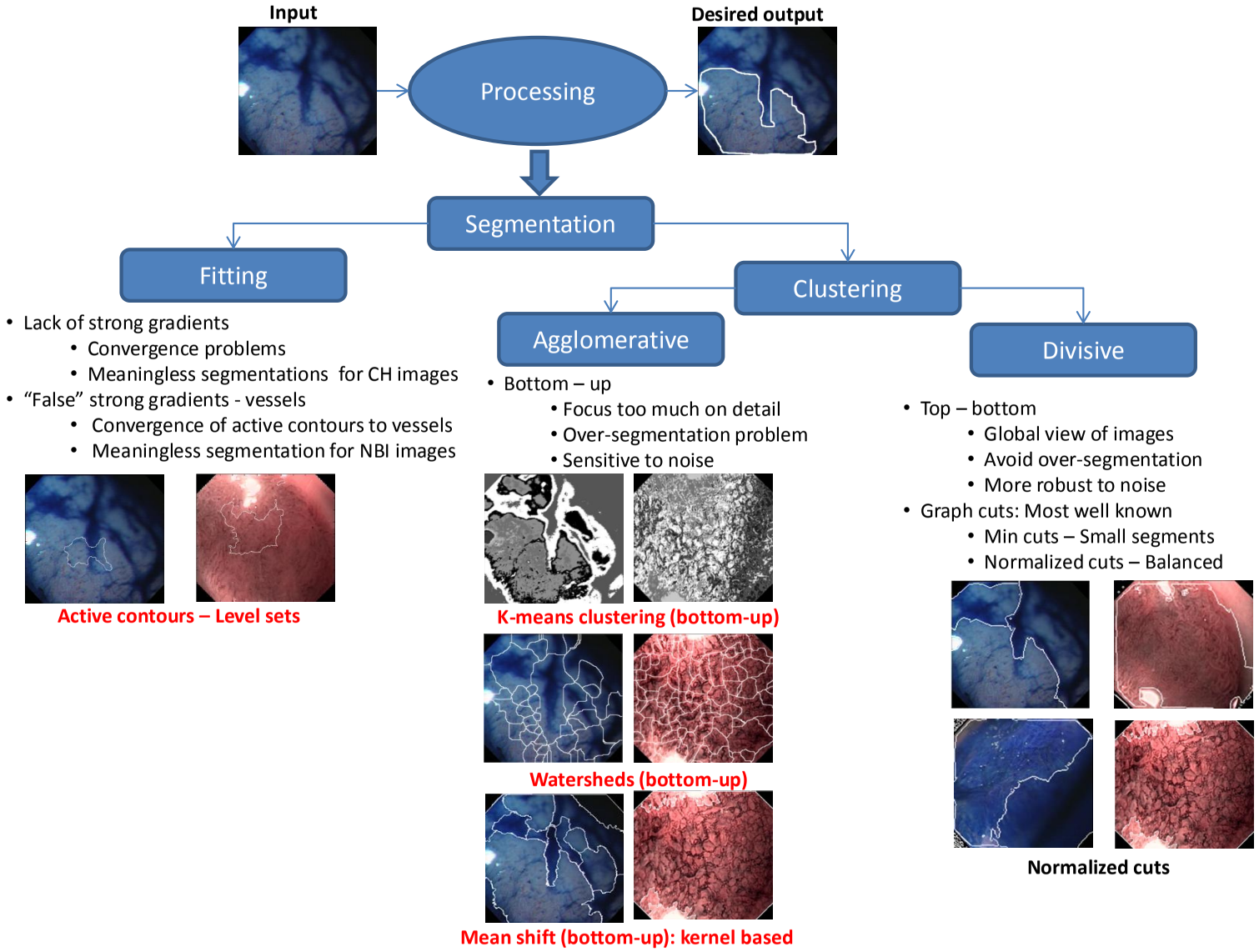


Figure 6.2: Segmentation using various methods. Active contours fail to converge to the boundaries of clinically relevant regions. Agglomerative clustering methods are sensitive to noise. Methods using kernels can perform well by a proper parametrization of the kernel based on image data. NCut is the most relevant choice given its more global approach to segmentation

So far, several graph cut based methods have been developed for image segmentation [43][94][17], we have used Normalized Cuts (NCut) for our experiments, which is a divisive clustering method. NCut can robustly generate balanced clusters in a graph and is superior to other spectral clustering methods such as average cuts and average association [119]. It addresses clustering in a global optimization framework and guarantees a globally optimal solution unlike other clustering methods which operate either on pixel level or kernels of some specific size. This makes NCut robust to noise [144][119] and in the meanwhile we also overcome the traditional problem of over-segmentation. Another important consideration is that NCut is less dependent on the parameters as compared to other clustering methods. It is however highly dependent on the images features as in any other segmentation method.

6.2 Normalized cuts (NCut)

The NCut method is a graph theoretic approach for solving the perceptual grouping problem in vision. It is a clustering based segmentation approach in which an image is represented as a weighted, undirected graph whose nodes correspond to individual pixels and graph weights are based on affinity between the pixels.

6.2.1 Graph based formulation of images

Given a graph $G = (V, E)$, where the nodes in the graph are the points in the feature space and an edge is formed between every pair of nodes (Fig. 6.3). The weight on each edge, w_{ij} is a measure of similarity between nodes i and j. In grouping, we seek to partition the set of vertices into disjoint sets $V_1, V_2, ..., V_m$ where by some measures the similarity among the vertices in a set V_i is high and across different sets V_i, V_j is low. Therefore, the following fundamental questions need to be addressed:

- What is the precise criteria of partition.
- How can the partition be computed efficiently.

In image segmentation and data clustering community, there has been much previous work using variations of the minimal spanning tree or limited neighbourhood set approaches. Although those use efficient computational methods, the segmentation

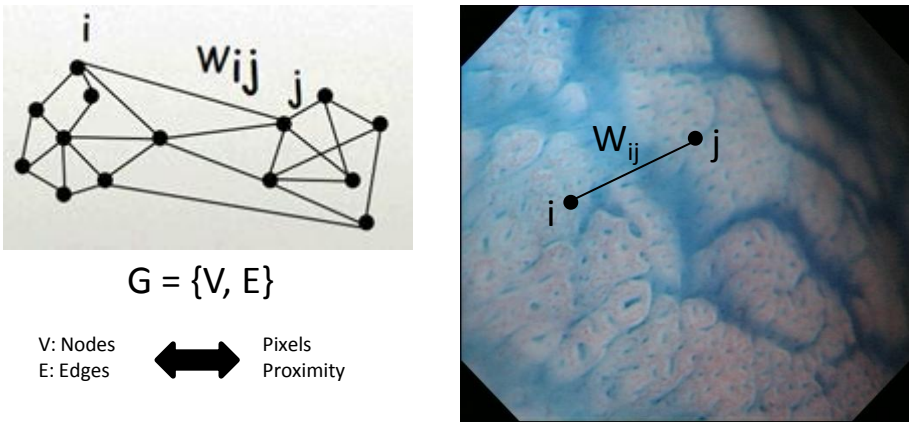


Figure 6.3: Graph based image segmentation methods

criteria used in most of them are used on local properties of the graph. Since perceptual grouping is about extracting the global impressions of a scene, this partitioning criterion falls short of the main goal [119].

We use a graph theoretic criterion for measuring the goodness of an image partition - *Normalized cuts* (NCut). It is a very popular method which has been very useful for segmentation of images from a wide variety of image categories. In the following subsections, we will discuss in detail about the partitioning strategy used in NCut. We also discuss about the methodology for the construction of affinity matrix using a variety of image features and use it for segmentation of GE images.

6.2.2 Grouping as graph partitioning

A graph $G = (V, E)$ can be partitioned into two disjoint sets, A and B where $A \bigcup B = V$, $A \bigcap B = \phi$, by simply removing the edges connecting the two parts. The degree of dissimilarity between these two pieces can be computed as the total weight of the edges that have been removed. In the graph theoretic language, it is called the *cut*:

$$cut(A, B) = \sum_{u \epsilon A, v \epsilon B} w(u, v) \tag{6.1}$$

The optimal bi-partitioning of a graph is the one that minimizes this cut value. Although there are exponential number of such partitions, finding the minimum cut of a partition is a well studied problem and there exist efficient algorithms for solving

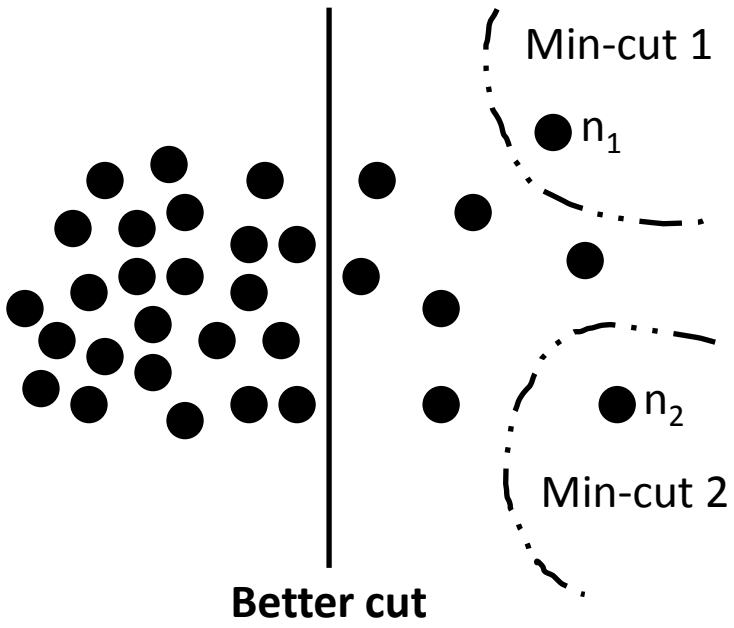


Figure 6.4: Illustration of Minimum cut. This results in the formation of isolated subgraphs due to too much focus on the individual graph edges. A better cut, yielding an optimal global bi-partition of the graph is also shown in the figure.

it. Wu and Leahy [140] proposed a clustering method based on this minimum cut criterion. In particular, they seek to partition a graph into k-sub groups, such that the maximum cut across the subgroups is minimized. This problem can be efficiently solved by recursively finding the minimum cuts that bisect the existing segments. As shown in Wu and Leahy's work, this globally optimal criterion can be used to produce good segmentation on some of the images. However as Wu and Leahy also noticed in their work, the minimum cut criteria favors cutting small sets of isolated nodes in the graph. This is not surprising since the cut defined in eq. 6.1 increases with the number of edges going across the two partitioned parts. Fig. 6.4 illustrates such a case. Assuming the edge weights are inversely proportional to the distance between the two nodes, we see the cut that partitions out node n_1 or n_2 will have a very small value.

In fact, any cut that partitions out individual nodes on the right half will have smaller cut value than the cut that partitions the nodes into the left and right halves. To avoid this unnatural bias for partitioning out small sets of points, a new measure of

disassociation between the two groups was proposed. Instead of looking at the value of total edge weight connecting the two partitions, this measure computes the cut cost as a fraction of the total edge connections to all the nodes in the graph. We call this disassociation measure the normalized cut (NCut) [119].

$$NCut = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)} \tag{6.2}$$

where $assoc(A,V)$ denotes the total connection from nodes in A to all nodes in the graph and $assoc(B,V)$ is similarly defined. With this definition of the disassociation between the groups, the cut that partitions out small isolated points will no longer have small NCut value, since the cut value will almost certainly be a large percentage of the total connection for that small set to all other nodes. Unlike the minimum cut criteria which has a bias towards cutting small sets of nodes, the NCut presents an unbiased criteria. Also an approximate optimal partition minimizing this criterion can be found as a solution to a generalized eigensystem.

6.2.3 Computing the optimal partition

Given a partition of nodes of a graph V into two sets A and B, let W be an $N \times N$ symmetric matrix with $W(u,v) = w_{uv}$. Let D be an $N \times N$ diagonal matrix with $D(u,u) = \sum_v W(u,v)$. Shi and Malik showed that minimizing NCut (eq. 6.2) can be reduced to minimizing a Rayleigh quotient [119]:

$$\min_x NCut(x) = \min_{\mathbf{y}} \frac{\mathbf{y}^t(D-W)\mathbf{y}}{\mathbf{y}^t D\mathbf{y}} \tag{6.3}$$

With the condition that $\mathbf{y} = \{a,b\}^N$ is a binary indicator vector specifying the group identity for each pixel i.e., $y_u = a$ if u belongs to group A and $y_v = b$ if pixel v belongs to B and N is the number of pixels. The solution to eq. 6.3 can be minimized by solving a generalized eigenvalue system of the form:

$$(D-W)\mathbf{y} = \lambda D\mathbf{y} \tag{6.4}$$

The optimal solution to this eigenvalue system corresponds to the second smallest eigenvector. Therefore, a graph can be subdivided by finding the second smallest eigenvector of eq. 6.4 and can subdivide the existing graphs, each time using the eigenvector with the next smallest eigenvalue (Please refer to [119] for full derivation).

6.2.4 Normalized cuts applied to image segmentation

Normalized cut criterion is applied to image segmentation by treating each pixel in the image as a point in some arbitrary feature space, represented as a node in the graph. The edge weights between pixel-pairs are specified based on some similarity between the pixels in terms of the features considered. Segmentation is achieved by partitioning the graph into coherent groups using the NCut criterion. So the grouping algorithm consists of the following steps:

- Given an image, set up a weighted graph $G = (V, E)$, and set the weight on the edges connecting two nodes being a measure of the similarity between the two nodes using some similarity measure or a set of measures. Form the matrices W and D.
- Solve $(D - W)\mathbf{y} = \lambda D\mathbf{y}$ for eigenvectors with the smallest eigenvalues.
- Use the eigenvector with second smallest eigenvalue to bi-partition the graph.
- Decide if the current partition should be sub-divided and recursively repartition the segmented parts if necessary. A threshold for NCut could be set so that the recursion stops when the normalized cut value between two partitions at any stage is greater than this threshold.

The quality of segmentation depends on the choice of the similarity measure chosen to represent edges in the graph and therefore, on the resulting weight matrix W in the normalized cuts procedure. Any of the low-level attributes like brightness, color, texture etc. at a pixel that are considered suitable for a particular class of images can be used to form W that can be input to the NCut grouping algorithm to yield hierarchical partitions sequentially. Using each attribute individually has its own advantages/disadvantages depending on the contents of the images being segmented.

6.3 Affinity Calculation

The graph $G = (V, E)$ is constructed by taking each pixel as a node and defining the edge weight w_{uv} between nodes u and v. This definition plays a decisive role in determining the overall performance of the image segmentation process. To calculate the edge weights, we have to first define an appropriate feature space, which can summarize the visual characteristics of our images. In general, features like color, texture,

statistical characteristics are useful for quantifying these characteristics (discussed in detail in the next section). The weight w_{uv} between two pixels can be defined as:

$$w_{uv} = e^{-\|F(u)-F(v)\|_2^2} \times \begin{cases} e^{-\|X(u)-X(v)\|_2^2} & if \|X(u)-X(v)\|_2 < r \\ 0 & otherwise \end{cases} \tag{6.5}$$

Where $X(u)$ is the spatial location of the node u and $F(u)$ is the feature vector. The criteria used for calculating the feature term in eq. 6.5 is a deciding factor in the resulting segmentation. Since we have a diversified range of images consisting of structured pits, highly textured localized variations and images showing excessive vasculature, the simplest visual features such as intensity are not adequate. We consider the following visual features very useful for segmentation of images in our scenario.

6.3.1 Multiscale edges

Given that the visual perception of various objects in the images is mainly accomplished by the boundaries of these objects, edge maps are one of the most important features for segmentation of an image. For calculating edgemaps, we have employed filters used by Stella [143]. These are designed to detect the edges at different scales and different orientations. These pairs of filters, denoted as $F_o(\sigma)$ and $F_e(\sigma)$ differ in their spatial phases. The odd-phased filters are first order derivatives whereas even-phased filters are second order derivatives, both smoothed with Gaussians specified by their standard deviations. An image I when filtered using this filter pair can be given as

$$E(\sigma) = (I * F_o(\sigma))^2 + (I * F_e(\sigma))^2 \tag{6.6}$$

Where $*$ is the convolution operator. It is important to note that $E(\sigma)$ is a function of scale of the filter. Typically the filter responses are calculated using filters at different scales and fusion of information from these responses is done using the following strategy

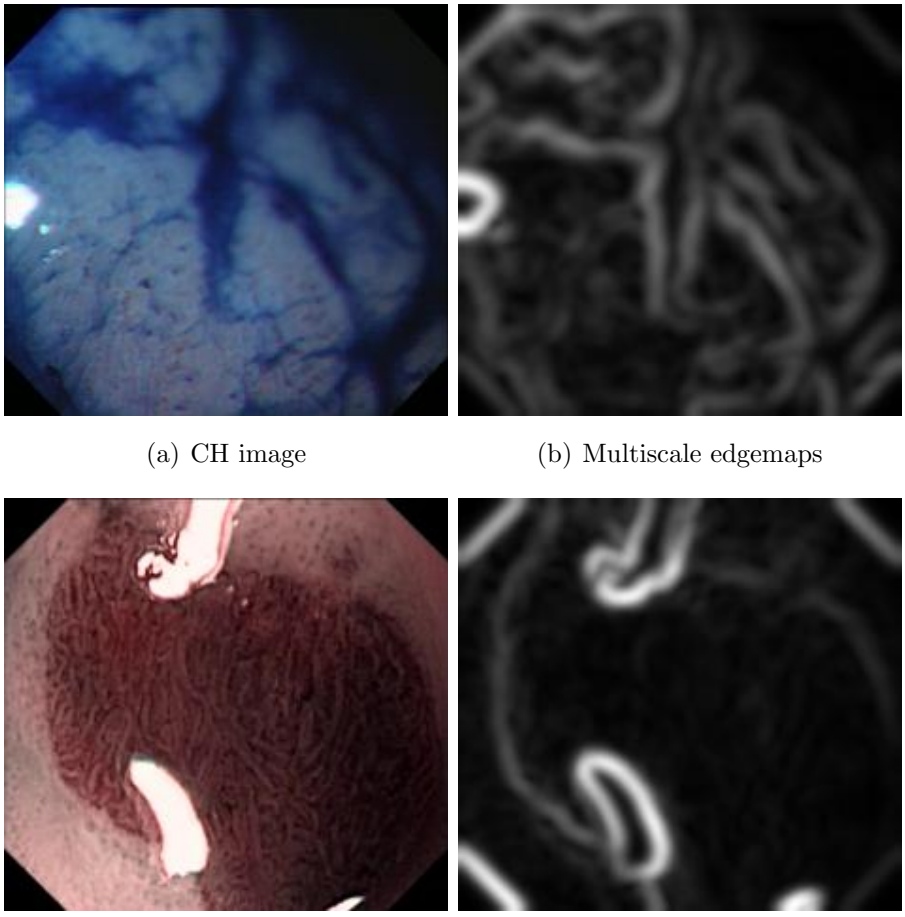
$$E^* = \max_{\sigma} E(\sigma) \tag{6.7}$$

Where E^* is the final energy from the filter having the highest energy. Responses

resulting from small scale filters respond strongly to local texture while those from high scale filters respond strongly to coarse features. Although criteria mentioned in equation 6.7 loses some information, it complements both local and global features in an image. The affinity from edgemaps is calculated using the following mapping [94]:

$$D_{E^*} = 1 - e^{(-E^*)} \tag{6.8}$$

This mapping of filter energy ensures that the values of affinity range between 0 and 1, indicating high or no affinity respectively.



(a) CH image (b) Multiscale edgemaps

(c) NBI image (d) Multiscale edgemaps

Figure 6.5: A visual illustration of multiscale edgemaps extracted from gastroenterology images

6.3.2 Multilocal Creaseness

Edges are one of the most traditional visual features for image segmentation. However in imaging scenarios where the images have rich texture content, extraction of edges can result in an enhancement of the local texture in the images. Therefore when these features are used for applications such as image segmentation, an over segmentation of images is a typical outcome which is undesirable in most cases. To tackle this issue, a more adequate descriptor is preferred, that can disregard the local texture in the images while emphasizing on the directional gradients that typically correspond to the boundaries of objects in the images. Creaseness features provide an intuitive solution to this problem and several measures of creaseness have the potential to serve the purpose [112][49][127]. In this paper, we have used the so-called multilocal level set extrinsic curvature with enhancement by structure tensor (MLSEC-ST) operator proposed by Lopez et al. [90] to enhance the ridge and valley patterns in the images. MLSEC-ST has the potential to deal with the heavily jagged texture efficiently and therefore shows superior performance compared with other algorithms on irregular and noisy landscapes [134][90]. It is therefore a more suitable candidate for predicting boundaries of clinically interesting regions in gastroenterology images.

Let $L(x)$ be the gray-level image being $x = (u, v)$ where u are columns in the images and v are the rows, $G_\sigma(x)$ be a 2D Gaussian of standard deviation σ and $s(x) = [\ \partial_u L_\sigma(x) \quad \partial_v L_\sigma(x)\ ]^t$ be the gradient vector field of the image smoothed by the Gaussian i.e. $L_\sigma(x)$. We define the structure tensor as follows:

$$S(x) = \begin{bmatrix} \partial_u^2 L_\sigma(x) & \partial_u L_\sigma(x)\partial_v L_\sigma(x) \\ \partial_v L_\sigma(x)\partial_u L_\sigma(x) & \partial_v^2 L_\sigma(x) \end{bmatrix} \tag{6.9}$$

Where $\partial_u L_\sigma(x)$ is the partial derivative of $L_\sigma(x)$ with respect to u and $\partial_u^2 L_\sigma(x)$ is the second order partial derivative of $L_\sigma(x)$ with respect to u. $\partial_v L_\sigma(x)$ and $\partial_v^2 L_\sigma(x)$ are similarly defined as derivative with respect to v. The structure tensor assumes that every point has a preferred orientation, which can be checked by defining a confidence measure: for each orientation, we associate a value $D_{Cr} \epsilon [0, 1]$, which can be computed from the eigen values of $S(x)$. Similarity of eigen values implies isotropy and as a result, D_{Cr} should be close to zero. A suitable function for D_{Cr} is as follows:

$$D_{Cr} = 1 - e^{-(\lambda_1(x) - \lambda_2(x))^2} \tag{6.10}$$

Where $\lambda_1(x) > \lambda_2(x) \geq 0$ are the eigen values of the structure tensor $S(x)$. Our

experiments show that creaseness features enhance the tissue boundaries significantly while suppressing the local texture of the regions in the images (Fig. 6.9). These features are therefore suitable for segmentation of tissue boundaries of gastroenterology images.

6.3.3 Color features

Although more complex color descriptors are available in the literature [96], we will focus on standard color spaces for feature extraction from the images. The most commonly used color space is RGB although it has an important limitation [115]: The inability of distance measures like Euclidean distance to capture the perceptual color differences in RGB space.

The CIE (Commission Internationale de l'eclairage) has defined two color spaces: L*u*v and L*a*b. Their most distinct advantage over other spaces is that they are designed to closely approximate the perceptually uniform color spaces [126]. In both of these, L* is the lightness component and defined in both spaces in the same way. These two spaces are different in their chromaticity components and in practice it is considered that there is no clear empirical advantage of using one over the other [126]. In this work, we employed L*u*v color space for extracting color features of the images (App. B.2). If c_u and c_v are colors of pixels u and v respectively, the difference between the colors can be represented as:

$$D_{Co} = 1 - e^{-||(c_u - c_v)||_2} \tag{6.11}$$

where $||.||_2$ is the vector norm operator.

6.3.4 Comparison of features for affinity

For generating the affinity matrix we need to calculate affinity measures between the sets of pixels in the images. In our implementation, we have used the intervening contours framework for calculation of affinity matrix [94]. Fig. 6.6 illustrates the intuition behind this idea: On the left is an image, the middle and right figure show a magnified part of the original image. In the middle and right image, it is clear that there is a strong edge separating p_3 from p_1 and p_2. Thus, we expect p_1 to be much more strongly related to p_2 than p_3. This intuition carries over in our definition of

dissimilarity between two pixels: if the straight line connecting any two points in an image is separated by a strong edge, they have lower affinity. Mathematically, affinity between two pixels can be written as:

$$w_{uv} = 1 - \max_{x \epsilon l_{uv}} D \tag{6.12}$$

Where w_{uv} is the affinity between pixels u and v, D is the magnitude of features extracted using edgemaps, creaseness or color differences and l_{uv} is the line connecting the pixels u and v.

6.3.5 Combining multiple visual features

As discussed previously, one of the strengths of NCut is that it readily admits various visual features for segmentation of the images. In our implementation, we have used the following strategy to combine affinities obtained using various visual features [94]

$$W = W_{D_{E^*}} \times W_{D_{Cr}} \times W_{D_{Co}} \tag{6.13}$$

where $W_{D_{E^*}}, W_{D_{Cr}}$ and $W_{D_{Co}}$ are affinites obtained using edgemaps, creaseness and color features respectively. This multiplication of affinities ensures that if one visual feature gives very low similarity between two pixels, the overall similarity between the pixels will be low.

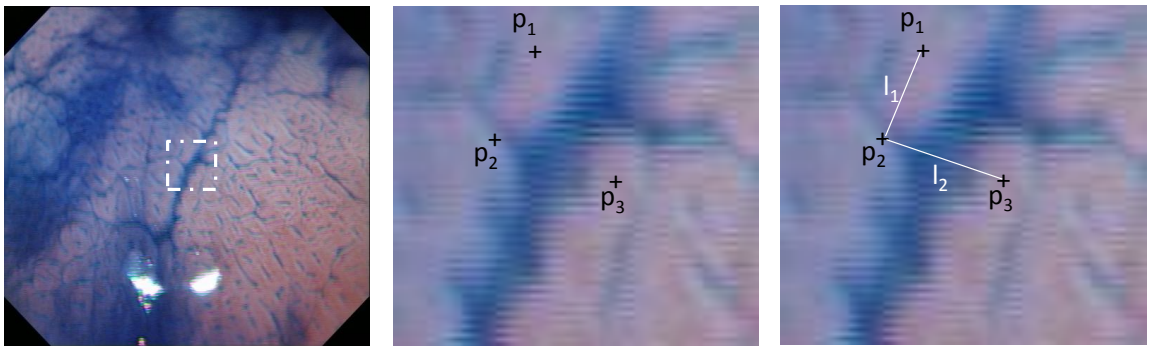


Figure 6.6: Intervening contours framework. Analysing an image (Left), three points are chosen for illustration. Points p_1 and p_2 have intuitively higher affinity as compared to any other pair of points.

6.4 Experiments

For our experiments, we have used the implementation of NCut by [1]Shi and Malik for our segmentation experiments. We have evaluated the performance of our proposed segmentation method for segmentation of lesions in GE images that were manually annotated by the physicians. We have used various visual features individually and as their combination to assess the variation in segmentation performance with the use of various visual features from the images.

6.4.1 Comparison metrics

The quantitative accuracy between the segments obtained by segmentation methods and the manual annotations has been computed using the well known Dice Similarity Coefficient (DSC) metric [113] (named after Lee Raymond Dice and measuring the information shared by two different sets). If we represent an annotation as A and an image segment as P, an overlap score between them can be represented as:

$$DSC = \frac{2(A \cap P)}{A + P} \tag{6.14}$$

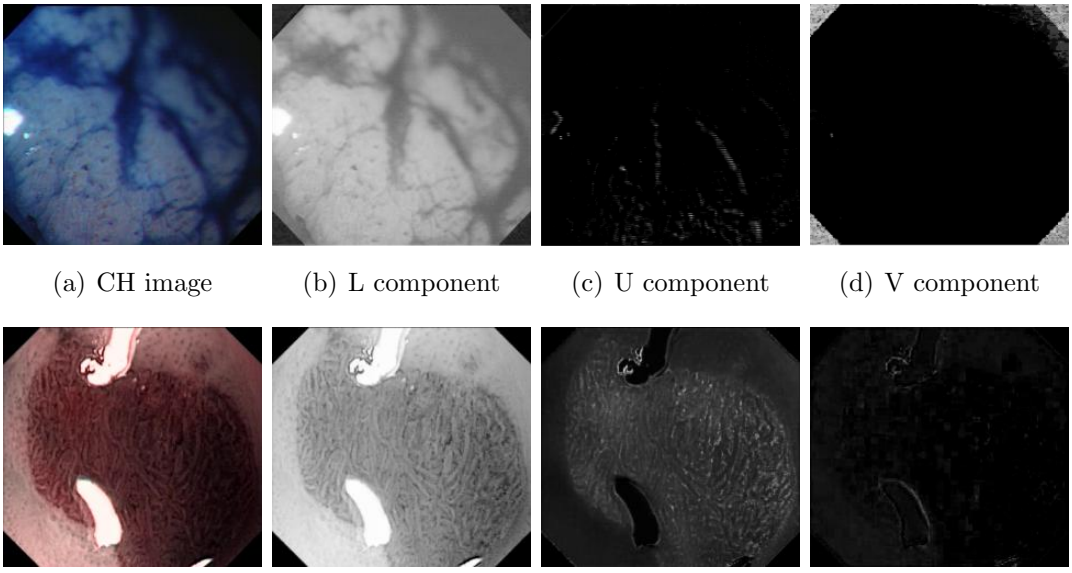
The values of DSC range between 0 and 1 for zero overlap and identical contours respectively, for annotated and segmented regions.

6.4.2 Segmentation using a single visual feature

When using individual image features, we get similar performance for image segmentation using either multiscale edgemaps or creaseness features (Table 6.1). However when using LUV color features, we get the worst segmentation performance. The relative conclusions are the same for both imaging modalities. To understand this, we will analyse the individual components of an LUV image. A visual illustration of individual channels shows that for these images the color components do not have much information (Fig. 6.7).

These components are mostly dark as typically there is not much color variation in gastroenterology images. These images consist of shades of similar color and thus the colors of most of the pixels in an image are confined to a smaller area in the color

[1]http://www.cis.upenn.edu/∼jshi/software/



(a) CH image (b) L component (c) U component (d) V component

(e) NBI image (f) L component (g) U component (h) V component

Figure 6.7: CH and NBI images with luminance (L) and chrominance components (U and V) visualized. The chrominance components are mostly dark thus elucidating on the fact that color features encapsulate less information in GE images.

space affecting the richness of the extracted color features in the images. Integrated in intervening contours, color features are expected to yield a homogeneous affinity matrix given that the image features in this case are local brightness distances which are sensitive to noise and local texture, and would be similar for most pixel pairs, thus creating bad partitions in the images.

For calculating Edgemaps, filters at 3 different scales were used (tuned to get the best segmentation results, relative conclusions using different parameters are the same). Edgemaps focus more on segmenting images based on richness of texture and therefore they focus on segmenting regions based on coarse or fine texture. This happens because for fine texture, the edge magnitudes are high and affinities are low but global optimization framework of NCut prevents formation of very small segments. On the other hand the clinically non-relevant regions (regions that do not allow a distinction between lesion and normal tissue) have typically low texture and thus the pixels in that region show high affinity, increasing the cost of the cut in those regions. The boundary therefore eventually settles at the spatial locations that segregate rich and coarse textured regions (Fig. 6.8). Creaseness features on the other hand tend to enhance the directional gradients while suppressing the local texture (Fig. 6.9).

This increases the cost of creating the cuts within the clinically relevant and non-

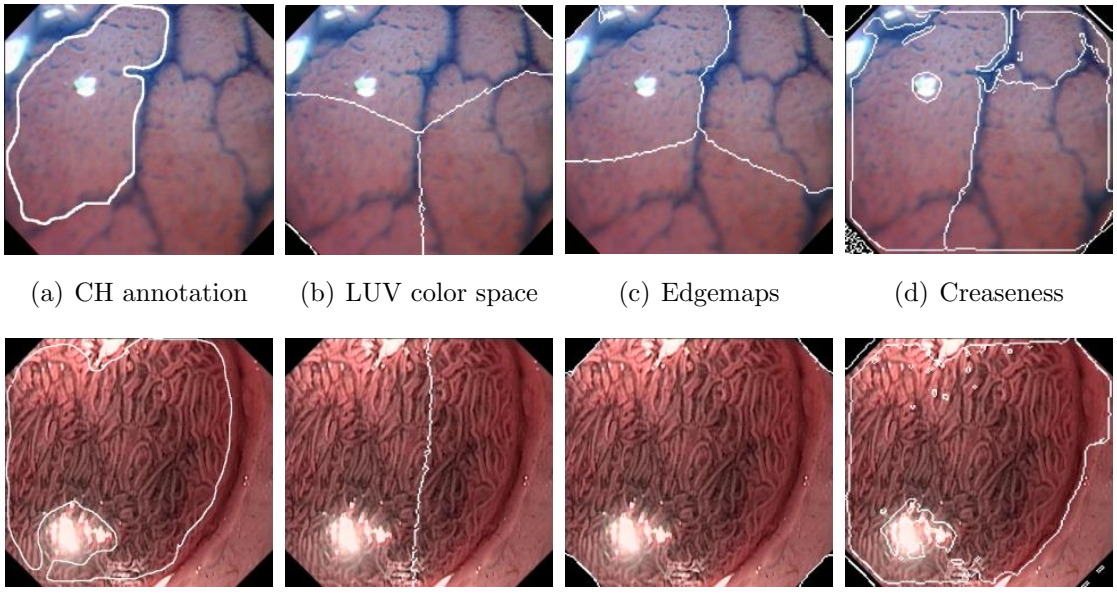
Table 6.1: Segmentation results for both gastrointestinal imaging modalities using a single visual feature.

Visual Features	CH	NBI
Color	0.53	0.64
Edgemaps	0.62	0.77
Creaseness	0.62	0.78

relevant regions. Integrated in NCuts segmentation framework, the boundaries are therefore established around the tissue whose directional gradient was enhanced as the cost of creating such a cut is minimum. From these results, we conclude that edgemaps and creaseness features have complementary objectives: Edgemaps focus on regions (richness or coaresness) while creaseness focuses on the boundaries (Fig. 6.9). It will therefore be interesting to analyse image segmentation when a combination of edgemaps and creaseness are used for feature extraction from the images.

6.4.3 Segmentation using multiple features

When using multiple features for segmentation, we get the best performance for segmentation of both CH and NBI images with reasonably high DSC (Table 6.2)



(a) CH annotation (b) LUV color space (c) Edgemaps (d) Creaseness

(e) NBI annotation (f) LUV color space (g) Edgemaps (h) Creaseness

Figure 6.8: Normalized cuts Segmentation using different visual features

when a combination of edgemaps and creaseness are used for feature extraction from the images. In particular, we get significant improvement in the segmentation results for NBI images (Fig. 6.10). For NBI images, rich texture is attributed to the exposure of local texture as well as vasculature in the images. Since there is rich homogeneous texture throughout the NBI images, edgemaps typically give higher overall edge magnitudes and thus lower affinities between the pixels throughout the image. The creaseness on the other hand enhance oriented texture such as the vasculature and suppress the local texture in the images. When used individually both these features create balanced partitions in the images by creating similar sized segments. When combined together, creaseness features complement the edgemaps with their low magnitudes in the image due to local texture suppression thus creating stronger affinites within the image. Moreover, excessive focus of creaseness on creases in the image (vasculature and pits) is balanced by the edgemaps which collapse the energy of edges in the images into various filters. Thus the overall energy of edgemaps is relatively much lower than creasesness. These complementary characteristics of edgemaps and creaseness when exploited together generate more relevant partitions of the images from a clinical perspective.

6.4.4 Comparison with other methods

Many state-of-the-art methods for image segmentation exist in the literature and based on our survey, we have compared the performance of NCut with two other segmentation methods, mean shift [34] and variational level sets [89]. These methods offer alternative conceptual formulation of the problem of image segmentation. Mean shift (MS) is a kernel based agglomerative clustering method in which the pixels are combined together in a spatial-range domain by concatenating their spatial and color values into a single vector. A kernel of specific size is then used to find the modes

Table 6.2: Segmentation results for both gastrointestinal imaging modalities using multiple features.

Visual Features	CH	NBI
Color + Edgemaps	0.6	0.70
Color + Creaseness	0.57	0.64
Edgemaps + Creaseness	0.64	0.85
Color + Edgemaps + Creaseness	0.63	0.84

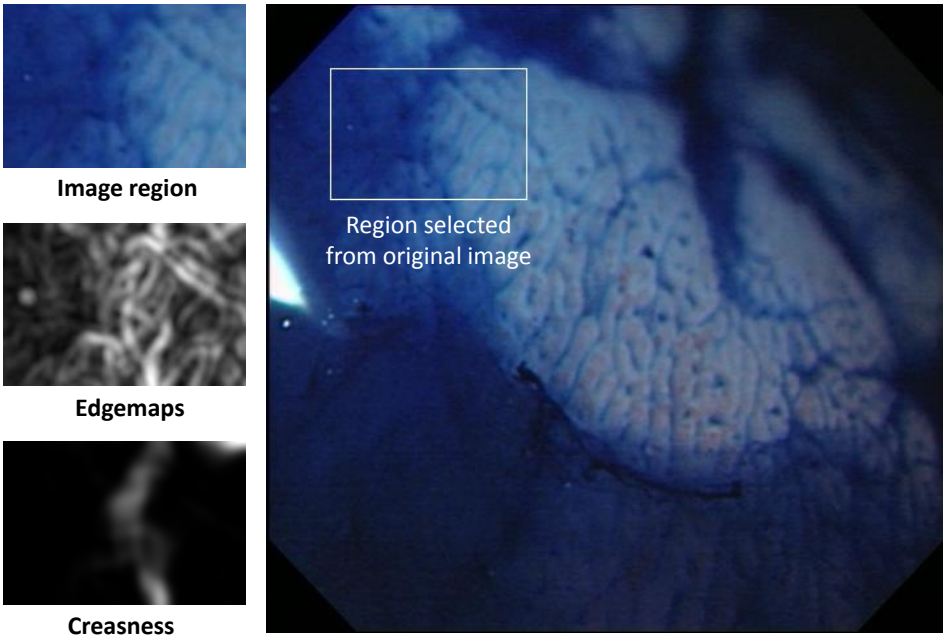


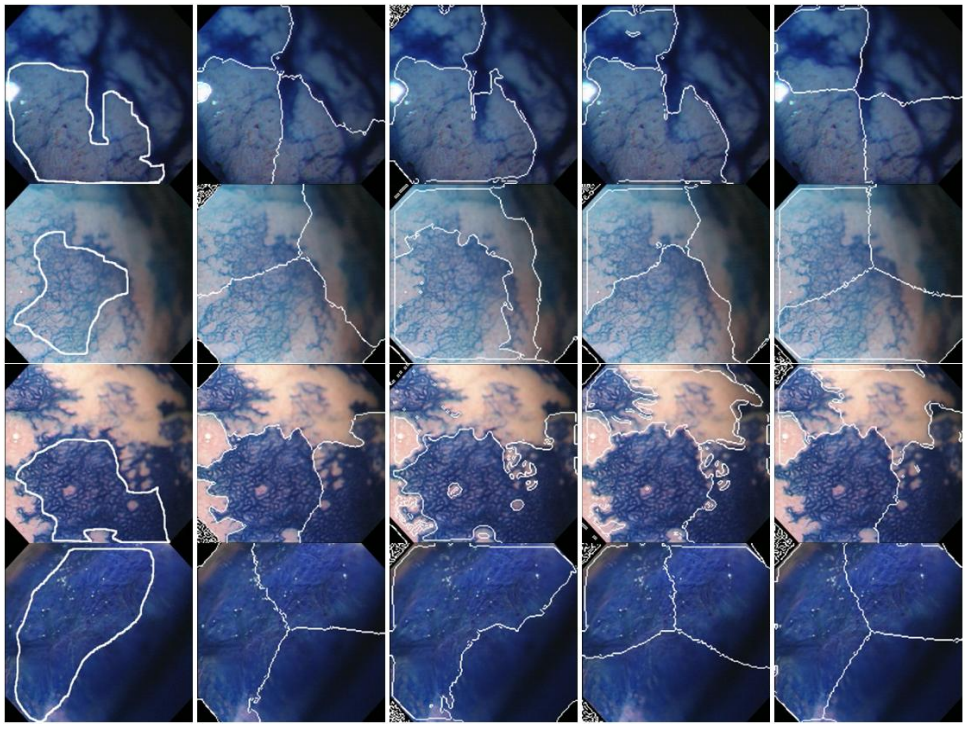
Figure 6.9: Comparison of image features obtained using edgemaps and creaseness.

of the distribution of the feature vectors in the feature spaces. This is a widely used segmentation method yielding very good performance for a wide variety of imaging applications. Its additional benefits also include publicly available implementation. We used [2]EDISON software which uses colors and spatial information of pixels as image features. The spatial and range parameters of the kernel in MS were selected by a manually searching those parameters which give the best DSC over the whole datasets.

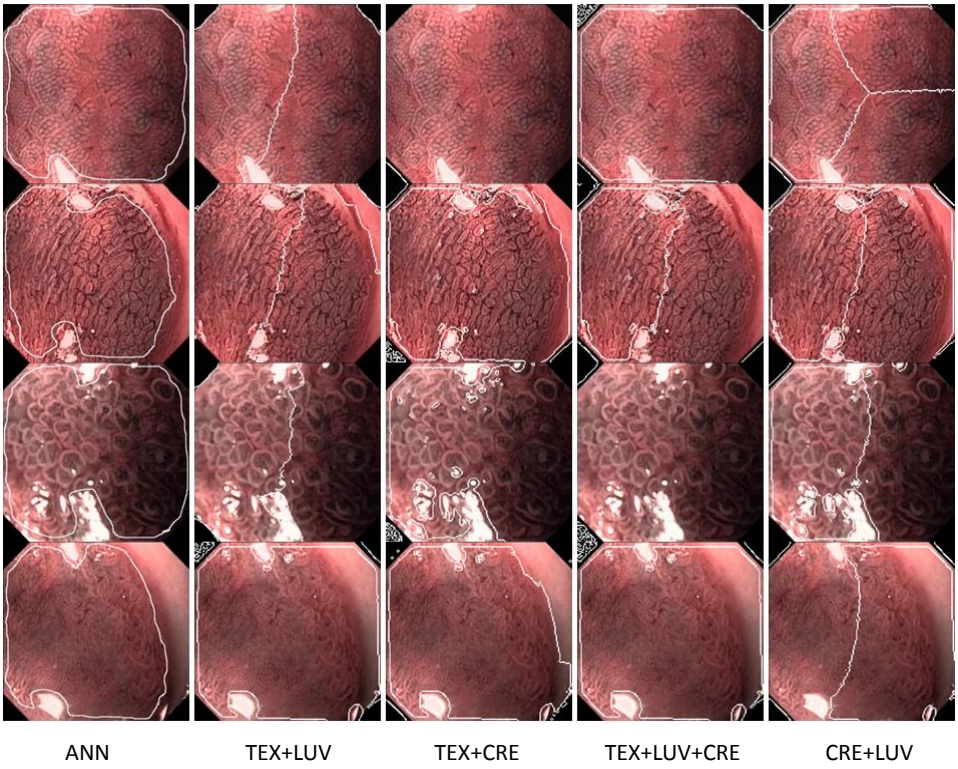
Level sets (LS) offer an alternative conceptual formulation of segmentation and is classified as one of the fitting models methods [47]. They are based on the minimization of cost functions based on image features. These are very high performance methods having a wide range of applications involving medical imaging [130], [142]. Their use for segmentation of CH images from colon has been previously reported [45]. These are iterative methods and are computationally expensive as compared with other methods considered in this paper. We used implementation of [3]Li et al. [89] for LS segmentation in which edge indicator function is used as image features for formulation of the minimization of external energy term.

Experiments show that for both CH and NBI images, NCut with affinity matrix derived

[2]EDISON System - http://coewww.rutgers.edu/riul/research/code/EDISON/
[3]Variational Level Sets - http://www.engr.uconn.edu/∼cmli/research/pals.html



(a) Segmentation using multiple features for CH images



(b) Segmentation using multiple features for NBI images

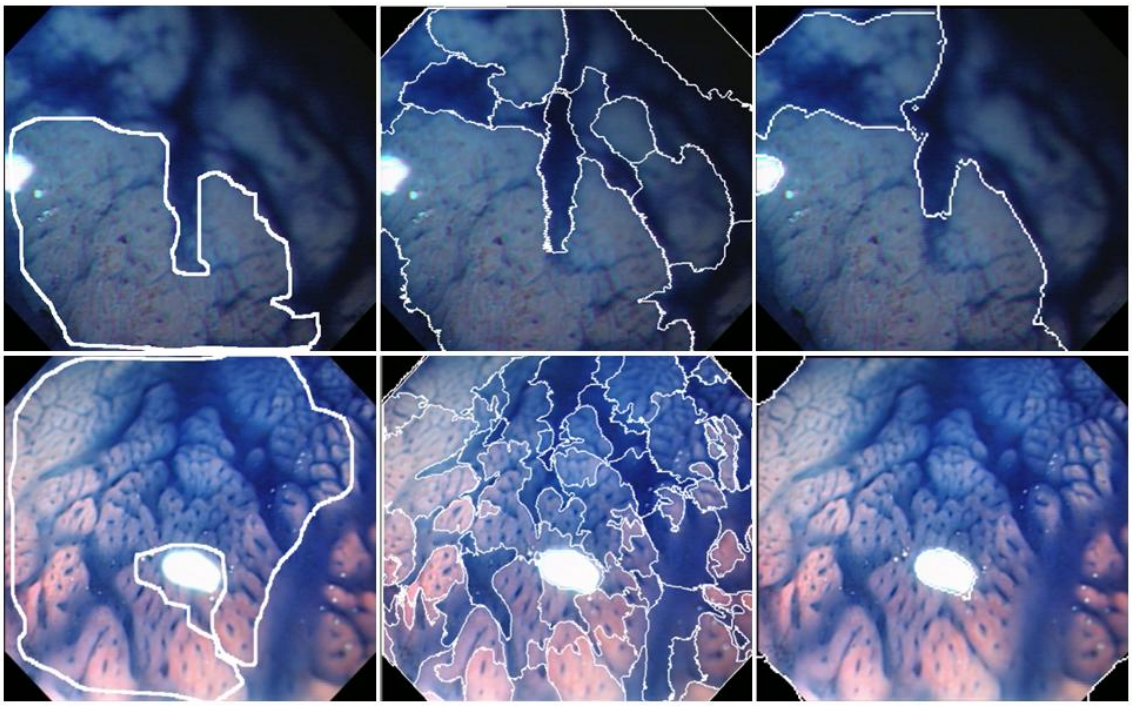
Figure 6.10: Visual illustration of manual annotations and segmentation obtained using combination of various features (ANN - Annotations; TEX - Edgemaps; LUV - CIELUV color features; CRE - Creaseness).

Table 6.3: Comparison for CH and NBI images using different segmentation methods (CH - Chromoendoscopy; NBI - Narrow-band imaging)

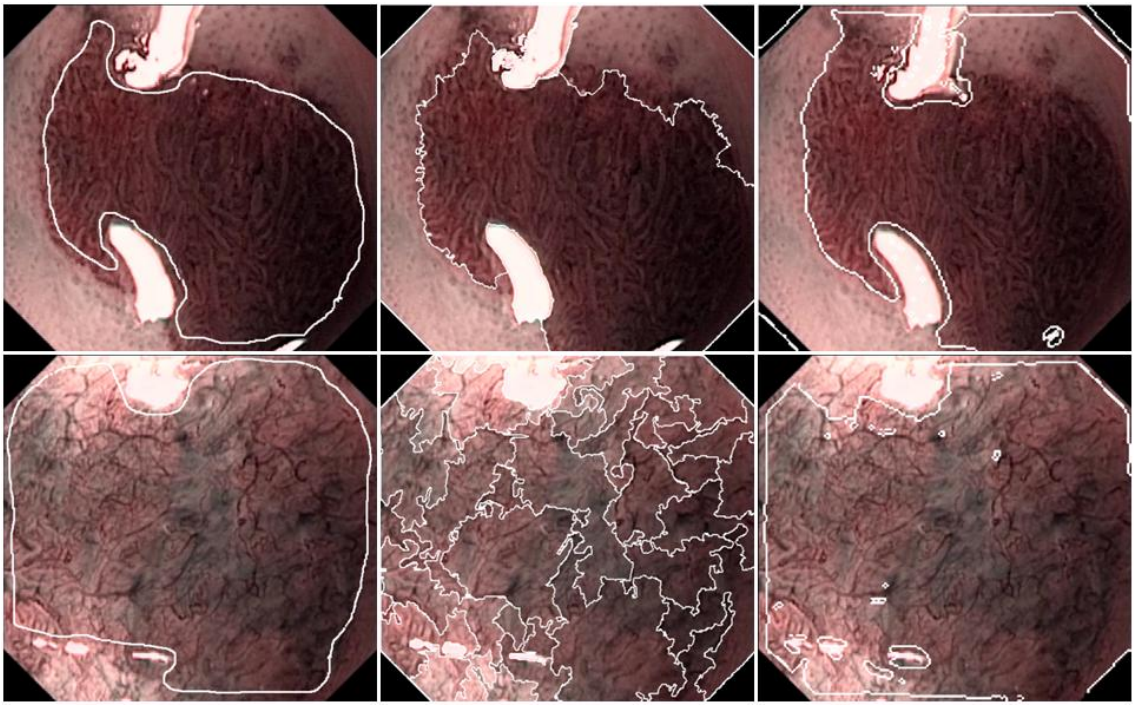
Segmentation methods	CH	NBI
Normalized Cuts	0.64	0.85
Mean shift	0.54	0.79
Level sets	0.55	0.78

from creaseness and edgemaps always outperforms other segmentation methods (Table 6.3). For CH images, NCut always outperforms MS and LS (unless color features are used to derive the affinity matrix). For NBI images, MS and LS show similar performance as compared to NCut when using individual features. It is important to note that even when no segmentation is done, using full images for NBI gives us a DSC of 0.76. This is due to the bigger size of image annotation for NBI images which are obtained from esophagus (tubular and rigid structure) unlike stomach (sac-like structure) resulting in a more coherent illumination of the visualized tissue, making most of the tissue relevant for clinical evaluation. The significant enhancement of segmentation performance is obtained when a combination of creaseness and edgemaps is used in the NCut framework giving us a very high DSC of 0.85.

MS method is highly dependent on the kernel bandwidths. It uses two different types of bandwidths: spatial and color. Both of these have a strong impact on the outcome of the resulting segmentation. This is because using MS, we often get over segmentation in images which have a high texture due to the presence of multiple modes in close proximities of the images giving a number of possible cluster centers which are later considered as image segments. This usually results in over-segmentation for images which are visually very rich. The problem is worse when using NBI images because they have a high texture because of the enhanced visibility of vasculature in the images. The diversity of the images enforces the need to calculate adapted kernel sizes based on image contents to be able to achieve maximum utilization of the mean shift method. Visual demonstration of over-segmentation yielded for both CH and NBI images is shown in Fig. 6.11. We can observe that the use of the same set of parameters for MS gives very diverse segmentation results therefore there is an inherent need to adapt the kernel parameters based on image content for MS to obtain desirable segmentations. For NCut, we have relatively stable results and no post-processing for region merging was required. Previous research for obtaining adapted kernel bandwidth has already been done for MS using ultrasonic images [57], however their application to gastroenterology imaging scenarios is yet to be explored. It is



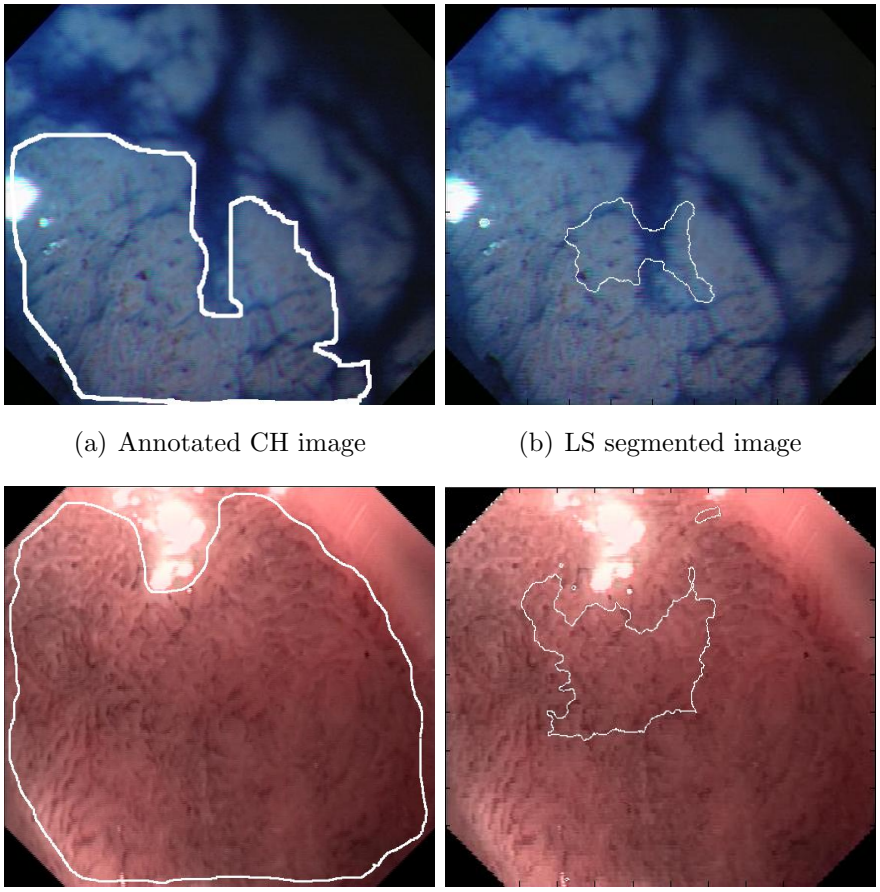
(a) Manual annotation (left column), mean shift segmented (middle column) and normalized cuts segmented (right column) CH image



(b) Manual annotation (left column), mean shift segmented (middle column) and normalized cuts segmented (right column) NBI image

Figure 6.11: Over-segmentation obtained using the same set of parameters on different images for MS in contrast with consistent performance of NCut.

an important research area as the adjustment of kernel parameters based on visual inspection of images has been observed to give good segmentation results for MS. For our implementation, in order to avoid the problem of over-segmentation, we fixed the output number of image patches obtained after segmentation and found the smallest bandwidth parameters which satisfied this criterion. We expect that a good parameter selection method is very vital to exploit the potential of MS methods to the maximum, potentially resulting in good segmentation.



(a) Annotated CH image (b) LS segmented image

(c) Annotated NBI image (d) LS segmented image

Figure 6.12: Illustration of CH and NBI Images being used for LS image segmentation. Lack of strong boundaries around the clinically relevant image region enforces the iterations to continue. The optimization never stops as the contour evolution continues even after 1000 iterations.

In image segmentation, active contours typically need strong edges to characterize the objects in the images. Therefore, for objects which have strong boundaries, they perform very well. In our scenario, we usually do not have strong edges characterizing the clinically relevant regions in the images. On the other hand, the presence of

vessels in the tissues creates false illusions about the presence of boundaries and thus variational level sets sometimes tend to converge to those vessels. For CH images where the vasculature is not prominent, there are relatively less edges in the images. In such cases, the lack of strong edges creates problems for the LS method to converge. As a result, we sometimes get image segments obtained in the middle of the image which are not meaningful (a visual illustration of this problem is shown in Fig. 6.12).

A possible solution of this problem could be the use of some other features for modeling the external energy term, which are physiologically relevant for the images in consideration. This could hopefully improve the performance of LS method. Use of such features for segmentation of CH images of colon for the detection of Aberrant Crypt Foci (ACF) have been reported in the literature [45]. Although the experimental validation is somewhat limited, good segmentation results have been obtained using physiologically relevant features which have been validated by the clinicians.

6.5 Conclusions

Graph based segmentation methods are very useful for segmentation of gastroenterology images. This is because using these methods, a good local-global interaction among the pixels in an image can be derived [55] resulting in meaningful segmentation of the images while being robust to noise and localized texture such as vasculature in the images. This results in overcoming the problem of over-segmentation which is typical using other state-of-the-art methods on gastroenterology images. Additionally, a good global partitioning criterion such as the one used in normalized cuts (NCut) complements the use of local features such as edges, creaseness etc. which if used as in state-of-the-art methods such as watersheds result in severe over-segmentation problems. We have used NCut for segmentation of gastroenterology images and obtained good segmentation results. We have used three different visual features for extraction of image contents: 1). CIELUV colors, 2). Multiscale edgemaps and 3). Multilocal creaseness. When individual image features are used, we get the best segmentation results for both imaging modalities using creaseness features. This happens due to the fact that creaseness features enhance the boundaries of clinically interesting regions, thus obtaining very good segmentation results. We also obtained good results using state-of-the-art multiscale edgemaps, which focus on region based texture characteristics of the images.

We obtain the best segmentation results when a combination of creaseness and multi-

scale edgemaps is used for segmentation. This happens because both these features are complementary: multiscale edgemaps focus on regions while creaseness focus on the boundaries. An affinity matrix obtained by a combination of both these features give us the best segmentation performance. The resulting segmentations are very close to manual annotations of the physicians. On the contrary, color features result in poor segmentation of gastroenterology images. This is because when using color (LUV) features, we are effectively using the luminance of the images for segmentation since the gastroenterology images have a narrow spectrum of colors available in them. A better solution for exploiting the use of color features could be using adapted color spaces. Such spaces, can be designed by the vector quantization of the traditional color spaces according to the distribution of colors in these spaces for images from specific scenarios. Color features from these adapted spaces would be more useful for segmentation of gastroenterology images.

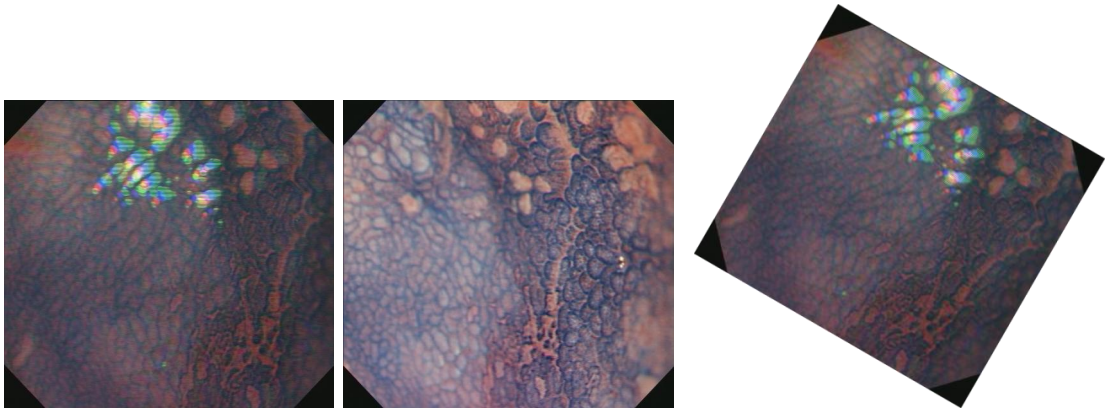
Normalized cuts outperformed two other segmentation methods considered in this chapter: Mean shift and Level sets. This choice is a good mixture of kernel based agglomerative clustering method, and segmentation by fitting method. For mean shift, the performance is highly dependent on the kernel size and therefore, an adapted choice of kernel bandwidth is vital for this method. Failing to do so, the choice of a fixed kernel bandwidth can result in over- and under- segmentation of the images based on their contents thus limiting its generic application to various imaging scenarios. Level sets are segmentation by fitting methods and are based on iterative based optimization of energy functions. They have been widely used for segmentation in various scenarios however lack of strong edges around manual annotations undermine their practical application for segmentation of gastroenterology images. We believe that good segmentations can be achieved using LS methods if the edge indicator functions in state-of-the-art implementations is replaced by some image features which are physiologically more relevant for gastroenterology images as done by Figueiredo et al. [45] for detection of Aberrant Crypt Foci (ACF) in colonoscopy images. Although our analysis has led us to study the behaviour of various segmentation methods, a more interesting comparison in future could be done using texture as image features for both MS and LS.

Chapter 7

Texture recognition for Gastroenterology

Texture recognition is an active area of research, that plays an important role in a wide variety of computer vision problems. Over the past three decades, many attempts have been made to propose various texture descriptors but the subjectivity of their design is based on parameters which are essential for specific applications. An important criterion for the design of texture descriptors for a wide range of applications is their invariance to rotation, scaling and illumination changes in the images. An example of such an application is feature extraction from endoscopic images for diagnosis of gastrointestinal (GI) cancer, which involves an analysis of pit patterns or vasculature in GI images. The need for invariance in this scenario originates from the decreased low level control of the camera due to inability to maintain a fixed distance between the endoscopic probe and the visualized tissue (affecting scale and illumination), and lack of sense of direction giving rise to varying perspectives of visualizing internal walls of the GI tract (Fig. 7.1). A reliable automated diagnosis system is expected to be able to cope up with these variations in the images, thus making the diagnosis robust to varying imaging conditions.

Our contributions in this chapter are motivated by the fact that conventionally in gastroenterology (GE), cancer diagnosis is done by visually finding some specific patterns in the images which lay the foundation of diagnosis. Currently, this involves a detailed visual inspection of the internal GI organs and is therefore inherently a function of human factors such as expertise and perception. Computer Assisted Decision (CAD) systems can be useful in improving the screening ability of the physicians for GI cancer by providing them with a second opinion, assisting them in situations where



(a) Original image (b) Magnified image (c) Rotated image

Figure 7.1: Dynamic imaging conditions in endoscopy. (a). A CH image captured during live exam, (b). CH image captured by bringing endoscopic probe closer to the tissue wall (scale and homogeneous illumination), (c). Artificially rotated image of same pattern.

their confidence level for diagnosis is low. Additionally, although the recognition of visual patterns is an essential factor in a gastroenterologist's ability to detect cancer, this is typically studied from books with pictures of these various patterns. The main problem of this atlas-based learning is that it is well-known in educational sciences that this type of book-based knowledge does not always migrate effectively to reality. This is particularly serious in gastroenterology since there is a strong contextual difference between the study room of a library and the gastroenterology exam room, in which there is the pressure of the clinical decision, and most importantly the divided attention between the image to be analyzed, the patient and the mechanical aspects of the procedure itself. It is not difficult to imagine that automatic classification algorithms will be essential building blocks of future interactive training systems that can positively impact a student's ability to detect cancer. Traditional texture analysis methods used for more conventional scenarios (e.g. content based image retrieval) are not expected to behave satisfactorily given the very specific nature and features of the GI images. In this chapter, we will propose novel texture descriptor which is robust to rotation, scale and illumination changes in the images. We will validate the performance of our novel descriptor by studying its invariance characteristics, and analyzing its performance for diagnosis of GI cancer.

Texture feature extraction has an extended history and a wide spectrum of methods are available in the literature. These, according to Tuceryan et al. [131] can be broadly divided into four main categories: geometrical, model-based, statistical and

filter based methods. Among these, statistical methods such as Gray Level Co-occurrence matrices (GLCM) [65], Local Binary Patterns (LBP) [103], Higher Order Local Autocorrelation (HLAC) [108] etc and filter based methods have been widely used. Statistical methods have been successfully used for feature extraction in several applications showing good performance however they share a limitation which is their inability to capture frequency content of the images. More recently, filter based methods have been preferred over these due to their space-frequency decomposition abilities [14]. In this chapter, we will focus on filter based methods given the diversity provided by their ability to combine micro- and macro- texture features in the form of multi-resolution (MR) analysis, which is a very important aspect while quantifying textures. Such methods can be grouped into three different categories [117]: 1) Spatial filtering, 2) Frequency filtering and 3) spatial-frequency filtering. A weakness shared by the former two methods is that they analyze the images at a particular scale but the latter gives us the additional feature of being able to analyze the images at different scales, paving the way for scale invariant texture features using MR analysis.

One of the most common approaches towards MR analysis is the use of wavelets [8] however they have poor directionality limiting their use for some applications [91]. Recently, curvelets [71] [51] [125] have been used for various application motivated by their ability to give a more precise (as compared to wavelets) description of the smooth curves by their piece-wise approximation using curvelets. However they are not good in representing point singularities. Our applied scenario of gastroenterology demands the detection of irregularity of mucosal structure and vessels which often require detection of point singularities (wavelets hold universality in representing both edges and point singularities [21]). In this chapter, we explore the use of Gabor filters given two of their unique characteristics that are not shared by any other existing method to the best of our knowledge: Optimal joint spatial-frequency resolution and their similarity with primary visual cortex of the mammals. In the context of invariant feature extraction, several attempts have been made to propose invariant features based on Gabor filters. K. Jafari-Khouzani et al. [73] used radon transform for rotation invariant texture features. R. Manthalkar et al. [98] proposed even symmetric Gabor filters for extracting rotation invariant features. J. Han et al. [62] proposed summation of filter energies over various orientations or scales to achieve rotation or scale invariance on Gabor features respectively. X. Xie et al. [141] used a scheme to obtain orientation normalization and a frequency searching procedure for rotation and scale invariant feature extraction. G. Healy et al. [60] proposed several micro- and macro- features using Gabor filters, which are invariant to image rotation. Kamarainen et al. [77] reordered the Gabor responses to achieve illumination, rotation and scale invariance

for applications such as matching. In most of these methods, either only rotation or scale invariant Gabor features have been proposed or the invariant recognition involves a one-to-one comparison, which is not suitable for feature space classifiers such as support vector machines.

In this chapter, we explain our motivation of using Gabor filters for feature extraction and explain the basic design of Gabor filters. Later, we discuss about the rotation and scaling properties of Gabor filters followed by our contributions to the design of invariant texture features:

- We propose novel rotation, scale and illumination invariant texture features (Autocorrelation Gabor Feature - AGF) based on Gabor filter responses. We theoretically explain invariance properties of Gabor filters to lay the foundation of our work. We empirically demonstrate the robustness of proposed features to rotation and scale changes in the images.
- A region based descriptor (Autocorrelation Homogeneous Texture - AHT) is proposed. Its robustness to image transformations is compared empirically with its state-of-the-art non-invariant counterpart, Homogeneous Texture (HT).
- We used the novel AGF texture features in a classification framework which is based on textons: texton-AGF. The use of invariant features gives us a rich dictionary which is invariant to rotations and scaling in the images. Empirical demonstration of the invariance of the dictionary is shown in the paper.
- Both, AHT and texton based descriptors are used for classification of GE images using two complementary imaging modalities (CH and NBI) from two different organs (stomach and esophagus respectively).

We conclude this chapter with a performance comparison of our novel image features with other state-of-the-art feature extraction methods.

7.1 Gabor filters

Gabor filters were introduced by Dennis Gabor in 1946 [48], while he was trying to formalize the mathematical function which can achieve optimal resolution of a signal in both frequency and time domain. The introduction of a 2D counterpart of the Gabor elementary functions by Granlund [56] laid the foundation of their use in image

processing. These functions can be used to find a spatial-frequency representation of the images which is found to have useful practical applications.

7.1.1 Motivation

Significant research on the use of Gabor filters for feature extraction for various pattern recognition applications has already been done, we based our selection of Gabor filters on their following important characteristics:

7.1.1.1 Similarity with visual cortex

One of the main motivations behind the use of 2D Gabor filters is their similarity with a simple cell in the mammalian visual cortex. The phenomenon of perceptual organization [44] enables humans to detect relationships among image elements such as parallelism, receptivity etc. This organization is captured by simple cells of the visual cortex of humans having receptive fields, which are restricted to small regions of space and are highly structured. There has been intensive research on the physiology of vision, but at least in the image processing community the most referenced paper seems to be the one by Daugman [37] where he presented a picture of the measured receptive field profiles of the cat's simple cells and showed the significant accuracy of the Gabor filter to embody the cat's simple cell. According to his research, the visual cortex has two important characteristics: band pass nature and direction selectivity, making them respond only to specific spatial frequencies and directions.

It is worth mentioning that these cells mostly exist in pairs, one cell in each pair having odd symmetry and the other one having an even symmetry. Studies on mathematical modeling of these cells elucidate their behavior as band pass filter bank structures, tuned to different orientations and scales just like Gabor filters. The Gabor filter is evidently a too simple model to explain all the details of the simple cells, but it does at least provide an approximate model to be used in simulations of the visual system. This remarkable similarity of Gabor filters with visual cortex makes their use very interesting in various image processing applications.

7.1.1.2 Optimal space-frequency localization

One of the basic principles in the classic Fourier analysis is its inability to achieve localization simultaneously in time and frequency domain. This can be precisely explained with the help of the uncertainty relation,

$$\triangle t \times \triangle f \geq \frac{1}{2} \tag{7.1}$$

According to this relation, there is a lower limit on the uncertainty which can be resolved for the joint representation of a signal in time and frequency domain. Dennis Gabor, while working on this uncertainty relation was concerned about finding the shape of the signal for which the product $\triangle t \triangle f$ assumes the smallest possible value i.e. for which, the inequality turns into an equality. He realized that a representation of signals which occupies minimum area ($\triangle t \triangle f = \frac{1}{2}$), can be constructed using a single building block: modulation of a sinusoid with a Gaussian function [48]. This remarkable discovery of Gabor has since been exploited very successfully in different areas of signal processing. The introduction of 2D counterpart of the Gabor elementary functions by Granlund [56] laid the foundation of their use in image processing to find a spatial-frequency representation of the images which is found to have very useful practical applications. Although traditional wavelets serve this purpose too, Gabor filters derive the representation of a signal which achieves the best possible localization of a signal in both frequency and space.

7.1.1.3 Multiresolution nature

One of the important characteristics of Gabor filters is that they allow a multiresolution analysis of the signals. For texture recognition, multiresolution features have been shown to provide significant performance gains in most of the applications. In our scenario, where we have a wide variety of images having different properties in terms of coarseness and fineness of the texture (classification proposals for chromoendoscopy and NBI images), techniques offering multiresolution feature extraction are expected to provide significant performance gains for final classification of the images as already indicated by our research on state-of-the-art on feature extraction for gastroenterology.

7.1.2 Gabor filter design

By definition, Gabor filters are obtained by modulating a sinusoid with a Gaussian function. Mathematically, they can be represented as [97]:

$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi jWx\right) \tag{7.2}$$

$$G(u,v) = \exp\left\{-\frac{1}{2}\left(\frac{(u-W)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right)\right\} \tag{7.3}$$

Where $\sigma_u = 1/2\pi\sigma_x$, $\sigma_v = 1/2\pi\sigma_y$, σ_x and σ_y are standard deviations of the Gaussian envelope along x and y dimensions and W is a constant representing the center frequency of the high frequency filter. Eq. 7.2 is the product of a Gaussian function with complex sinusoid. This forms a bandpass filter in the frequency domain, where the bandwidth and center frequency of the filter are controlled by the standard deviation of the Gaussian function and the frequency of complex sinusoid respectively. Let us now consider a filter bank which is composed of a number of self similar filters. If $g(x,y)$ is the mother wavelet, a self-similar wavelet dictionary is obtained by appropriate dilations and translations of $g(x,y)$ as follows:

$$g_{mn}(x,y) = a^{-m}g(x',y') \tag{7.4}$$

where

$$x' = a^{-m}(x\cos\theta + y\sin\theta)$$

$$y' = a^{-m}(-x\sin\theta + y\cos\theta).$$

and $m = 1,2,...,S$ and $n = 1,2,...,K$. S and K are number of scales and number of orientations respectively, and $\theta = n\pi/K$. This set of functions forms a non-orthogonal basis of functions for multiresolution decomposition. The non-orthogonality implies that there is redundant information in the filtered images. Manjunath used a strategy to reduce this redundancy [95] by designing the Gabor filters that ensure that the half-peak amplitudes of the filters touch each other in the Fourier spectrum (Fig. 7.2). This results in the following formulas for computing the filter parameters.

$$a = \left(\frac{U_h}{U_l}\right)^{\frac{1}{S-1}} \tag{7.5}$$

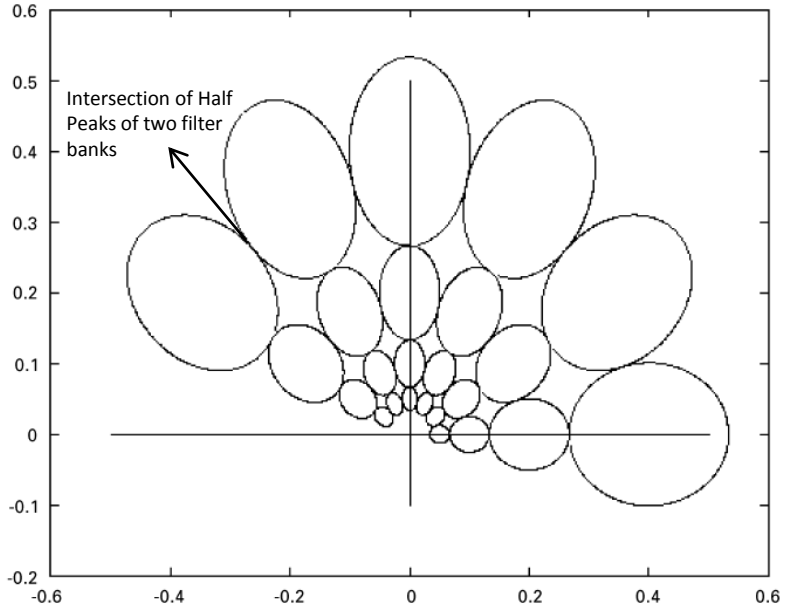


Figure 7.2: Frequency spectrum of a Gabor filter bank intersecting at half peak amplitudes (adapted from [95]).

$$\sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2\log 2}} \tag{7.6}$$

$$\sigma_v = \tan\left(\frac{\pi}{2k}\right)\left[U_h - 2\log 2\left(\frac{\sigma_u^2}{U_h}\right)\right] \times \left[2\log 2 - \frac{(2\log 2)^2\sigma_u^2}{U_h^2}\right]^{\frac{-1}{2}} \tag{7.7}$$

Where U_l and U_h are the center frequencies of filters with smallest and largest bandwidths respectively. A Gabor filter bank having a number of bandpass filters, with varying center frequencies, bandwidths and orientations is controlled by the parameters of Gabor wavelets. An input image, $\xi(x, y)$ when filtered by the set of Gabor wavelets is given as:

$$R_{mn}(x, y) = \int_{x_1}\int_{y_1} \xi(x, y) g_{mn}(x - x_1, y - y_1) dx_1 dy_1 \tag{7.8}$$

7.1.3 Invariance properties of Gabor filters

There are three important properties of Gabor filters, which help us in establishing the invariant feature extraction for pattern recognition.

7.1.3.1 Rotation property

The most obvious use of time-frequency space is the ability to detect time varying changes in the frequency content of the signal. For image based application, these changes correspond to the rotation of the images. Let us considered an image $\xi_1(x, y)$ filtered by Gabor filters $g(x, y, f, \theta)$, where f is the frequency of the filter and θ is its orientation. The response of the filter is given as

$$r_{\xi_1}(x, y, f, \theta) = \int_{x_\tau} \int_{y_\tau} g(x_\tau, y_\tau, f, \theta)\xi_1(x - x_\tau, y - y_\tau)dx_\tau dy_\tau \tag{7.9}$$

For $\xi_2(x, y)$, which is the rotated version of $\xi_1(x, y)$ by an angle α around (x_0, y_0) the resulting filter response can be given as

$$\xi_2(x, y) = \xi_1(\hat{x}, \hat{y})$$

$$\hat{x} = (x - x_0)\cos\alpha + (y - y_0)\sin\alpha + x_0$$

$$\hat{y} = -(x - x_0)\sin\alpha + (y - y_0)\cos\alpha + y_0$$

Substitution of the rotated image to equation 7.9 and integrating it over the transformed axes $(\hat{x}, \hat{y})$ at a location (x_o, y_o) gives us

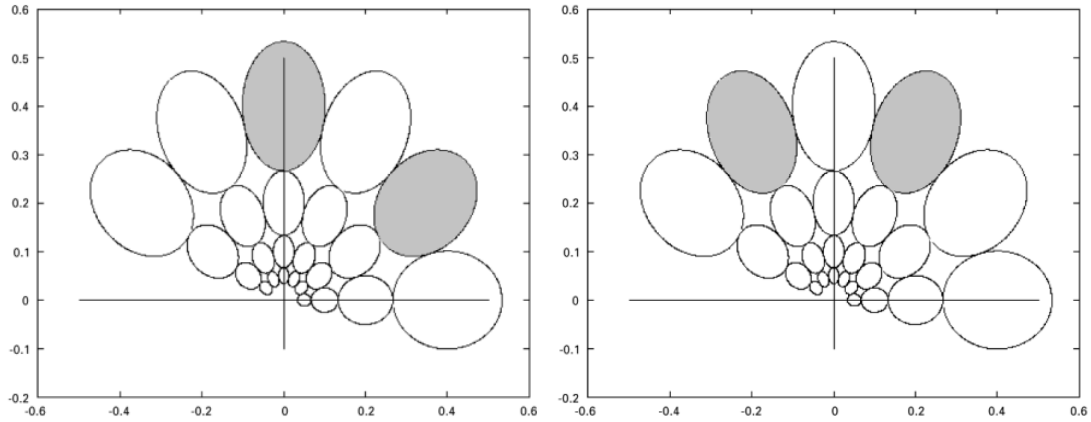
$$r_{\xi_2}(x, y, f, \theta) = \int_{\hat{x}_\tau} \int_{\hat{y}_\tau} \hat{g}(x_\tau, y_\tau, f, \theta)\xi_1(x - \hat{x}_\tau, x - \hat{y}_\tau)d\hat{x}_\tau d\hat{y}_\tau \tag{7.10}$$

For $\hat{g}(x_\tau, y_\tau, f, \theta) = g(\hat{x}_\tau, \hat{y}_\tau, f, \theta - \alpha)$, equation 7.10 can be written as

$$r_{\xi_2}(x, y, f, \theta) = \int_{\hat{x}_\tau} \int_{\hat{y}_\tau} g(\hat{x}_\tau, \hat{y}_\tau, f, \theta - \alpha)\xi_1(x - \hat{x}_\tau, y - \hat{y}_\tau)d\hat{x}_\tau d\hat{y}_\tau \tag{7.11}$$

$$r_{\xi_2}(x, y, f, \theta) = r_{\xi_1}(x, y, f, \theta - \alpha)$$

This property shows that the Gabor response for an image which has been rotated by some angle α is the same as the response of a correspondingly rotated filter for the original image (Fig. 7.3). The conclusion therefore is that for a rotated image, the filter response in the filter bank would be represented by the filter which has been rotated by the same angle as that of the image.



(a) Response to an image (b) Response to rotated image

Figure 7.3: Effect of image rotation on Gabor filter responses. The same response is represented by a filter that is rotated by the same angle

7.1.3.2 Scale property

The scale property was not recognized initially within the Gabor functions since strategies involving a number of self similar functions with consistent sharpness, and constant durations of the filter were dominating in the time-frequency methods. The rectangular lattice proposed by Gabor [76] has had a strong impact on the development of Gabor analysis [76].

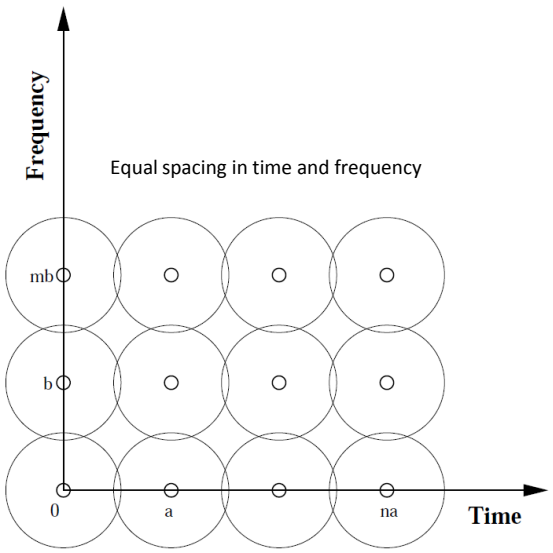


Figure 7.4: Time-frequency representation of Short Time Fourier Transform - STFT (adapted from [76]).

If the sharpness of Gaussian functions in the Gabor function in eq. 7.3 is selected to be constant and not dependent on the frequency, the envelope size of the Gabor filter is the same regardless of the frequency. Filters in this case do not represent scale information and frequency cannot be interpreted as the scale. That is the case with most short-time Fourier transform (Fig. 7.4) approaches and also the original form of the Gabor expansion. Before a scale property of the Gabor filters can be utilized in scale invariant detection, the rectangular lattice structure is typically discarded and spacing of the frequencies must be selected in a more proper manner. The rectangular structure is discarded by the selection of varying standard deviations of the Gaussian functions which are spaced appropriately in the frequency domain (Fig. 7.5). For proper spacing of the filters, the center frequencies of the filters are logarithmically established:

$$f_k = c^{-m} f_{max} \tag{7.12}$$

Where f_{max} is the center frequency of the highest frequency filter, c is the frequency scale factor which is a constant and m is defined in eq. 7.4. This establishes the logarithmic spacing of the frequencies and therefore results in interpretations of frequencies of filters as scales. This typically results in a lattice with non-linear spacing of filters.

Let us now assume that the image $\xi_1(x, y)$ has been scaled homogeneously by a constant factor c. The new transformed image can be written as $\xi_3(x, y) = \xi_1(cx, cy)$. It is easy to show that the filter response of ξ_3 can be given as

$$r_{\xi_3}(x, y, f, \theta) = \int_{x_\tau} \int_{y_\tau} \hat{g}(x_\tau, y_\tau, f, \theta)\xi_3(x - x_\tau, y - y_\tau)dx_\tau dy_\tau$$

$$r_{\xi_3}(x, y, f, \theta) = \int_{x_\tau} \int_{y_\tau} \hat{g}(x_\tau, y_\tau, f, \theta)\xi_1(cx - cx_\tau, cy - cy_\tau)dx_\tau dy_\tau \tag{7.13}$$

Substituting

$$cx_\tau = \hat{x}, cdx_\tau = d\hat{x}$$

$$cy_\tau = \hat{y}, cdy_\tau = d\hat{y}$$

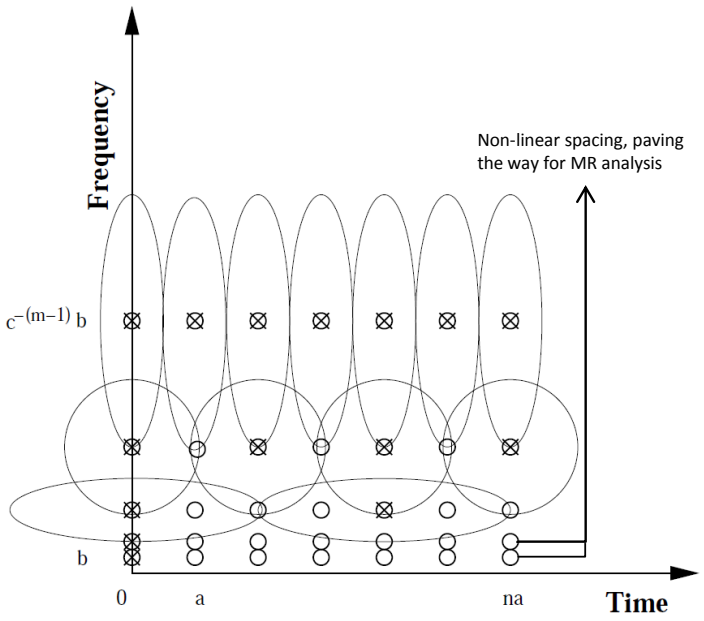


Figure 7.5: Variation of center frequencies of bandpass filters according standard deviation of Gaussian functions give the notion of multiresolution analysis (adapted from [76]).

in eq. 7.13 and simplifying gives us

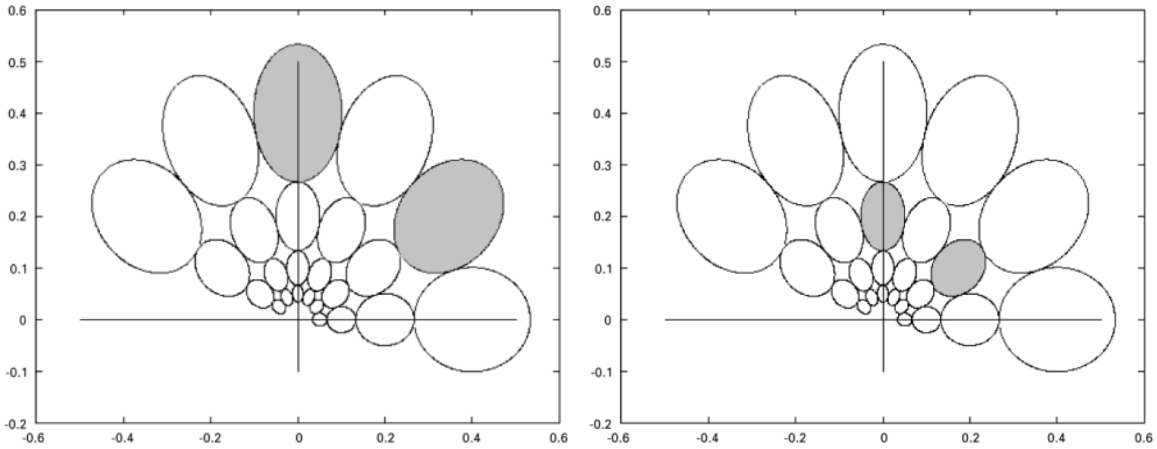
$$r_{\xi_3}(x, y, f, \theta) = \frac{1}{c^2} \int_{\hat{x}} \int_{\hat{y}} \hat{g}(x_\tau, y_\tau, f, \theta) \xi_1(cx - \hat{x}, cy - \hat{y}) d\hat{x} d\hat{y} \tag{7.14}$$

For $\hat{g}(x_\tau, y_\tau, f, \theta) = g(\hat{x}, \hat{y}, \frac{f}{c}, \theta)$, eq. 7.14 can be written as

$$r_{\xi_3}(x, y, f, \theta) = \frac{1}{c^2} \int_{\hat{x}} \int_{\hat{y}} g(\hat{x}, \hat{y}, \frac{f}{c}, \theta) \xi_1(cx - \hat{x}, cy - \hat{y}) d\hat{x} d\hat{y}$$

$$r_{\xi_3}(x, y, f, \theta) = \frac{1}{c^2} r_{\xi_1}(cx, cy, \frac{f}{c}, \theta) \tag{7.15}$$

Thus image scaling results in changing the frequency contents of the images, which are captured by another Gabor filter representing the new frequency contents. The scale property in the Gabor filters therefore implies that the response for a scaled image by a constant factor c is equal to the response of a correspondingly scaled Gabor filter used for filtering the original image (Fig. 7.6). Therefore for a scaled image, the filter response would be represented by the filter which has been scaled by the same factor as that of the image.



(a) Response to an image (b) Response to scaled image

Figure 7.6: Effect of scaling on Gabor filter responses

7.1.3.3 Illumination property

Let us now assume that the image $\xi_1(x, y)$ undergoes a uniform illumination change. This change can be modeled by multiplying the image with a constant i.e., $\xi_4(x, y) = c\xi_1(x, y)$. Where c is a constant factor.

$$r_{\xi_4}(x, y, f, \theta) = \int_{x_\tau} \int_{y_\tau} g(x_\tau, y_\tau, f, \theta)\xi_4(x - x_\tau, y - y_\tau)dx_\tau dy_\tau$$

$$r_{\xi_4}(x, y, f, \theta) = \int_{x_\tau} \int_{y_\tau} g(x_\tau, y_\tau, f, \theta)c\xi_1(x - x_\tau, y - y_\tau)dx_\tau dy_\tau$$

$$r_{\xi_4}(x, y, f, \theta) = cr_{\xi_1}(x, y, f, \theta) \tag{7.16}$$

Therefore the convolution property shows that a uniform illumination change in the image has the effect of inducing this change linearly on the resulting filter responses. In scaling property, attenuation of filter responses as a result of scaling an image results in homogeneous illumination change, which can also be modeled by this convolution property.

7.1.4 Simple Gabor Feature Space (SGFS)

Let us now consider an image which has been filtered by a Gabor filter bank. Kamarainen [83] proposed that the response at i^{th} pixel can be represented as:

$$R_i(m,n) = \begin{bmatrix} r_i(x_i,y_i;f_0,\theta_0) & r_i(x_i,y_i;f_1,\theta_0) & \cdots & r_i(x_i,y_i;f_{S-1},\theta_0) \\ r_i(x_i,y_i;f_0,\theta_1) & r_i(x_i,y_i;f_1,\theta_1) & \cdots & r_i(x_i,y_i;f_{S-1},\theta_1) \\ \vdots & \vdots & \ddots & \vdots \\ r_i(x_i,y_i;f_0,\theta_{K-1}) & r_i(x_i,y_i;f_1,\theta_{K-1}) & \cdots & r_i(x_i,y_i;f_{S-1},\theta_{K-1}) \end{bmatrix} \tag{7.17}$$

Where $r_i(x_i, y_i, f_m, \theta_n)$ is the response of the filter bank at i^{th} pixel with center frequency f_m and orientation θ_n. m and n follow the same definition as in equation 7.8.

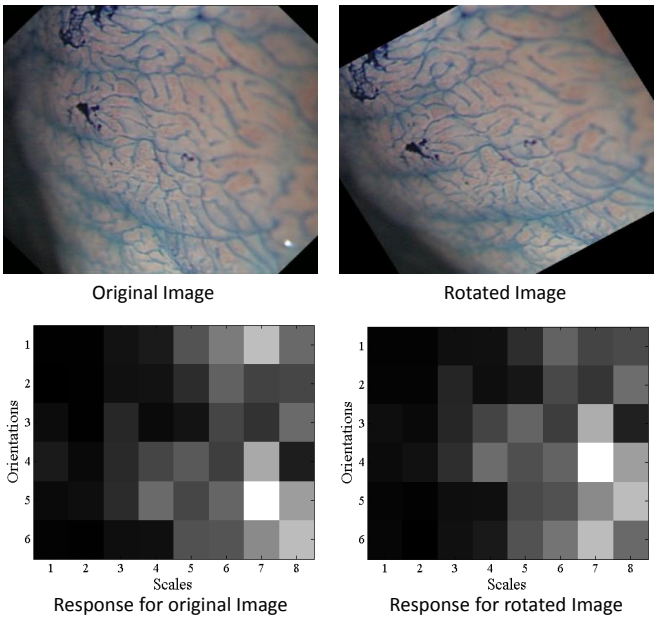


Figure 7.7: A visual demonstration of the fact that in $SGFS$, rotation of an image corresponds to the circular shift of rows

This representation of the pixel responses known as the $SGFS$ has advantages over a conventional vector representation of the Gabor filter responses mainly concerning invariant recognition. This is because an image rotated by α degrees and scaled by a factor c is represented as a shifted matrix i.e., $R'_i(m,n) = R_i(m-c, n-\alpha)$.The shift is linear for scale changes and circular for rotations.

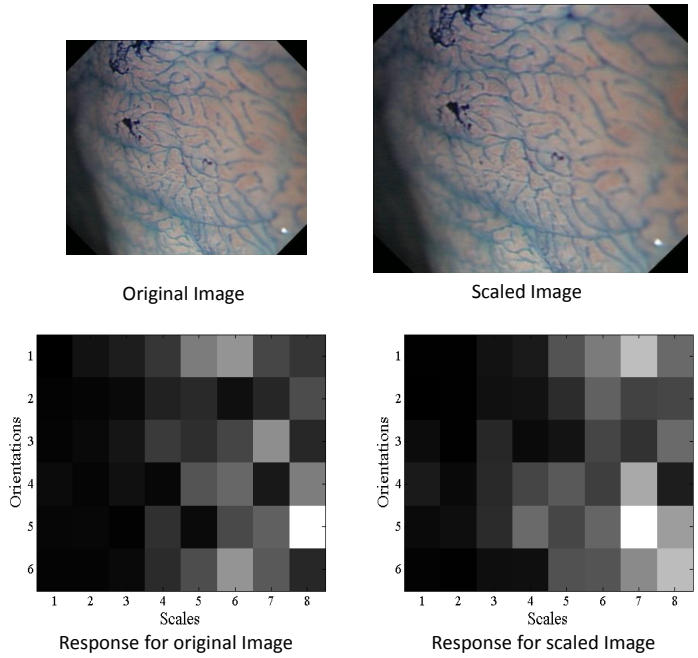


Figure 7.8: A visual demonstration of the fact that in $SGFS$, scaling of an image corresponds to the linear shift of columns

Therefore using this representation, operations for rotation and scale invariant searches of objects can be defined as the row wise circular shift of the response matrix corresponding to the rotation of the object (Fig. 7.7) around the location (x_0, y_0) and a column wise shift corresponds to the scaling of an object (Fig. 7.8). An illumination invariance of the matrix can be achieved by normalizing this feature matrix. For applications involving object matching etc., the reference images are available so row and column shifts can be applied to the $SGFS$ of the query object to match it with the reference for invariant matching. For applications such as classification which do not concern matching, $SGFS$ restricts the use of classifiers such as k-nearest neighbors for invariant classification. This restriction on the use of classifiers to conserve invariance undermines the true potential of Gabor features for classification related applications as feature space classifiers which are expected to give very good performance such as Support Vector Machines (SVM) cannot be used to achieve invariant recognition using $SGFS$.

7.1.5 Autocorrelation Gabor Features -AGF

Invariant recognition using $SGFS$ is accomplished using column wise or row wise shifts of $R_i(m, n)$. Given the difficulty of this space in classification related tasks, we

have proposed a novel feature extraction method which makes use of autocorrelation function to achieve invariance of the feature space without the need of any search along any of the dimensions in the resulting matrix based on its shift invariance property. Autocorrelation of a continuous function $x(t)$ is given as:

$$A(\tau) = \int_{-\infty}^{+\infty} x(t)x(t+\tau)dt \tag{7.18}$$

where τ is a random shift applied to $x(t)$. As can be seen, autocorrelation of $x(t)$ is independent of t and is a function of τ only. We use this well known shift invariance property of autocorrelation [66] to achieve rotation and scale invariant features. We know that rotation and scale transformations induce shifts in the matrix $R_i(m,n)$. Autocorrelation of $R_i(m,n)$ gives us a matrix, which eliminates the dependence of autocorrelation on m and n and thus on scales and rotation respectively. This matrix is composed of the following autocorrelation coefficients at i^{th} pixel,

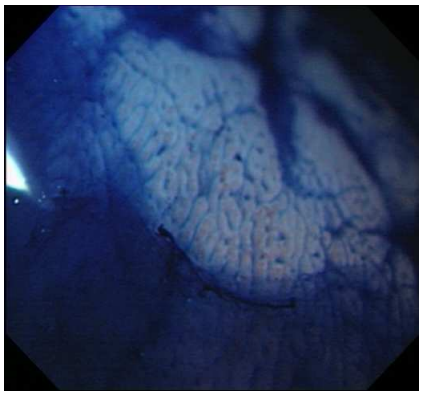
$$z_i(s,k) = \sum_{m=0}^{S-1}\left[\sum_{n=0}^{K-1} R_i(m,n)R_i(m+s,n+k)\right] \tag{7.19}$$

Where $s = 0, 1, \cdots, 2S-1$ and $k = 0, 1, \cdots, 2K-1$. As can be seen in the equation, $z_i(s,k)$ is independent of the original variables i.e., m and n and instead depends on the lags s and k. It is independent of shifts in the rows and columns of $R(m,n)$ which results from the rotation and scaling of the image respectively. All coefficients in $z_i(s,k)$ can be arranged as a vector,

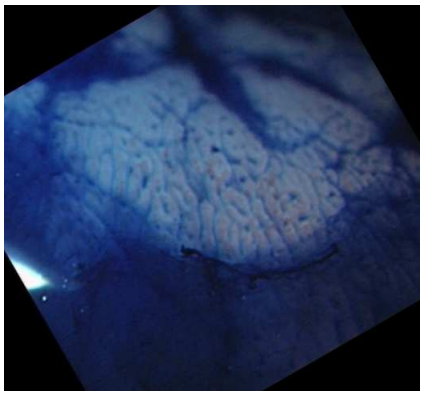
$$Z_i(d) = \begin{bmatrix} z_i(0,0) \\ z_i(0,1) \\ \vdots \\ z_i(1,0) \\ z_i(1,1) \\ \vdots \\ z_i(2S-1,2K-1) \end{bmatrix} \tag{7.20}$$

Where $d = (2S-1)(2K-1)$. The symmetric property of autocorrelation function implies that half of the coefficients of the vector $Z_i(d)$ can be dropped thus giving us

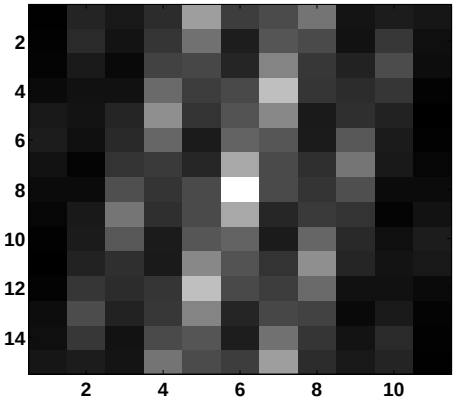
$$AGF_i^t = \begin{bmatrix} Z_i(1) & Z_i(2) & \cdots & Z_i(d/2) \end{bmatrix}^t \tag{7.21}$$



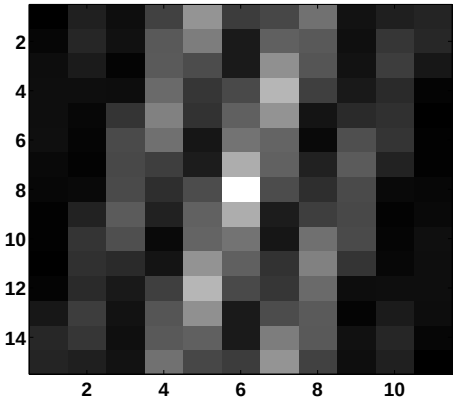
(a) Original CH image



(b) Rotated CH image



(c) AGF of original image



(d) AGF of rotated image

Figure 7.9: Demonstration of AGF extracted from the center of a gray scale CH image. Visual inspection of the AGF show its robustness to image rotations, thus giving us rotation invariant features.

AGF_i is the feature vector at i^{th} pixel in the image. This response vector is independent of row or column shifts of the matrix $R_i(m, n)$ and therefore independent to scale and rotation changes in the original image. Illumination invariance can be achieved by the normalization of AGF. The resulting AGF can be interpreted as having a good discrimination capability between shape and size characteristics of various patterns in the image as the Gabor filters do. This implies from the capability of autocorrelation coefficients to represent Gabor responses as a relation between different orientations ($\triangle n = n_2 - n_1$) and scales ($\triangle m = m_2 - m_1$), thus helping in the determination of various characteristics of the patterns in the images such as shape, size, regularity etc., as a relative measure between different orientations and scales and not as an absolute

measure, as happens in Gabor filters. Even if the images are transformed, relative characteristics between various scales and orientations do not change. Therefore by the use of AGF, we retain most characteristics of the Gabor filters while providing invariance of the vectors to scale, rotation and homogeneous illumination in the images as an additional feature. AGF are therefore expected to be able to handle dynamic imaging conditions in a more efficient way.

7.2 Classification using AGF

Although AGF can be directly used for classification purposes, the high dimensionality of the AGF limits their practical use directly for pattern recognition tasks. We therefore integrate them in two different classification frameworks for pattern recognition motivated by conceptually alternative views on texture characterization.

- **The texton framework: texton-AGF -** Assuming that the images are composed of some basic visual primitives, known as *textons*.
- **Autocorrelation Homogeneous Texture -** Assuming that visual properties of texture are relatively constant over a spatial region.

7.2.1 The texton framework: texton-AGF

In psychophysics, Julesz (1981) [75] and colleagues discovered that pre-attentive vision is sensitive to some basic image features while ignoring other features. His experiments measured the response time of human subjects in detecting a target element among a number of distracters in the background. For example, Fig. 7.10 shows two pairs of elements in comparison. The response time for the left pair is instantaneous (100 to 200 ms) and independent of the number of distractors. In contrast, for the right pair the response time increases linearly with the number of distractors. This discovery was very important in psychophysics and motivated Julesz to conjecture a pre-attentive stage that detects some atomic structures, such as elongated blobs, bars, crosses, and terminators (Julesz, 1981), which he called textons for the first time. In practice, there can be numerous ways to formulate this conjecture. One of them is filter banks.

Representation of textures using filter banks is at the same time very versatile and overly redundant. This is because they have slowly varying spatial characteristics

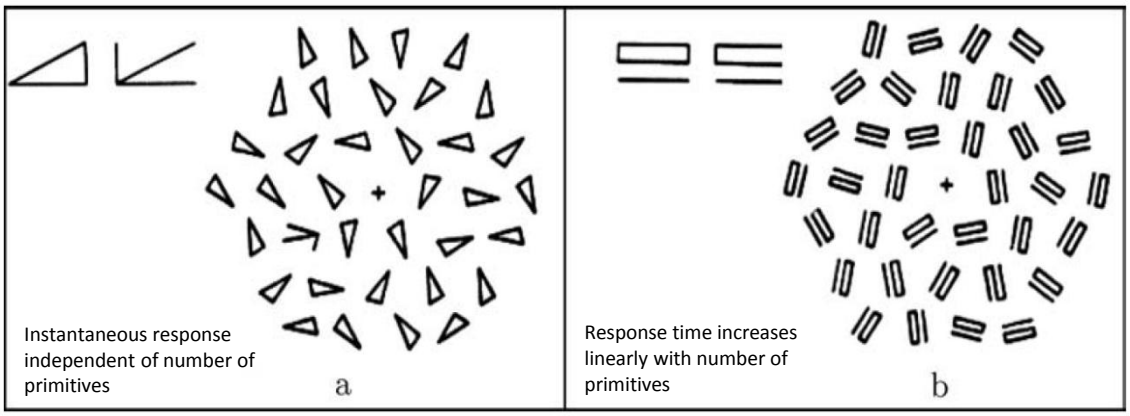


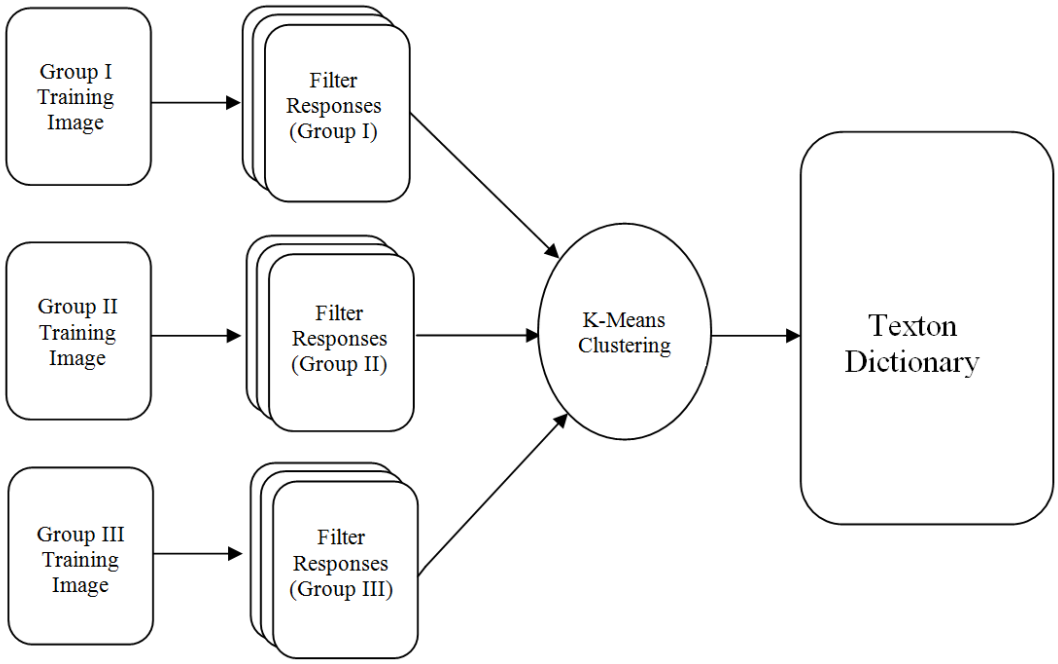
Figure 7.10: Two typical examples of searching a target element among a number of background distractors. The search time for the left pair is constant independent of the number of distractors, while it increases linearly with the number of distracters for the right pair (adapted from [75]).

and we do not expect filter responses to be completely different from one another at different spatial locations. Therefore, we can get several distinct filter response vectors while all others can be seen as compositions of their noisy variations. This intuition leads us to inspect the filter responses in a 'bag of words' approach giving us a number of prototype response vectors, known as textons [146]. In such a framework, the classification algorithm can be divided into two stages as done by Varma [133]: Learning of texton dictionary and classification of novel images (Fig. 7.11).

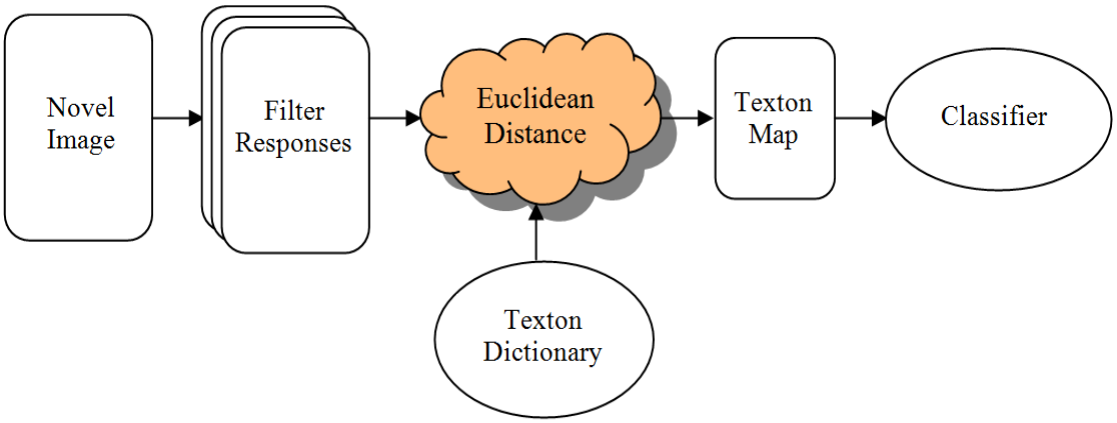
In the texton learning stage, training images are selected from the dataset to generate features at each pixel. K-means clustering [129] is used to obtain a specific number of textons: Texton-AGF. All textons obtained from different training images are collected into a single dictionary. In the testing stage, the AGF at each pixel in the novel image are quantized to the closest texton-AGF in the dictionary. We quantify the distance between two responses by Euclidean distance. Histograms showing texton occurrence frequencies corresponding to the novel images are later generated. These histograms can now be input to state-of-the-art classifiers for classification purposes.

7.2.2 Autocorrelation Homogeneous Texture - *AHT*

An alternative view of the texture characteristics of an image is that the visual properties of the texture are relatively constant over a spatial region in an image i.e., the local texture in an image is homogeneous. The characterization of such regions can



(a) Learning of texton dictionary



(b) Classification of images using textons

Figure 7.11: A texton based framework for classification of images.

be done using the statistical properties of image regions. Manjunath [95] proposed that assuming that the local texture regions in an image have spatially homogeneous visual characteristics, the statistics of an image region can be summarized using first order statistics (mean and standard deviation) of magnitude of the Gabor filter responses of those regions. These statistics are accumulated in a feature vector to represent the characteristics of an image region, also known as Homogeneous Texture (HT). Although this is a powerful texture descriptor, it is not robust to image rotations and

scaling. Assuming homogeneousness of a texture region, we propose a novel texture descriptor (Autocorrelation Homogeneous Texture - AHT) for an image region using AGF:

$$\mu_{z_q} = \frac{1}{N}\sum_{i=0}^{N} z_i(q) \tag{7.22}$$

$$\sigma_{z_q} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Z_i(q) - \mu_{z_q})^2} \tag{7.23}$$

$$AHT = \left[\mu_{z_1}, \sigma_{z_1}, \mu_{z_2}, \sigma_{z_2}, \cdots, \mu_{z_{d/2}}, \sigma_{z_{d/2}}\right] \tag{7.24}$$

Where $q = 1, 2, \cdots, d/2$. μ_{z_q} and σ_{z_q} are the mean and standard deviation of the q^{th} component of AGF respectively and N is the total number of pixels in the image. Since the foundation of AHT lies in autocorrelation of $R(m, n)$ which removes dependence on scales and orientations, we expect AHT to exhibit invariance to the said transformations in the images. We encompass the empirical demonstration supporting this intuition in the subsequent sections.

Finally the classification of images using features obtained from both the texton-AGF (texton histograms) and AHT can be done using various state-of-the-art machine learning methods.

7.3 Experiments

We aim to analyze the performance of our novel descriptors for classification of Chromoendosopy (CH) and Narrow-band imaging (NBI) images. In this section, we will empirically demonstrate invariance of our novel descriptor (AGF) and compare it with its state-of-the-art counterpart i.e., Gabor filters (GF). We will later use AGF for classification of Gastrointestinal (GI) images into their respective classes. For all experiments, Gabor filtering was done using 6 orientations and 8 scales (i.e., $K = 6$ and $S = 8$). The selection of these parameters was done empirically by analysing the performance filters over a range of possible parameters and then selecting the suitable ones for our data-sets. Nonetheless, the relative results concerning both invariance and classification are consistent irrespective of filter parameters.

7.3.1 Invariance testing

Given that an important motivation of this work was the design of texture descriptors invariant to rotation, scale and illumination changes it is important to analyze the invariance characteristics of the proposed features. In our experiments we are using the images captured during live endoscopic examinations. For testing invariance of the descriptors to rotation and scale changes, it is very difficult to acquire transformed images during real examinations. Since the endoscopic probe is manually controlled and we acquired the magnification endoscopy images, if an attempt to make a small rotation of the probe for acquiring rotated image changes the distance between the tissue and the probe by millimeters, we will get a very different (magnified or otherwise) view of the same tissue. On the other hand, it is impossible to decouple the scale and illumination changes in the images because they both depend on the distance between the camera and the tissue. Given this, we have decided to induce artificial rotation and scale changes to real endoscopic images so that a single parameter is changed by a precise amount. We have used [1]Matlab routines to obtain rotated (*imrotate* Matlab function) and scaled (*resample* Matlab function) versions of real endoscopic images.

7.3.1.1 *AGF* vs. *GF*

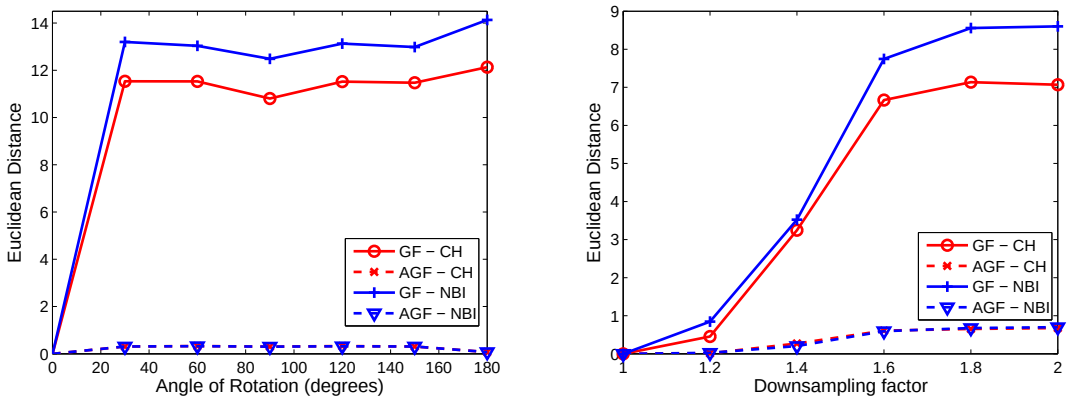
To make a comparison at every pixel, of an image and its geometrically transformed version, one-one mapping between pixels in both (original and transformed) images is obtained. For an image, which has been rotated by an angle θ and scaled by a factor Z, an approximate one-one mapping between the pixels can be obtained using the following equation:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = Z \left(\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \times \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \right) \tag{7.25}$$

Where x_0 and y_0 are coordinates of the pixel around which the image has been rotated. We compare the average Euclidean distance between the responses of original and transformed image at individual pixels. This average distance tells us, how much did the features changes when the images are transformed:

$$ED = \frac{1}{N} \sqrt{\sum_{i=0}^{N-1} (F_i - F'_i)^2}$$

[1]Matlab - http://www.mathworks.com/products/matlab/



(a) Effect of rotation on feature stability

(b) Effect of scaling on feature stability

Figure 7.12: Average Euclidean distance between feature vectors extracted from original images and images which have been subjected to geometric transformations i.e., rotation and scaling (GF - Gabor Filters; AGF - Autocorrelation Gabor Filters; CH - Chromoendoscopy; NBI - Narrow-band imaging).

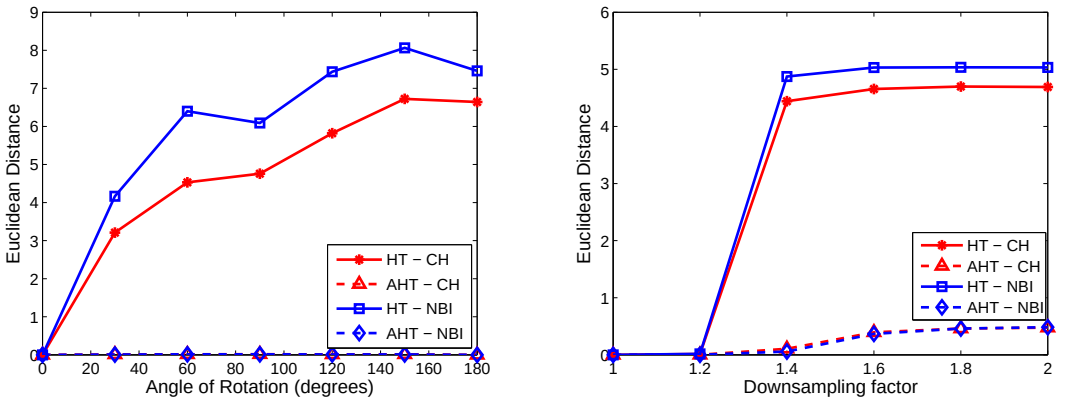
Where F_i is the feature vector at i^{th} pixel in original image and F'_i is the feature vector at i^{th} pixel of the transformed image and N is the number of pixels in the image. Before making a comparison, we normalize the GF between 0 and 1 to ensure a fair comparison, as our novel features are scaled between 0 and 1 for providing illumination invariance.

Our experiments show that the AGF features are stable for both rotation and scale changes in the images (Fig. 7.12). The results are consistent for images from both, CH and NBI images. We can therefore conclude that the AGF provide the much needed stability of the image features for very dynamic imaging conditions, which is vital for gastroenterology images.

7.3.1.2 *AHT* vs. *HT*

HT features [95] have been widely used for texture recognition tasks. They are also a part of the MPEG-7 visual descriptors and show very good performance as texture feature extractors. However one of the problems of this descriptor is that it is sensitive to the rotation and scaling of the images. We have compared the stability of our novel proposed features, AHT to the HT for various image transformations (rotations and scaling). For this purpose, we calculate both of these features from original CH and NBI images and their transformed (rotated and scaled) counterparts and compare

both HT and AHT for their stability to the said transformations by calculating the Euclidean distances between the features from original and transformed images.



(a) Effect of rotation on feature stability

(b) Effect of scaling on feature stability

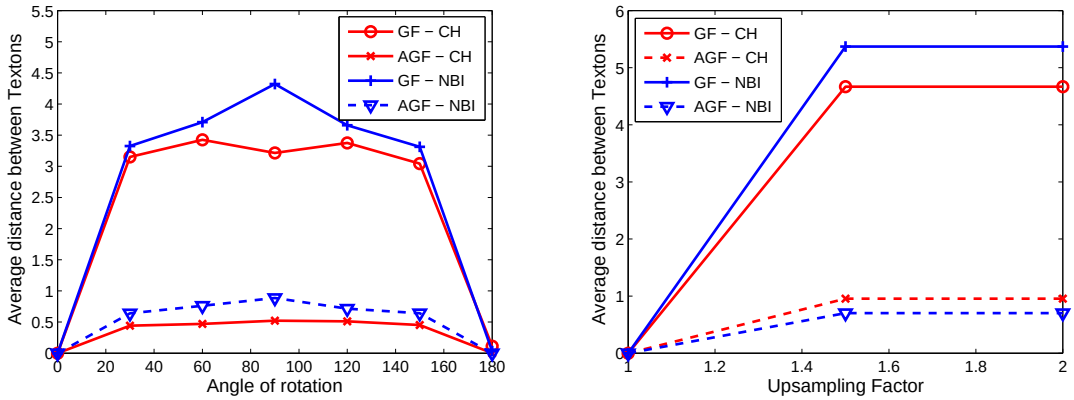
Figure 7.13: Average Euclidean distance between feature vectors extracted from original images and images which have been subjected to geometric transformations i.e., rotation and scaling (HT - Homogeneous Texture; AHT - Autocorrelation Homogeneous Texture; CH - Chromoendoscopy; NBI - Narrow-band imaging).

Our experiments show that AHT features are stable for both rotation and scaling thus these are rich features and are suitable to be used for feature extraction from images which suffer diverse imaging conditions (Fig. 7.13).

7.3.1.3 Texton-AGF vs. Texton-GF

Traditionally for texton-based classification frameworks using filter banks, the main problem is the non-invariance of the filter responses. Due to this problem, the words in the dictionary obtained using an image differ from the words obtained using the rotated or scaled version of the same image. This causes the dictionaries obtained using a set of images a function of the image orientations and scales. Since our AGF are robust to image transformations (Fig. 7.12), we should inspect if they can be used to construct a texton dictionary which is robust to image transformations. We have carried out an experiment to empirically demonstrate the invariant characteristics of the dictionary of texton-AGF. For this purpose, we obtained dictionaries using both AGF and GF for original images by k-mean clustering using a set of initialized points in the feature space. Dictionaries by k-means clustering using the same set of starting points were obtained for transformed images by inducing rotation and scale changes in

the real endoscopic images from our dataset. The average Euclidean distance between words in the dictionaries using original and transformed images was calculated. Our experiments show that the dictionaries obtained using texton-AGF are robust to image transformations as compared to dictionary obtained using state-of-the-art texton-GF (Fig. 7.14).



(a) Effect of rotation on feature stability

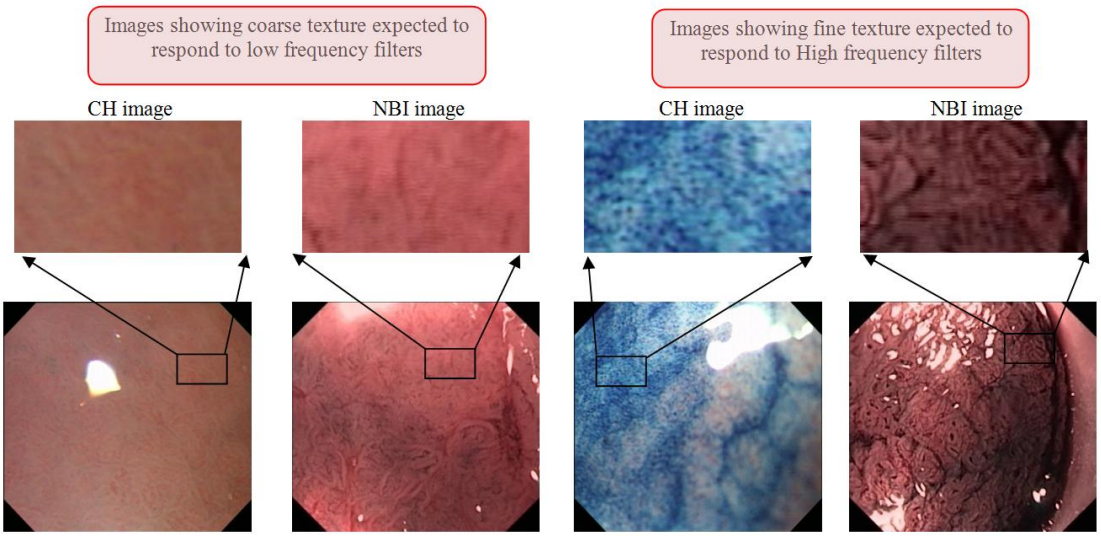
(b) Effect of scaling on feature stability

Figure 7.14: Average Euclidean distance between feature vectors extracted from original images and images which have been subjected to geometric transformations i.e., rotation and scaling (GF - Gabor Filters; AGF - Autocorrelation Gabor Filters; CH - Chromoendoscopy; NBI - Narrow-band imaging).

It is pertinent to mention that the novel (AGF) features are robust to the parameters of the Gabor filters. Therefore even when the parameters of Gabor filters are changed, our results show good robustness of our novel features to such transformations in the images.

7.3.2 Classification

Our objective in classification is to diagnose cancer in the patients or to identify patients, which are at a higher risk of developing cancer in future. We have used images from two different organs using two different imaging modalities for this purpose. Although the broader objective is the same: cancer diagnosis, there are some specific classification objectives which have to be met for both imaging modalities. For CH images, the classification of gastric mucosa is established by dividing images into 3 groups based on their color and texture features (Fig. 5.3). The main difference between the groups is the regularity and shape of the gastric mucosa. The deter-



(a) Typical images having low texture (b) Typical images having high texture.

Figure 7.15: A rich multi-resolution analysis of images is needed to correctly differentiate between gastric images which have coarse, medium and high texture

mination of regularity of gastric mucosa requires a broader spatial analysis. On the other hand, the absence of regularity or distortion of pit patterns results in a high texture of the images, which requires a fine analysis of the images. For NBI images, the classification for Barrett's esophagus is accomplished by dividing images into 4 groups (Fig. 5.9). This is mainly based on shape and regularity of pit patterns as well as vasculature in the images. For determining regularity, a broader spatial analysis is required but a more fine analysis is required for determining the distorted pit patterns and vasculature.

The main difference between both CH and NBI images is that the NBI images have a high vasculature and poor color resolution. The novel features proposed in the thesis exhibit invariant behavior when facing rotation, scale and illumination changes in the images irrespective of the parameters of Gabor filters. Multi-resolution analysis is an important aspect for analysis of gastroenterology images. This is because the images belonging to different classes have varying spatial characteristics, implying that they should be analyzed at various resolutions (Fig. 7.15). Normal images are characterized by the regularity of vessels and mucosal patterns and this characteristic requires a broader spatial analysis of the images.

The pre-cancerous (especially) and cancerous images, however, have predominantly distorted patterns and this characteristic spreads over a much smaller spatial neighborhood. This requires a more fine analysis of the images in order to effectively capture

such features. A higher number of scales at which we analyze the images would give us a richer multi-resolution analysis. This fact was experimentally verified by selecting higher number of scale for Gabor filters, which gave us better and more interesting classification results. In all our experiments, following filter parameters were used: $U_h = 0.5$, $U_l = 0.03$, $S = 8$, $K = 6$. The selection of these parameters was done empirically by analyzing the performance filters of over a range of possible parameters and then selecting the suitable ones for our data-sets. For image classification, we used only manually annotated image patches for feature extraction. Support Vector Machines (SVM) in the [2]Weka data mining tool were used in the classification experiments presented here. All results were obtained using 10-fold cross validation.

For texton based methods, we select four images per class from each imaging modality to create the texton dictionaries. While learning the dictionaries, we set the number of clusters obtained using K-means clustering to 10. This is a good compromise between a higher number of clusters which would create difficulties for K-means to converge and also result in sparse features due to high dimensionality of data, and low number of clusters which would affect the richness of the dictionaries. Therefore for CH images we obtain a a dictionary of 120 textons whereas for NBI images we obtain a dictionary of 160 textons (40 textons per class, 3 classes for CH and 4 classes for NBI). The AGF obtained at every pixel (eq. 7.21) in novel images is quantized to one of these textons in the texton dictionary and consequently, we generate 120- and 160- dimensional histograms as feature vectors for each CH and NBI image respectively. For AHT and HT, the feature vector obtained for images were used directly for classification. Performance comparison is done using overall classification rates and area under ROC curves (A_z).

7.3.2.1 Overall classification

Experiments show that texton-AGF yield good classification results for both CH and NBI images (Table 7.1). It is a robust method which consistently gives higher classification accuracies for both imaging modalities which are complementary in nature. This shows their superiority as generic descriptors which can handle dynamic imaging conditions while giving good classification results. In our experiments, the average classification accuracy for texton-AGF is the same as texton-GF for CH images however for NBI images the former gives about 5% higher average classification accuracy. This shows the higher suitability of texton-AGF for different imaging

[2]Weka - http://www.cs.waikato.ac.nz/ml/weka/

scenarios presenting complementary challenges from a computer vision perspective. AHT gives better results as compared to HT for both CH and NBI images showing the superiority of invariant features over their non-invariant counterparts. For CH images, we get relatively lower classification rates as compared to NBI images. This is because, by using NBI we compromise on the color characteristics of the images but get an enhanced visualization of vasculature. This enhancement introduces strong edges in the images that are reflected as shifting of filter responses from the filters having low frequency to the ones having higher frequencies. This results in richness of texture features in NBI images. On the other hand, CH images are characterized by two visual cues: texture and color, since full visible spectrum of light is used for tissue illumination and die is applied for mucosal structure enhancement. Since we use only texture features for feature extraction from the images, a full description of images does not include color based clinical interpretations for tissue discrimination, thus compromising on the performance achieved using CH images. Although this is beyond the scope of this chapter, given that our focus is invariance of texture features and its impact on classification performance, we believe that if color features are also used for CH along with texture features (as done in [123]), classification results for CH can be improved further. We also evaluated the performance of texton-GE (Texton-Gabor Energy), in which the magnitudes of Gabor filter responses are used for processing of the data. Texton-AGF outperforms these also for both imaging modalities by a significant margin.

	CH images		NBI images	
	A_z	Accuracy	A_z	Accuracy
Texton-AGF	0.852	82.3%	0.945	88.4%
Texton-GF	0.855	82.3%	0.90	83%
Texton-GE	0.761	71.5%	0.882	79.8%
AHT	0.856	83%	0.91	80.3%
HT	0.827	80.7%	0.85	72.1%
LBP	0.857	83%	0.822	68.7%
RILBP	0.785	76.1%	0.742	57.7%

Table 7.1: Overall performance of various descriptors (AGF - Autocorrelation Gabor Features; GF - Gabor Filters; HT - Homogeneous Texture; AHT - Autocorrelation Homogeneous Texture;CH - Chromoendoscopy; NBI - Narrow-band imagimg; LBP - Local Binary Patterns; RILBP - Rotation Invariant LBP; Az - Area under ROC curve).

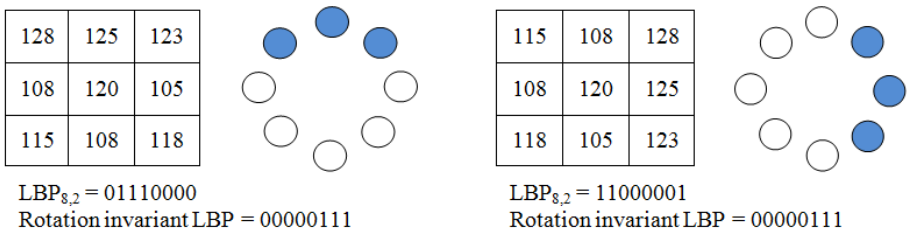


Figure 7.16: Two different patterns having same Rotation Invariant Local Binary Patterns.

Given that there is evidence in state-of-the-art for the use of Local Binary Patterns (LBP) for classification of CH images, we also included them in our analysis. For LBP, we used uniform patterns with neighborhood of 8 points in a radius of 2 for feature extraction followed by the generation of a histogram of LBP. The histogram was later used as a feature vector for classification using SVM. An important observation (Table 7.1) is that when using rotation invariant LBPs, the classification performance depreciates significantly as compared to that of uniform LBPs. This happens because LBPs focus too much on local variations. Usually these variations are calculated on the pixel level. This makes it hard for LBPs to differentiate between different patterns (Fig. 7.16). This happens due to a circular shift of the pattern (to make them rotation invariant) until smallest decimal number of the binary pattern is obtained. This causes difficulties for rotation invariant LBPs to differentiate between perceptually different patterns such as the ones in Fig. 7.17. In NBI images, the texture is very rich which is mainly caused by excessive vasculature and LBPs fail to capture the structure of the vessels due to their spatially local characteristics, which causes the classification performance to degrade when using rotation invariant LBPs.

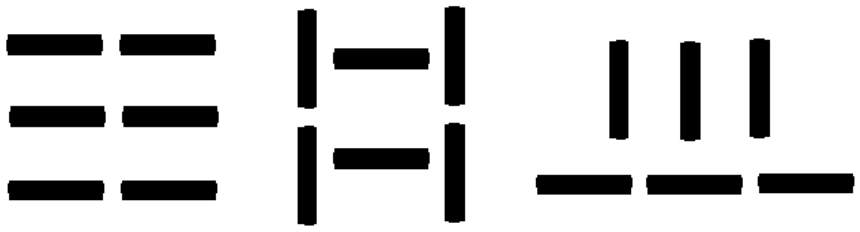


Figure 7.17: Too much focus on local variations for Local Binary Patterns (LBP) makes it hard to differentiate between visually different patterns. These patterns give the same rotation invariant LBP and hence same histograms for the three distinct patterns. A descriptor doing a broader spatial analysis will give different descriptors for these patterns, thus differentiating between them effectively.

7.3.2.2 Texton-AGF

Since texton-AGF shows good performance generically for both imaging modalities, we analyzed its confusion matrices in detail. For CH images we achieved an overall classification rate of 82.3%, the results are reasonably good for images belonging to Group I and II but for Group III images however, the results are not promising given the high number of false negatives (Table 7.2).

		Automatic Classification		
		Group I	Group II	Group III
Manual Classification	Group I	35	4	1
	Group II	5	69	1
	Group III	5	7	3

Table 7.2: Confusion matrix for classification of CH images using texton-AGF

There is a strong possibility that the misclassification in Group III images is partially caused by the fact that we only have 13.4% Group III images in the whole dataset, implying an inadequate training of the classifier. Even with this inadequate training, it is important to note that 7 of the false negatives occur in Group II and there is a very low false positive rate (only two, out of which one image actually belongs to Group II), thus hinting at the suitability of using texton-AGF for computer assisted decision (CAD) systems. This is because for a CAD system, an alarm will be raised by the system and it is expected to be tackled appropriately with the intervention of the physicians. Group II images have the highest classification rates among all individual classes.

		Automatic Classification			
		Group I	Group II	Group III	Group IV
Manual Classification	Group I	45	6	1	1
	Group II	4	64	0	1
	Group III	0	0	6	6
	Group IV	1	3	1	69

Table 7.3: Confusion matrix for classification of NBI images using texton-AGF

For NBI images we achieved an overall classification rate of 88.4%, which are better overall classification rates for all the groups as compared to CH images. The proposal for classification of NBI images by Singh et al. [121] is perceived as a purely texture based task taking into account two important aspects from a clinical perspective:

gastric mucosa and superficial vascular structures in the images. This makes the classification of such images a very hard task. From the confusion matrix (Table 7.3), we observe high true positive rates for Group II (pre-cancerous) and IV (cancerous). The only Group for which there is high misclassification is Group III (also pre-cancerous) however we only have about 7% images in whole NBI dataset belonging to this group which implies inadequate training of the classifier. For a CAD system this is not serious as the images are misclassified as cancer and an alarm will be raised for follow up by the physician.

7.4 Conclusions

In this chapter, we have introduced a novel texture descriptor, the Autocorrelation Gabor Features (AGF), which are invariant to rotation, scale and illumination changes in the images. Gabor filters are employed to the images and a specific rearrangement of the filter responses is used that enforces rotation and scale changes as shifts in a matrix. The shift invariance property of autocorrelation function is exploited to normalize the shifts in the matrix, giving us rotation and scale invariant texture features. Normalization is used to achieve illumination invariance. These AGF are integrated into two separate feature extraction approaches. A bag of words approach is used exploiting the redundancy of filter banks to create texton dictionary using some training images and texton histograms are generated as features, which can be integrated with feature based classifiers to perform classification task. Another usage of AGF follows Manjunath et al. [95] assuming homogeneousness of images to calculate first- and second-order statistics of individual features of AGF for the whole image - Autocorrelation Homogeneous Texture (AHT). Empirical demonstration of invariance of AGF, texton dictionary (texton-AGF) and AHT is provided. An applied scenario of diagnosis of gastrointestinal (GI) cancer is considered to validate the performance of the proposed texton-AGF and AHT. A comparison of these descriptors with their respective state-of-the-art counterparts i.e., Gabor filter textons (texton-GF) and homogeneous texture (HT) is done. Experiments show that the proposed descriptors perform well as compared to their counterparts, thus hinting at the superiority of invariant texture analysis in the applied scenario of GE. Another important result is the superior performance of the narrow-band images (NBI) as compared to chromoendoscopy (CH) images. We attribute this result to the fact that NBI increases the visibility of the vessels, enhancing texture in the images. Consequently the filter responses shift to higher frequencies. Diagnosis in CH imaging however incorporates

color as an important visual cue, which we do not use in this chapter. Usage of color features is expected to give better performance for CH images.

Although our proposed features are invariant to rotation, scaling and homogeneous illumination, another important challenge will be to deal with illumination gradients in the images. In the future, we intend to study the behavior of AGF to illumination gradients and somehow incorporate this additional challenge to our novel descriptor. In GE, there is evidence that color descriptors are complementary for vital stained magnification endoscopy. As an extension of our novel features on the applied scenario of GE, we intend to combine our novel texture descriptors with color descriptors to improve the diagnosis of cancer.

Chapter 8

Conclusions

In this chapter, we will revisit the thesis objectives and assess the computer vision tools used for achieving the objectives in scope of this thesis. These objectives led us to propose some novel methods. We will briefly discuss the results that we achieved through our novel methods. We will conclude this chapter with a short discussion of generic and specific objectives, guided by this thesis.

8.1 Discussion on objectives

In chapter 1, we proposed a list of objectives for the research carried out in this thesis. Our objectives led us to the following research path.

8.1.1 Computer vision and clinical semantics

One of the objectives of this thesis was to identify the computer vision challenges for designing computer assisted decision (CAD) systems for gastroenterology (GE). In this context, various challenges pertaining to imaging dynamics were identified and the design of adequate computer vision methodologies to address these challenges was targeted. For our research, two specific target scenarios were considered and a careful review of the clinical flow of the diagnosis was done. The most important conclusion was that the visual information encompassed by the tissue mimics the state of the patient. It was therefore considered vital to revisit the clinical taxonomies which result in a particular diagnosis for the target scenarios considered in this thesis. This study revealed the need to analyze or detect some particular structure which can be found

in the images. The following question was therefore answered: How can computer vision mimic various clinical observations while robustly incorporating the dynamics of imaging conditions in GE imaging scenarios?

8.1.2 Pattern recognition

In chapter 3, we proposed a pattern recognition (PR) system for computer vision module in a CAD system. This PR system consists of three main building blocks: segmentation, feature extraction and classification. Segmentation was motivated by a clinicians particular interest in a region for finding a pattern exhibiting relevant clinical information for diagnosis. Consequently, one of the most vital tasks was to perform image segmentation. Image segmentation is followed by feature extraction. This is motivated by the fact that a feature space is needed that can enhance the separability of various clinical taxonomies. This is achieved using the feature extraction stage, which exploits the coarse and fine image characteristics to extract the features from the images. Once the patterns have been mapped onto a feature space having enhanced differentiation capabilities, classification is needed. This classification finally decides the clinical state of the patient which can be either of the following three: normal, pre-cancerous or cancerous.

8.1.3 Evaluation study

In this thesis, two target scenarios were considered.

- Chromoendoscopy (CH) images for diagnosis of stomach cancer.
- Narrow-band imaging (NBI) for diagnosis of Barrett's esophagus.

These scenarios pose complementary challenges both from a technological and anatomical perspectives. From the technology point of view, CH images benefit from shape and color of the mucosal patterns whereas the NBI images benefit from the shape and vasculature of the images. The NBI images have very poor color resolution but they enrich the view of capillaries thus assisting in making diagnosis in a complementary way. These distinct characteristics pose complementary challenges from a computer vision perspective.

On the other hand, the complementary nature of our dataset also comes from the sites in the GI tract. NBI images were acquired from the esophagus which is a tube like structure. CH images on the were acquired from stomach which is a sac-like structure. This allows for a wide range of angular view of the same tissue as compared to esophagus, demanding more robustness of the image descriptors. The evaluation study was therefore thorough and can be used to draw generic conclusion on the ability of the PR system.

8.2 Discussion on results

8.2.1 Image segmentation

We have found the divisive image segmentation methods to be more interesting in gastroenterology (GE) imaging scenarios. This is mainly because GE images have very high texture and divisive clustering methods present a more global view of the images, which is important to avoid oversegmentation which is typical for agglomerative clustering methods. We used normalized cuts for our experiments. The integration of creasness features in normalized cuts (NCut) image segmentation framework is very useful as creasness tends to enhance the oriented gradients in the images even when they are weak, which is very typical for gastroenterology images. A combination of multiscale edgemaps and creasness outperforms all other visual cues and their combinations. Color features were also used for the analysis and they did not give good results for segmentation. This is because of the typically reduced color spaces in endoscopic images. Most of the color space is therefore redundant. Effective use of color features is possible by using adapted color spaces which are a true representative of the color contents in the images. The performance comparison of NCut is done with two other popular segmentation methods: mean shift (MS) and level sets (LS). These methods are complementary: mean shift is a kernel based method agglomerative clustering methods and is highly dependent on the parameters of the kernel. Level sets on the other is a segmentation by fitting method and typically requires strong gradient around the clinically relevant region, which is not the case for endoscopic images. Our novel image segmentation strategy integrates creasness features in NCut framework to obtain the resulting image segmentations which are better as compared to segmentation done by other methods.

8.2.2 Feature extraction

One of the most fundamental requirements in the feature extraction of endoscopic images is the robustness of image features to rotation, scale and illumination changes in the images. This requirement follows from dynamic imaging conditions which are attributed to the lack of low level control of the camera. A robust characterization of the tissues means that a a tissue of correctly classified into a particular class irrespective of the lighting conditions. Gabor filters were chosen for feature extraction motivated by their optimal spatial-frequency resolution, multiresolution nature and similarity with primary visual cortex. Gabor filter banks provide a detailed analysis of the images at various scales and orientations. A detailed study of invariance characteristics of the Gabor filters was carried out. This study led to a novel approach to obtain invariant features was proposed - Autocorrelation Gabor Features (AGF). The invariance of the novel features was empirically demonstrated on the available dataset. These AGF were later used for classification based on two different approaches: Bag of words approach (texton-AGF) and Autocorrelation Homogeneous Texture (AHT). Experiments were carried out to compare the novel descriptors with other state-of-the-art approaches.

As a conclusion, robustness of the AGF to rotation, scale and illumination changes was evident and higher classification rates for classifying images from our dataset were achieved. The results also showed that methods such as local binary pattern (LBP) which restrict the image analysis to only a small spatial neighbourhood are not useful as generic descriptors for a wide range of imaging scenarios.

8.2.3 Future work

The research carried out in this thesis has broadened the horizon and opened more research areas in various modules for CAD systems. These can be broadly divided into the following categories:

8.2.3.1 Specific objectives

Although we have provided rotation, scale and illumination invariant texture descriptors, there is one more contribution to invariance which is vital for feature extraction and is of particular interest for capsule endoscopy i.e., illumination gradient. If the visual descriptors are invariant to this gradient, they will be even more useful as they will provide another added advantage that is vital for several endoscopic

imaging scenarios. It is clear that our focus in this thesis has been the proposal of texture descriptors which are invariant to certain transformations. For systems having the objective of achieving higher classification accuracy, using only texture features is not adequate. State-of-the-art research in gastroenterology shows that color descriptors can be useful for achieving higher classification accuracy in various scenarios. However, the shortcoming of the color spaces rests in the fact that they are designed to encompass the full spectrum of visible wavelengths. Some spaces such as CIEXYZ even encompass that part of the color space which is not in the visible spectrum. It is therefore vital to use adapted color spaces for extraction of color features from the images. These can give more rich descriptors which can eventually help in increasing the overall classification performance for assisted decision making, which is one of the important objectives for a reliable system.

An assumption made for image classification in our work is that we have 'perfect segmentation' of our images. Although this is adequate for research purposes in assessing which visual descriptors work well for the specific scenarios, a system deployed in the hospital used for assisted decision making is eventually expected to perform image segmentation in an automated manner followed by feature extraction and classification of the image. It is therefore important to create a pattern recognition system as discussed in Chapter 3 by the integration of the above mentioned modules. This integration poses some novel challenges for further research such as the determination of the most relevant image patch from a segmented image. It will also be interesting to test the algorithms presented in this thesis to several other imaging modalities such as capsule endoscopy.

8.2.3.2 Generic objectives

Generically, the main objective of the CAD system from a computer vision perspective should be to give a high classification accuracy. Currently, this an evolving area of research so such systems are currently limited to assisted decision making. Therefore, it is acceptable for the time being not to have perfect diagnosis. In specific, it is acceptable to have false positives to raise false alarms. In assisted decision making, the physician is expected to analyse such cases manually to confirm the diagnosis. However it is preferable not to have false negatives as otherwise, the system cannot be used for screening of population. A fully automatic system is far fetched however, generically in the future, it is vital to improve the overall classification rates as much as possible for developing a fully automatic system for decision making. Integration of

a superior information access system and human computer interaction are necessary for full automation of the system.

As a final remark, the higher incidence rate of cancer of the gastrointestinal tract currently is taking its toll on a high number of human lives. The increase in life expectancy from GI cancer is linked to an early detection of cancer. Therefore, significant amount of work and resources are needed to address this concern to save millions of human lives.

Appendix A

Acronyms

ACF	Aberrant Crypt Foci
AGF	Autocorrelation Gabor features
AHT	Autocorrelation homogeneous texture
CAD	Computer Assisted Decision
CE	Capsule Endoscopy
CH	Chromoendoscopy
CIE	Commission Internationale de l'Eclairage
CT	Computed Tomography
CV	Computer Vision
DSC	Dice Similarity Coefficient
DCD	Dominant Color Descriptor
DTCWT	Dual-Tree Complex Wavelet Transform
DWT	Discrete Wavelet Transform
EHD	Edge Histogram Descriptor
FP	False Positive
FN	False Negative
GE	Gastroenterology
GF	Gabor filters
GI	Gastrointestinal
GLCM	Gray Level Co-occurrence Matrices
HSI	Hue-Saturation-Intensity
HSV	Hue-Saturation-Value
HT	Homogeneous texture

LBP	Local Binary Patterns
LOOCV	Leave-one-out Cross Validation
LS	Level sets
MR	Multiresolution
MRI	Magnetic Resonance Imaging
MS	Mean shift
NBI	Narow-band imaging
NCut	Normalized Cuts
PR	Pattern recognition
RGB	Red-Green-Blue
ROI	Region of interest
ROC	Receiver operating characteristics
SBI	Suspected Blood Indicator
SGFS	Simple Gabor Feature Space
SVM	Support vector machine
TN	True Negative
TP	True Positive
VE	Virtual Endoscopy

Appendix B

Color Conversion

To use an adequate color spaces for particular applications, conversion of RGB images to the relevant color spaces is needed.

B.1 RGB to HSV conversion

In each cylinder of an HSV space, the angle around the central vertical axis corresponds to *hue*, the distance from the axis corresponds to *saturation* and the distance along the axis corresponds to *value*. This axis represents the gray scale points. The conversion formulas are as follows [122]:

$$H = \arccos\left\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{\sqrt{(R-G)^2+(R-B)(G-B)}}\right\} \tag{B.1}$$

$$S = 1 - \frac{3}{R+G+B}[\min(R,G,B)] \tag{B.2}$$

$$V = \max(R,G,B) \tag{B.3}$$

B.2 sRGB to CIELUV conversion

Assuming that the input from a display is sRGB, it is first transformed to linear sRGB as follows:

$$C_{linear} = \begin{cases} \frac{C_{srgb}}{12.92}, & C_{srgb} \leq 0.04045 \\ (\frac{C_{srgb}+a}{1+a})^{2.4}, & C_{srgb} > 0.04045 \end{cases} \tag{B.4}$$

Where C is R, G or B. Linear RGB values can be transformed to and from CIEXYZ using a three-by-three matrix transformation as follows [124]:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7151 & 0.0721 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R_{linear} \\ G_{linear} \\ B_{linear} \end{bmatrix} \tag{B.5}$$

From CIE XYZ, the following transformation is used to convert to CIELUV color space [105]:

$$L^* = \begin{cases} 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 & , \ \frac{Y}{Y_n} > 0.008856 \\ 903.3\frac{Y}{Y_n} & otherwise \end{cases} \tag{B.6}$$

$$u^* = 13L^*(u' - {u_n}') \tag{B.7}$$

$$v^* = 13L^*(v' - {v_n}') \tag{B.8}$$

where u', v' and ${u_n}', {v_n}'$ are calculated from

$$u' = \frac{4X}{X + 15Y + 3Z}, v' = \frac{9Y}{X + 15Y + 3Z} \tag{B.9}$$

$${u_n}' = \frac{4X_n}{X_n + 15Y_n + 3Z_n}, {v_n}' = \frac{9Y_n}{X_n + 15Y_n + 3Z_n} \tag{B.10}$$

where X_n, Y_n and Z_n are constants of values 0.9504, 1.0 and 1.09 respectively.

Appendix C

List of Publications

[1]. F. Riaz, F. Silva, M. Ribeiro, M. Coimbra, "Impact of Visual Features on The Segmentation of Gastroenterology Images Using Normalized Cuts", *IEEE trans. Biomedical Engg. (submitted)*, 2012.

[2]. F. Riaz, F. Silva, M. Ribeiro, M. Coimbra, "Invariant Gabor Texture Descriptors for Classification of Gastroenterology Images", *IEEE trans. Biomedical Engg. (second review round)*, 2012.

[3]. F. Riaz, F. Silva, M. Ribeiro, M. Coimbra, "Towards Rotation, Scale and Illumination Invariant Texture Features for Gastroenterology Images", *IEEE International Conference on Pattern Recognition (submitted)*, 2012.

[4]. F. Riaz, F. Silva, M. Ribeiro, M. Coimbra, "How Well Do Current Segmentation Algorithms Produce Clinically Relevant Segmentations for Gastroenterology", *IEEE EMBC (submitted)*, 2012.

[5]. F. Riaz, M. Areia, F. Vilarino, M. Dinis-Ribeiro, M. Coimbra, "Identifying potentially cancerous tissues in Chromoendoscopy images", in *proc. of Iberian Conference on Pattern Recognition and Image Analysis*, Gran Canaria, Spain, Jun 2011.

[6]. F. Riaz, M. Areia, F. Baladaque, P. Nunes, M. Dinis-Ribeiro, M. Coimbra, "Gabor Textons for Classification of Gastroenterology Images", in *proc. of IEEE International Symposium on Biomedical Imaging*, Chicago, USA, Apr 2011.

[7]. M. Coimbra, F. Riaz, M. Areia, F. Baladaque, M. Dinis-Ribeiro, "Segmentation for Classification of Gastroenterology Images", in *proc. of IEEE EMBC, Buenos Aires*, Argentina, Sep 2010.

[8]. M. Dinis-Ribeiro, F. Riaz, M. Coimbra, "Computer-assisted analysis of gastric magnification chromoendoscopy images: feasibility and potential applications", (abstract) in *World Congress of Gastroenterology*, London, UK, Nov 2009.

[9]. F. Riaz, M. Dinis-Ribeiro, M. Coimbra, "Quantitative Comparison of Segmentation Methods for In-Body Images", in *proc. of IEEE EMBC*, Minneapolis, USA, Sep 2009.

[10]. F. Riaz, A. Sousa, P. Pimentel-Nunes, M. Areia, M. Coimbra, M. Dinis-Ribeiro, "Analise assistida por computador de imagens de cromendoscopia gastrica: praticabilidade e aplicabilidade", (abstract) in *29th National Meeting in Gastroenterology*, Porto, Portugal, 2009.

[11]. F. Riaz, M. Dinis-Ribeiro, M. Coimbra, "A Review of Current Computer Aided Diagnosis Systems for Polyp Detection in Virtual Colonoscopy", in *proc. of Conftele*, Santa Maria da Feira, Portugal, May 2009.

[12]. F. Riaz, M. Dinis-Ribeiro, M. Coimbra, "Semantic relevance of current image segmentation algorithms", in *proc. of IEEE WIAMIS*, London, UK, May 2009.

References

[1] CIELUV color model. *http://dba.med.sc.edu/price/irf/Adobe_tg/models/cieluv.html*.

[2] The electromagnetic spectrum. *http://bigfootproof.com/groups/visible-invisible-d2-ranges.html*.

[3] HSV color space. *https://www.e-education.psu.edu/geog486/l2_p9.html*.

[4] Mathworks GLCM. *http://matlab.izmiran.ru/help/toolbox/images/enhanc15.html*.

[5] A. Albiol, L. Torres, and E. J. Delp. Optimum color spaces for skin detection. *IEEE International conference on image processing*, 1, 2001.

[6] G. K. Anagnostopoulos, K. Yao, C. J. Hawkey, and K. Ragunath. Novel endoscopic observation in Barrett's oesophagus using high resolution magnification endoscopy and narrow band imaging. *Aliment Pharmacol Ther*, 26, 2007.

[7] M. Areia, P. Amaro, M. Dinis-Ribeiro, M. A. Cipriano, C. Marinho, A. Costa-Pereira, C. Lopes, L. Moreira-Dias, J. M. Romaozinho, H. Gouveia, D. Freitas, and M. C. Leitao. External validation of a classification for methylene blue magnification chromoendoscopy in premalignant gastric lesions. *Gastrointestinal Endoscopy*, 2008.

[8] S. Arivazhagan and L. Ganesan. Texture classification using wavelet transform. *Pattern Recognition Letters*, 24, 2003.

[9] J. Baillie. *Gastrointestinal endoscopy: Basic principles and practice*. Butterworth-Heinemann (Oxford and Boston), 1992.

[10] S. Bejakovic, R. Kumar, T. Dassopoulos, G. Mullin, and G. Hager. Analysis of crohn's disease lesions in capsule endoscopy images. *IEEE International Conference on Robotics and Automation*, 2009.

[11] G. Berci and M. Paz-Partlow. Electronic imaging in endoscopy. *Surgical Endoscopy*, 2(4), 1988.

[12] J. Bernal, F. J. Sanchez, and F. Vilarino. A region segmentation method for colonoscopy images using a model of polyp appearance. *5th Iberian Conference on Pattern Recognition and Image Analysis*, 6669, 2011.

[13] S. Beucher. The watershed transformation applied to image segmentation. *Conference on Signal and Image Processing in Microscopy and Microanalysis*, 1991.

[14] M. H. Bharati, J. J. Liu, and J. F. Macgregor. Image texture analysis: methods and comparisons. *Journal of Chemometrics and Intelligent Laboratory Systems*, 72(1), 2004.

[15] B. Block, G. Schachschal, and H. Schmidt. *Endoscopy of the upper GI tract: A training manual*. Thieme publishers, 2004.

[16] A.C. Bovik, M. Clarke, and W.S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE transactions on Pattern Recognition and Machine Intelligence*, 12(1), 1990.

[17] Y. Boykov and G. Funka-Lea. Graph cuts and efficient n-d image segmentation. *International Journal on Computer Vision*, 70(2), 2006.

[18] M. J. Bruno. Magnification endoscopy, high resolution endoscopy, and chromoscopy; towards a better optical diagnosis. *British Medical Journal*, 52(4), 2003.

[19] F. W. Campbell and J. G. Robson. Application of fourier analysis to the visibility of gratings. *Journal of Physiology*, 197, 1968.

[20] C. Candan, M. A. Kutay, and H. M. Ozaktas. The discrete fractional fourier transform. *IEEE transactions on Signal Processing*, 48(5), 2000.

[21] E. J. Candes and D. L. Donoho. *Curvelets - A surprisingly effective nonadaptive representation for objects with edges. Curves and Surfaces*, C. Rabut, A. Cohen, and L. L. Schumaker, Eds. Nashville, TN: Vanderbilt Univ. Press, 2000.

[22] M. I. Canto. Staining in gastrointestinal endoscopy: The basics. *Endoscopy*, 31(6), 1999.

[23] M. I. Canto. Vital staining and barrett's esophagus. *Gastrointestinal Endoscopy*, 49(3), 1999.

[24] E. C. C. Cauberg, D. M. Bruin, D. J. Faber, T. G. Leeuwen, J. J. Rosette, and T. M. Reijke. A new generation of optical diagnostics for bladder cancer: Technology, diagnostic accuracy, and future applications. *European Urology*, 2009.

[25] T. Chan and L. Vese. An active contour model without edges. *Lecture notes in computer science*, 1999.

[26] B. B. Chaudhuri, N. Sarkar, and P. Kundu. Improved fractal geometry based texture segmentation technique. *IEE Proceedings-E of Computers and Digital Techniques*, 140(5), 1993.

[27] C. H. Chen, L. F. Pau, and P. S. P. Wang. *Handbook of Pattern Recognition and Computer Vision*. World Scientific Publishing Co., 1998.

[28] Y. Cheng. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE transactions on Information Theory*, 21(1), 1975.

[29] Y. Cheng. Mean shift, mode seeking and clustering. *IEEE transactions on Pattern Recognition and Machine Intelligence*, 17(8), 1995.

[30] V. Christos, L. J. Hadjileontiadis, C. N. Liatsos, C. C. Mavrogiannis, and G. D. Sergiadis. Abnormal pattern detection in wireless capsule endoscopy images using nonlinear analysis of rgb color space. *in proc. of IEEE EMBC*, 2010.

[31] H. Cichoz-Lach and K. H. Celinski. Modern methods of endoscopic diagnosis of gastrointestinal tract. *Journal of Physiology and Pharmacology*, 58(3), 2007.

[32] M. T. Coimbra and J. P. Cunha. Mpeg-7 visual descriptors - Contributions for automated feature extraction in capsule endoscopy. *IEEE Trans. on Circuits and Systems for Video Technology*, 16(5), 2006.

[33] Children Hospital Colorado.

[34] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE transactions on Pattern Recognition and Machine Intelligence*, 24(5), 2002.

[35] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603 –619, may 2002.

[36] A. H. Dachman and H. Yoshida. *Gastroenterology and hepatology: the modern clinicians guide*. Elsevier Science, 2004.

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

[38] C. Demir and B. Yener. Automated cancer diagnosis based on histopathological images: a systematic survey. *Rensselaer Polytechnic Institute, Troy, NY*, 2006.

[39] R. C. Dubes and A. K. Jain. Random field models in image analysis. *Journal of Applied Statistics*, 16, 1989.

[40] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and sons Inc., 2000.

[41] F. Emura, Y. Saito, and H. Ikematsu. Narow-band imaging optical chromocolonoscopy: Advantages and limitations. *World Journal on Gastroenterology*, 14(31), 2008.

[42] O. D. Faugeras and W. K. Pratt. Decorrelation methods of texture feature extraction. *IEEE transactions on Pattern Recognition and Machine Intelligence*, 2(4), 1980.

[43] P. Felzenszwalb and D. Huttenlocher. Efcient graph-based image segmentation. *International Journal on Computer Vision*, 2004.

[44] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *IEE Journal of Radio and Communication Engineering*, 4(12), 1987.

[45] I. N. Figueiredo, P. N. Figueiredo, G. Stadler, O. Ghattas, and A. Araujo. Variational image segmentation for endoscopic human colonic aberrant crypt foci. *IEEE Transactions on Medical Imaging*, 29(4), 2010.

[46] M. Filip, S. Lordache, A. Sftoiu, and T. Ciurea. Autofluorescence imaging and magnification endoscopy. *World Journal of Gastroenterology*, 17(1), 2011.

[47] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2003.

[48] D. Gabor. Theory of communication. *IEE Journal of Radio and Communication Engineering*, 93(26), 1946.

[49] J. M. Gauch and S. M. Pizer. Multiresolution analysis of ridges and valleys in grey-scale images. *IEEE trans. Pattern Analysis and Machine Intelligence*, 15(6):635–646, 1993.

[50] J. J. Gibson. *The Perception of the Visual World*. Boston: Houghton Mifflin, 1950.

[51] F. Gmez and E. Romero. Rotation invariant texture characterization using a curvelet based descriptor. *Pattern Recognition Letters*, 32(16), 2011.

[52] K. Goda, H. Tajiri, M. Ikegami, A. Dobashi, and N. Yoshimura. Clinical impact of narrow-band imaging magnifying endoscopy for 'basal layer type squamous cell carcinoma' in the esophagus. *Digestive Endoscopy*, 23(1), 2011.

[53] M. Goetz, M. Vieth, S. Kanzler, P. R. Galle, P. Delaney, M. F. Neurath, and R. Kiesslich. In vivo confocal laser laparoscopy allows real time subsurface microscopy in animal models of liver disease. *Journal of Hepatology*, 48(1), 2008.

[54] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital Image Processing*. Prentice hall, 2008.

[55] L. J. Grady and E. L. Schwartz. Isoperimetric graph partitioning for data clustering and image segmentation. *Technical report, Department of Cognitive and Neural Systems, Boston University*, 2003.

[56] D. H. Granlund. In search of a general picture processing operator. *Computer Graphics and Image Processing*, 8, 1978.

[57] T. Grenier, C. R. Muller, F. Davignon, O. Basset, and G. Gimenez. Variable bandwidth mean shift for smoothing ultrasonic images. *Proc. of EUSIPCO*, 2005.

[58] M. Hafner, L. Brunauer, H. Payer, R. Resch, A. Gangl, F. Wrba, and A. Vecsei. Computer-aided classification of zoom endoscopical images using fourier filters. *IEEE trans. on Info. Tech. in Biomed.*, 14(4), 2010.

[59] M. Hafner, R. Kwitt, A. Ulh, F. Wrba, A. Gangl, and A. Vecsei. Computer-assisted pit-pattern classification in different wavelet domains for supporting dignity assessment of colonic polyps. *Journal on Pattern Recognition*, 42(6), 2008.

[60] G.M. Haley and B.S. Manjunath. Rotation-invariant texture classification using a complete space-frequency model. *IEEE Trans. Image Processing*, 8, 1999.

[61] M. T. Hallissey, W. H. Allum, A. J. Jewkes, D. J. Ellis, and J. W. Fielding. Early detection of gastric cancer. *British Medical Journal*, 301, 1990.

[62] J. Han and K. K. Ma. Rotation-invariant and scale-invariant gabor features for texture image retrieval. *Image and Vision Computing*, 25(9), 2007.

[63] A. K. Hara, J. A. Leighton, V. K. Sharma, R. I. Heigh, and D. E. Fleischer. Imaging of small bowel disease: Comparison of capsule endoscopy, standard endoscopy, barium examination, and ct. *RadioGraphics*, 25, 2005.

[64] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5), 1979.

[65] R. M. Haralick, K. Shanmugan, and I. Dinstein. Texture features for image classification. *IEEE transactions on Systems, Man and Cybernatics*, 3(6), 1973.

[66] J. Heikkila. A new class of shift invariant operators. *IEEE Sig. Proc. Letters*, 11(6), 2004.

[67] P. S. Hiremath, B. Dhandra, R. Hegadi, and G. Rajput. Abnormality detection in endoscopic images using color segmentation and curvature computation. *in proc. of International Conf. on Neural Info.*, 3316, 2004.

[68] D. P. Hurlstone, D. S. Sanders, A. J. Lobo, M. E. McAlindon, and S. S. Cross. Indigo carmine-assisted high-magnification chromoscopic colonoscopy for the detection and characterisation of intraepithelial neoplasia in ulcerative colitis: A prospective evaluation. *Endoscopy*, 37(12), 2005.

[69] S. Hwang and M. Celebi. Polyp detection in wireless capsule endoscopy videos based on image segmentation and geometric feature. *Internation Conf. on Acoustics Speech and Sig. Proc.*, 2010.

[70] S. Hwang, J. H. Oh, J. Cox, S. J. Tang, H. F., and Tibbals. Blood detection in wireless capsule endoscopy using expectation maximization clustering. *Proc. of SPIE on Med. Imaging*, 6144, 2006.

[71] M. Islam, D. Zhang, and G. Lu. Rotation invariant curvelet features for texture image retrieval. *IEEE Int. Conf. on Multimedia and Expo.*, 2009.

[72] D. W. Jacobs, D. Weinshall, and Y. Gdalyahu. Classification with nonmetric distances: Image retrieval and class representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6):583–600, June 2000.

[73] K. Jafari-Khouzani and H. Soltanian-Zadehl. Radon transform orientation estimation for rotation invariant texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6), 2005.

[74] A. K. Jain and D. Zongker. Representation and recognition of handwritten digits using deformable templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(12):1386–1390, 1997.

[75] B. Julesz. Textons, the elements of texture perception, and their interactions. *Nature*, 290, 1981.

[76] J. K. Kamarainen. Feature extraction using gabor filters. *PhD thesis, Lappeenranta University of Technology*, 2003.

[77] J.-K. Kamarainen, V. Kyrki, and H. Kamarainen. Invariance properties of gabor filter based features - overview and applications. *IEEE Trans. on Image Process*, 15(5), 2006.

[78] A. Karargyris and N. Bourbakis. Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos. *IEEE trans. Biomedical Engg.*, 58(10), 2011.

[79] S. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras. Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Trans. Infor. Tech. Biomed.*, 7(3), 2003.

[80] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), 1988.

[81] L. Kaufman and P. J. Rousseeuw. *Finding groups in data: An introduction to cluster analysis.* Wiley series in probability and mathematical statistics, John Wiley and Sons Inc., Newyork, 1990.

[82] R. Kumar, Z. Qian, S. Seshamani, G. Mullin, G. hager, and T. Dassopoulos. Assessment of chron's disease lesions in wireless capsule endoscopy images. *IEEE trans. Biomedical Engg.*, 59(2), 2012.

[83] V. Kyrki, J. K. Kamarainen, and H. Kalviainen. Simple gabor feature space for invariant object recognition. *Pattern Recognition Letters*, 25(3), 2003.

[84] A. Laine and J. Fan. Texture classification by wavelet packet signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11), 2000.

[85] B. Li, M. Meng, and J. Lau. Computer-aided small bowel tumor detection for capsule endoscopy. *Artificial Intelligence in Medicine*, 52(1), 2011.

[86] B. Li and M. H. Meng. Computer-aided detection of bleeding regions and capsule endoscopy images. *IEEE trans. Biomedical Engg.*, 56(4), 2009.

[87] B. Li and M. H. Meng. Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. *Journal of Computers in Biology and Medicine*, 39(2), 2009.

[88] B. Li and M. H. Meng. Texture analysis for ulcer detection in capsule endoscopy images. *Image and Vision Computing*, 27, 2009.

[89] C. Li, C. Xu, C. Gui, and M.D. Fox. Level set evolution without re-initialization: a new variational formulation. volume 1, pages 430 – 436, June 2005.

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

[91] J. Ma and G. Plonka. The curvelet transform. *IEEE Signal Processing Magazine*, 27(2), 2010.

[92] M. Mackiewicz, M. Fisher, and C. Jamieson. Bleeding detection in wireless capsule endoscopy using adaptive color histogram model and support vector classification. *in Proc. of SPIE*, 6914, 2008.

[93] A. Maistrou. Level set methods - overview. *Technical report, Computer Aided Medical Procedures, TUM*, 2008.

[94] J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *International Journal on Computer Vision*, 2001.

[95] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8), 1996.

[96] B. S. Manjunath, J. R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE trans. on Circuits and Systems for Video Technology*, 11(6), 2001.

[97] B. S. Manjunath, P. Wu, S. Newsam, and H. D. Shin. A texture descriptor for browsing and similarity retrieval. *Elsevier, Signal Processing: Image Communication*, 16(1), 2000.

[98] R. Manthalkar, P. K. Biswas, and B. N. Chatterji. Rotation invariant texture classification using even symmetric gabor filters. *Pattern Recognition Letters*, 24(12), 2003.

[99] J. Mao and A. K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition*, 25(2), 1992.

[100] T. Nakamura and A. Terano. Capsule endoscopy: past, present, and future. *Journal of Gastroenterology*, 43(2), 2008.

[101] M. M. Navas. Capsule endoscopy. *World Journal of Gastroenterology*, 15(13), 2009.

[102] V. Nedovic. Visual feature extraction for content analysis. 2006.

[103] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29, 1996.

[104] S. Osher and J. A. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79(1), 1988.

[105] D. Park, J. Park, and J.H. Han. Image indexing using color histogram in the cieluv color space. *Proceedings of 5th Japan-Korea workshop on computer vision*, 1999.

[106] A. A. Pentland. Fractal based description of natural scenes. *IEEE transactions on Pattern Recognition and Machine Intelligence*, 6, 1984.

[107] S. D. Pitlik, V. Fainstein, D. Garza, L. Guarda, R. Bolivar, A. Rios, R. L. Hopfer, and P. A. Mansell. Human cryptosporidiosis: spectrum of disease. report of six cases and review of the literature. *Arch. Intern. Med.*, 143(12), 1983.

[108] V. Popovici and J. P. Thiran. Pattern classification using higher-order local autocorrelations coefficients. *Pattern Recognition Letters*, 25(10), 2004.

[109] R. Porter and N. Canagarajah. Robust rotation-invariant texture classification: wavelet, gabor filter and gmrf based schemes. *IEE proceedings vision, image and signal processing*, 144(3), 1997.

[110] F. Riaz, M. Areia, F. B. Silva, M. Dinis-Ribeiro, P. P. Nunes, and M. Coimbra. Gabor textons for classification of gastroenterology images. *Proc. of IEEE ISBI*, 2011.

[111] M. D. Ribeiro. Clinical, endoscopic and laboratorial assessment of patients with associated lesions to gastric adenocarcinoma. *Faculdade de Medicina da Universidade do Porto, PhD thesis*, 2005.

[112] J. H. Rieger. Topographical properties of generic images. *International Journal of Computer Vision*, 23(1):79–92, 1997.

[113] C. J. Van Rijsbergen. *Information Retrieval.* Butterworth-Heinemann, Newton, MA, USA, 1979.

[114] J. C. Russ. *The Image Processing Handbook.* CRC Press London, 1994.

[115] M. Sarifuddin and R. Missaoui. A new perceptually uniform color space with associated color similarity measure for content-based image and video retrieval. *Proc. ACM SIGIR Workshop on Multimedia Information Retrieval*, 2005.

[116] S. M. Savaresi, D. L. Boley, S. Sittanti, and G. Gazzaniga. Cluster selection in divisive clustering algorithms. *in proc. of SIAM International Conference on Data Mining*, 2002.

[117] S. Selvan and S. Ramakrishnan. Svd-based modeling for texture classification using wavelets transformation. *IEEE transactions on Image Processing*, 16(11), 2007.

[118] P. Sharma, A. Bansal, S. Mathur, S. Wani, R. Cherian, D. McGregor, A. Higbee, S. Hall, and A. Weston. The utility of a novel narrow band imaging endoscopy system in patients with Barretts esophagus. *Gastrointestinal Endoscopy*, 64, 2006.

[119] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888 –905, Aug 2000.

[120] C. Signorelli, F. Villa, E. Rondonotti, C. Abbiati, G. Beccari, and R. Franchis. Sensitivity and specificity of the suspected blood identification system in video capsule enteroscopy. *Endoscopy*, 37, 2005.

[121] R. Singh, G. K. Anagnostopoulos, K. Yao, H. Karageorgiou, P. J. Fortun, A. Shonde, K. Garsed, P. V. Kaye, C. J. Hawkey, and K. Ragunath. Narrow-band imaging with magnification in Barrett's esophagus: validation of a simplified grading system of mucosal morphology patterns against histology. *Endoscopy*, 40, 2008.

[122] A. Sousa. Analysis of color and texture features of vital stained magnification endoscopy images for computer assisted diagnosis of precancerous and cancer lesions. *Phd Thesis, Faculdade de Engenharia da Universidade do Porto*, 2008.

[123] A. Sousa, M. Dinis-Ribeiro, M. Areia, and M. Coimbra. Identifying cancer regions in vital-stained magnification endoscopy images using adapted color histograms. *Proc. of IEEE ICIP*, 2009.

[124] M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta. *A standard default color space for the internet - sRGB.* HP and Microsoft Tech. Report, 1996.

[125] I. Sumana, M. Islam, Z. Dengsheng, and L. Guojun. Content based image retrieval using curvelet transform. *IEEE Internat. Workshop Multimedia Signal Process.*, 2008.

[126] W. Tao, H. Jin, and Y. Zhang. Color image segmentation based on mean shift and normalized cuts. *IEEE trans. Systems, Man and Cybernatics*, 37(5):1382–1389, 2007.

[127] J. P. Thirion and A. Gourdon. Computing the differential characteristics of isointensity surfaces. *Computer Vision and Image Understanding*, 61(2):190–202, 1995.

[128] G. J. Tortora and B. H. Derrickson. *Principles of anatomy and physiology.* Wiley Inc., 2000.

[129] J. T. Tou and R. C. Gonzalez. *Pattern Recognition Principles.* Massachusetts: Addison-Wesley, 1974.

[130] A. Y. Tsai, A. J. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky. A shape-based approach to the segmentation of

medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2), 2003.

[131] M. Tuceryan and A.K. Jain. Texture analysis. 1993.

[132] V. Vapnik. *Estimation of Dependences Based on Empirical Data*. Springer Verlag, Newyork, 1979.

[133] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1), 2005.

[134] E. Vazquez, R. Baldrich, J. Weijer, and M. Vanrell. Describing reflectances for color segmentation robust to shadows, highlights, and textures. *IEEE trans. Pattern Analysis and Machine Intelligence*, 33(5):917–930, 2011.

[135] A. Vilanova, A. Konig, and E. Groller. VirEn: a virtual endoscopy system. *Machine Graphics and Vision*, 8(3), 1999.

[136] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE transactions on Pattern Recognition and Machine Intelligence*, 13(6), 1991.

[137] J. D. Waye, D. K. Rex, and C. B. Williams. *Colonoscopy: Principles and practice*. Wiley-Blackwell, 2011.

[138] H. S. Winter, M. S. Murphy, J. F. Mougenot, and S. Cadranel. *Pediatric gastrointestinal endoscopy*. Hamilton, Ont., London, 2006.

[139] B. J. Wood and P. Razavi. Virtual endoscopy: A promising new technology. *American family physician*, 66(1), 2002.

[140] Z. Wu and R. Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1101–1113, Nov 1993.

[141] X. Xie, Q. Dai, K. M. Lam, and H. Zhao. Efficient rotation- and scale-invariant texture classification method based on gabor wavelets. *Journal of Electronic Imaging*, 17(4), 2008.

[142] A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tennanbaum. A geometric snake model for segmentation of medical imagery. *IEEE Transactions on Medical Imaging*, 16(2), 1997.

[143] S. S. Yu. Segmentation using multiscale cues. *IEEE Conference on Computer Vision and Pattern Recognition*, 1:247–254, Jul 2004.

[144] S. X. Yu and J. Shi. Multiclass spectral clustering. *in IEEE International Conference on Computer Vision*, 1:313–319, 2003.

[145] Q. J. Zhou, J. M. Yang, B. Y. Fei, Q. S. Xu, W. Q. Wu, and H. J. Ruan. Narrow-band imaging endoscopy with and without magnification in diagnosis of colorectal neoplasia. *World Journal of Gastroenterology*, 17(5), 2011.

[146] S. Zhu, C. Guo, Y. Wu, and Y. Wang. What are textons. *European Conference on Computer Vision*, 2002.