2D/3D Endoscopic image enhancement and analysis for video guided surgery

Bilel Sdiri

Thesis submitted to:
Norwegian University of Science and Technology, &
Université Paris 13, Sorbonne Paris Cité

for the degree of Doctor of Philosophy respectively in Computer Science and Signal & Image processing

UNIVERSITÉ PARIS 13

USPC
Université Sorbonne Paris Cité

NTNU
Norwegian University of Science and Technology

2018

2D/3D Endoscopic image enhancement and analysis for video guided surgery

Thesis submitted for the cotutelle degree of
DOCTOR OF PHILOSOPHY in
COMPUTER SCIENCE [1], and in
SIGNAL & IMAGE[2]

Defended by:

BILEL SDIRI

on February 6^{th}, 2018.

Jury:

Reviewers:	Prof. Kacem Chehdi, Université de Rennes 1, France.
	Prof. Mihai Mitrea, Institut Mines-Télécom Paris, France.
Examiners:	Prof. Pierre Gouton, Université de Bourgogne, France.
	Prof. Anissa Mokraoui, Université Paris 13, France.
	Dr. Mounir Kaaniche, Université Paris 13, France.
Supervisors:	Prof. Azeddine Beghdadi , Université Paris 13, France.
	Prof. Faouzi Alaya Cheikh, NTNU in Gjøvik, Norway.
	Prof. Ole Jakob Elle, Oslo University Hospital, Norway.
	Prof. Edwin Bjørn , Oslo University Hospital, Norway.

[1]from the Norwegian University of Science and Technology (NTNU)
[2]from Université Paris 13, Sorbonne Paris Cité.

Learn from yesterday, live for today, hope for tomorrow.
The important thing is not to stop questioning.

(Albert Einstein)

Declaration of Authorship

I, Bilel Sdiri, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed:

(Bilel Sdiri)

Date: 10/01/2018

Summary

Minimally invasive surgery has made remarkable progress in the last decades and became a very popular diagnosis and treatment tool, especially with the rapid medical and technological advances leading to innovative new tools such as robotic surgical systems and wireless capsule endoscopy. Due to the intrinsic characteristics of the endoscopic environment including dynamic illumination conditions and moist tissues with high reflectance, endoscopic images suffer often from several degradations such as large dark regions, with low contrast and sharpness, and many artifacts such as specular reflections and blur. These challenges together with the introduction of three-dimensional (3D) imaging surgical systems have prompted the question of endoscopic images quality, which needs to be enhanced. The latter process aims either to provide the surgeons/doctors with a better visual feedback or improve the outcomes of some subsequent tasks such as features extraction for 3D organ reconstruction and registration. This thesis addresses the problem of endoscopic image quality enhancement by proposing novel enhancement techniques for both two-dimensional (2D) and stereo (i.e. 3D) endoscopic images.

In the context of automatic tissue abnormality detection and classification for gastro-intestinal tract disease diagnosis, we proposed a pre-processing enhancement method for 2D endoscopic images and wireless capsule endoscopy improving both local and global contrast. The proposed method expose inner subtle structures and tissues details, which improves the features detection process and the automatic classification rate of neoplastic, non-neoplastic and inflammatory tissues. Inspired by binocular vision attention features of the human visual system, we proposed in another work an adaptive enhancement technique for stereo endoscopic images combining depth and edginess information. The adaptability of the proposed method consists in adjusting the enhancement to both local image activity and depth level within the scene while controlling the inter-view difference using a binocular perception model. A subjective experiment was conducted to evaluate the performance of the proposed algorithm in terms of visual quality by both expert and non-expert observers whose scores demonstrated the efficiency of our 3D contrast enhancement technique. In the same scope, we resort in another recent stereo endoscopic image enhancement work to the wavelet domain to target the enhancement towards specific image components using the multiscale representation and the efficient space-frequency

localization property. The proposed joint enhancement methods rely on cross-view processing and depth information, for both the wavelet decomposition and the enhancement steps, to exploit the inter-view redundancies together with perceptual human visual system properties related to contrast sensitivity and binocular combination and rivalry. The visual quality of the processed images and objective assessment metrics demonstrate the efficiency of our joint stereo enhancement in adjusting the image illumination in both dark and saturated regions and emphasizing local image details such as fine veins and micro vessels, compared to other endoscopic enhancement techniques for 2D and 3D images.

Title: 2D/3D Endoscopic image enhancement and analysis for video guided surgery.
Key-words: Image enhancement, endoscopic/medical imaging, stereo images, image analysis, 3D.

Résumé

Grâce à l'évolution des procédés de diagnostiques médicaux et les développements technologiques, la chirurgie mini-invasive a fait des progrès remarquables au cours des dernières décennies surtout avec l'innovation de nouveaux outils médicaux tels que les systèmes chirurgicaux robotisés et les caméras endoscopiques sans fil. Cependant, ces techniques souffrent de quelques limitations liées essentiellement l'environnement endoscopique telles que la non uniformité de l'éclairage, les réflexions spéculaires des tissus humides, le faible contraste/netteté et le flou dû aux mouvements du chirurgien et du patient (i.e. la respiration). La correction de ces dégradations repose sur des critères de qualité d'image subjective et objective dans le contexte médical. Il est primordial de développer des solutions d'amélioration de la qualité perceptuelle des images acquises par endoscopie 3D. Ces solutions peuvent servir plus particulièrement dans l'étape d'extraction de points d'intérêts pour la reconstruction 3D des organes, qui sert à la planification de certaines opérations chirurgicales. C'est dans cette optique que cette thèse aborde le problème de la qualité des images endoscopiques en proposant de nouvelles méthodes d'analyse et de rehaussement de contraste des images endoscopiques 2D et 3D.

Pour la détection et la classification automatique des anomalies tissulaires pour le diagnostic des maladies du tractus gastro-intestinal, nous avons proposé une méthode de rehaussement de contraste local et global des images endoscopiques 2D classiques et pour l'endoscopie capsulaire sans fil. La méthode proposée améliore la visibilité des structures locales fines et des détails de tissus. Ce prétraitement a permis de faciliter le processus de détection des points caractéristiques et d'améliorer le taux de classification automatique des tissus néoplasiques et tumeurs bénignes. Les méthodes développées exploitent également la propriété d'attention visuelle et de perception de relief en stéréovision. Dans ce contexte, nous avons proposé une technique adaptative d'amélioration de la qualité des images stéréo endoscopiques combinant l'information de profondeur et les contours des tissues. Pour rendre la méthode plus efficace et adaptée aux images 3D le rehaussement de contraste est ajusté en fonction des caractéristiques locales de l'image et du niveau de profondeur dans la scène tout en contrôlant le traitement inter-vues par un modèle de perception binoculaire. Un test subjectif a été mené pour évaluer la performance de l'algorithme proposé en termes de qualité visuelle des images générées par des observateurs experts et non experts dont les scores ont démontré l'efficacité de notre technique 3D d'amélioration du contraste. Dans cette même optique, nous avons développé une autre technique de rehaussement du contraste des images endoscopiques stéréo basée sur la décomposition en ondelettes. Ce qui offre la possibilité d'effectuer un traitement multi-échelle et d'opérer une traitement sélectif. Le schéma proposé repose sur un traitement stéréo qui exploite à la fois l'informations de profondeur et les redondances inter-vues, ainsi que certaines propriétés du système visuel humain, notamment la sensibilité au contraste et à la rivalité/combinaison binoculaire. La qualité visuelle des images traitées et les mesures de qualité objective démontrent l'efficacité de notre méthode qui ajuste l'éclairage des images dans les régions sombres et saturées et accentue la visibilité des détails liés aux vaisseaux sanguins et les textures de tissues.

Titre: Rehaussement de contraste et analyse des images endoscopiques 2D/3D pour des chirurgies assistées par vidéo.
Mots-clés: Traitement d'images, images médicales , image endoscopiques , images stereo, 3D.

Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisors in both France and Norway, namely Professor *Azeddine Beghdadi* in Université Paris 13, Professor *Faouzi Alaya Cheikh* in the Norwegian University of Science and Technology, and Professors *Ole Jakob Elle*, and *Bjørn Edwin* in the Intervention Centre (IVS) at Oslo University Hospital. Thank you for your invaluable feedback, patience, motivation, support in overcoming numerous challenges, and opportunities given to me throughout my research work. I have learnt a lot from all of you, not only academically but also as a person. Beside my advisors, I would like to thank the committee members, namly Professor *Mihai Mitrea*, and Professor *Kacem Chehdi* for accepting to review my thesis work and for their insightful comments and questions.

Being in a co-tutelle PhD program has exposed me to two different educational and cultural norms at The Norwegian Colour and Visual Computing Laboratory (Colorlab) in Gjøvik, the Laboratory of Information Processing and Transmission (L2TI) in Paris, and the IVS in Oslo. It has been such a great experience to work in two different environments and interact with colleagues I met along the way, who became friends. A special mention to *Mounir Kaaniche* and *Rafael Palomar* for their constructive feedback, suggestions and support during our collaborations. I would like to thank as well my fellow doctoral students, colleagues, and friends in both labs and in the IVS, namely Hilda, Thomas, Rahul, Mohib, Louise, Sony, Magnus, Ismael, Mustapha, Linda, Walid, Nidhal, and many more. In addition, I would like to express my gratitude to the administration in both Colorlab and L2TI, namely Amel, Hilda, Jingjing, Rachael, the L2TI director Professor *Anissa Mokraoui*, and Galilée doctoral school director Professor *Dominique Ledoux* for their understanding, support and help in overcoming the different paperwork and logistic issues.

A very special gratitude goes out to my family members that I am very blessed to have in my life: My father *Salem*, my mother *Fatma*, and my sister *Dorra* who represent the core of my motivation to strive for excellence and reward them for their patience and struggle. Thank you for your sincere love, for your patience, prayers, and for your constant support throughout my entire life. My sincere thanks goes also to my aunt *Salma*, her husband *Mustapha* and her daughters who have been my family in Paris during the past six years since my arrival to France. Thank you for being there for me whenever I needed you, for your rides to the airport whenever I have a travel, for your listening, support, and sincere love.

Last but not the least, I would like to thank any person who have somehow motivated me, supported me, or taught me any thing starting from my primary school until my university period in Sfax, Paris, Gjøvik and Oslo.

Abbreviations

2D	Two-dimensional
3D	Three-dimensional
AFI	Autoflorescence imaging
AHE	Adaptive histogram equalization
AMBE	Absolute mean brightness error
AME	Absolute measure of enhancement
ARWR	Adaptive random walk with restart
BJND	Binocular just-noticeable difference model
BOF	Bag of words
BP	Brightness reservation
BPDFHE	Brightness preserving dynamic fuzzy histogram equalization
BPDHE	Brightness preserving dynamic histogram equalization
CCD	Charge-coupled devices
CE	Contrast enhancement
CEE	Contrast enhancement evaluation
CLAHE	contrast limited adaptive histogram equalization
CLE	Confocal laser endomicroscopy
CNNs	Convolution neural networks
CT	Computed tomography
CWT	Complex wavelet transform
DC	Disparity compensation
DCT	Discrete cosine transform
DE	Disparity estimation
DL	Deep learning
DOG	Difference of gaussians
DSCQS	Double-stimulus continuous quality scale
EBCE	Edge-based contrast enhancement method
EC	Edge content
EIE	Endoscopic image enhancement
EME	Measure of enhancement
FICE	Fuji intelligent chromoendoscopy
FT	Fourier transform
GHE	Global histogram equalization

GI	Gastro-intestinal
HCC	Hepatocellular Carcinoma
HD	High definition
HDR	High dynamic range
HE	Histogram equalization
HJND	Hybrid just noticeable difference
HVS	Human visual system
IEM	Image enhancement measure
IR	Infrared
IVS	The Intervention Centre
JJND	Joint just noticeable difference
JND	Just noticeable difference
JNDD	Just noticeable difference in depth
LDR	Low dynamic range
LGB	Lateral geniculate body
LS	Lifting Scheme
MB	Mean brightness
MBI	Multi-band imaging
ME-NBI	Magnified narrow band imaging
MIS	Minimally invasive surgery
ML	Machine learning
MRI	Magnetic resonance imaging
MSR	Multi-scale Retinex
MST	Medial superior temporalis
MVD	Multi-view video plus depth
NBI	Narrow band imaging
NCC	Normalized Cross-Correlation
NRCIR	Natural rendering of color image based on retinex
NTNU	Norwegian University of Science and Technology
OCT	optical-coherence tomography
OS	Open surgery
PDE	Partial differential equation
PUP	Predict-Update-Predict
RC	Region content
RGB	Red, Green, and blue (colorspace)
ROI	Region of interest
SAD	Sum of Absolute Differences
SDME	Second derivative like measure of enhancement
SE	Surface enhancement
SEI	Stereo endoscopic images
SIFT	Scale-invariant feature transform
SROCC	Spearman rank order correlation coefficient
SSD	Sum of square differences
SVD	Singular value decompositio
SVM	Support vector machines

TE	Tone enhancement
UM	Unsharp masking
VLS	Vector lifting scheme
WCE	Wireless capsule endoscopy
WLE	White light endoscopy

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> You can't connect the dots looking forward; you can only connect them looking backwards.
>
> STEVE JOBS

1.1 Medical context and motivation

Conventional open surgery (OS) has been used for hundreds of years as the main and more intuitive surgical solution for patients. It consists on making large incisions in the skin using a scalpel and separating the underlying tissues to get a direct access to the surgical target. This process implies remarkable large scars, a long recovery time and hospital stay to follow up the patient status, and considerable post-operative pain in some cases. The medical and technological progress achieved in the last decades has made a huge leap forward in the modern surgical practice by developing minimally invasive surgery (MIS) which became a popular diagnosis and treatment tool, especially with the recent innovations including stereo laparoscopes, wireless capsule endoscopy (WCE), etc.

In contrast with open surgery, MIS is performed through small incisions to attenuate the surgical trauma and morbidity. The abdomen is insufflated with a specific dose of gas (CO_2) in order to create a working volume through which surgical instruments can be inserted via ports. Since direct viewing of the surgical scene is not possible, a long thin tube equipped with miniature camera, called an endoscope, is inserted to the abdomen to assist the surgeon's navigation and tasks. The captured endoscopic scene is displayed on a monitor to provide visual feedback of the anatomical structures and the surgical instruments. As a comparison example taking the appendectomy case, according to the Society of American Gastrointestinal and Endoscopic Surgeons, a typical open appendectomy requires an incision of 4 inches long while in most laparoscopic appendectomies, surgeons operate through three small incisions between quarter to half inch each.

MIS offers many advantages starting by reducing the unwanted patient analgesia side effects since local treatments require less analgesia compared to OS. Furthermore, avoiding large wounds lead to decreasing blood loss, discomfort, and post-operative pain. In the same vein, not exposing the abdomen delicate tissues to external environment contaminants decreases the probability of having post-operative complications, especially wound related problems such as infection, cellulitis , dehiscence, and incisional hernia [96]. In addition to that, performing the operation within the body cavity can avoid drying, cooling and excessive retraction and handling of internal endo-organs, which are associated often to open surgery techniques. All these benefits lead to accelerating the recovery process and reducing the hospitalization time, which implies less risks of having complications related to long inactivity rest periods in bed such as urinary retention or muscle atrophy. Furthermore, the patients tends to prefer having less and smaller scars than large ones occurring with OS, and that seems to reduce the post-operative anxiety related to self-image degradation [2].

One of the main challenges facing the surgeons during the endoscopic surgical training and exercise is to adapt their visual system to operate with a limited two-dimensional (2D) field of view. Indeed, the images of the three-dimensional (3D) endoscopic scene including different organ structures and various surgical instruments are transmitted and projected onto a 2D screen. This dimensional shift reduces the perceptual cues for the surgeons, making it more complicated to determine the depth within the scene and distinguish the relative positions of the different anatomical structures and instruments. An additional difficulty is that the MIS field of vision is reduced compared to the open surgery case, which implies the necessity of special training, perceptual and mental skills [260] to work with limited spatial resolution screen images. This lack of depth perception in addition to the absence of tactile feedback can influence the surgeon's performance and, consequently, affect the accuracy of the surgical procedure.

To overcome this depth sensory loss, industry has demonstrated an understanding to this problem by developing 3D vision surgical systems. The first solution appeared with *Da Vinci Surgical System* [78, 107] (Intuitive Surgical, Sunnyvale, CA, UBC) which integrate a stereo laproscope and offer other features such as a virtual reality training simulator and more stabilization via electromechanical tools damping the vibrations due to machinery or shaky surgeon hands. Later, several independent stereo endoscopes for non-robotic MIS were proposed such as *Endoeye Flex 3D* [142] (Olympus Medical Systems, Shinjuku, Tokyo, Japan), the *Viking 3D-HD Vision System* [85, 297] and the *EndoSite 3Di Digital Vision System* [20] (Viking Systems, Westborough, MA, USA). The later device generates spectral depth improved images of the scene displayed using an ergonomic head-mounted display.

This technological evolution allowing the convergence to 3D vision surgical systems highlighted, however, various issues related to endoscopic images quality which is already a challenge for the 2D case. Indeed, the special characteristics of the endoscopic environment including dynamic illumination conditions, moist tissues with high reflectance, and motion due to the patient's breathing and the intra-operative surgeons tasks, can lead to several artifacts such as smoke, blur, and specular reflections which degrades the quality of endoscopic images and, consequently, can affect the surgeon's work accuracy and efficiency. Therefore, it is crucial to enhance endoscopic images in order to provide the surgeons with a good quality visual feedback of the endoscopic scene or to improve the output of some subsequent tasks such as features extraction for instruments tracking or 3D organ reconstruction. As will be detailed in the following section presenting the thesis practical context and workflow, our research work focuses on the enhancement issue for both 2D and 3D endoscopic images.

1.2 Research objectives and contributions

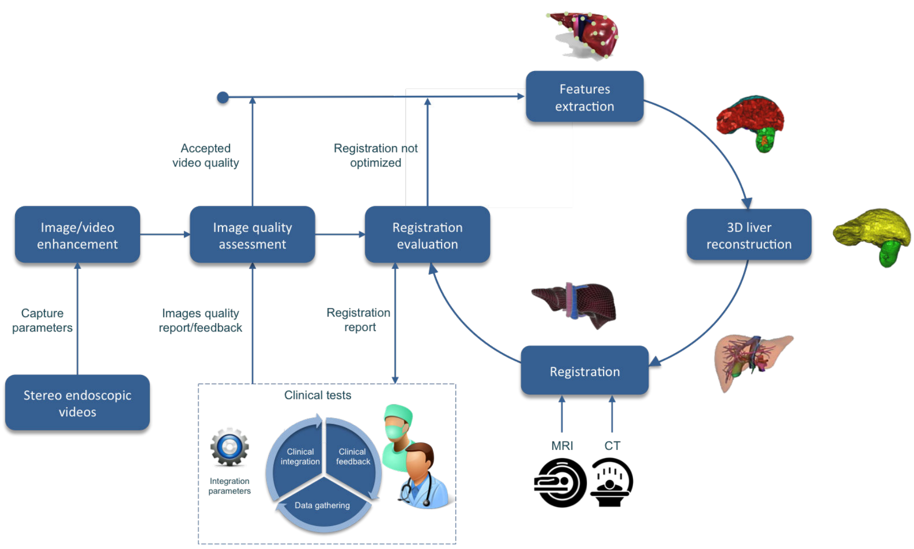


Figure 1.1: Research workflow context of the thesis

This thesis is a part of a Norwegian project entitled *HyPerCept*, gathering fourteen partners including France. This research has been funded by both the Research Council of Norway through project no. 221073 *HyPerCept Colour and quality in higher dimensions* and Université Paris 13. The project scope is composed of four different sub-programs:

1. Improved image quality for the visually impaired.
2. Image quality in higher temporal and spatial dimensions.

3. Imaging technology for health, security, and art: Functional image quality.

4. Color and quality in imaging devices – value creation in the imaging industry.

As a part of the second sub-program, our research work has been carried out between the Laboratory of Information Processing and Transmission[1] (Laboratoire de Traitement et Transport de l'Information-L2TI) at Université Paris 13[2] and the Norwegian Colour and Visual Computing Laboratory [3] at the Norwegian University of Science and Technology (NTNU) in Gjøvik[4]. A close collaboration is also established with The Intervention Centre[5] (IVS) at Oslo University Hospital for medical and clinical aspects of our research. More precisely, we work on the clinical case dealing with the treatment of hepatic neoplasms, which are considered as the second most common cause of cancer death worldwide [122].

Liver resection, referred to also as Hepatectomy, is the treatment of choice for patients diagnosed with localized Hepatocellular Carcinoma (HCC) or metastatic colorectal cancer [26]. The goal of such operation consists in removing the tumoral cells along with surrounding liver tissue within a safety margin to be optimized in order to preserve as much healthy liver tissues as possible [239]. A successful surgical liver resection can increase the patient's chances to live longer with a 5-year survival rates of about 10% to 58% depending on the tumor size [215].

During the past decade, computer-assisted surgical systems have been very helpful to surgeons and clinicians to make the right decisions related to the surgical interventions planning and navigation. More precisely in our liver resection case, such systems providing various patient specific models are more frequently incorporated in the clinical practice since they improve the resection plan accuracy/optimization and facilitate the navigation planning. These models are based mainly on pre-operative magnetic resonance imaging (MRI) and computed tomography (CT) and can be either geometric, functional, mechanical, or simulations. In particular, it has been demonstrated that using patient specific 3D geometric organ reconstructed models improves not only the localization of the tumor and the surgery planning precision [156, 161, 89], but also the confidence and orientation of surgeons during the operation [157].

While the liver resection procedure may be performed through a traditional OS or using MIS, the surgeons at IVS prefer the latter intervention due

[1]http://www-l2ti.univ-paris13.fr
[2]https://www.univ-paris13.fr
[3]https://www.ntnu.edu/colourlab
[4]https://www.ntnu.no/gjovik
[5]http://www.ivs.no

its advantages previously presented. In order to increase the depth feeling in the perceived endoscopic scene and facilitate the navigation and surgeon's tasks, the traditional 2D laparoscope has been substituted with a stereo endoscope composed of two cameras with slightly different perspectives. Furthermore, in order to optimize the resection plan and execution, a 3D liver model is reconstructed and registered with the patient specific data in order to guide the surgeons in the pre-operative planning stage and during the surgical procedure.

Figure 1.1 summarizes the research workflow related to this thesis and its clinical context. The main objectives of this study are twofold:

- Enhance the perceived quality of both monoview and stereo endoscopic images by exhibiting tissue details and abnormalities in order to improve the diagnosis accuracy and reliability. The intra-operative visual feedback quality is also of great importance for the surgeons since it can affect the efficiency of their tasks. Indeed, endoscopic videos should highlight significant details such as tissues texture, organs boundaries and blood vessels. The latter, for instance, should be blocked before performing a resection in order to avoid launching a bleeding, which can compromise the patient safety. If not properly controlled, the bleeding may oblige the doctors to convert to an OS, in which case the patient gets the worst of both surgical exercises (i.e. increasing the operating time without any post-operative advantage).

- Enhance the endoscopic images to improve the outcomes of subsequent post-processing tasks such as features extraction for WCE abnormalities classification and 3D liver reconstruction from the stereo images provided by the endoscope. For our liver resection surgery context, the latter task is prerequisite to register the patient-specific data, establish the resection and navigation plans, and provide the surgeon with an efficient control of robotic-assisted surgical systems.

1.3 Publications

Throughout the span of this study, six publications in international conferences and two journal papers have been developed, see the list below.

International journal paper:

- Bilel Sdiri, Mounir Kaanich, Faouzi Alaya Cheikh, Azeddine Beghdadi, and Ole Jakob Elle, "Joint enhancement methods for stereo endoscopic images based on 3D wavelet decomposition and binocular combination", submitted to IEEE Transactions on Medical Imaging.

- Bilel Sdiri, Azeddine Beghdadi, Faouzi Alaya Cheikh and Ole Jakob Elle, "Adaptive contrast enhancement method for stereo endoscopic images combining depth information and binocular visibility map", To be submitted to EURASIP Journal on Image and Video Processing.

International conference papers:

- Azeddine Beghdadi, Muhammad Ali Qureshi, Bilel Sdiri, Faouzi Alaya Cheikh, Mohamed Deriche, "A database for image contrast enhancement evaluation, Submitted to the International Conference on Quality of Multimedia Experience (QoMEX 2018), Sardinia, Italy.
- Bilel Sdiri, Mounir Kaaniche, Faouzi Alaya Cheikh, Azeddine Beghdadi, and Ole Jakob Elle, "Joint enhancement of stereo endoscopic images based on vector lifting scheme",In conference of the international society for medical innovation and technology (SMITT), November 2017, Torino, Italy.
- Muhammad Ali Qureshi, Azeddine Beghdadi, Bilel Sdiri, Mohamed Deriche, Faouzi Alaya Cheikh, "A comprehensive performance evaluation of objective quality metrics for contrast enhancement techniques", In European Workshop on Visual Information Processing (EUVIP), October 2016, pp.1-5, Marseille, France.
- Bilel Sdiri, Azeddine Beghdadi, Faouzi Alaya Cheikh, Marius Pedersen and Ole Jakob Elle, "An adaptive contrast enhancement method for stereo endoscopic images combining binocular just noticeable difference model and depth information.", In Electronic Imaging, February 2016, no. 13, pp.1-7, San Francisco, USA.
- Bilel Sdiri, Faouzi Alaya Cheikh, Kushtrim Dragusha and Azeddine Beghdadi, "Comparative study of endoscopic image enhancement techniques for endoscopic stereo matching and wireless capsule endoscopy images classification.", In Colour and Visual Computing Symposium (CVCS), August 2015, pp. 1-5. IEEE, 2015
- Bilel Sdiri, Azeddine Beghdadi and Faouzi Alaya Cheikh. "Brief overview on specular reflections removal techniques for endoscopic images." In European Workshop on Visual Information Processing (EUVIP), December 2014, Paris, France.

Other

- Online publicly available ***Contrast Enhancement Evaluation Database (CEED2016)***[6] on Mendeley web platform. The database includes 30 images with various color distributions, textures, contrast variations, and different realistic contrast enhancement artifacts; the associated contrast enhanced versions using representatives of the most common CE techniques in the literature; the corresponding subjective evaluation scores; objective assessment measures; and statistical correlation analysis data.

1.4 Thesis outline

This thesis is organized as follows.

In Chapter 2, we review the idea of depth perception with the fundamentals of human binocular vision. Second, we emphasize the relationship between the two views of the stereo image from a geometrical point of view. This chapter is motivated by the fact that the knowledge of stereo vision is prominent to understand the main issue of this thesis. Moreover, we review the most important methods that aim to estimate disparity information which is crucial to exploit the cross-view redundancies that can benefit the developed algorithms. Finally, we outline the relationship between depth and disparity maps.

In Chapter 3, we outline the main challenges and problems related to endoscopic images. After presenting the main artifacts and noise that can occur for endoscopic images, we present and categorize the different existent technologies and methods for endoscopic image enhancement.

Chapter 4 deals with the contrast enhancement (CE) issue from two different perspectives. The first part of the chapter focuses on the evaluation of contrast enhancement by addressing the lack of a benchmark allowing the assessment of different CE evaluation metrics. In the second part of the chapter, we propose a 2D endoscopic image enhancement technique improving both the global and local contrast together with a chromatic processing, for stereo-matching and WCE classification.

In Chapter 5, we present an adaptive enhancement technique for stereo endoscopic images combining depth data and the binocular just noticeable difference model. Chapter 6 describes proposed joint enhancement methods for stereo endoscopic images based on 3D wavelet decomposition and binocular combination.

Finally, some relevant conclusions and possible extensions of this research are addressed in Chapter 7.

[6] MENDELEY data plateform: Contrast Enhancement Evaluation Database (CEED2016)

Chapter 2

Basics of stereo vision

Our life is endless in the way that
our visual field is without limit.

LUDWIG WITTGENSTEIN

2.1 Introduction

Historically, stereo-vision has played an important forefront role in the evolution of man. Starting with the prehistoric man, whose survival depended mainly on his capability to perceive the distance between him and a life-threatening factor such as a predator, until the modern man who exploits this capacity to develop tools allowing him to have an increasing control of his environment. Binocular vision is strongly linked to various perceptual mechanisms, considerable amounts of information, as well as physiological and psychological aspects [262]. The visual stimuli/information can be related either to the surrounding environment such as, the relative position of objects, the respective sizes, the lighting condition, or to some human visual system (HVS) features such as eyes convergence and retinal disparity. Once the retinal left and right images are captured with slightly shifted perspective, the HVS starts performing complex analysis of these monocular two-dimensional (2D) views. This analysis followed by cortex processing leads to the *binocular fusion*, which allows us to perceive depth and relative distance to objects/obstacles within the environment.

The increasing understanding of the HVS and its perception features, over the past centuries, has paved the way to the development of the first image acquisition device and the invention of stereoscopes, followed later by a great technological evolution leading to the various contemporary 2D and 3D capture and display systems. The wide success of these technologies has prompted very active research topics such as computer vision, image compression, image enhancement, and quality assessment for both two-dimensional and stereoscopic images.

In order to understand the research work and the different technical concepts introduced in this thesis, it is first necessary to grasp, or at least outline, the basic principles involved in the capture of stereo retinal images and the subsequent processing steps leading to depth perception. Therefore, this chapter is devoted to presenting the fundamentals of the stereo HVS perception and some basic aspects related to the stereo image acquisition process.

The reminder of this chapter is organized as follows. Section 2.2 presents the principle of stereo-vision as well as the different HVS characteristics that affect the perceived image quality, namely the binocular combination and rivalry. Section 2.3 overviews the depth cues related to the perceived environment allowing interpretation of the third dimension. In Section 2.4, we present some basic concepts related to stereo acquisition systems including the geometrical relation between camera views and stereo matching. Finally, Section 2.5 concludes the chapter.

2.2 Stereo vision aspects related to the HVS

2.2.1 A journey though the HVS: from a photon to depth perception

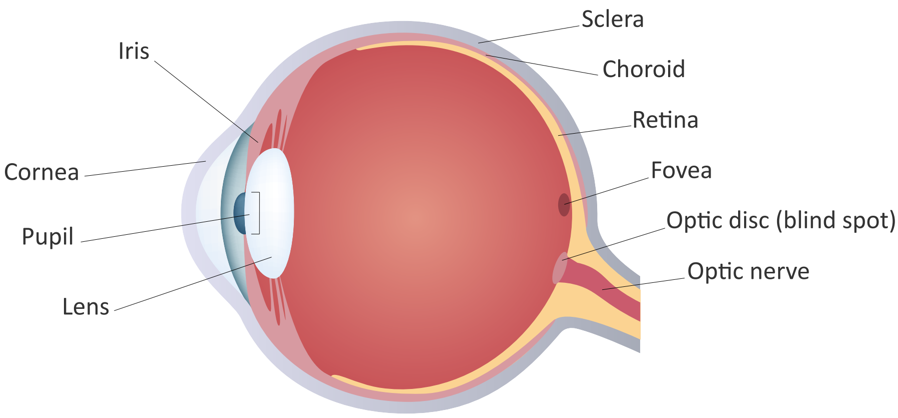


Figure 2.1: Schematic of human eye anatomy.

The fact that we are born with two eyes is the first clue showing that, since its creation, the sensory human visual apparatus is supposed to ensure a three-dimensional perception. This is the reason why people having one-eyed vision, due to either a congenital birth defect, illness, or accident, have poor relief perception and appreciation of distance and relative positions. Indeed, the frontal horizontal disposition of the eyes symmetrically located with respect to the sagittal plane, with a 50-70 mm disparity for humans, produces two slightly different images of the world. Without this

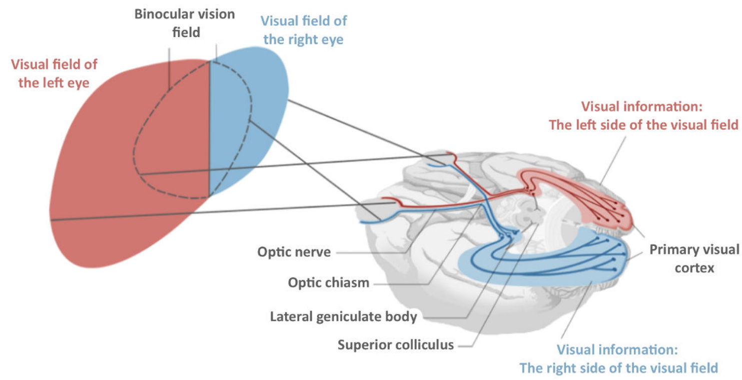


Figure 2.2: Binocular vision of the human visual system

slight perspective difference of retinal views, the brain cannot interpret the 3D information. Furthermore, each eye exposes a monocular vision field approximated to 150 degrees with a wider temporal angle than the nasal side (due to the existence of the nasal bone). For humans and predators, the right and left visual fields overlap in an important area surrounding the fixation point on which the eyes converge. This zone of about 120 degrees corresponds to the binocular vision field, which is prerequisite to the binocular vision and depth perception.

The perception process is launched in the eye, which anatomy is depicted in Figure 2.1. Once the visual information is captured by both eyes, it is then transmitted to the brain where a complex process is performed to allow the interpretation of the three-dimensional environment. The light rays penetrating through the pupil undergo a refraction by the cornea and the lens which focuses them on the retina creating a reversed image of the perceived world. If we compare the eye to a camera, the retina can be considered as the film. This extremely fine neuronal tissue is the starting point of the visual information journey to the cortex, as illustrated in figure 2.2.

The retinal photoreceptor cells convert the light energy of the optically projected images into electric signals. This process, called also *visual phototransduction*, is performed by three types of photoreceptor cells: rods, cones and photosensitive ganglion cells. Being triggered by very few or single photon [95, 14], the highly sensitive rod cells function mainly at night and dim lightening conditions providing a gray-scale vision. On the other hand, cones require significantly brighter light levels to allow colour perception thanks to its sensitivity to specific wavelengths of light (i.e. colours). The photosensitive ganglion cells are considered as a third class photoreceptor cells since they don't contribute directly to vision, but rather in hormonal regulation and circadian rhythm maintenance. Their existence has been noticed for the first time in 1923 during an experiment in which coneless and rodless mice responded to light stimulus. Advanced research on these cells, however, didn't begin until the 1980s and reported that, beside rods and cones, a small group of cells demonstrated a sluggish response to long term exposition to light [322], and they were called photosensitive ganglion cells.

The ganglion cells are the last neurons that process the retinal signal before leaving the retina. Thereafter, the visual information is carried by the optic nerve, through the optic chiasm along the optic tract, via the lateral geniculate body (LGB) to reach finally the visual cortex. Although the LGB in both left and right hemispheres receive visual input from both eyes, each LGB receives retinal information of only one half of the visual field. This can be explained by the decussation of the nasal side optic fibers to the opposite side of the brain in optic chiasma, as shown by Figure 2.2. Consequently, the right hemisphere receives visual information of the left side of the visual field, and vice versa.

Note that up to this stage, the visual information has been undergoing chromatic and contrast pre-prosessing in both hemispheres but it is still monocular. Once arrived to the primary visual cortex V1 at the back of the brain, the visual information is processed by the the simple and complex cells, which are sensitive to several attributes such as the orientation, spatial and temporal frequency, and the direction of motion. The main aim of this process is to filter the visual input in order to enhance the edges and contours [222, 223].

From V1, two possible visual pathways can be distinguished: a ventral stream and a dorsal stream [70, 229, 93]. The ventral stream, referred to also as the "what" pathway, is principally involved in the discrimination of colors and shapes and contains areas such as V2, V4, and the inferior temporal cortex [186, 45]. The dorsal stream, called also the "where/how" pathway, including the medial temporal (MT: V5) and medial superior temporalis (MST) areas, is predominantly responsible of spatial information processing such as object motion detection, relative position, and depth perception (i.e. stereopsis) [205]. These two pathways share the progressive processing of the visual inputs via forward and backward connections leading to the combination and the fusion of the two retinal images into one binocular signal. Several complex properties of the HVS are involved in the production of the binocular 3D image, namely binocular combination, suppression and rivalry. Because of their extremely important impact on both the processing of the visual information and the finally perceived output, these binocular HVS features are detailed in the following sections.

2.2.1.1 Binocular combination

Binocular combination refers to the fundamental feature of the HVS by which two different and completely separate neural signals are seamlessly combined to produce a single cyclopean perception. The latter term, imaginatively invented by *Bela Julesz* [126], is inspired from the Greek/Roman mythology of Cyclopes who are acquired with a single centered eye in the forehead. Literally, the idea behind it is the way stereo sighted humans and animals perceive the center area of their fused left and right visual fields, as if perceived by another imaginary cyclopean eye centered between the real eyes.

The research regarding binocular combination can be divided into two main branches. The first category of studies focuses on localizing the exact section of the brain performing this operation [12, 106, 287, 330], which has been considered for a long time as "somewhere in the visual cortex". The second branch of research has been devoted to model this phenomenon. Many physiological and psychophysical experiments have investigated how i) two identical monocular patterns merge to produce a single cyclopean view [169, 194, 243], ii) two slightly different monocular patterns engender the perception of depth [201], iii) two different enough monocular patterns result in binocular rivalry [10].

Ding and Sperling [62] are the first to propose a measure of appearance for cyclopean images resulting from binocular combination of sine-wave gratings with different contrasts and phases but identical frequency for both eyes. Their contrast-gain control model [62] has demonstrated a successful performance in modeling phase perception in binocular normal vision, which inspired further studies [102].

Among the popular developed models, we can find the probability summation [243], the quadratic summation [28], power summation [169], twin summation [194], two-stage gain control [207], and binocular normalization [219] for both close and supra-threshold binocular contrast combination.

2.2.1.2 Binocular rivalry and the suppression theory

Binocular rivalry is a remarkable visual phenomena in which the perception alternates between dissimilar signals. When each eye is presented with different images/scenes (referred to also as dichoptic presentation) or facing an ambiguous visual information, the perception tends to frequently alternate between left and right "rivalling" views. For instance, if one eye views horizontal lines and vertical lines are presented to the other eye simultaneously, the HVS will perceive only one of the images for few seconds, then the other, and so on as long as the observer cares to see [318]. Intuitively, this process may imply that the visual cortex suppress alternately one monocular signal and allows only one to be perceived. This is the main reason why binocular rivalry has been associated for many years with the suppression theory (*Porta*-1593, *Gassendi*-1658, *Dutour*-1760), which prompted intensive discussions and split of the scientific perceptual research community between supporting and disapproving studies [100, 101]. According the suppression theory, whether left and right views are similar or dissimilar, rivalry is the only binocular process and interaction being performed by the HVS. The reported explanation is that when the monocular signals are very similar, we can not notice the alternations. It is only when perceiving two different images that this natural alternation is revealed. This idea set the fundamentals of the singleness of vision theory which stipulates that we always perceive "only one thing from everything" (i.e. one signal among two visual signals).

The first clear definition of rivalry was proposed later by *Charles Wheatstone*. He supported the fusion theory, which is the alternative of the vision singleness theory, and demonstrated the relationship between depth perception and binocular disparity (i.e. stereopsis). According to *Wheatstone*, rivalry is a special case occurring when the binocular fusion and combination is impossible [319]. Indeed, visual cortex processing is always trying to perform a binocular combination/fusion, however, when the inter-view difference exceeds a certain level, the process fails giving rise to binocular rivalry. This explains why we briefly perceive some unstable composite of the two retinal signals in the rivalry alternation transitions when facing two extremely different images.

2.2.2 Depth cues related to the HVS

2.2.2.1 Ocular information

The ocular information is related to the eye movement and the state of its components. Adjusting the focal length of the eye lens, called accommodation, enables our visual system to focus on objects situated in different spots. When the angle between the lines of sight of both eyes change, we talk about eyes vergence.

Accommodation In order to adjust the visual focus on the objects of the scene, situated in different distance, the ciliary muscles adjust the shape of the lens. When they are relaxed, the eye lens become thin, thus its focal length increases and distant objects are perceived. When the ciliary muscles contract, the curvature of the eye lens is increased and the eye lens becomes thick. Consequently, its focal length decreases and we see the nearby objects, as illustrated in Figure 2.3.

The fact that the HVS adapts dynamically the tension applied on the lens implies that it has primary absolute information regarding the distance of the object to be focused on. Beyond approximately two meters, the sensory depth information provided by the accommodation decreases as the ciliary muscles reach the maximum relaxation. In [313], the authors reported that in spite of its limitation and weak depth signal, the accommodation remains a useful source of depth information. This usefulness is not related to the distance judgment accuracy, but to the evaluation of objects size, which is an important depth cue presented in Section 2.3.

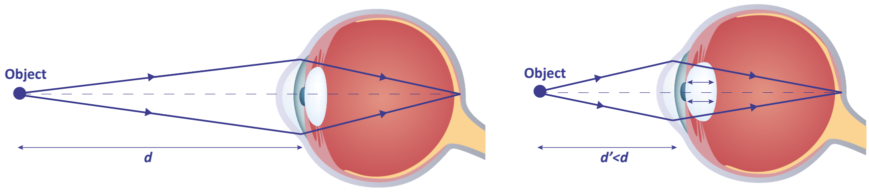


Figure 2.3: Accommodation of the eye's lenses with distance to the observer.

Vergence An ocular vergence movement is the simultaneous rotation inwards of both eyes so that the axes of sight converge at a particular point allowing the fixation of an object on this location. In [25], the authors define the vergence as the angle in degree between left and right sight axes. The aim of eye fixation is to project the perceived object in a central area of the retina characterized with the highest concentration of cone cells. This special retinal spot, called *fovea*, is responsible of sharp vision.

The eyes can either converge to fixate on a closer target $O1$, or diverge to gaze at a farther object $O2$, as illustrated in Figure 2.4. We can notice that the vergence angle varies according to the distance as α_1 is larger than α_2. The HVS exploits this information to interpret the depth of the objects $O1$ and $O2$. Indeed, the information is provided by the nerve impulses in the ocular muscular tissue that controls the eyes movement. Thus, the vergence angle is strongly correlated to distance and varies as a function of it. In addition to that, a change of the vergence angle is often followed by an accommodation adjustment, especially if the perceived object is at close distance. Thus, accommodation and vergence are two strongly interdependent depth cues providing absolute distance information relative to a fixation point, at close distance from the observer.

2.2.2.2 Binocular disparity

In the previous section, we have seen that the vergence and accommodation are two significant interconnected mechanisms allowing to fixate an object. Once the eyes rotate and the focal length is adjusted by the lens to focus on the target object, the latter is projected in the foveae of both eyes becoming the center of visual attention. Binocular disparity occurs when the same perceived object is projected to different spatial positions in left and right retinae. Considered as one of the primary depth cues, this shift in retinal coordinates is due to the interpupillary distance between left and right eyes,

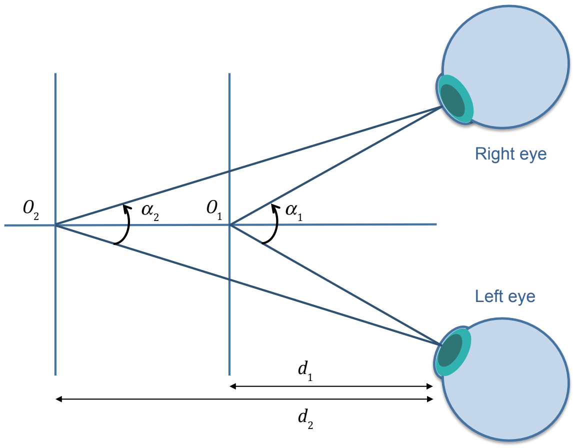


Figure 2.4: Eye's vergence.

resulting in a parallax of the perceived object. It is measured as the retinal angles difference of both projections of the same object in each eye. For the sake of example, let us consider Figure 2.5 in which the observer is fixating the object F, consequently its image is projected on the foveae of his left and right eyes. Two other points $(P1, P2)$ are also projected on both retinae yielding two angles (α, β) with left and right foveae respectively. The disparity created by the points $P1$ and $P2$ is given by the equations (2.1 , 2.2) respectively.

$$\eta_{P1} = -\alpha - (-\beta) = \beta - \alpha \tag{2.1}$$

$$\eta_{P2} = -\alpha - (+\beta) = -(\alpha + \beta) \tag{2.2}$$

We can notice that $\eta_{P1} > 0$ while $\eta_{P2} < 0$. The HVS interpret this information as the relative positions of the points $P1$ and $P2$ with respect to the fixation point F. If the disparity is negative, then the visual cortex interpret that the related object is behind the fixed point and vice versa. The *horopter*, invented by *Aguilonius* in 1613, is one of the main benchmarks of the retinal disparity [60]. In 1826, *Vieth-Mullerfor* introduced the circular horopter (see Figure 2.6), known also as *Vieth-Müller circle* which define the binocular disparity as a function of the eyes vergence. According to the position of a point inside the horopter, the disparity can be either *crossed* or *uncrossed*. Indeed, if the perceived point is between the observer and the fixed object, then the optic sight lines cross inside the horopter yielding a crossed disparity (i.e positive angle). At the opposite, the sight axes cross outside the horopter since the perceived point is behind the fixed object, which implies an uncrossed disparity in this case (negative angle).

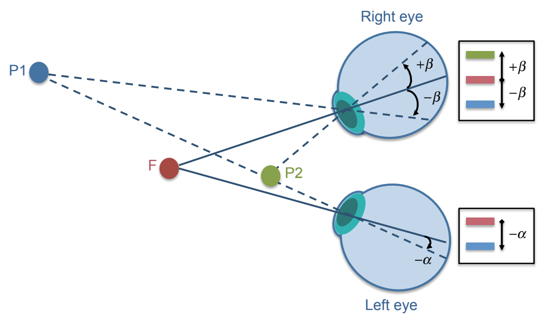


Figure 2.5: Horopter: Vieth-Müller circle.

The cortex gathers and combines these angluar disparity feedbacks to interpret the depth and generate a 3D perceived scene, as illustrated in Figure 2.7. The perception of distance to objects (depth) based on parallax implies intuitively that the HVS deals with the correspondence problem between the left and the right retinal views. This complex process raises a significant question: how can the visual cortex determine which feature in a retinal image matches with the same feature in the other view ? The first visual theories assumed that the correspondences information is obtained by a shape/edges analysis performed before the depth perception, which was later demonstrated to be the opposite by stereogram experimental tests [201].

2.3 Stereoscopic aspects related to the perceived environment

One of the fundamental tasks linked to stereovision is to provide an accurate interpretation of the depth and distance between the different objects of our visual world. This functionality is based, not only on the existence of two slightly different retinal images, but also on various depth cues related to the physiology of the HVS or incorporated in the perceived scene. The association of these different perceptual cues, on the basis of binocular parallax, allows an accurate reconstruction of the third dimension. Table 2.1 summarizes the different information sources [234, 48] allowing the depth perception, categorized according to five characteristics: monocular versus binocular, optical versus ocular, absolute versus relative, dynamic versus static and quantitative versus qualitative.

We can notice from Table 2.1 that one eyed vision can generate sometimes a depth perception. This can be explained by the importance of monoc-

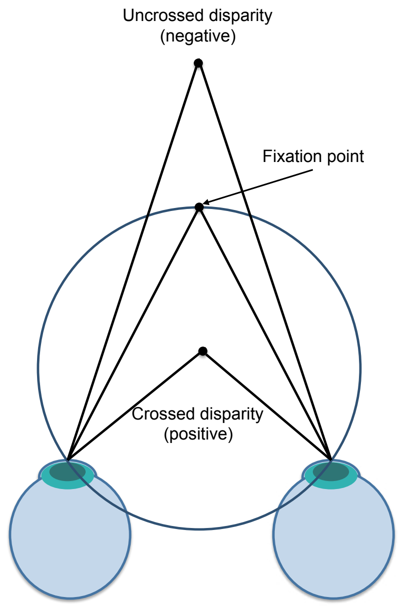


Figure 2.6: Binocular disparity.

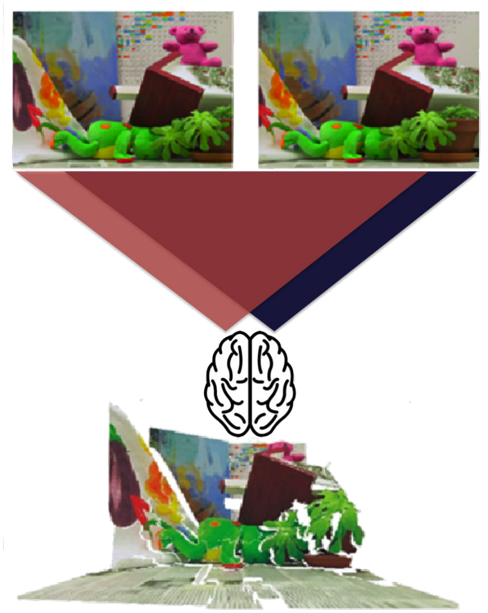


Figure 2.7: Stereopsis: depth perception using the binocular fusion of two retinal 2D images.

	Binocular / Monocular	Ocular / Optical	Relative / Absolute	Static / Dynamic	Qualitative / Quantitative
Binocular disparity	**binocular**	optical	relative	static	quantitative
Convergence	**binocular**	**ocular**	**absolute**	static	quantitative
Accommodation	monocular	**ocular**	**absolute**	static	quantitative
Texture accretion/del.	monocular	optical	relative	**dynamic**	quantitative
Motion Parallax	monocular	optical	relative	**dynamic**	quantitative
Position/ Horizon	monocular	optical	relative	static	quantitative
Relative size	monocular	optical	relative	static	quantitative
Familiar size	monocular	optical	**absolute**	static	quantitative
Texture gradient	monocular	optical	relative	static	quantitative
Convergence of parall.	monocular	optical	relative	static	**qualitative**
Edge interpretation	monocular	optical	relative	static	**qualitative**
Shading and shadows	monocular	optical	relative	static	**qualitative**
Aerial perspective	monocular	optical	relative	static	**qualitative**

Table 2.1: Information sources for depth perception [234, 48].

ular depth cues in assisting the powerful visual cortex processing to interpret the depth within the perceived scene. Indeed, taking a look on Figure 2.8, we can feel a depth sensation with both monocular and stereo vision even though the image is 2D. This is thanks to the various depth cues that the image includes such as relative size, texture gradient, and shadows. The following sections present an overview of the different extraneous cues allowing the depth perception.

2.3.1 Dynamic depth cues: Motion

We are rarely immobile while perceiving or exploring our environment and the different elements composing it. Motion is even considered as a foundation of human vision [164]. The displacement of the observer, his environment, or the fixed objects results in a movement of the retinal image of objects over time, providing consequently the HVS with dynamic perceptual depth information.

Motion parallax occurs when objects, located in different relative distances with respect to the observer, are perceived moving with different retinal velocities, directions and rates according to the HVS viewpoint displacement.

When the observer is moving, the fixated point is maintained in the same retinal position because of the eyes vergence and accommodation compensating for the displacement. The retinal images of objects situated in front of the fixated point are, however, moved to the opposite direction of the observer movement. On the other hand, the objects behind the fixation point

(a) "*Jour de pluie à Paris*", painting by Gustave Caillebotte, 1877

(b) "*Un dimanche après-midi à l'Île de la Grande Jatte*" painting, by Georges-Pierre Seurat, 1884

Figure 2.8: Illustration of monocular depth cues in 2D images: relative size, texture gradient, and shadows. (source: *Wikipedia*.)

are moved in the same direction of the displacement. Note that the closer an object to observer is, the faster it moves.

2.3.2 Pictorial depth cues

The perception of a 2D image can yield a compelling sense of depth even though the observer is motionless or perceiving through only one eye. This can be explained by the monocular static cues that the image contain, which are independent of any motion or stereopsis. Since hundreds of years, many artists have realized this fact and exploited such visual indicators to create some stunning and famous art works giving a depth feeling/illusion, such as shown in the paintings of Figure 2.8. Before the Italian Renaissance, no artwork including pictorial depth cues were found. The first principle and laws of perspective were defined by *Leonardo da Vinci (1452–1519)*. Pictorial depth cues are related mainly to the way and the shape with which objects of the scene appear giving visual information about their relative and absolute distance. The 3D pictorial information can be either geometric (the perspective, blur, relative size, texture, occlusion), photometric (aerial perspective, shadow, transparency), or cognitive (familiar size). In the following, some main pictorial depth cues will be presented.

2.3.2.1 Occlusion

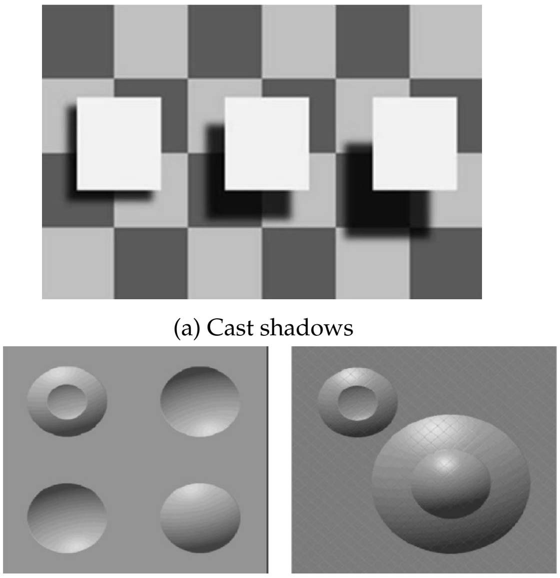
Occlusion stipulates simply that if an object A is covering or obscuring another object B, then A is necessarily closer to the observer than B. Although it only allows to create an order of the perceived objects according to their relative positions (i.e. no indication about the relative or absolute distance), some researchers consider it as one of the most reliable and pervasive depth cues [278].

Several studies reported that it is one of the primary visual depth cues comprehended by infants shortly after their birth [285, 137]. For this reason, compared to other depth information sources, occlusion is considered as a simple visual information, not requiring complex processing by the HVS to be interpreted.

2.3.2.2 Shades and Shadows

This cue is related to the way by which light falls on an object, or get reflected from its surface to another object or to the space. The shape, the location and the direction the shadow can provide us with important information about the object's position with respect to the light source, other objects of the scene, and to the observer.

Perception studies [197, 97] has divided shadows into two types: cast shadows and attached shades, as illustrated in figure 2.9. The first type of shadows allow the observer to perceive the relative position of an object in depth (Figure 2.9(a)), while attached shadows (Figure 2.9(b)) gives visual feedback about whether the object has inwards relief, outward relief, or flat (i.e. no relief).



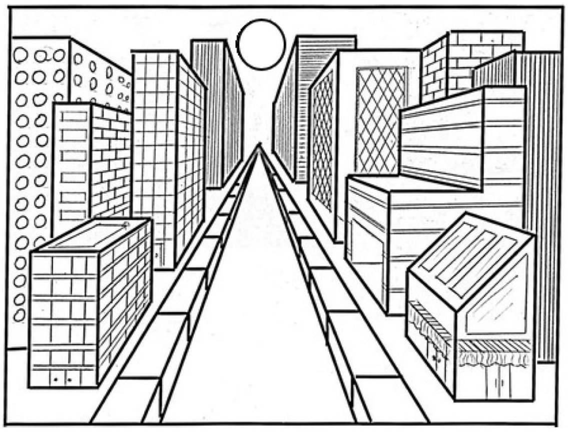
(a) Cast shadows

(b) Attached shadows

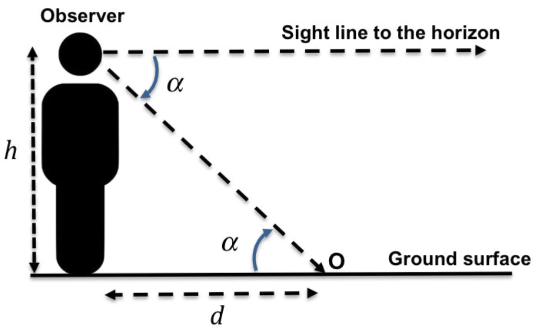
Figure 2.9: Shadow and shades in depth perception.

2.3.2.3 Perspective

This depth cue is linked to two different aspects: visual and geometric. The visual aspect stipulates that two parallel lines converge in the horizon to a vanishing point, as illustrated in figure 2.10(a). In other words, as the distance to the observer increases, the distance between the parallel lines decreases. A good example to illustrate the convergence of parallel lines is when observing the rails on railroad tracks. The geometric aspect is the relative position to the horizon of a surface. The distance d of any object O on the surface to the observer can be determined based on the horizon angle α (the angle between sight axis to the horizon and the sight axis when fixating the object) and the perpendicular high h of the observer's eyes with respect to the surface. Since the latter information is known by the HVS, the angular distance can be computed intuitively by the visual cortex ($d = h \times \arctan(\alpha)$)(see Figure 2.10(b)).



(a) Linear Perspective: two parallel lines converge in the horizon to a vanishing point. (source: *Wikipedia*)



(b) Relative position to the horizon of a surface.

Figure 2.10: Perspective depth cue.

2.3.2.4 Relative size

For identical objects, the HVS interprets the smaller sized one as the farther away from the observer. This can be explained by the fact that a distant object is projected into the retina with smaller size image than the closer similar object, as illustrated in Figure 2.11. The distance d to the observer can be expressed in this case as follow:

$$d = \frac{s}{r} \times f \approx \frac{s}{r}, \tag{2.3}$$

where f is the focal length of the eye, r is the retinal image size, and s is the size of the object. The latter information (i.e. the size s) can be confusing and ambiguous because the HVS cannot determine clearly whether a retinal image size corresponds to a smaller object that is closer to the observer, or a larger farther object. To overcome this uncertainty, the visual cortex performs heuristic processing assuming that both retinal object images have the same absolute size, which implies that the relative distance is determined based on the relative retinal images sizes.

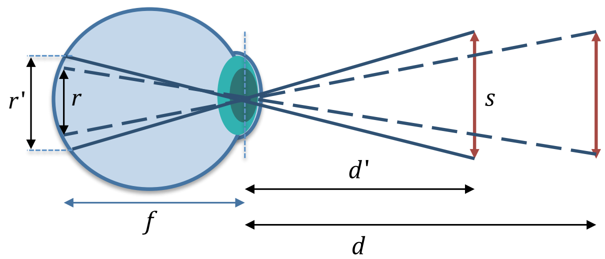


Figure 2.11: Relative size: the size of the retinal images of the same object change according to the distance from the observer.

2.3.2.5 Familiar size

Although relative size can give an visual information about how objects are relatively far/close to each others, it cannot provide, on its own, an accurate information about how distant any of the objects is from the observer. However, if we know "*a priori*" the real size of the object, our brain can resolve the equation 2.3 and determine the absolute distance. The term familiar implies simply that most of objects have a range of size or characteristic size with which people are familiar and used to. For instance if the Eiffel tower and a person appear having the same size, then our visual cortex will interpret that the person is much closer than the Eiffel tower.

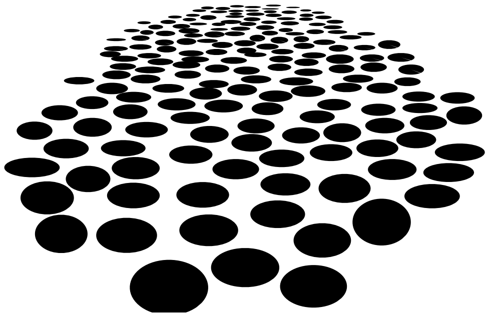


Figure 2.12: Texture gradient illustrated in white/black circles

	0 - 2	2 - 20	Above 30
Convergence	✓		
Accommodation	✓		
Motion	✓	✓	
Occlusion	✓	✓	✓
Relative size	✓	✓	✓
Relative height	✓		
Atmospheric perspectif			✓

Table 2.2: Range of effectiveness of different depth cues (in Meters) [48].

2.3.2.6 Texture gradients

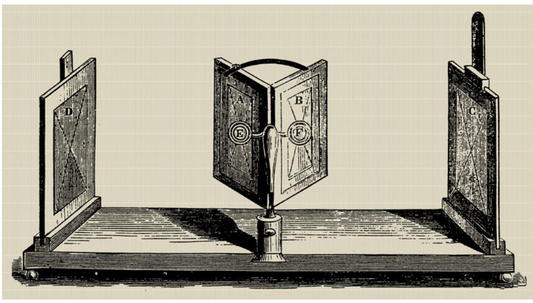
Texture gradient is considered as an important depth cue by many researchers and artists, and has been the focus of several perceptual studies [44, 79, 317]. As distance from the observer increases, the shape changes, the size decreases and objects appear denser as illustrated in Figures 2.12, 2.8(a). Linear perspective and texture gradients are often associated and very correlated with the relative size depth cue (i.e. no linear perspective nor texture gradients depth cue without relative size change).

2.3.2.7 Atmospheric Perspective

The atmospheric perspective, called also aerial perspective, provide additional depth cues in some particular conditions. Despite the empty appearance of the atmosphere, there are plenty of unnoticeable molecules scattering a bit the sunlight. The farther away the perceived objects are, the more the particles and pollutants scatter light. Consequently, far away objects appear "washed out", i.e with less sharpness and lower contrast. Table 2.2, present the effectiveness range of distance-sensitive depth cues detailed in this section.

2.4 Stereoscopy: stereo acquisition systems

The increasing understanding of the different basics and mechanisms related to the HVS has paved the way for the development of visual tools and devices attempting to simulate the binocular stereo vision. In 1838, while studying the binocular fusion theory, *Wheatstone* has created an optical tool made of two mirrors at 45 degree angles with respect to the observer's eyes as illustrated in Figure 2.13(a). He demonstrated using drawings and this device that presenting two slightly different images to the right and left eyes stimulate the HVS to combine them, yielding a depth perception. This key tool used to prove the fusion theory is the first invented and patented stereoscope. With his invention, *Wheatstone* has laid the cornerstone for the development of stereo systems allowing a depth perception from two slightly different 2D images, known also as *stereoscopy*.



(a)

(b) (c)

Figure 2.13: The evolution of stereoscopes: Wheatstone stereoscope (a) and lenticular stereoscopes of (b) *Holmes* - 1861, and (c) *Brewster* - 1870 (source: *Wikipedia*)

The invention of photography one year later, as well as the evolution of optics made stereoscopy a very active research field. This led to new improved versions of the stereoscope, namely the portable lens-based stereoscopes of *Holmes* (Figure 2.13(b)) and *Brewstern* (Figure 2.13(c)), and the creation of the stereogram which is a photography-based stereoscope using a pair of 2D photographs.

Nowadays, with the evolution of electronic systems, as well as the rapid computer science progress, stereoscopic acquisition systems and displays are widely spread in different application domains such as education, entertainment, defense, television and cinema, surgical and medical tools... In other words, almost all the applications available in 2D are tending to converge to the third dimension.

It is worth mentioning that the term "3D" is sometimes confused between two meanings. While in many contexts it refers mainly to the depth perception, the Computer Graphics and Vision communities use it as well to refer to the reconstruction and rendering of 3D geometric shapes and meshes. For the sake of clarity, the term 3D in this dissertation will designate the perception of relief. The 3D rendering of meshes and volumes will be referred to as "3D reconstruction".

To produce a depth perception, a stereoscopic system should present slightly different images to each eye in order to simulate the binocular disparity and trick the HVS. Therefore the system should include two cameras rotated to each others to create a binocular field and capture two slightly different view angles of the scene. The following paragraphs present some main properties and aspects of stereoscopic acquisition systems.

2.4.1 Epipolar geometry

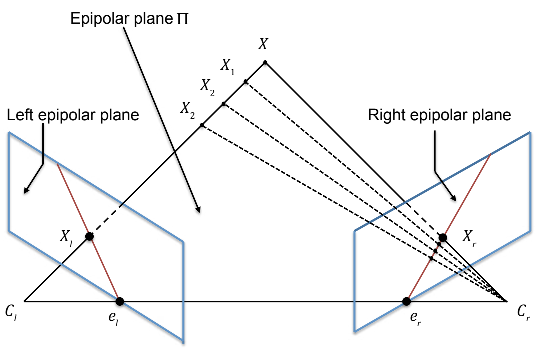


Figure 2.14: Epipolar geometry between a pair of images.

Stereoscopic systems consists in capturing the same three dimensional scene using two cameras with different view angles. This conversion of the 3D scene to 2D images, referred to also as *perspective projection*, yields geometric projections that are very correlated mathematically. The epipolar geometry describes the relation between 3D points of the real scene and their correspondent projections in left and right camera views. In Figure 2.14, left and right cameras are respectively illustrated by their focal centers C_l and C_r. P_l and P_r represent their correspondent perspective projection planes. Technically, the image planes in real cameras are behind the focal center. However, for better illustration, we place these virtual image planes in front of the optical/focal center of each camera.

The points X_l and X_r are the projections of the 3D point X in the left and right captured camera views respectively. The segment $[C_l, C_r]$ joining the focal centers represents the *baseline*, which intersects respectively the left and right image planes in two points e_l and e_r, called *epipolar points* or *epipoles*. The points X, C_l and C_r form the so-called *epipolar plane* π which intersect respectively left and right image planes in lines (e_l, X_l) and (e_r, X_r), referred to as *epipolar lines*. Notice that (C_l, X) is perceived by the right camera capture as a line while the left camera sees it as a point since it is aligned with its lens focal center, and vice versa for (C_r, X). Lets suppose now that given the point X_l, we want to search for its corresponding point in the other image plane. The epiploar geometry provides *epipolar constraints* that allow to resolve this problem. Indeed, we can notice that the corresponding point X_r to X_l must lie on the epipolar line of the right plane image. Similarly, the corresponding point to X_r is aligned with the epipolar line of the left view. As X varies, all its projections are situated on the epipolar line. These epipolar constraints can be also formulated using the following equation:

$$X^T F X' = 0, \tag{2.4}$$

where F is the fundamental matrix [190], X and X' are the corresponding points in left and right images. The matrix F is sized 3×3 and depends on both extrinsic and intrinsic parameters of the two cameras. In [189, 305], the authors present and compare the different techniques allowing to estimate the fundamental matrix F. This geometry implies as well that it is possible to know whether two planar projections correspond to the same 3D point. Using the epipolar constraints, we can determine the 3D point X if we know the two camera planar projections X_r and X_l, a process known also as *triangulation*. Indeed, knowing X_r and X_l implies that we can determine their optical axes, which should intersect in the 3D point X if both projections correspond to the same 3D point. This will be further discussed when detailing the relationship between the disparity and depth in section 2.4.3.3.

To summarize, the epipolar constraints reduce the complexity and the research field of a corresponding pair of points from the whole 2D image planes to 1D lines: the epipolar lines. In practice, this problem can be simplified further if both cameras are aligned in the same plane. The process in this case is called *epipolar rectification*.

2.4.2 Epipolar rectification

In the previous section, we studied the general *converging camera configuration* case in which both left and right cameras are rotated towards each others to have a common field of view. This angle variation yields an inclination of both epipolar lines. The other possible camera disposition in computer vision is the *coplanar camera configuration* in which the cameras are placed according to the same plane, thus, having two parallel focal axes not intersecting with the image planes. In this case the epipolar points lie in the infinity.

The main advantage of this configuration is that all corresponding epipolar lines become collinear to the scanning lines and with each other, which facilitates the research of homologous points between the planar images. Indeed, the corresponding points are located in this case on the same horizontal line, parallel to the baseline. In other worlds, both points have the same vertical coordinates, as illustrated in Figure 2.15. Many rectifications methods have been proposed in the literature such as the planar rectification [76] and the cylindrical rectification [176]. A survey of the different existing rectification methods can be found in [92, 76, 237].

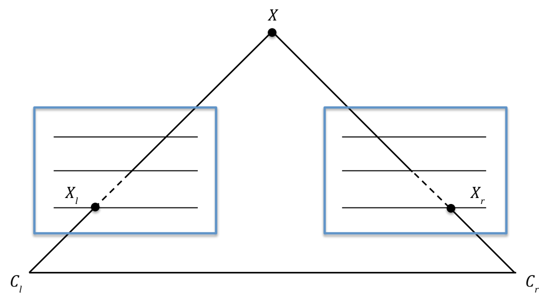


Figure 2.15: Epipolar rectification: the epipolar lines become collinear and parallel to the image scanlines.

2.4.3 Stereo matching

Although that epipolar geometry reduce the research field for homologous points, it does not give any information about the exact location of both corresponding points. Indeed, once the epipolar geometry is computed, the research of two points (P_l, P_r) projecting the same 3D point P can be constrained to a 1D research area: the epipolar line. Now, given a projection pixel $P_l(x_l, y_l)$ in the left view, the process allowing to find its exact corresponding pixel $P_r(x_r, y_r)$ in the right image is called stereo matching. In practice, the output of this step is the spatial displacement between the coordinates of both left and right homologous projection pixels. Stereo matching is one of the most challenging tasks in computer vision that has been extensively studied in the last decades. Besides the epiloar constraint, several other common assumptions and simplifications allow to greatly facilitate the matching process [49, 147, 294]. The first category of assumptions are the ordering and uniqueness constraints which can be applied only if specific geometric and photometric properties of the captured objects are fulfilled. The other set of constraints is based on disparity values assumptions and can be related to the disparity continuity, the disparity gradient limits and the absolute disparity. The stereo matching problem can be formulated mathematically as follows:

$$\vartheta : \mathbb{R}^2 \rightarrow \mathbb{R}^2,$$

$$(x_P^r, y_P^r) \rightarrow (x_P^r, y_P^r) = (x_P^l - x_P^r, y_P^l - y_P^r), \tag{2.5}$$

where ϑ is a function associating a disparity vector to each pixel of the right (index r) view allowing to find its correspondent in the left (index l) image. After rectification, the research is further simplified since conjugate points are on the same horizontal scanline and, consequently, have the same vertical coordinates. Equation 2.5 becomes in this case:

$$\vartheta_x : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$(x_P^r, y_P^r) \rightarrow \vartheta_x(x_P^r, y_P^r) = (x_P^l - x_P^r) \tag{2.6}$$

Stereo-matching algorithms are composed generally of four main steps: 1-matching cost computation, 2-cost aggregation, 3- disparity computation and/or optimization, and 4- disparity refinement [267]. Based on the actual sequence of the processed steps, the scan area (support), and the research features, the literature distinguishes two main classes of stereo matching algorithms: global and local.

2.4.3.1 Local stereo matching methods

Given a reference pixel in one view (i.e. either left or right view), the local methods consist in constraining the search of its corresponding pixel on a small window/block in the other image, as illustrated in Figure 2.16. For the sake of illustration, let us take for instance the left image as the reference of matching. This means that for each pixel $P_l(x_l, y_l)$ in the left image, the algorithm will search its corresponding pixel $P_r(x_r, y_r)$ in the right view.

The first step consists in initializing a window surrounding the pixel of interest in both views and defining a search area S in the right image. In

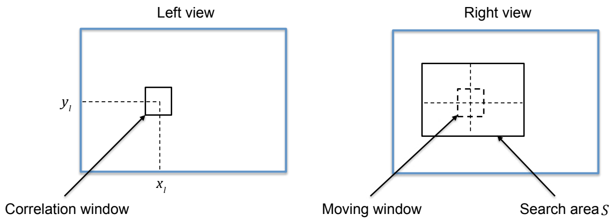


Figure 2.16: Correlation-based stereo matching methods.

the following step, the window in the right view is moved searching for a match with the left static window. As the dynamic window shifts, a cost function between the left static and the right current block is computed. The cost computation is based on similarity measures such as the Sum of Square Differences (SSD), the Sum of Absolute Differences (SAD) and the Normalized Cross-Correlation (NCC), which are the most popular. The dynamic window yielding the maximum NCC or minimum (SSD or SAD) is the considered as the best match.

These correlation-based methods present, however, some limitations related mainly to the choice of the window shape and size. Indeed the basic rectangular shape with a fixed size fails detecting a match in brisk image contrast changes such as in object edges and boundaries. Many approaches addressed this limitation by proposing dynamic shape and size windows which adapt in real-time to the image pixels intensity variation [192, 310, 323].

2.4.3.2 Global stereo matching methods

Global methods consists in making explicit smoothness assumptions incorporated into an optimization problem to be resolved. Such methods typically skip the first stereo-matching aggregation step and search for the optimal disparity vectors that minimize a global cost function $E(\vartheta)$. The latter is made as a combination of a data term $E_{data}(\vartheta)$ measuring the displacement between homologous pixels (step 1) and a smoothness-constrained regularization term $E_{smooth}(\vartheta)$ enforcing the disparity to be smooth. Thus, the global cost function is defined as follows:

$$E(\vartheta) = E_{data}(\vartheta) + \alpha E_{smooth}(\vartheta), \tag{2.7}$$

where α is a positive weight constant controlling the impact of both terms. The final disparity map is the optimal solution minimizing the cost function of the Equation 2.7. The used minimization algorithm represents the main difference between the various global stereo matching methods.

The most common minimization techniques proposed in the literature are dynamic programming [143, 232, 311], graph cuts [24, 148], belief propagation [71, 146, 292], and variational methods [150, 213, 214, 286].

Dynamic programming algorithm is based on smoothness and ordering constraints in order to optimize the stereo correspondences in each scan line. It is made of two main steps: the first consists in performing a forward computation of the path costs matrix associated to each possible correspondence pixel candidate, while in the second backward step, the algorithm selects only the matching pixel yielding the minimum path cost. In spite of its advantage of constraining the disparity smoothness along the epipolar line, this method suffers from inter-scanlines inconsistency.

To overcome this problem, graph cut algorithms were proposed. The basic idea is to transform the matching process to a pixel labeling problem in which the best match corresponds to the minimal cut within the constructed graph. On the other hand, belief propagation is a message passing algorithm in which the probability that a receiver pixel corresponds to a match is evaluated in each iteration based on the sender pixel information. As for variational methods, they consist in minimizing the energy function of Equation 2.7 by solving the corresponding non-linear Euler-Lagrange equation [4].

Despite their accurate results, the global stereo matching algorithms suffer from a high computational cost requiring heavy memory and processing resources. Consequently, this heavy process includes as well long execution delay that, sometimes, the generation of a single stereo pair disparity map can take several minutes. To address this problems, parallel implementations on multi-core architecture have been proposed in [214]. The algorithm was implemented according to the processor capacities and in a way that multi steps of the technique can be processed simultaneously. Finally, it is also important to mention that setting a proper value of the weight parameter α can be a very challenging task.

2.4.3.3 The relationship between depth and disparity

The stereo matching allow to associate to each pixel of a stereo view, an homologous in the other view. Let us recall that each pair of matched left and right points (p^l, p^r) represent actually the planar 2D projections of a single 3D point $P(X, Y, Z)$ of the captured scene by both cameras, which explains the spatial 3D coordinates (X, Y, Z) . Therefore, based on these projections, it is possible to determine the three dimensional point using a process known as triangulation, illustrated in Figure 2.17. Using the similarity of the triangles having as hypotenuses the segments $[C^l p^l]$ and $[C^l P]$, we can deduce the following:

$$\frac{x_P^l}{f} = \frac{X}{Z} \tag{2.8}$$

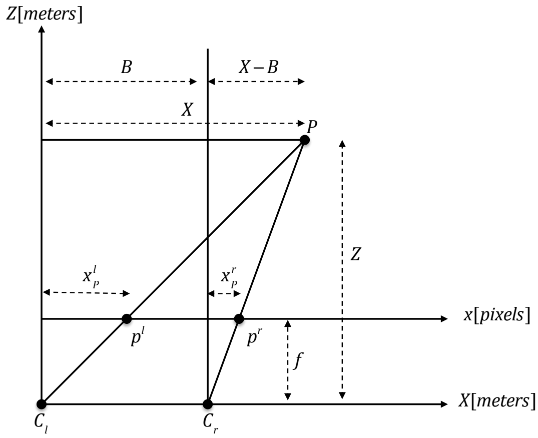


Figure 2.17: Relationship between depth information and the disparity.

Similarly, we can notice also that:

$$\frac{x_P^r}{f} = \frac{X - B}{Z} \tag{2.9}$$

where f is the focal camera length and B denotes the baseline. The final step consists in computing the disparity using the post-rectification definition of disparity in Equation 2.6 to obtain:

$$\vartheta_x(x_P^r, y_P^r) = (x_P^l - x_P^r) = \frac{B \times f}{Z} \tag{2.10}$$

Thus, we can conclude, as follows, the depth information which is represented by the coordinate Z reflecting the distance between the point and the cameras plan:

$$Z = \frac{B \times f}{(x_P^l - x_P^r)} \tag{2.11}$$

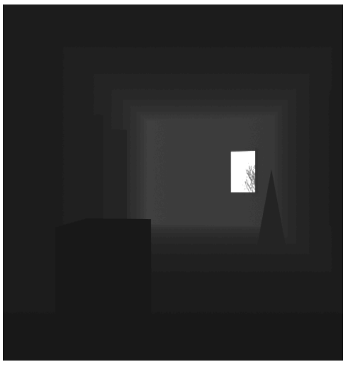
Figure 2.18 illustrates the "room3D" stereo image and its corresponding ground truth disparity map and depth map computed based on the camera parameters. We can notice that from Figure 2.18(c) and Figure 2.18(d) that the depth and the disparity are inversely proportional.

2.5 Conclusion

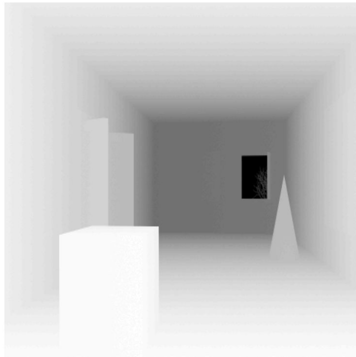
This chapter presented the stereo-vision principle and the different physiological and extraneous cues leading to the depth perception. As a projection

(a) Left image

(b) Right image



(c) Depth map



(d) Disparity map

Figure 2.18: Depth and disparity maps computed from the "room3D" stereo image based on the camera parameters.

to stereopsis, an image can be generated either by an eye or by a camera. The aim of both systems is to transform the visual signal capturing the surrounding 3D environment into 2D images. Inspired by the binocular disparity, the acquisition of 3D visual data require two cameras displaced with slight inwards rotations forming the equivalent of the binocular vision field. The two images captured with slightly different perspectives of the same scene, are the input to the HVS processing, such as the retinal images, to recover the 3D information. The second part of the chapter focused on the geometrical relationship between the 3D scene and the captured 2D views projected on the camera sensing planes. This relation includes as well inter-view matching constraints and assumptions which aim is to find corresponding points within a stereo-pair image, a process leading to the estimation of the disparity map. We have consequently recalled the most important techniques to estimate the disparity map. The increasing understanding of different important features of the HVS such as the binocular combination/rivalry has inspired many stereo image processing concepts and approaches such as asymmetric compression [181, 208], cyclopean-image-based quality assessment technique [34] and the binocular just noticeable difference model [333].

Chapter 3

Endoscopic image enhancement

> The improvement of understanding is for two ends. First, our own increase of knowledge; secondly, to enable us to deliver that knowledge to others.
>
> JOHN LOCKE

3.1 Introduction

During the last decades, endoscopic image enhancement (EIE) has become a very popular research field due to the success of minimally invasive interventions and the continuous innovation of new treatment and diagnosis tools such as robotic surgery systems, stereoscopic laparoscopes, and wireless capsule endoscopy. Based on the literature study, we distinguish two main aims for EIE. The first goal consists in improving the perceived endoscopic image quality by exhibiting tissue details and abnormalities in order to improve the diagnosis accuracy and reliability. The intra-operative visual feedback quality is also of great importance for the surgeons since it can affect the efficiency of their tasks. Indeed, endoscopic videos should highlight significant details such as tissues texture, organs boundaries and blood vessels. The latter, for instance, should be blocked before performing a resection in order to avoid launching a bleeding, which can compromise the patient safety. If not properly controlled, the bleeding may oblige the doctors to convert to an OS, in which case the patient gets the worst of both surgical exercises (i.e. increasing the operating time without any post-operative advantage).

The second possible purpose of EIE is to improve the outcomes of subsequent post-processing tasks such as features extraction for 3D organ reconstruction and registration [238]. For some laparoscopic surgeries, these tasks are prerequisite to register the patient-specific data, establish a navigation plan, and provide the surgeon with an efficient control of robotic-assisted surgical systems.

Thanks to the technological advances, endoscopic imaging systems (EIS) have been undergoing a profound development change that diagnostic endoscopy is evolving from simply diagnosing evident diseases to detecting subtle abnormalities. The recently developed endoscopic imaging technologies have been attempting to produce a high quality visual data for diagnosis, which was not formerly available before performing a biopsy followed by a histological interpretation. With high definition (HD) technologies integrating charge-coupled devices (CCD) three times higher than the conventional endoscopes/displays, the new EIS allow early detection and classification of lesions, leading thus to the treatment optimization. The main purpose of the advanced diagnostic EIS include:

- Improving the detection of mucosal alterations and minute lesions having a potential neoplasia.
- Providing enough data to characterize the tissue of interest and decide whether it is neoplastic, non-neoplastic or inflammatory.
- Performing an *in-vivo* histology and optical biopsy.

The characteristics of the endoscopic site, including dynamic illumination conditions and special artifacts, together with the necessity to highlight certain tissues features has made endoscopic image enhancement a very active research field [120]. Depending on whether the enhancement approach is software/hardware-based, and on the related surgical use-case stage (i.e. pre-operative for diagnosis or intra-operative), endoscopic image enhancement techniques can be categorized into three main classes:

- Image-enhanced endoscopic technologies developed by industry to allow the examination of the entire gastrointestinal (GI) tract lumen, referred to also as *field enhancement* techniques. Such systems encompass visual image quality enhancement techniques that are based on dye, and in-chip technologies including optical, and/or electronic methods. Among this category, autoflorescence imaging (AFI) [269], narrow band imaging (NBI) [153] *(Olympus Medical Systems, Shinjuku, Tokyo, Japan)*, i-SCAN [120, 99] *(PENTAX Medical)*, and Fuji intelligent chromoendoscopy (FICE) [120, 151, 288] *(Fuji Intelligent Color Enhancement, Fujinon, Saitama, Japan)* are the most popular systems. These technologies are mainly based on light-tissue interactions to improve the early diagnosis of gastrointestinal tract neoplasia.
- Virtual histology, called also *point enhancement*, for intra-operative *in vivo* histological survey. Among these techniques, confocal laser endomicroscopy [33, 140], endocytoscopy [113], and optical coherence

tomography [103, 72] produce real-time cellular and sub-cellular images for optical biopsy.

- Image processing techniques that operate directly on endoscopic images by adjusting some features such as its histograms or the pixel intensities/frequencies, without requiring any additional in-chip technology.

In this chapter, we present the different existent technologies and methods for endoscopic image enhancement. Although that the scope of this thesis focuses on image processing techniques for EIE, it is important to present the first two categories for the sake of completeness. Another reason to briefly describe the field and point enhancement methods is that some EIE methods include some notions related to these technologies. Before that, it is however crucial to outline the main challenges and problems related to endoscopic images. Therefore, we start with Section 3.2 which gives an overview of the main artifacts/noise that can occur for endoscopic images. Section 3.3 highlights the evolution of endoscopy video acquisition technologies, namely the significant improvement of image resolution and magnifying techniques, which are primary basic factors that can impact the endoscopic image quality. Sections 3.4 investigates the most popular endoscopic imaging technologies developed by the industry, which are currently being used in the clinical practice. Section 3.5 gives a brief overview of the virtual histology tools. Section 3.6 study the different image processing approaches and techniques for endoscopic image enhancement, and section 3.7 concludes the chapter.

3.2 Endoscopic images noise

Digital images can be degraded during the various capture, processing and transmission steps which might yield some errors, refereed to as noise or artifacts. Since endoscopic images are generally displayed in real-time without any network transmission or compression (except for WCE), the noise mainly occurs during either the acquisition process or the intra-operative exercise. Indeed, endoscopic videos have quite special characteristics related to their content (geometric characteristics of the endo-human organs, dynamic illumination conditions, color properties of the tissues, etc.) and the capture/processing procedure (special camera specifications and calibration, reduced and distorted field of view, display system). Thus, endoscopic image enhancement algorithms have to deal with specific problems with respect to the inherent artifacts. In this section, we list and focus on the possible artifacts and problems related to endoscopic visual content enhancement.

3.2.1 Specular reflections

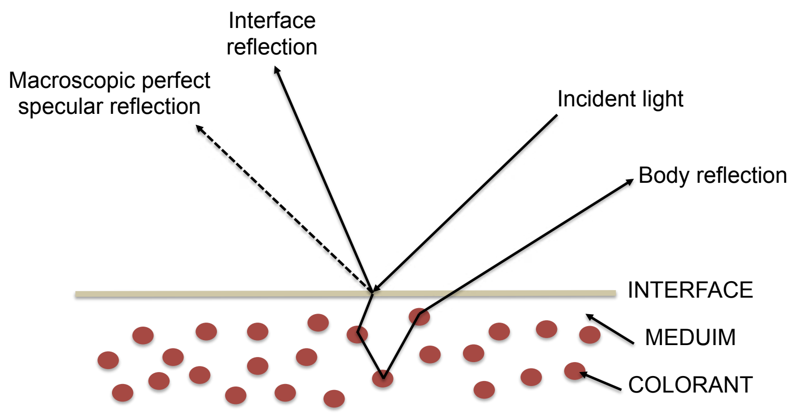


Figure 3.1: Creation of specular reflections in an endoscopic environment

Due to moist endoscopic tissues, specular reflections appear as white glare or light-colored glare that get brighter when the surface normal bisects the angles between the vision direction (endoscope camera) and the incident light. The literature distinguish two types of reflections: the specular reflections and the diffuse reflections. The first type is the light reflected at the interface between the air and the tissues surface and it is governed by the well known *Fresnel* law [273]. This latter relates the specular reflectance to the index of refraction of the material, the angle of incidence and the polarization of the incoming illumination. The diffuse reflection represents the light that penetrates though the tissue interface and passes through the medium, where it undergoes and get scattered by the colorant. Thus, the incident light is either transmitted through the endoscopic tissue (when the latter is not opaque), absorbed by the colorant, or reflected through the same interface by which it entered. In the last case the reflected light generates the specular component. Figure 3.1 summarizes the different types of specular reflection created in an endoscopic environment.

Many approached have been proposed to separate the specular components and removes the highlights. Color analysis [9, 268] methods rely on understanding the distribution of the diffuse and specular components in a color image and using this information in the separation process. In the same vein, other techniques combine the color information with a classifier to decompose the input image into shading and reflectance images [302]. The inpanting techniques [296, 209, 86] fills in the highlights regions by propagating information from region boundaries making use of the spatial correlation. Similarly, the multi-flash methods [73, 3] generate an image with reduced highlights using an image sequence taken by a fixed camera for varying flash light positions. The partial differential equation (PDE) based technique [196] achieves the separation by iteratively eroding the specular component at each pixel. For complex textured surfaced, using a single input image does not remove the specularities correctly. Therefore, other methods [277, 329] generate a pseudo-diffuse component image called specular-free image. This provides a partial separation of the specular component, which is later used to complete the reflection component separation of the original image. In a quiet different approach based on the assumption that specular components tends to be polarized, polarization methods [191, 308, 105] consists in placing a polarization filter in front of the sensor in order to measure the light reflected by the surface and then use this data to recover the original signal. In [7], *Artusi et al.* present a survey of the different existing specular reflections removal techniques.

Several factors may effect the applicability of specularities removal methods in a laparoscopic surgery context because of the limitation of the endosurgical site and the real-time constraints. These factors can be the number of required input images, light constraints, the reflection model used, and whether the process is fully automated or semi-automated. Requiring a large number of input images reduces the feasibility of the method making it complicated to implement and resulting a heavy processing. This is the case for example for the histogram-based approach [38] which requires 200 images to generate satisfactory results. Furthermore, a large number of images may increase the acquisition time which is quiet limited for laparoscopic surgeries having real time constraints. Indeed it is not practical for the surgeon to interrupt his work for several seconds each time a specular reflection appears. On the other hand, methods relying on a single input image do not often remove specularities correctly. Therefore, using a limited number of images (2-60) can ensure a good compromise between the feasibility and efficiency. Methods requiring hardware assistance such as the polarized filter for polarization methods and the flash system for the multi-flash techniques can not be applied in this case for two reasons. First, because of the limitation of the surgical site. Second, because introducing such equipments can reduce the visibility for the surgeons and can prompt incompatibilities with the applicable standards in terms of equipments and sterilization. This implies that reproducing the setup is quite impossible in the operating theatre.

In addition to that, requiring a user interaction reduces considerably the applicability of a method in endoscopic imaging. This is the case for instance for color analysis methods and inpainting techniques that require performing a manual segmentation as it is difficult to achieve accurate automatic segmentation in the presence of specularities. Furthermore, the assumptions made by several reflection models may not be satisfied in some cases, which make the technique unable to detect highlights properly. One typical assumption is that the spectral distribution of the incident light does not change after the reflection process (i.e. it is the same as that of the specularities). In [5], the authors have demonstrated that the error introduced by this approximation is not negligible. This is a typical assumption of most of the methods based on the dichromatic reflection model. Another assumption related to this model stipulates that the diffuse reflection follows the *Lambertian* model and that the illumination color is uniform throughout the highlight. This is not th case for endoscopic human tissues. In fact, it is valid only when the highlight is generated by a single source light, however, it can be generated by multiple inter-reflections making the colors non-uniform.

In practice, the new generation of endoscopes and robotic surgery systems offer more flexibility to attenuate or avoid specular reflections by controlling the light intensity level or slightly shifting the flexible endoscopic arm to move away the highlights from the surgical target spot.

3.2.2 Surgical smoke and lens fogging

Surgical smoke and lens fogging are considered as two major intra-operative problems that can reduce the visibility for surgeons and impact their performance. Indeed, in practice their work is frequently interrupted in order to clean lens fog or evacuate smoke/gas that parasitizes their field of view.

The smoke from electrosurgical units, commonly called also as *plume*, can occur by any procedure aiming to cut, ablate, desiccate, coagulate, fulgurate, or vaporize the endoscopic tissues using the various electronic surgical devices. Beside reducing the visibility, it can increase the risk of complications for the patients and yield a negative impact on the surgical stuff health with long exposure according to some studies [21, 31]. Surgical smoke is also one of the main challenges for real-time tracking of surgical instruments in endoscopic videos [289, 23].

While several proposed image processing techniques [303, 325] attempted to remove perceptually the surgical smoke and improve the vision in the endoscopic surgical site, smoke evacuation apparatus are becoming popular because of their high capture velocity and efficiency. Nowadays plume evacuation pencils and filters such as *PlumePen Elite*[1] and *ClearFlow Smoke Evacuation System*[2] can be plugged easily to the surgical unit or directly into the trocars providing a real-time absorption and evacuation/filtration of the odours and chemical toxins. The success of the smoke evacuators can be explained by their efficiency, the easy integration in the surgical setup and the fact that such systems resolve two problems simultaneously: maintain

[1]Buffalo Filter^{TM}: www.buffalofilter.com/
[2]LaproSurge^{TM}: www.laprosurge.com

a clear field of vision, and mitigate health risks and complications for both the patient and the surgical stuff.

3.2.3 Defective pixels errors

One of the major concerns that faced the end-users in the past was to resolve the problem of defective pixel errors which occurs in the camera sensors. These errors can happen when a pixel fails to sense or reproduce light levels correctly. Nowadays, thanks to the optical and technological advances, cameras are equipped with automatic dynamic detection/correction processes allowing to remove continuously in real-time such errors. Particularly, the new generation of endoscopes includes chip-in-scope and chip-in-tip technologies which provide further algorithms to overcome optical issues such as vignetting and distortion [228]. All technological and optical advances focusing on the endoscopic video/image quality issues attempt continuously to reduce the various possible artifacts and noise, providing the surgeons and the medical stuff with a relatively high quality visual content. The following sections are devoted to present the different tools and hardware/ software solutions allowing to optimize the endoscopic image/video quality and exhibit some prerequisite features for the diagnosis and treatment of patients.

3.3 High resolution and magnifying endoscopy

The first fiber endoscope was commercialized in 1961 marking the beginning of a huge leap in the early diagnosis of GI diseases. The following development achievements in terms of endoscopic technologies have led to a large replacement of fiber-optic endoscopes by electronic video-endoscopoes. Conventional endoscopes are made of CCD chips integrating 100K to 300K cells, which implies that each image is represented by a pixel density of 100.000 to 300.000 pixels. This technical feature is of great importance since it indicates the image resolution and, consequently, it affects the ability to distinguish the images details or two closely approximated objects/points. Thus, as the pixel density increases, the image resolution gets better and tissue abnormalities can be distinguished and detected easier and earlier with more sensitivity to noise. Nowadays, the new generation of endoscopes are equipped with 400K CCD chips, and has reached recently 850K to more than one million pixel density [19]. In digital imaging systems, endoscopes and other capture/display devices equipped with such high pixel density are referred to as high definition (HD) technologies.

It is important to note that the term HD is sometimes confused with magnifying technology when talking about endoscopes. Indeed, some endoscopes are equipped with an optical adjustable focus mechanism made of a dynamic motor driven lens attached to the scope's tip. The focal distance can be adjusted by moving the scope closer towards the mucosal surface providing a zoomed image with a magnification capacity that range from 15 to 150 times [141, 19]. Such magnification ability allows to discriminate very fine objects and details that are only 10 to 71 microns in diameter [230, 19]. These special endoscopes providing electronically larger images, as an alternative to optical magnification, are known as magnifying endoscopes. Although electronic magnifying technology provides a zoomed view of the examined image region with more details, its performance is limited to a certain magnification level beyond which image pixels become visible (i.e. pixelization), which implies that the image quality decreases drastically. This can be explained by the fact that as the zoom level increases, the number of pixels forming the displayed magnified image spot decreases. Indeed, unlike the flexible dynamic magnification levels, the number of pixels in an image spatial resolution is constant. The high pixel density of HD technologies compensate the spatial resolution loss during the magnification and ensures a sufficiently high number of pixels to produce a relatively good quality image with sharp details, which mitigate/remove the blur and pixelization artifacts.

The latest generation of 4K resolution has been also recently integrated in clinical endoscopy, although very active discussions are being held about the effectiveness and advantages that such expensive technology can offer. The endoscopic towers nowadays are generally equipped with a 27-32 inch monitors. In order to perceive and distinguish 4K resolution details with a 32in monitor, the observer must be around 60 cm distant from the screen, which is often not practical in the operating room settings. Indeed, without a movable screen support, the monitor should to be placed in the sterile area above the patient. Furthermore, the 4K technology is not effective if the full endoscopic vision system chain, from the endoscope camera sensors to the monitor, are not 4K made. Therefore, this feature is considered at the moment as a high-end technology for customers that are able to afford a 50+ inch screens in the surgical theater. It is also noteworthy mentioning that an extended BT2020 colorspace substituting the BT709 norm will be combined with 4K technologies for endoscopic procedures [228].

3.4 Field enhancement imaging technologies

Chromoendoscopy [94] has been used for years as one of simplest techniques to examine the superficial tissue structure and inspect pre-malignant symptoms, small cancerous lesions, as well as estimating the submucosal[3] invasion and the histological type. The technique relies on the reaction of

[3] Submucosal : a supporting layer of loose connective tissue directly under a mucous membrane

the GI mucosa [4] to the application of dyes or stains (generally either methylene blue or indigo carmine) during the endoscopy. Despite its safety and reduced cost, the failure of standardization and the labor intensive nature of chromoendoscopy impeded its integration in the clinical daily routine.

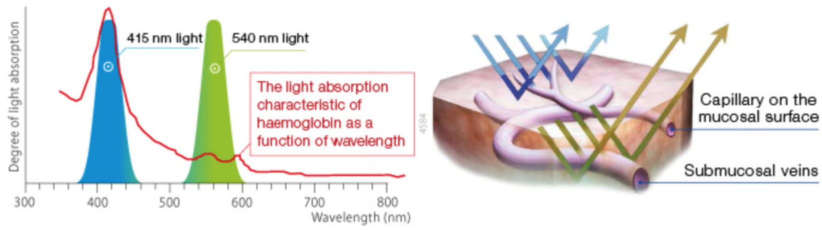
The major flaw with classical white light endoscopy (WLE) is that the detection efficiency is limited to how grosss and obvious the morphological changes are. The diagnosis is mainly based on either biopsy sampling of clear endo-macroscopic features, or "*blind*" biopsy sampling of random tissue spots within normal appearing mucosa. WLE is generally not able to identify fine alterations in the flat lesions of the mucosa, which hinders early detection of GI malignancies in the early resectable and curative stages.

The need for a better quality endoscopic visual feedback via more efficient and flexible tools has raised the following question: How can we evolve or substitute conventional chromoendoscopy and WLE which require biopsy ? The answer has been provided with the optical and technological advances, which led to the innovation of the "*digital chromoendoscopy*" including NBI, i-SCAN, and FICE systems.

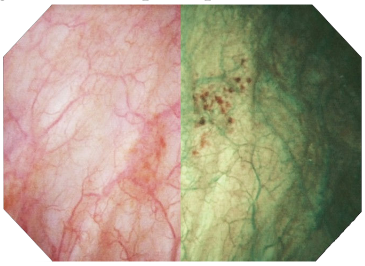
NBI is an optical filter technology that enhances mainly the images of structural mucosal patterns and mucosal/submucosal vessels. Unlike conventional chromoendoscopy, NBI offers more functional flexibility since it can be switched on and off using a button on the endoscope. The system consists in interposing narrow bandpass filters in front of a white light source, composed of equally mixed wavelength, to adjust the illumination wavelength of the perceived tissues. The currenty used NBI systems are composed of two filters separating the light spectrum to a the blue illumination (415 nm) and green illumination (540 nm). Both wavelength values correspond respectively to the main and a secondary hemoglobin absorption peaks. While the short wavelengths penetrate only the superficial mucosal layers exposing capillaries in brown color, the second wavelength penetrates deeper within mucosal and submucosal layers to be absorbed by blood vessels which appear as cyan, as illustrated in figure 3.2. Some studies [288, 226] reported that NBI performance can be improved when combined with magnification (ME-NBI). This combination can ameliorate the analysis accuracy of both abnormal surface architecture of neoplastic lesions (pit-pattern) and the epithelial crests of the mucosa. The most relevant contribution is, however, the clear visualization of the vascular mucosa network, which is useful especially for the abnormal neoangiogenesis process evaluation in high-grade dysplasia or early cancer [153, 264]

FICE is a flexible spectral imaging color enhancement technique, taking advantage of the recent imaging technology advances allowing the estima-

[4]Mucosa: a membrane rich in mucous glands that lines body passages and cavities (as of the digestive, respiratory, and genitourinary tracts) which connect directly or indirectly with the exterior (mucous membrane)



(a) NBI light reflection principle on mucosal tissues



(b) Tissue illuminated with a conventional white light (left side) and with NBI (right side)

Figure 3.2: NBI optical filtering, *source: www.olympus.co.uk*

tion of specific light wavelengths forming the image (i.e. spectral estimation technology). Since FICE exploits a wide range of spectral combinations, it is also often referred to as multi-band imaging (MBI) or optimal band imaging. MBI is a digital image-processing technology that produces an enhanced appearance of mucosal surface structures based on specific light wavelengths forming virtually reconstituted images. Similarly to NBI, it is possible to couple FICE with an optical (zoom in/out) or electronic magnification to improve the perception of mucosal details. However, while NBI uses optical filters for narrow spectral transmittance, FICE is software driven technique integrating image-processing algorithms that exploits spectral estimation methods [217]. The process is composed of 3 steps. First, the images captured using a standard color CCD endoscope are transmitted to a processing circuit within the video processor where a spectral estimation matrix is computed. Using the estimated spectra, a single-wavelength virtual image is then reconstructed. Finally, the red, green and blue (RGB) monitor inputs are affected with three selected single-wavelength images that compose the final color-enhanced FICE image displayed in real-time. There are only few published researches studying the efficiency of MBI for differentiation or detection of GI tract lesions, in spite of its reported better performance compared to WLE.

Similarly to FICE, i-SCAN technology modulates the reflected light from endoscopic surfaces using computer algorithms to enhance the captured images. it is made of three different components: Surface enhancement (SE), contrast enhancement (CE), and tone enhancement (TE). It is possible to apply more than one mode at the same time since they are arranged in serie. SE improves the contrast between dark and illuminated areas of the surface. An algorithm is applied on the luminance component pixel intensities in order to visually enhance the mucosal surface structure and lesions boundaries details while preserving the natural color distribution. The aim of CE is to allow better details perception of delicate irregularities surrounding the surface and mucosal vascular pattern morphology. This enhancement is achieved by simply adding slight blue color distribution to dark areas. Thus, this mode results only in a bluish-white staining effect without modifying the image brightness or colorfulness. In the TE mode, a specific tone mapping is applied separately to each component of the RGB color space according to five different processing modes depending on the operated fragment: pit pattern, vascular pattern, Barrett's, oesophagus, stomach and colon modes. In contrast with NBI, red remains the predominant blood vessels color in all i-SCAN modes.

In a different approach, endoscopic AFI consists in generating real-time pseudo-color images using the interaction between tissue fluorophores and a light source having a specific wavelength. The primary purpose of this imaging technique is to provide a red flag technology allowing rapid examination of relatively large surface of the GI mucosa for image guided biopsy and early detection of any possible dysplastic region. When a short wavelength light illuminates the tissues, endogenous fluorophores (adenine, dinucleotide, nicotinamide, collagen, flavin and porphyrins) are stimulated to re-emit fluorescent light.

The autofluorescence reaction and characteristics differ between normal, neoplastic and inflamed tissues according to the concentration, distribution and type of different fluorophores together with the perfusion characteristics and biochemical composition [90]. Since malignant tissues are linked with emission of a longer wavelength (i.e., with a different color than the absorbed light), it is possible to visually distinguish them and perform a diagnosis based on the generated colors. Generally, normal tissues are pseudocolored as green, blood vessels are represented as dark green, while hyperplastic/neoplastic and hypertrophic fundic mucosa of the stomach regions appear as magenta. An AFI-positive lesion (suspected to be in a neoplasia stage), is defined as a region having different color from the surrounding mucosa with a defined circumferential margin [134]. This system is able to reveal early stage cancers that are not detectable by conventional WLE [91, 307].

While many studies demonstrated the AFI potential to help detecting premalignant lesions (dysplasia) beside early cancers, they revealed as well several significant limitations of this technique, namely the large number of false-positive results. The consequent low positive predictive value can be improved using complementary methods such as optical-coherence tomography (OCT), confocal laser endomicroscopy [316] or ME-NBI. Another reported weakness related to AFI is the instability of the color tone due to the presence of mucus, air insufflation or instrument angulation. In [47, 132, 133], the authors reported that using an endoscopic system incorporating trimodal endoscopic imaging technologies combining AFI, ME-NBI and WLE can ameliorate the diagnostic accuracy for early cancer and high-grade dysplasia.

Based on the observation that infrared (IR) rays disperse relatively little and are slightly absorbed by the human body, consequently can reach deeper tissues than the visible rays [202, 231], other studies [83, 114, 203, 116, 145, 115] focused on testing and developing IR techniques. Using the indocyanine green as an IR ray absorber, this method exploits the technical fact that conventional endoscopes based on CCDs are receptive to both visible and invisible rays. The studies demonstrated that the deep IR penetration provides better contrast and visualization of deeper lesions reaching even the submucosal layer.

3.5 Virtual histology : point enhancement

The endoscopic imaging and technological advances achieved in the last decades have made it now possible to perform microscopic examination at cellular level, which helps to improve the tissue characterization of different types of neoplastic and non neoplastic lesions of the GI tract. In the following sections, we present the two leading optical techniques allowing the observation of such details level, namely confocal laser endomicroscopy (CLE) and endocytoscopy.

3.5.1 Confocal laser endomicroscopy

Confocal laser endomicroscopy (CLE) is an on-site endoscopic histological tool allowing microscopic examination of the GI mucosa using a fluorescent agent. Since its development, approximately fifteen years ago [138], CLE has been studied for a steadily growing number of diseases [98, 15, 139]. Due the reported high correlation rates with histopathology [43, 314], the trials has been extended from simple feasibility studies in different portions of the GI tract, to inflammatory and pre-neoplastic/neoplastic diseases that are commonly diagnosed using only random biopsies. In the latter case, CLE is considered as a "*smart biopsy*" tool, targeting specific regions of interest via intravital microscopy. This feature extended even more the usefulness of CLE towards intravital microscopy of the liver, pancreas, or the biliary tract [80].

As for the technical aspect, CLE consists in focusing a low power laser onto a single interest point within the tissues. The reflected light from this point is concentrated to a detector through a pinhole, while the 2D imaging plane is being scanned by a laser raster. In order to avoid blurring effect, the reflected light from any surface outside the focally illuminated spot is geometrically excluded. This allows having a visual feedback not only at the tissue surface, but below the suface as well without any need to physical disruption of the tissue integrity.

The first development of the technology suggested an endoscope-integrated system (e-CLE) in which a miniature scanner is integrated into a conventional colonoscope *(EC3870K, Pentax Medical, Japan)* offering different applications to the upper and lower endoscopy. The second technology, which is probe-based CLE system (p-CLE) *(Cellvizio, Mauna Kea Technologies, France)*, is made of a fibre bundle having the flexibility to be passed through the working channel of most endoscopes. Being the only commercially available system nowadays, the p-CLE can be used to other GI tract parts and present the advantage of faster image acquisition and real-time microscopic video processing.

Despite the flexibility to different clinical applications and the high accuracy, CLE is still not completely adopted in the clinical routine. This is due to several reasons namely the expensive cost of the procedure, the luck of reimbursement and codification in some countries, the absence of standardization, the need to train physicians for accurate image interpretation, the role of pathologists in the process, and medico-legal issues.

3.5.2 Endocytoscopy

Endoctyoscopy [154, 283] is an ultra-high endoscopy magnification technique (exceeding 450 times magnification) allowing spot-oriented cellular assessment of the GI tract using an pre-procedural staining to enhance the cellular structures. The endoctyoscopic technology is based on a high-power objective lens with fixed focus and can be either integrated into the endoscope or designed as a prob-based system. As in a conventional endoscope, a light source and a video processor are connected to the endocytoscope. The real time processing and visualization of endoscopic images require, however, the simultaneous functioning of two processors.

Before the endoctyoscopic magnification, the target region must be stained using a double stain technique simulating the combination of hematoxylin and eosin staining of conventional histology. The first staining step consists in treating the target region with N-acetylcysteine (10%), which is a mucolytic agent. In the second step, methylene blue (1%) and crystal violet

(0.1%) are applied to stain the nucleus and the cytoplasm. The endocytoscope probe is then passed through the working channel of a regular endosope and maneuvered gradually until reaching a direct contact of the tip with the stained mucosal surface of the target spot.

In spite of the reported high correlation rates with histology [131], many challenges have to be addressed before accepting this novel technology in the clinical routine. These challenges include the sensitivity of the system to blood, mucosa and cardiac/respiratory movements of the patient, which can lead to an obscure or unclear view. An additional weakness is the incapacity of mapping and observing the entire area of interest since only a small spot is visible at the same time.

In the light of the increasing technological advances, all field and point imaging techniques presented in this section have great potential for further improvements; which can lead according to some ambitious reports to an "a priori" cancer development detection [228]. There are, however, many challenges and problems ahead. As for the emerging newer technologies, the limitations include weak reproducibility and the luck of standards and norms. Furthermore, additional systematic training may be required for endoscopists to update and improve their skills in cellular/sub-cellular images interpretation. This implies also the need to further study and adapt some medical diagnosis positions related especially to the role of pathologists and endoscopists. Finally, the relatively high cost and possible legal/ethical issues can affect the integration and reimbursement by health-care systems in some countries. Therefore, with the significant development of image processing techniques, many studies were devoted to enhance endoscopic images quality using these less-constraining methods which result relatively well enhanced images.

3.6 Image processing techniques

In the previous sections, we presented the different technological and in-chip systems/techniques for endoscopic image enhancement. In spite of its relatively good performance, such tools yield many integration challenges and limitations related mainly to the expensive cost, the power consumption and the impact on the surgical settings and protocols. One efficient alternative solution is to make use of the different image processing techniques that can be implemented generally without requiring any additional hardware or devices. Therefore, this section is devoted to review the various image processing methods used to enhance endoscopic images. Before dealing with technical details, it is however firstly necessary, for the sake of completeness to recall the different basics and approaches for image processing and enhancement.

The literature distinguishes two main approaches for 2D image processing, namely spatial domain based methods and transform domain methods. The term spatial refers to the flat two dimensional plane of the image, which corresponds to the matrix containing discreet sampled change of the pixels intensity. Although this conventional approach is the earlier explored and often the simpler, it is not convenient when performing a targeted enhancement of a specific level of details within the image. The transform domain methods, on the other hand, operate on the transformed image coefficients using for instance Fourier, wavelet, or discrete cosine transforms (DCT).

3.6.1 Spatial domain image enhancement

Depending on the processing purpose and the number of pixels involved simultaneously in the enhancement, the existing spatial-domain image processing techniques can be classified into two categories [244, 118] : spatial filtering and intensity transformations (termed also point transformations). The latter aims either to perform a thresholding or adjust the image contrast by operating separately on every single pixel. Spatial filtering, on the other hand, involve the surrounding area of the currently processed pixel in the enhancement to perform for example an image smoothing or sharpening. The spatial domain enhancement can be denoted by the following equation:

$$I^{enh} = T\left[I^{org}\left(x, y\right); W\left(x, y\right); \Omega\right] \tag{3.1}$$

where I^{org} refers to the input original image, I^{enh} is the enhanced output image, and T is an operator applied on I^{org} over a window W centered in the pixel of spatial coordinates (x, y) according to the property Ω (local or/and global descriptors/attributes) which determines whether the processing is local or global. T is referred to as well as spatial filter, *spatial mask*, or *kernel*. The type of the neighborhood processing determined by the kernel defines the nature of filtering. In the extremely smallest possible kernel size (1×1), the processing depends only on the value of the window center, which corresponds to the pixel intensity at spatial coordinates (x, y). In this case, Equation 3.1 becomes an *intensity transformation* that varies as a function of the currently processed point intensity only, and can be re-written as follows:

$$o = T(i) \tag{3.2}$$

where i and o refer respectively to the pixel intensity at any point having spatial coordinates (x, y) in I^{org} and I^{enh}. This is the reason why intensity processing is called also *point/single-pixel processing*. Figure 3.3 illustrates three main classes of frequently used intensity transformation functions for image enhancement [118, 244], namely: i) linear function including both identity and negative transformations, ii) power-law function containing the *n*th power and the *n*th root mapping, and logarithmic transformations including *log* and the *inverse-log* functions. The variable L designates the number of intensity levels. For instance, an 8-bits coded image have $L = 2^8$ gray-scale levels with intensity values in the range $[0,\ L-1]$.

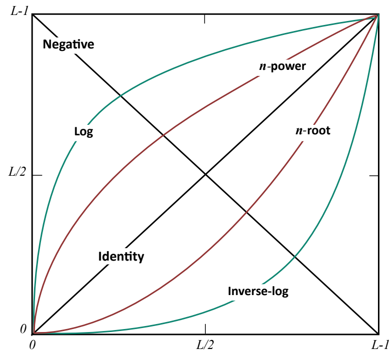


Figure 3.3: Some basic intensity transformation functions

Table 3.1 presents the mathematical formulations of the different transformations listed above. The identity function is the simple particular case in which the output intensity values are the same as the input. The negative function consists in inversing the intensity of each pixel to obtain an output similar to the negative image of a photographic plastic transparent film. This transformation is useful to expose details which are embedded in black or dark regions of the image, especially if such areas dominate the spatial image resolution. Similarly, the *log* function [67, 195] enlarges low intensity levels to a wider range while compressing the high level intensity values. This transformation is very useful especially when dealing with images having large variations in pixel intensity values, such as for Fourier transformed images having wide dynamic range. The *inverse-log* function performs the opposite transformation. These compression and spreading features of intensity values can be performed, generally, by any function having the same shape as the log transformation, including the *power-law function* [118, 244] which is more flexible and appropriate for this purpose. Indeed, a fractional power value (γ) maps a narrow range of low tonal levels (dark regions) to a larger output range values. In the opposite case, higher γ values compress the dynamic range. The flexibility of this function can be noticed by varying the fractional value of γ which produces other curves of the same family, unlike the log function which has a unique transformation curve. Thus, the *power-low* function is also useful to adjust the contrast of both dark and washed out images. It is important to note that several acquisition and display systems integrate a power-low transformation, which might produce darker or bleached out images depending on the device-characteristic

Transformations	Mathematical formulation
Negative	$o = L - 1 - i$
Log	$o = c \times \log(1 + i)$
Power-Low	$o = c \times i^{\gamma}$

Table 3.1: Main intensity transformation functions for spatial domain image enhancement

γ value. To compensate this power transformation and display an image that is closer visually to the input signal, an inverse transformation should be applied before displaying the image. This process having the mathematical form of $o = i^{\frac{1}{\gamma}}$ is conventionally known as *gamma correction* [104, 246].

A complementary family of transformations consists in using piecewise functions which offers much more flexibility than the previously mentioned methods since its shape can be defined adaptively and arbitrary according the targeted mapping together with the nature of input images. This family includes the contrast stretching functions and the intensity-level/Bit-plane slicing. The first category permits spanning the pixel values to cover the full range of possible intensity levels. Intensity-level slicing is a suitable transformation when the aim is to highlight a particular range of intensity levels. The are two main possible ideas to do this. The first is to saturate the region of interest (ROI) pixels intensities while darkening the other pixels, which generates a *binary image*. The second way is to darken (or brightens) the targeted pixel while preserving the intensity values of outside the ROI. Finally, Bit-slicing consists in exposing the targeted regions based on the contributions of specific bits to the overall perceived image, instead of adjusting intensity level ranges. The definition of piecewise transformations, however, requires the specification of significant user parameters, which might prompt challenging questions regarding how to determine such inputs.

Inspired by the contour detection feature of the HVS [135, 46], other spatial-domain methods [17, 123] are based on local contrast measures and edginess information by combining *Gordon*'s method [82] and the theory of contour detection [200]. In the same vein, the *Retinex* theory (RT) has derived many efficient techniques for local/global dynamic range compression and color enhancement/constancy [160, 124, 125, 144, 178, 304]. The term retinex is actually the combination of two words : *retin*e and cort*ex*. The idea behind this name is that the retinex theory models the HVS perception of colors and lightness which are processed by the retina and the visual cortex. Based on the color consistency phenomenon, the theory stipulates that each image can be decomposed into two different components: an illumination image and a reflectance image. The color constancy refers to a feature of the HVS color perception which identify and match any color under different illumination conditions. Thus, according to this property, it is possible to enhance a given image by extracting the illumination component and generating an illumination-independent image. The retinex techniques are, however, constrained by the *gray-world assumption* which states that the average of the three image color channels (RGB: Red, Green and Blue) have to average out to a common gray value [68]. The violation of this assumption produces a "graying" effect caused by a color saturation diminution. This is the case for endoscopic and WCE images where the red color dominate the spectral channels. Additionally, setting the MSR parameters is a difficult task as it can lead to partial enhancement of the image (i.e., enhance only specific regions within the image) [254].

An old process that used to be applied in the printing and photographic industry to increase the sharpness of images consisted in subtracting a smoothed version of a signal (unsharp version) from the input original image. The unsharp mask containing the high frequency edge information is then weighted and added to the original image. This process is known also as *unsharp masking* (UM). In practice, the highpass filtered component can be obtained directly by applying a spatial filtering of the original image using a negative *Laplacian* kernel. Despite the simplicity and the satisfactory obtained results in many applications, UM techniques suffer from two major limitations. i) The processing focus mainly on high contrasted regions of the image, which may result in unpleasant "overshoot" artifacts especially in smooth ares. ii) Basing the enhancement on high frequency filtering makes the method very sensitive to noise. To overcome these problems, various UM techniques relying mainly on nonlinear filtering have been proposed [256, 247, 236, 257].

In a different approach, many spatial-domain enhancement methods are based on processing the intensity levels distribution of an image which is given by the image *histogram*. For an image which pixel intensities ranges in $[0,\ L-1]$, the histogram can be computed using a function $h\left(r_k\right) = n_k/(M \times N)$ where r_k is the kth intensity level, n_k is the number of pixel having the kth intensity level r_k, and M and N stand respectively for the image hight and width expressed in terms of number of pixels. Histogram manipulation has been extensively studied, producing numerous histogram-based enhancement techniques. Among them, histogram equalization (HE) is on the most popular methods. It consists on expanding uniformly the pixels distribution to cover the entire range of all possible intensity levels. Despite its simplicity and low computational cost, HE can produce an over-enhancement or exhibit noise [136]. Additionally a uniform distribution does not produce necessarily always the best image quality as it can yield a loss of image details [136, 17]. Therefore, it is useful sometimes to specify the shape of the processed output image histogram. In this case, the process is called *histogram matching*. Many other HE based techniques have been proposed in the literature such as the adaptive histogram equalization (AHE) [245], the contrast limited adaptive histogram equalization (CLAHE) [335], and brightness preserving dynamic fuzzy histogram equalization (BPDFHE) [275]. Unlike the conventional HE techniques, AHE applies several local histograms on distinct small blocs of the image to redistribute locally the lightness values. The BPDFHE is an ameliorated version of the brightness preserving dynamic histogram equalization (BPDHE) [109] in terms of computational complexity. It consists in representing and processing the images in the fuzzy domain to adjust the contrast without changing the brightness. CLAHE controls the contrast enhancement of AHE to avoid an over-amplification of noise, which can be achieved by limiting the slope of the cumulative distribution function.

3.6.2 Transform domain image enhancement

One of the major fundamental challenges in image processing is to determine the most suitable data representation that optimizes the subsequent image analysis and processing. In spite of their simplicity and the relatively satisfactory results, spatial domain enhancement techniques do not allow efficient selective processing of specific image components/content or addressing a certain level of details, which is an important feature when dealing with medical and endoscopic images. In addition to that, an efficient spatial multi-resolution enhancement (e.g. pyramidal decomposition [149]) of images with rich and complex content requires often the combination of several point transformation and filtering methods increasing, thus, the computational cost/time and the implementation complexity.

One way to overcome this problem is to perform a decomposition/ transformation of the input signal, which allows better targeting of specific image components together with controlling the level of enhancement due the efficient spectral separation. Moreover, a transformation permits an efficient energy compactness which facilitate the transmission and storage of images. This is the reason why many standard compression techniques such as JPEG and JPEG2000 perform a transformation prior to the images compression. The principal workflow-chart of any transform-domain image enhancement technique can be summarized in three main steps as illustrated in Figure 3.4. The general idea is to perform the processing on a transformed version of the input image then apply the inverse transform to obtain the enhanced spatial-domain output image.

Historically, Fourier transform (FT) has been one of the most important decompositions that revolutionized and dominated the signal and image processing fields. FT has been widely used in image enhancement for filtering, smoothing, and sharpening. In a compression and data transmission context, processing images in the DCT domain has also attracted significant attention of researchers due to its energy compactness comparable with KLT (which is the reference) and its integration in some compression stan-

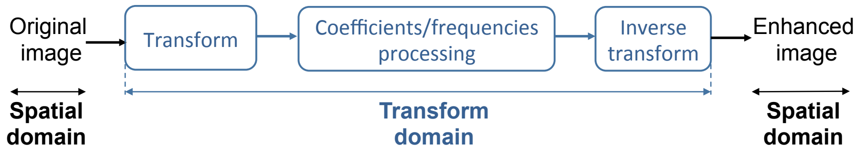


Figure 3.4: General flow-chart of transform domain image enhancement approach

dards such as MPEG and JPEG. In [224], the authors map the DC coefficients within each 8×8 block to adjust the local background illumination. The same scale factor is then tuned and applied to the AC values to preserve the local contrast. In [166], *Lee* exploits the Retinex theory [160] and a proposed spectral content measure based on HVS perception to compress the dynamic range of the image and enhance its contrast. The DCT enhancement-based methods can, however, produce visible blocking artifacts especially if the processing is block-independent. Addressing this problem implies the cost of more computational complexity and memory requirements.

To address the aforementioned shortcomings of DCT, wavelets are an analysis and processing tool that has been extensively studied and adopted in many image processing fields such as compression [27, 61], denoising [32], face recognition [321] and feature extraction [180]. This success can be explained by its intrinsic characteristics, which are namely, the efficient space-frequency localization and the multi-scale representation. In contrast with DCT decomposition which transforms images from the spatial to the frequency domain, wavelets produce a good representation of the input image in both scale and space domains. These two advantages are very important when processing endoscopic images, which are characterized by complex contents and fine details such as blood vessels or subtle tissue abnormalities together with dynamic brightness and wide dark regions. Therefore, we will focus on the wavelet transform based image enhancement techniques in what follows.

3.6.2.1 Wavelet based image enhancement

Unlike the sinusoidal basis functions of the Fourier Transform, wavelet decompositions are based on little waves, referred to also as *wavelets*, having limited duration and varying frequency. The choice of wavelet transform when processing 2D images in general, and medical/endoscopic images in particular, can be justified by two main reasons. First, the transformation decomposes the image into low and high frequency subbands. The compact information of the scaling subband, referred to also as approximation, represents the global image structure, brightness and contrast information.

Each of the approximation coefficients contains a concentrated data about the local spatial distribution of illumination. Wavelet subbands, on the other hand, contain the image details such as edges corresponding to significant coefficients, and trivial coefficients that represent either texture information or noise. Therefore, such efficient separation facilitates the design of an enhancement technique combining brightness image adjustment, details exhibition (e.g. edges and textures), and noise suppression. Second, the mutli-resolution property allows deep analysis of the image by isolating detail components at different scales. The appeal of such property is simple: an image feature that is not detectable at a certain resolution, can be easily detected and processed after increasing the decomposition to a higher/deeper level.

Similar to any transform-domain image processing, the wavelet-based image enhancement is composed of three main steps. First, the image is decomposed into L levels using a wavelet transform. The choice of the decomposition level depends on the image structure complexity and the targeted level of details to enhance. For instance in a medical context, some image details such as fine veins and tissues might be unaccessible after one decomposition. The image decomposition results in four subbands per level referred to as low-low subband (LL) which is called also approximation, low-high suband (LH), high-low suband (HL), and high-high suband (HH). Secondly, the wavelet coefficients are adjusted according to the targeted image components and features to be enhanced. Indeed, tuning the approximation coefficients will modify the global image brightness while processing the wavelet subbands adjust the image details. If necessary, a denoising process can be applied as well in this step to mitigate the trivial coefficients of the HH subband depending on the noise type and its estimated variance. Finally, an inverse wavelet transformation is applied on the processed subbands to obtain the enhanced image.

Wavelets have been widely used in 2D images enhancement. In a medical context, Laine *et al.* [155] propose a contrast enhancement method for mammography images using a nonlinear mapping function with an adaptive thresholding for wavelet coefficients shrinkage. The later step aims to denoise the images while preserving edges, which exhibits the micro-calcifications associated to breast tumors. The technique proposed in [39] selects automatically the wavelet bases and parameters that produce the best contrast enhancement. In [41, 42], the authors propose a global and flattening mapping functions to estimate the enhanced scaling coefficients. The scale factor of an approximation pixel value is then tuned to its corresponding wavelet coefficient depending on local information and the decomposition scale. A shrinkage factor is also added for noise suppression. In a medical context, *Li et Kang. [175]* applied a soft threshold processing to wavelets coefficients without modifying the low frequency values.

In [299, 300, 298], the authors enhance the image contrast based on a contrast measure in the wavelet domain. The latter techniques are inspired by the fact that the HVS detects the details of the scene according to the ratio between high-frequency and low-frequency information. In a different approach [272], the singular value decomposition (SVD) is used for image enhancement and denoising by truncating the lower singular values of the wavelets coefficients. The SVD of the low frequency subband is also used in [258, 57] to enhance the image brightness. In the following paragraphs, we will detail the main principle of both the approximation and detail subbands enhancement according to the different published research works.

Processing of scaling subband The dense information compact in the approximation subband represents the overall image complexity and structure. Therefore, modifying the scaling coefficients has a profound impact on the global image brightness and contrast. An enhanced approximation coefficient $\hat{s}$ is generally obtained by weighting the original scaling intensity s with a proper scale factor κ_s, as follows :

$$\hat{s} = \kappa_s.s = f(s).s \tag{3.3}$$

where $f(.)$ is a monotonically increasing mapping funtion. In [42], the authors proposed the following continuous flexible mapping function:

$$f(\bar{s}) = \begin{cases} 1 + \alpha_1(\beta - \bar{s}) \exp\left(-\frac{|\bar{s}-\beta|^2}{\sigma_1^2}\right), & 0 \leq \bar{s} \leq \beta \\ 1 + \alpha_2(\beta - \bar{s}) \exp\left(-\frac{|\bar{s}-\beta|^2}{\sigma_2^2}\right), & otherwise \end{cases} \tag{3.4}$$

$$\bar{s} = \frac{s}{2^L I_{max}} \tag{3.5}$$

$$\hat{s} = f\left(\frac{h(s)}{2^L I_{max}}\right).s \tag{3.6}$$

where $\bar{s}$ is the normalized scaling coefficient computed as a function of the level of decomposition L and the maximum image intensity value I_{max} as defined in Equation 3.5. The flexibility of this mapping function is achieved via the parameters that are set automatically based on the statistical characteristics of the image to control the mapping shape and make it adaptive for different types of images. The amplitude constants α_1 and α_2 determine the range of scale factor possible values. The parameters σ_1 and σ_2 control the shape of the scaling up and down, and the balance control parameter β is between 0 and 1. In order to avoid any ringing effect that could result from direct weighting of scaling coefficients, local neighboring correlation can be considered by applying a smoothing function on a window centered by the processed approximation pixel. Thus, Equation 3.3 can be rewritten as Equation 3.6 where $h(.)$ is a smoothing convolution filter (e.g. Gaussian filter, Box filter). The flexible mapping function $f(.)$ can be replaced by the Twicing funtion [216], the function used in [166], or the S-function [51]. The mathematical expression of these functions is given respectively by Equations 3.7, 3.8 and 3.9 and the respective plots are illustrated in Figure 3.5.

$$\tau(x) = x(2-x), \quad 0 \leq x \leq 1 \tag{3.7}$$

$$\eta(x) = \frac{\left(x^{\frac{1}{\gamma}} + \left(1-(1-x)^{\frac{1}{\gamma}}\right)\right)}{2}, \quad 0 \leq x \leq 1 \tag{3.8}$$

$$\psi(x) \begin{cases} n\left(1-(1-\frac{x}{m})^{p1}\right) & 0 \leq x \leq 1 \\ n+(1-n)\left(\frac{x-m}{1-m}\right)^{p2} & \\ m \leq x \leq 1 & 0 \leq m \leq n \leq 1,\ p1, p2 > 0 \end{cases} \tag{3.9}$$

Processing of wavelet subbands Enhancing the global image brightness and contrast by adjusting the scaling coefficients ,without modifying accordingly the wavelet subbands, can reduce the image details. Therefore, it is important to perform a local contrast enhancement by tuning the wavelet coefficients in order to preserve or sharpen the image details (e.g. textures and edges). Similarly to the approximation coefficients adjustment, a wavelet coefficient can be enhanced by multiplying its intensity value by a proper scaling factor κ_w as follows:

$$\hat{w} = \kappa_w.w \tag{3.10}$$

One of the simplest ideas for estimating κ_w is to consider the scaling factor of the spatially corresponding approximation coefficient. Such modification can lead, however, to over-enhancement or undesirable distortions especially when dealing with a high decomposition level. This can be explained by the simple fact that the number of wavelet coefficients is much more than the number of approximation coefficients, and increase according to the number of decompositions levels. Moreover, it is not convenient to enhance global image brightness information and local details data using

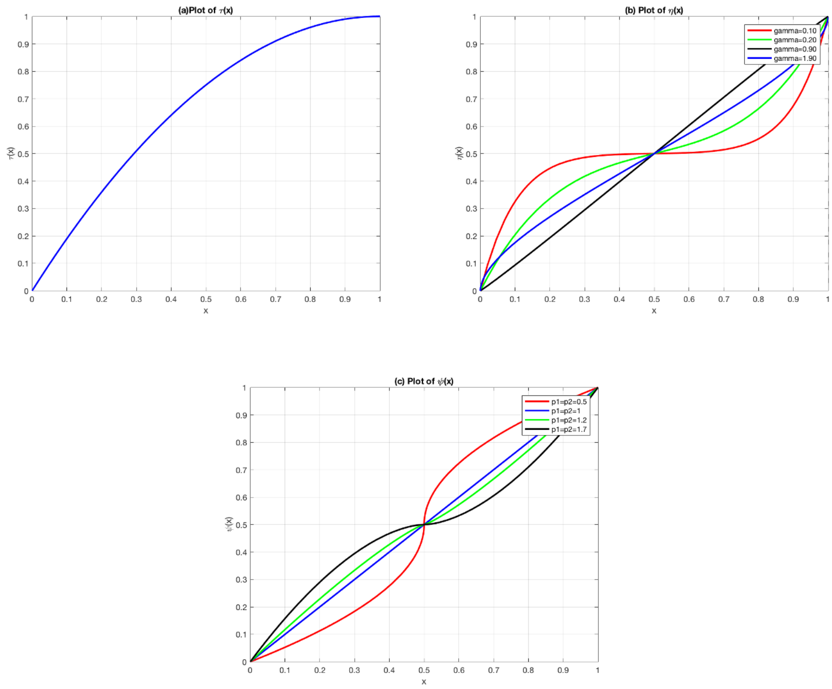


Figure 3.5: Plots of the mapping functions: (a) $\tau(x)$, (b) $\eta(x)$, (c) $\psi(x)$ with $m = n = 0.5$

the same scale factor because they represent different spectral components of the same image. Therefore, the scale factor of each wavelet coefficient should be adapted based on local information reflecting the included details and/or noise. Furthermore, a denoising process can be considered in this step since wavelet subbands contain plenty of trivial coefficients that might represent noise. In [42], the wavelet coefficients scale factor κ_w is computed as follows:

$$\kappa_w = \kappa_s.\gamma_w.\lambda_w \tag{3.11}$$

where γ_w is a locality factor based on the wavelet coefficient w and computed using either the mapping function of Equation 3.4 or the functions proposed in [39, 290]. λ_w is a shrinkage term to suppress the small wavelet noise coefficients which are emphasized by the multiplication of the first two terms. Depending on the type of noise, the value of λ_w can be computed using on the technique in [40] and should be between 0 and 1.

3.6.3 Stereo image enhancement

The evolution of 3D vision laparoscopic systems has revealed new issues related to the endoscopic image enhancement and quality. The first challenge is to adapt the enhancement to the particular characteristics of the endoscopic domain including special textures, colors, artifacts, and dynamic illumination conditions. The second challenge lies in the enhancement of 3D images itself. Indeed, applying conventional 2D enhancement techniques separately on each view does not yield often optimal results as it neglects two main factors:

i) the stereo image characteristics such as the inter-view content redundancies/correlation and depth features (e.g. disparity, depth map);

ii) the perception functioning of the human visual system (HVS) and its binocular vision features (e.g. binocular rivalry and suppression) that have significant impact on the perceived quality of a stereo image, as described in Chapter 2.

Therefore, many researchers tried to address the stereo image enhancement challenge from a different perspective, by exploiting the depth information and binocular vision properties. In [291], the authors conducted subjective experiments to study the relation between three main factors, namely the sharpness, blur and depth. Based on the experimental results and the idea that closer objects of the scene stimulate more the sensitivity of the HVS, *Subedar et Karam.* [291] proposed an enhancement technique that controls the sharpness improvement as a function of the disparity values. In [87] the authors combined depth information and local edges to increase the image contrast. The disparity map is first segmented to distinguish the different objects of the scene. Then, each object boundaries are enhanced depending on its depth level to promote the nearest objects according to the observer perception. This method depends, however, on the accuracy of the disparity map estimation and the segmentation process, especially that the authors deduce the objects of the right view from the segmented objects of the left image using the disparity map. In [177], *Lim et al.* perform a color correction using a similarity based color matching. The spectral content of the image is first decomposed into illumination and reflection using the Retinex theory [160]. A color correction is then performed by matching each corresponding pair of block histograms between the reference and a target image using the structural similarity index measure.

The aforementioned methods, however, do not take into account efficiently the inter-view correlation and content redundancies between the left and the right images. An independent intra-views enhancement can increase the inter-views difference within stereo images, which can lead to visual fatigue and discomfort of the observer since the HVS is very sensitive to the inter-view difference [158]. To address this question, the inter-view unnecessary increase of the illumination intensities is removed in [128] using the binocular just-noticeable difference model (BJND) [333]. The image sharpness is first enhanced using an adaptive unsharp masking technique [247], then an optimization function with additional constraint is solved to suppress the inter-view change exceeding the BNJD threshold. .

3.6.4 Endoscopic image enhancement

After introducing the different notions and basics related to the image processing approach, we can present and discuss the different endoscopic image enhancement techniques proposed in the literature.

In [233, 55], the authors propose a real-time Retinex enhancement technique for gastric endoscopy images. The illumination component is estimated using a variational model algorithm to reduce the computational complexity and, thus, allow the real-time processing. To avoid any-over-enhancement, the estimated luminance is gamma corrected then used to reconstruct the reflectance (color) component. Similarly, *Wu et al.* [324] propose an enhancement framework composed of three main steps. Firstly, adjust the image contrast using multi-scale Retinex (MSR) [124], followed by a gamma correction to avoid any over-enhancement. Second, a brightness diversity process is performed to generate a simulation of several low-dynamic-range (LDR) images with multiple exposures. The resulting images are, finally, fused using a pseudo high dynamic range (HDR) image synthesis process based on bilateral filtering [179] to generate the final enhanced image. As discussed previously, Retinex-based techniques are constrained by the gray-world assumption [68] which violation in endoscopic images (because of the red color domination) can produce a "graying" effect degrading the image quality. Moreover, determining the MSR parameters for the whole image is a challenging task since a specific setting may yield partial enhancement of only some regions of the image.

In a different context of small intestine examination, the authors of [110, 111] propose a color reproduction and enhancement method for wireless capsule endoscopy images. The image is first decomposed into three 2D spectral components using the Fuji Intelligent Color Enhancement (FICE) technique. The image having the maximum entropy is then selected for the color reproduction process. In this step, the chrominance map of the reference RGB image is added to the enhanced spectral gray level image after a luminance/texture matching based on a statistical neighborhood method. In the same vein, a space-variant color reproduction method for endoscopy images is proposed in [112]. The luminance component is first processed using an adaptive sigmoid function having two controlling parameters. Using a texture based transformation, the chrominance of the reference image is adjusted to generate a new chrominance map. Finally, the resulting luminance and chrominance components are converted to RGB color space. Both of these color reproduction and enhancement techniques exhibit some of the vascular and mucosal tissues abnormalities and characteristics as well as pit patterns in polyp/lesion.

As a part of a study about the impact of an enhancement pre-process on features tracking [271], *Selka et al.* propose two simple real-time enhancement methods for endoscopic images. The first method consists in converting the RGB original image to the HSV color space and apply an histogram equalization on the Saturation (S) and Value (V) channels. The second method focus more on enhancing the local contrast to sharpen the image details. A morphological contrast operator based on neighborhood processing is generated by computing the black and white top hat of the image. The latter is then added to the original image to adjust bright regions contrast, while the black top hat is substrated from the obtained output to increase the brightness of dark objects. In a similar work [8], wireless micro-ball endoscopic images are first processed using a moving median filter to highlight the image details. An histogram information correction is then performed to adjust the dynamic range of the image. In a stereo processing context , *Hai et al.* [88] propose an enhancement method for 3D endoscopic images based on contrast limited adaptive histogram equalization. After equalizing the histogram for both left and right views, the image having higher average illumination data is selected as a reference for the color matching step. This process consists in adjusting the color channels coefficients using a mapping function deduced according to the cumulated histograms of both views.

In a context of WCE imaging, other techniques relied on anisotropic diffusion to adjust the images brightness and filter noise while preserving local features such as edges and textures. These methods, however, do not enhance efficiently the chromatic information of endoscopic images. Furthermore, the computational complexity increases as a function of the number of iterations, while degrading the image quality. Table 3.2 summarizes and compare the different literature enhancement techniques for both endoscopic and WCE images. We can notice that, except the recent method of [88], no enhancement algorithm is adapted to stereo endoscopic images and their 3D properties including interview redundancies and depth information. Furthermore, all the enhancement techniques process the endoscopic images in the spatial domain, which is not appropriate if we want to target specific features or level of details. All these discussed points will be further discusses and addressed in the following chapters which present our proposed endoscopic image enhancement methods.

	Enhancement Techniques	Local / Global	Image component	Real-time	Hardware	Limitations
Retinex	[55, 233]	Global	L / R	Y	Y	Computational complexity, heavy processing, hardware architecture
	[324]	Global	L	N	N	Enhancement parameters adapted for specific images, heavy HDR procesing
Color reproduction	[110, 111]	Global	L / C	N	Y	Depends on FICE technology, computational complexity
	[112]	Global	L / C	N	N	
Filtering	[271]	Global	L / C	Y	Y	Noise amplification, does not preserve edges and local details
	[255]	Global	L / C	Y	N	
Anisotropic diffusion	[171, 173, 274]	Global	L / C	Y	N	Image quality degrades as iterations number increases
HE	[271]	Global	L	Y	Y	Over-enhancement
	[8]	Global	L	Y	N	
	[88, 279, 309, 220]	Local : tiles	C	Y	N	Image dependency of the contrast gain limit, noise amplification in smooth regions

Table 3.2: Enhancement techniques for WCE and endoscopic images: Y=yes, N=No, L=Luminance, R=Reflectance, C=Chrominance, B=Brightness

3.7 Conclusion

In this chapter, we presented the different existing enhancement methods for endoscopic images. The first part was devoted to outline the endoscopic imaging technologies which rely on special devices and hardware to enhanced the image quality. Despite its relatively good performance, many challenges and limitations have to be addressed before integrating such technologies in the clinical routine. The challenges include the weak reproducibility, the luck of standards and norms, legal/ethical issues, and the necessity for additional systematic training for the surgeons and the medical stuff to use the devices and interpret the output enhanced signals, which raises questions related to the role of pathologists/endoscopists that need to be revised.

The second part of the chapter investigated the different image processing techniques that allow to enhance endoscopic image quality and exhibit some important features without using any additional devices. After giving an overview of the various image processing approaches and some relevant basic concepts, we presented the different image-processing based techniques for endoscopic image enhancement. As the surgical vision systems are nowadays converging to the third dimension by providing a 3D endoscopic visual feedback, we have also discussed the state of the art techniques to enhance stereoscopic images.

While reviewing all techniques presented in the latter part, we could notice that the increasing understanding of the HVS and some various perception features presented in Chapter 2 has inspired many image enhancement techniques such as edge based contrast enhancement methods [17, 123], the Retinex techniques [160, 124, 125, 144, 178, 304], the frequency-based contrast measures of [299, 300, 298], the enhancement of stereo images based on the depth level of objects within the scene [291, 87], and the enhancement techniques controlling the inter-view differences using the visibility BJND thresholds [128]. Finally, it is worth mentioning that many image enhancement methods rely on the combination of different smoothing or sharpening techniques in both spatial and transform domains [81, 58], which extended some spatial domain techniques such as the UM [184] to the frequency domain.

Chapter 4

2D Spatial-domain image enhancement and assessment

> There is no quality in this world that is not what it is merely by contrast. Nothing exists in itself.
>
> HERMAN MELVILLE

4.1 Introduction

Retinal photo-receptors and ganglion cells play a leading role in determining the sensitivity of the HVS to difference in brightness and colors of the perceived scene [13, 119, 59, 163]. In image processing, the functioning of these cells allowing to distinguish different luminance degrees in a 2D static image describes the well-known contrast sensitivity. The latter measure demonstrated the importance of contrast as an image feature having a significant impact of the overall perceived image quality. Therefore, many contrast enhancement (CE) techniques emphasizing salient images features have been proposed in the literature. In endoscopic imaging, such techniques are relevant to exhibit fine tissues details and abnormalities, which is useful to both enhance the perceived quality for doctors and improve the output of contrast-sensitive subsequent tasks such as features extraction. Image contrast is therefore an important feature for our endoscopic enhancement context as well.

This chapter deals with 2D images contrast enhancement issue from two different perspectives. The first part of the chapter (Section 4.2) focuses on CE assessment to address an investigated need for both contrast evaluation and enhancement research communities: the lack of a benchmark allowing the assessment of different CE evaluation metrics. In the second part of the chapter, i.e. Section 4.3, we propose a 2D endoscopic image enhancement technique improving both the global and local contrast together with a chromatic processing, for stereo-matching and WCE classification.

4.2 Performance assessment of objective quality metrics for CE techniques

Due to the retinal-photoreceptor and ganglion cells properties, the HVS is very sensitive to the contrast of the surrounding observed world. Particularly for images, this psychophysical feature has a significant impact on the overall perceived quality. It is considered generally as the difference of brightness or color between two neighboring regions/points. Due to the context-dependent nature of contrast, various definitions have been proposed in the literature to model this visual property allowing to distinguish objects and determine how sharp image details are. Most of these definitions are mainly based on the ratio between difference and average of luminance intensity values within a window. The idea behind this relation is that a small luminance difference is not noticeable if the average intensity value is high, whereas the same difference effect will increase as the average luminance intensity decrease. As an example of well-known contrast definitions based on this assumption, we can find *Weber*'s and *Michelson*'s contrasts [212, 241, 240, 295].

To enhance the quality of poor or weak contrasted images, many techniques have been proposed in the literature in both spatial and transform domains and for different types of images (i.e. nature, medical, satellite imaging, etc.). The main aim of these algorithms is to ameliorate visual images quality by emphasizing special salient features without degradation the global natural appearance of brightness and colors. Evaluating the resultant enhancement is, however a complex task as it depends on numerous factors and parameters such as image content, naturalness, colorfulness, pleasantness, etc. Since subjective quality assessment is time consuming, requires special devices and following specific evaluation protocols, many efforts have been devoted to design objective image quality evaluation techniques. However, despite the numerous proposed assessment metrics, there is no standard conventional benchmark allowing the evaluation of the different contrast enhancement assessment methods themselves. To overcome this problem, only few studies [282, 312, 185] have been performed.

In [312], the authors investigate the efficiency of degradation/distortion measure metrics in evaluating the enhancement of processed images. These latter are assessed using a subjective test then analysis are performed to determine the accuracy of the used degradation quality metrics in predicting objective scores reflecting the HVS appreciation. In [282], Simone *et al.* perform a psychophysical experiment to evaluate a proposed contrast measure based on multi-scale pyramidal analysis inspired by [261]. The difference with the latter contrast measure developed by Rizzi *et al.* [261] is twofold. First, the local contrast in each pixel is computed using pyramidal difference of gaussians (DOG) instead of the 8-neighborhood mean calculation. Second, the overall contrast is a weighted combination of three different measures deduced from the chromatic channels in the LAB colorspace, instead of using only the luminance channel as performed in [261]. Their study revealed that an assessment based on Region-Of-Interests does not improve the fidelity of contrast measures and that the latter should account for the global image impression. In a more recent work, *Liu et al.* [185] propose a new dataset (CID:IQ) containing reference images, corresponding processed versions with different distortions and the associated subjective scores.

It could be noticed that most of the previously presented databases aim to evaluate image quality assessment metrics designed mainly to distorted/degraded images, without any focus on contrast enhanced images. With the large number of CE techniques for different application domains, developing a dataset incorporating subjective ground-truth evaluation data together with original and contrast enhanced images can be useful for the image-processing research community focusing on enhancement techniques, especially CE. Therefore, as a part of a collaboration with *King Fahd University of Petroleum and Minerals* (*Saudi Arabia*), we proposed a database containing: i) original images, ii) associated contrast enhanced versions using different conventional CE techniques, iii) corresponding subjective evaluation scores, and iv) objective assessment data. To achieve this, we performed the following steps:

- select our database images according to specific criteria to cover different content types, chromatic information, and level of details,
- prepare and perform a psychophysical experiment to collect subjective evaluation scores which will be considered as ground-truth data reflecting the HVS judgment of perceived images quality,
- enhance each image of our constructed database using different conventional CE techniques,
- assess the quality of enhanced images using various CE evaluation metrics,
- analyze the fidelity of CE assessment metrics based on the correlation between obtained objective scores and the ground-truth data provided by the observers ranking.

Using the correlation results, we can comment on the consistency of the tested CE assessment metrics with the HVS appreciation of the perceived contrast enhanced images. The database can useful as well to evaluate newly proposed CE techniques and metrics. In the following, we detail the different steps mentioned above.

4.2.1 Contrast enhancement techniques

With the great attention devoted to study and improve 2D/3D images contrast, numerous contrast enhancement techniques have been proposed in the literature. The existing CE algorithms can be categorized into six main classes namely, pixel-based, mathematical morphology-based, histogram-based, transform-based, and HVS-based techniques. In our work, one representative method is selected from each of the six previously mentioned categories of CE techniques. The selection includes the edge-based contrast enhancement method (EBCE) [17], contrast-limited adaptive histogram equalization (CLAHE) [335], global histogram equalization (GHE), discrete cosine transform based algorithm (DCT) [224], Top-Hat transformation (TOPHAT) [225], and Multi-Scale Retinex (MSR) [124, 253, 211]. Since the aim of this work is to study CE assessment metrics rather than evaluating or improving CE algorithms, the default parameters of the techniques listed above are used in our experimental tests without any adjustment.

4.2.2 Contrast enhancement assessment metrics

In spite of the increasing number of CE techniques, only few efforts have been devoted to assess the contrast enhancement quality. This can be explained by three major reasons: i) the fact that each CE method is adapted to a specific application domain, ii) the subjective judgment of the perceived images that differs from a person to another, and iii) the difficulty to find objective measures accounting for high level vision features and interaction with low-level image analysis when evaluating the perceived quality of enhancement algorithms [16]. In the same vein, determining the most appropriate and significant visual features to be considered when designing a quantitative enhancement measure is a challenging task. This can explain the reason why the few proposed measures for CE evaluation are not well correlated with the perceived quality of the enhanced images. Among the various CE measures, we have selected for our study six well-known metrics, namely the absolute measure of enhancement (AME) [235], absolute mean brightness error (AMBE) [37], edge content (EC) [265], second derivative like measure of enhancement (SDME) [235], measure of enhancement (EME) [235], and image enhancement measure (IEM) [121]. The mathematical expressions of the these metrics is provided by Table 4.1 where $I_{ij}^{\max}$, $I_{ij}^{\min}$, and I_{ij}^{center} denote respectively the maximum, minimum, and center pixel intensity values within a block centered in spatial coordinates (i, j); I_r and I_e designate respectively the original and enhanced images; M and N are the height and width of the images; L refers to the number of grayscale levels; and ϵ is a small positive constant to avoid the division by zero.

The AME and EME metrics are inspired respectively from *Michelson*'s and *Weber*'s contrast measures, which are not designed for natural images.

CE assessment no.	Mathematical formulation
1	$EME = \frac{1}{B_1 \times B_2} \sum_{i=1}^{B_1} \sum_{j=1}^{B_2} 20 \ln \left(\frac{I_{ij}^{\max}}{I_{ij}^{\min} + \epsilon} \right)$
2	$AME = \frac{-1}{B_1 \times B_2} \sum_{i=1}^{B_1} \sum_{j=1}^{B_2} 20 \ln \left(\frac{I_{ij}^{\max} - I_{ij}^{\min}}{I_{ij}^{\max} + I_{ij}^{\min}} \right)$
3	$SDME = \frac{-1}{B_1 \times B_2} \sum_{i=1}^{B_1} \sum_{j=1}^{B_2} 20 \ln \left(\frac{I_{ij}^{\max} - 2I_{ij}^{center} + I_{ij}^{\min}}{I_{ij}^{\max} + 2I_{ij}^{center} + I_{ij}^{\min}} \right)$
4	$EC = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \lvert \Delta I\left(i,j\right) \rvert$
5	$IEM = \frac{\sum_{i=1}^{B_1} \sum_{j=1}^{B_2} \sum_{n=1}^{8} \lvert I_{i,j}^{e,c} - I_{i,j}^{e,n} \rvert}{\sum_{i=1}^{B_1} \sum_{j=1}^{B_2} \sum_{n=1}^{8} \lvert I_{i,j}^{r,c} - I_{i,j}^{r,n} \rvert}$
6	$AMBE = \lvert \mathrm{E}\left(I_r\right) - \left(I_e\right) \rvert$

Table 4.1: Mathematical expressions of the CE evaluation measures used in our experimental work. $I_{ij}^{\max}$, $I_{ij}^{\min}$, and I_{ij}^{center} denote respectively the maximum, minimum, and center pixel intensity value with a block centered in spatial coordinates (i, j); I_r and I_e designate respectively the original and enhanced images; M and N are the height and width of the images; L refers to the number of grayscale levels; and ϵ is a small positive constant to avoid the division by zero.

Indeed, the Min/Max operators incorporated in these measures may over-estimate or under-estimate the contrast in noisy image regions. To overcome this problem, both AME and EME are computed based on small windows and the final overall score is deduced as a combination of all blocks-based measures. We can notice also that both metrics are very sensitive to noise due to the Minimum and Maximum operators incorporated in their mathematical expressions. In order to reduce the sensitivity to noise, the center pixel intensity value within each block is integrated in the estimation of the SDME score. The default block size is set to either 3×3 or 5×5. In practice, all EME, AME and SDME are computed on non-overlapping windows. The main difference is the blocks size. Indeed, both EME and AME scores are calculated based on 8×8 blocks while SDME is computed using a 5×5 window.

The EC metric is based on the local image gradient. In the mathematical expression of EC, ΔI refers to the gradient of the gray-scale image computed using a Sobel operator. When performing a contrast enhancement, it is generally recommended to preserve the global original brightness of the input image. Based on this idea, the AMBE metric describes the deviation of the mean intensity value before and after the enhancement to measures the brightness reservation (BP) related to CE techniques. In its mathematical formulation, E designates the statistical expectation. Thus, lower AMBE scores implies a better brightness preservation, consequently a better enhancement. A very low score, however, is linked to a weak CE performance. Generally, a median value of the intensity levels range indicates a good brightness preservation and enhancement. It is important to mention that BP in this case does not imply the preservation of image naturalness. The IEM measure computation is performed on non-overlapping blocks sized 3×3. The overall score is calculated as the ratio of sum of absolute differences within each block between the center pixel and its neighbors in the same window of the output enhanced images and the corresponding block in the input image. This absolute difference of intensity values between the center pixel and its block neighbors within each window from the original and the enhanced image describes the sharpness and contrast change. Note that higher scores of EC, IEM and RC, and lower values of AME, AMBE and SDME indicate a better enhancement.

4.2.3 Experimental setup

The final aim of creating and processing images is generally to display the output signal to the visual sensory apparatus of observers, which is the ultimate judge of the perceived quality. Since the HVS is very sensitive to image contrast, the accuracy and fidelity of CE assessment metrics is measured in terms of correlation with the HVS appreciation. Therefore, we conducted a psychophysical experiment to collect ground-truth data containing the subjective judgment of each enhanced image using the CE techniques presented in 4.2.1.

One of the most important parameters in such subjective experiments is the creation of images database. In order to cover a wide range of various image types with different content, contrast variation, and chromatic distribution, three measures are used to select our dataset images namely Colorfulness (CF) [320], Spatial Information (SI) [320], and Global Contrast Factor (GCF) [204]. The initial dataset contained 30 color images having the same spatial resolution (512×512) and selected either from conventional datasets commonly studied by the image/signal processing community or from captures made by our lab members. The six CE techniques presented in section 4.2.1 were then applied to generate the enhanced versions for each original image of the constructed dataset.

The subjective tests were performed in a specially designed test room in the laboratory of Information Processing and Transmission (L2TI) at Université Paris 13. To allow the observer to focus more on the perceived images and the different global and local contrast features, the test room was kept in low ambient illumination level as in [185]. The images were displayed on an *EIZO ColorEdge CG242W* screen with (1920×1200) resolution (24.1 inch). The latter was calibrated using the *Eye-One Match 3* by *X-rite* and the *i1*-spectrophotometer which profiles the monitor. The final screen calibration parameters are summarized in Table 4.2.

Display calibration parameters	Parameter value
Luminance	119 cd/m2
Color-space	sRGB
Color-temperature	6500K
Gamma -correction	2.2
Frame-rate	60Hz
Contrast	80

Table 4.2: Display calibration parameters

A graphic interface, illustrated in Figure 4.1, was specially designed for the experiment to display each original image centered horizontally between two random contrast enhanced versions on a grays-scale background color. Such chromatic tone neutralize the visual effect of the surrounding background and focus the observer attention on the perceived images. Additionally, the spatial screen disposition of the reference image between the enhanced versions facilitates the after-effects analysis for the HVS. Fifteen expert and non-expert observers having different gender, age, and with normal or corrected-to-normal visual acuity were invited to evaluate the displayed images quality using the pairwise preference based ranking protocol (Condorcet method). The distance from the monitor was fixed to 80 cm with no possibility to go further due to the space limitation using the test room wall, which is also painted in gray. Figure 4.2 illustrates the experimental settings namely the test room, the screen, the graphic interface of the subjective test, and the observer's position.

None of the observers was informed about image quality features or the contrast definition. They were asked to simply give their preference based on the perceived quality by choosing the enhanced image which appears better than the original version according to their visual subjective judgment. In case of equivalent or equal perceived quality degree, the observers were allowed to assign the same rank for both processed images. The *Kendall*'s coefficient of concordance [280] was used to control both the ranking significance and the agreement of all observers on the attributed rankings. Based on our primary statistical analysis, we noticed a significant ranking disagreement on four images. These latter were considered as outliers and, thus, were discarded from our images dataset for the rest of results analysis.

4.2.4 Results and discussion

From the ground-truth data obtained by the subjective test, we have extracted the preferences scores indicating how many times an CE technique

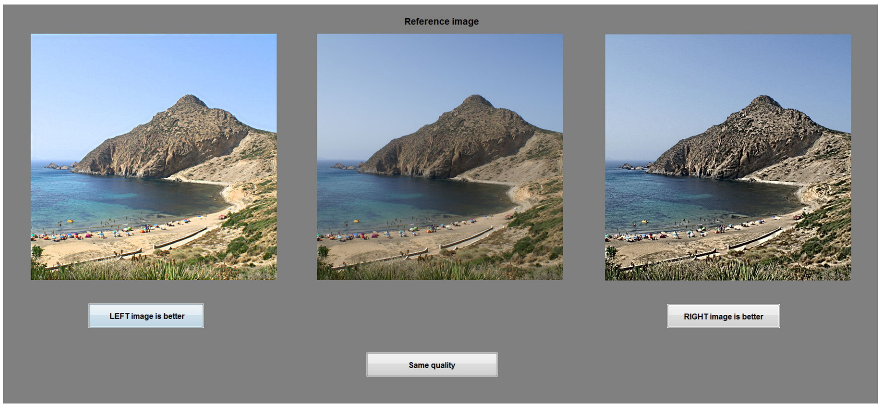


Figure 4.1: Graphic interface of the subjective test

Figure 4.2: Subjective test settings

was preferred over the others. Table 4.3 presents the sum of preference scores per CE method. We can notice that the observers prefer the enhancement quality of respectively CLAHE, RETINEX and EBCE algorithms while TOPHAT and GHE techniques are selected only 200 times combined (4% of all preferences). The objective scores of the six CE assessment metrics applied on each enhanced version of the remaining 26 dataset images using the six CE methods are also computed and averaged (see Table 4.3). According to the objective scores, the images enhanced using GHE and TOPHAT present the best quality, which is in contrast with the HVS judgment obtained by the objective preference rankings.

	TOPHAT	**EBCE**	**CLAHE**	**DCT**	**MSR**	**GHE**
AMBE	1.27	0.81	4.08	47.59	22.26	24.77
SDME	-0.24	-0.12	-0.06	0.02	-0.10	-0.15
AME	-0.38	-0.20	-0.08	0.08	-0.10	-0.31
IEM	2.29	1.72	1.18	1.31	1.57	1.81
EME	1.32	0.49	0.20	-0.14	0.43	1.32
EC	1.07	0.43	0.17	0.27	0.53	0.72
Subjective ranking	150	1388	1422.5	560	1410	50

Table 4.3: Summary of objective CE metrics scores and subjective rankings for the different studied CE techniques

To derive the correlation between objective measures and the ranking based on the HVS appreciation, the Spearman Rank Order Correlation Coefficient (SROCC) is computed between the subjective ranking and the objective scores of each CE technique. SROCC is a non-parametric measure describing the statistical dependence between two ranking data. The final SROCC score ranges in $[-1, 1]$ with a stronger correlation as the value tends to 1. In the opposite case, a final SROCC score value tending to -1 implies weak correlation or no correlation at all. Given I_i an original image $(i = 1, 2, .., n_I)$ and $I_{i,j}$ its processed version using an enhancement technique E_j $(j = 1, 2, .., n_J)$, the SROCC is computed as follows:

$$\rho_{i,k} = 1 - \frac{6d_{i,j}^2}{n_{J(\ n_I^2 - 1)}}, \ i = 1, 2, .., \ n_I \tag{4.1}$$

where $d_{i,j}$ denotes the difference between subjective ranks preferences and objective scores obtained by assessment metric k for each image as illustrated in Figure 4.3. Note that for our case $n_I = 26$ and $n_J = 6$. Median values of SROCC scores related to each metric are finally calculated as presented in Table 4.4. Based on Figure 4.3, we can clearly observe that AME and SDME metrics yield the best correlations with respect to the subjective ranking. The negative EME metric correlation can be explained by its high scores assessing the images processed using GHE and TOPHAT, which result in the worst image quality according to the observers preference. Finally, we can notice also from the SROCC correlation values in Figure 4.3 that most of the CE assessment metrics are not well correlated with the HVS judgment of the perceived quality, which implies the need for more extensive studies providing better fidelity CE assessment scores.

Metric	**SROCC** (ρ)
AMBE	-0.2571
SDME	0.4000
AME	0.4286
IEM	-0.3143
EME	-0.4000
EC	0.2857

Table 4.4: Median SROCC scores for the studied CE assessment metrics

4.3 Enhancement of endoscopic and WCE images

In order to prevent the development of cancer, it is of extreme importance to detect tissue abnormalities such as lesions, bleeding, inflammations, and polyps which might appear in the early stage of the disease. For instance, in a colon cancer context, colonoscopy is a conventional technique which allows early pre-cancerous pathologies detection by inserting a long thin tube (i.e. colonoscope) into the rectum. Despite its simplicity and relative efficiency, this technique suffers from many limitations such as the labor intensive nature of the task requiring a time-consuming process that can yield discomfort of the patient.

Wireless capsule endoscopy is one of the most recent diagnosis technological innovations that substituted successfully the conventional colonoscopy according to several studies [199, 227]. This technique is based on a camera sized and shaped similarly to a pill swallowed by the patient. Once inside the human body, the capsule starts capturing and transmitting wirelessly the endoscopic scene frames to a data server. To perform the diagnosis, the images should be firstly downloaded then examined separately by the doctors/physicians to analyze and determine the abnormalities, which implies a manual time-consuming process. To improve the diagnosis operation, several automatic abnormalities detection and classification techniques have been proposed in the literature [199, 165, 11, 172].

The specifications of the capsule endoscopy acquisition process beside the endo-humain domain characteristics disclosed, however, new challenges

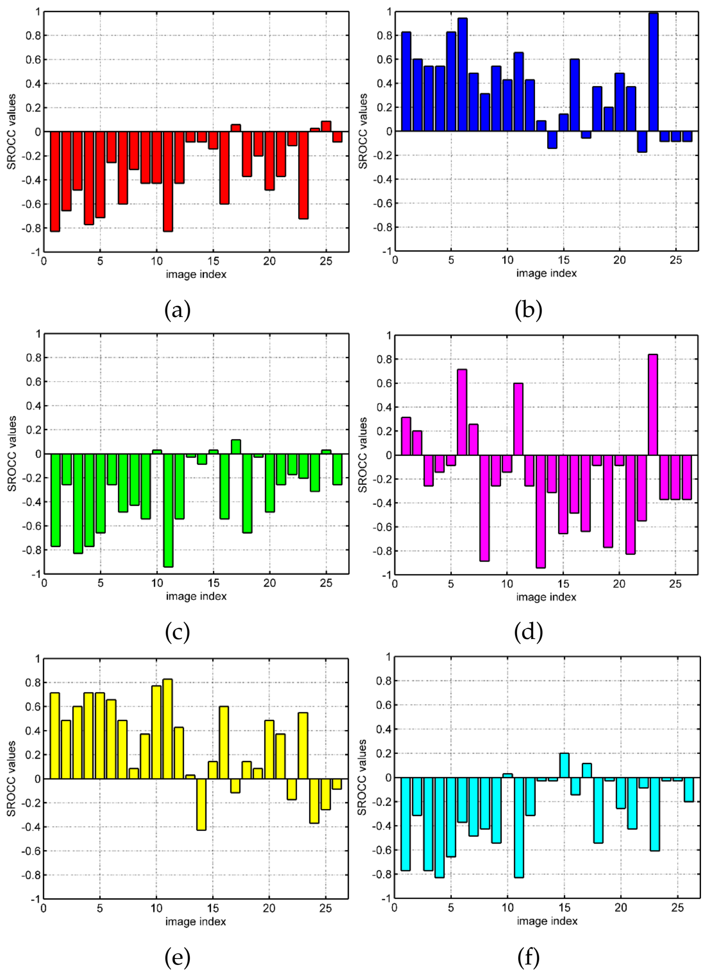


Figure 4.3: SROCC plots illustrating the correlation between CE assessment metrics scores and subjective preference rankings

and artifacts related to endoscopic image quality. Indeed, the motion of the patient and the navigation of the capsule may cause a blurring effect in some frames. Furthermore, there is no ambient light in the bowel except the on-board light source which obey the capsule motion, yielding dynamic illumination conditions with wide dark areas and artifacts such as specular reflections and vignetting [198].

Therefore, performing a processing step on the captured WCE images is very critical for both: i) improving the visual quality of the displayed images for the doctors; and ii) exhibiting the inner subtle structures and tissues details, which represent important image characteristics for the detection and classification processes based mainly on features detections/matching. The latter task can be linked as well to the 3D liver reconstruction and registration task (illustrated by our research application-domain diagram in Chapter 1), which requires features detection and matching between the left and right views of endoscopic images. Therefore, we present in this section an enhancement method that processes the sharpness, color distribution and brightness of both WCE and stereo endoscopic images.

4.3.1 Proposed enhancement technique

Color naturalness [328, 52] is a very important image property that needs to be carefully considered when designing an enhancement technique. Indeed, dramatic modifications of the color distribution may confuse the observer interpretation about the perceived scene. Particularly in an endoscopic diagnosis or treatment context, tissues and organs colors are key elements to determine and detect abnormalities which are often symptoms of dysfunctions or diseases, namely cancer. For instance, a color variation within a tissues may reveal neoplastic or inflamed tissues. Similarly, inverting the color appearance of color-sensitive filed/point imaging technologies seen in Chapter 3 (e.g. green and magenta colors modification in AFI-generated images) can confuse doctors and miss-lead their diagnostic interpretation. Therefore, preserving the color naturalness is crucial when enhancing endoscopic images since it can affect the medical work and, consequently, compromise the patient safety.

Retinex based methods, such as the iterative or center/surround algorithms [75, 253], are one of the most popular processing techniques which account for color constancy, as discussed in Section 3.6. The heavy processing of these techniques can lead, however, to several undesired effects such as amplifying image artifacts, changing the color temperature, or giving a false impression that other light sources exist in the scene. In [35], the authors address these limitations by proposing a one-filter Retinex [211] based rendering technique for natural color images. In spite of its simplicity and stability with different types of scenes, the authors reported poor performance for medical and unnatural images. Additionally, the rendering of the method does not have a significant impact on image sharpness and local image details, which are very important features for our endoscopic imaging context. The latter problem was addressed later in [36] by incorporating a high-frequency components enhancement step based on the multi-channel decomposition of Cortex Transform [50].

Inspired by the natural rendering of color image based on retinex (NRCIR) [35], we propose an enhancement technique for WCE/endoscopic images combining One-filter Retinex [211] for brightness adjustment, edge-based contrast enhancement [17] for local CE, and histogram rescaling [210]. The processing targets three main image components namely, brightness, color distribution, and local contrast. The aim of the proposed method, beside improving the images visual quality, is to study the effect of a pre-processing step on endoscopic stereo matching and abnormalities classification in WCE images. Figure 4.4 illustrates the block diagram of the proposed enhancement technique, which is composed of nine steps. In the following paragraphs, we will describe the different tasks composing the enhancement pre-processing.

4.3.1.1 Global tone mapping

This step depends on a key-image value representing the dominant illumination intensity level in the image. Such information gives an idea about the overall image brightness, which is a useful data especially to adapt the mapping according to whether the processed image is bright or include wide dark regions. In [306, 259, 152, 210], the authors approximate the key-image value as follows:

$$key = e^{\sum_i \sum_j [\log(L(i,j))+\epsilon]/M \times N} \tag{4.2}$$

where $L(i,j)$ denotes the luminance component pixel intensity value at spatial coordinates (i,j), M and N designate the width and height of the input original image I, and ϵ is a constant to avoid computing the $\log$ of zero. Note that the darker an image is, the lower its corresponding key value gets.

The mapping can be performed using any of the functions presented in Section 3.6.1, namely the power-low transformation. Particularly, the gamma-correction seems to be very useful for endoscopic images to adjust the brightness in wide dark areas since it maps narrow ranges of low intensity levels to larger output range values. The significant mapping gain of low intensity levels by gamma-correction can, however, expose unnoticeable imperfections especially for compressed images. This is the case for WCE frames which are compressed during the wireless transmission process. Thus, applying a gamma-correction mapping can reveal the ringing

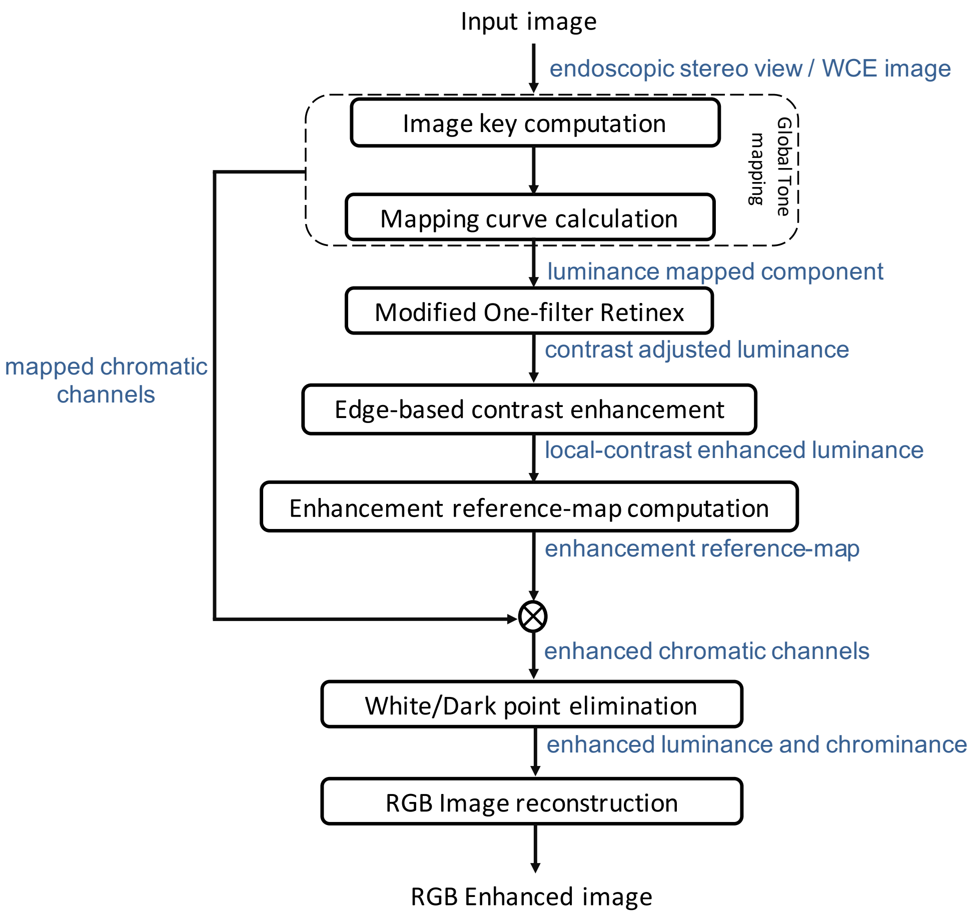


Figure 4.4: Flow diagram of the proposed enhancement method

and blocking artifacts of the lossy coding, which degrade the visual image quality and adversely affect the features detection and classification of abnormalities. Therefore, in our work, we use a mapping function having a circular curve to avoid over-enhancement of WCE images. In order to automatically adapt the mapping to each image brightness level and content, the radius of the circular curve is computed as a function of the image-key value as in [35]. Thus, the radius is calculated as follows:

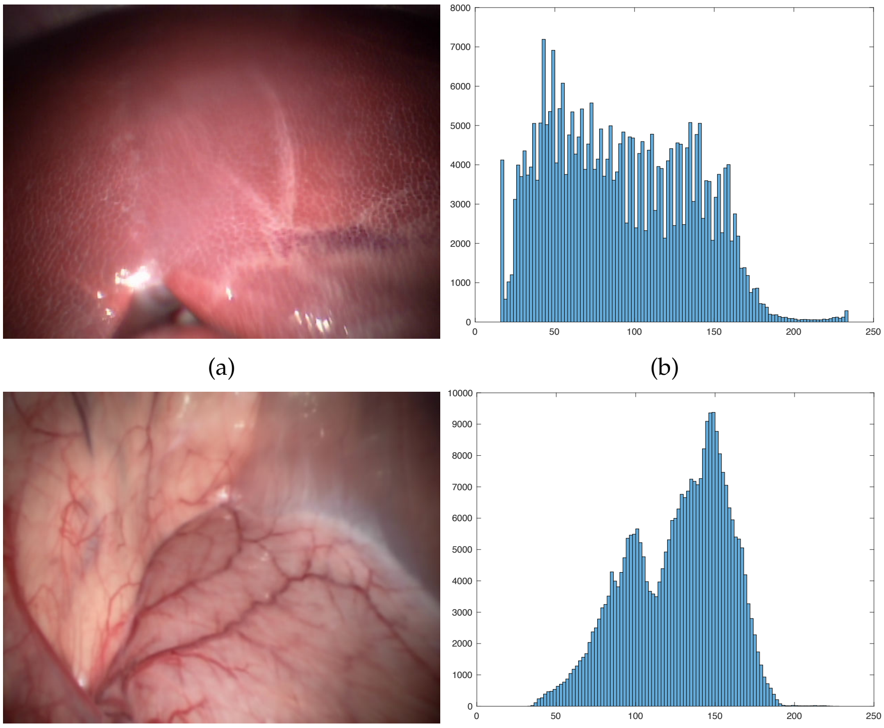
$$r = \begin{cases} 3 \times \log\left(\frac{key}{10} + \epsilon\right) & if\ key \leq \alpha_{dark} \\ 3 \times \log\left(10 - \frac{key}{10} + \epsilon\right) & if\ key \geq \alpha_{bright} \end{cases} \tag{4.3}$$

where $\epsilon > 0$ designates a constant added to avoid the logarithm of zero problem and α_{dark} α_{bright} denote two constants determined based on the brightness level of the dataset images, which is reflected by its key values. Figure 4.5 shows two sample images with its corresponding histograms and key values. It can be noticed that the image including wide dark regions and unbalanced brightness has a low key value, in contrast with the image of Figure 4.5(c) which presents a high brightness level, consequently a high key value. Additionally, we can also notice the correlation between image histograms and the corresponding key values in terms of brightness level. The mapping is then performed according to Equation 4.4 where I_{map} denotes the mapped output image, I_{orig} is the original input image, (i_0, j_0) refer to the spatial coordinates of the mapping circle center, and α_{dark} α_{bright} are deduced empirically based on our dataset images to the values 80 and 140 respectively.

$$I_{mapped} = \begin{cases} i_0 + \sqrt{r^2 - (I_{orig} - i_0)^2} & if\ key \leq \alpha_{dark}, \\ j_0 - \sqrt{r^2 - (I_{orig} - i_0)^2} & if\ key \geq \alpha_{bright}, \\ I_{orig} & otherwise, \end{cases} \tag{4.4}$$

4.3.1.2 Brightness and contrast enhancement using One-filter Retinex and edginess information

This step consists in enhancing the global contrast of the luminance component, which is computed generally as the ratio between center pixel intensity value and its surrounding background information. In [159, 206], the authors demonstrated, however, that the perceived contrast depends not only on the direct neighbors but also on the local contrast of distant surrounding pixels. Based on this idea, the MSR [124, 253, 211] uses the following mathematic formulation to derive a filter allowing the estimation of the local background information for each pixel:



(a) (b)

(c) (d)

Figure 4.5: Correlation between images histogram and the corresponding key values: (a) $keyvalue = 79$, (b) $keyvalue = 121.39$

$$L_{mask} = L_{input} * F_{MSR}, \tag{4.5}$$

$$F_{MSR} = \sum_{i=1}^{\lfloor \log 2(K) \rfloor} e^{-\frac{\left(x^2+y^2\right)}{2^{2i}}}, \tag{4.6}$$

$$K = \frac{\max\left(size\left(L_{input}\right)\right)}{8}, \tag{4.7}$$

where F_{MSR} is the filter mask, L_{input} is the input luminance component, (x, y) denotes the spatial coordinates of the pixel mask window center which values vary from 1 to K, and L_{mask} is the mask image containing the estimated background information. Once the latter is computed, the Retinex image can be obtained by ratio between the input luminance channel L_{input} and the image mask L_{mask}, as follows:

$$L_{retinex} = \frac{L_{input}}{L_{mask}} \tag{4.8}$$

Note that the mask image intensity values, which represent the denominator of Equation 4.8, have a significant impact on the output Retinex image. Indeed, as a region in the mask image gets darker, its pixel intensities tend to lower values, which yields significantly higher pixel intensity levels in the Retinex image (i.e. over-enhancement). For heavily compressed images, this implies as well that the blocking effect in dim regions will be emphasized and more visible in the output Retinex component. On the other hand, the high intensity values in bright regions of the mask image will result in relatively small pixel intensity levels after the ratio operation. Visually, the impact of such transformation is a graying effect in the concerned areas illustrated with a washed-out appearance, which is also known as the Halo effect.

To tackle this undesired over-enhancement problem, *Chen et Beghdadi.* [35] process the denominator image mask component with a logarithm function to increase the intensity values in dark regions and reduce it in bright areas (i.e. dynamic range compression). In our work, we adopt this solution to avoid any over-enhancement that can hinder features extraction either for stereo-matching or WCE abnormalities classification. Thus, Equation 4.8 can be re-written as follows:

$$L_{retinex} = \frac{L_{mapped}}{\log_{10}\left(L_{mask} + \epsilon\right) + \epsilon} \tag{4.9}$$

where L_{map} is the globally mapped luminance component from the previous step in Section 4.3.1.1 and $\epsilon > 0$ is a constant to avoid both the division and logarithm of zero problems. The following sub-task consists in more exhibiting the tissues details and boundaries by exploiting the edginess information and the linked local contrast measure defined in [17]. Further description and discussion of this method will be provided in Chapter 5 where we adopt this technique for 3D endoscopic images. Enhancing the regular and micro tissues edges will improve and facilitate the local features detection process which is very sensitive to image gradient.

4.3.1.3 Reference-map computation and color enhancement

As explained previously, color perception is a key property in endoscopic and WCE images since it can assist doctors in their diagnosis/treatment tasks. Therefore, it is important to enhance chromatic image components accordingly to the luminance processing. One simple idea to perform this is by applying the previous contrast enhancement step separately on each RGB component. Such processing of the very inter-dependent RGB color channels yields, however, often hue shift and unnatural color rendering. The reason for this result is the high correlation between Red, Green, and Blue image components in RGB color-space since each channel contain brightness information of the captured scene. Therefore, before applying the previously presented contrast enhancement, we convert the RGB color-space of the processed image to the L*a*b color-space in oder to allow a decorrelation and separation of the brightness data from chromatic information. Only the luminance channel (L) is then selected for contrast processing of Section 4.3.1.2.

In order to perform an balanced chromatic processing, it is important to have an enhancement information that serves as a reference-map allowing a proportional color adjustment to the luminance channel enhancement. Therefore, once the contrast processing of the previous step is achieved, the enhanced luminance component is used to compute the reference map as follows:

$$Ref_{map} = \frac{L_{enhanced}}{L_{map} + \epsilon} \tag{4.10}$$

where $L_{enhanced}$ designates the $L_{retinex}$ luminance processed by the edge-based contrast enhancement in the previous step, and L_{map} denotes the mapped luminance component after step 1 (i.e. luminance of the mapped image directly after color-space conversion from RGB to YCbCr). Once Ref_{map} is computed, the mapped blue and red difference chroma components C$_i$ $(i = Cb, Cr)$ from step 1 are processed accordingly to the luminance channel enhancement using the reference map as follows:

$$Cenh_i = C_{mapped} \times Ref_{map}, \; (i = Cb, Cr) \tag{4.11}$$

where C_{mapped} is the chromatic component mapped in Section 4.3.1.1. An histogram clipping operation is then applied, such explained in [183, 168], to remove the possible extreme white/black points that might occur due to the ratio operations of Equations 4.9 and 4.10. Instead of simply selecting the maximum intensity value, the highlights are determined based on a set of white pixels to estimate the intensity range that covers the bright points. To remove these dark/bright intensities, the ranges below 1% and exceeding 99% are clipped-off from the enhanced image histogram. Thus, we attenuate the effect of possible existing specular reflections in endoscopic images and we avoid any visible over-enhancement artifacts.

4.3.1.4 Experimental results and discussions

To test the enhancement impact on both stereo-matching of 3D endoscopic images and WCE abnormalities classification, the previously presented processing steps are applied on two different datasets. The 3D endoscopic dataset (3D-END) is composed of two stereo videos provided by [221]. The first video is characterized by a static laparoscope and liver deformation due to respiration (0.15 GB each 26 second) while the second video includes a dynamic configuration with motion of both the stereoscope and liver (0.36 GB each 3.5 minutes). The WCE images dataset (2D-WCE) is provided by the *Intervention Center* (IVS) at *Oslo University Hospital* with a tag file including the manual classification labels of the frames to assist the automatic classification process.

The enhancement performance evaluation is based on three main criteria:

- the visual quality of the enhanced images compared to the original ones,
- the number of detected feature points before and after the enhancement,
- increasing the number of features is not always useful since their spatial distribution within the image is also important (i.e. not useful to detect more points that are concentrated in a small spot). Additionally, in a stereo matching process, many of the detected features in each view may not match with their corresponding interest points in the other view, thus, get discarded. Therefore, another measure is added to better assess the efficiency and the usefulness of the enhancement in terms of abnormalities classification improvement. For 3D endoscopic images, the number of stereo-matched features will be calculated. The features distribution and the usefulness are further investigated in WCE by computing the classification ratio before and after the enhancement.

We consider the natural rendering of color image based on Retinex (NRCIR) [35] and the Edge-based contrast enhancement (EBCE) [17] techniques as a baseline for comparison since they cover both chromatic and contrast/sharpness enhancement. Additionally, since histogram equalization techniques, namely the contrast-limited adaptive histogram equalization method (CLAHE) [335] has been widely used to enhance the quality of endoscopic images [88, 279, 309, 220], we consider CLAHE together with EBCE and NRCIR for the performance comparison of our processing.

Figure 4.6 illustrates the enhanced images of a left stereo endoscopic view from the 3D-END 4.6(a) dataset using CLAHE (Figure 4.6(b)), EBCE

(Figure 4.6(c)), and the proposed technique (Figure 4.6(d)). Figures 4.7 and 4.8 show the enhancement performance on sample images from the 2D-WCE dataset. For all figures, we can notice that CLAHE relatively over-enhances the images despite improving the contrast. Additionally, the chromatic information is slightly improved but a small graying effect covers the tissues due to over-enhancement. The latter can be also visible by the blocking effect related to the JPEG compression of the WCE images, which was hidden before the CLAHE processing. The EBCE technique, on the other hand, improves the local contrast without any clear brightness or color enhancement and may yield overshooting or over-enhancement (e.g. halo effect). As for the performance of the proposed method, we can notice that the processing adjusts well the image brightness, contrast and chromatic channels without changing the original tonal distribution. Additionally, the liver structures/tissue are more exhibited without introducing any over-enhancement. The blur effect in the original image due to the patient and camera-pill motion is also reduced thanks to the local contrast enhancement. Similarly, for 2D-WCE dataset images, the proposed method enhances well the image brightness reducing the dark areas and exposing the inner structures of the bowel wall (i.e. fine vessels and veins). As for the chromatic processing, the colors are clearly more saturated while preserving the naturalness appearance of colors. Although that a JPEG compression have been applied to the WCE images during the wireless transmission process, no side artifacts such as blocking or halo effects are noticed in the enhanced images using our proposed method.

In terms of stereo matching performance, Tables 4.5 and 4.6 show that the number of detected and matched points has increased considerably after each enhancement technique. Compared to the separated contrast processing (i.e. CLAHE method) and sharpness enhancement (i.e. EBCE technique), the best improvement of features detection and matching is provided by our proposed method which combines chromatic processing together with both global and local contrast enhancement. This features stereo matching improvement should yield a better reconstructed 3D liver model. However, since we don't have an optimized operational 3D reconstruction workflow to confirm the impact of the features stereo matching improvement after the enhancement, the efficiency of the detected interest points and their distribution is further investigated using the WCE classification process which is very sensitive to these parameters. The well-known scale-invariant feature transform (SIFT) technique [187] is used to extract WCE images features due to its robustness to scale and rotation variations. Each frame is decomposed into small patches on which SIFT algorithm is applied to generate a collection of descriptor vectors having the same size (128). To optimize the accuracy of feature points descriptors computation, the SIFT is combined with "bag-of-words" model (BOF) [250] which is a popular visual descriptor designed initially for documents classification. It consists in counting the words frequencies in a text, which is equivalent to making an histogram of the document vocabulary.

In our context dealing with abnormalities classification of WCE images, the bags of features built using the descriptors of SIFT points are clustered using the *k-means* algorithm [193, 22] to reconstruct the visual vocabulary. The latter serves to distinguish two types of images: i) abnormal WCE frames containing an inflammation or polyps and ii) normal image that are free from the cancer lesions. In the training process, a machine learning (ML) approach based on the support vector machines (SVM) algorithm is used with two inputs: a vector containing the training dataset class labels, and another vector including all the training dataset SIFT descriptors. Once the training is finished, the obtained abnormalities classification model is tested on our WCE database using the *IVS* tag file including the manual classification labels which are considered as a ground truth classification data. Table 4.7 shows that the classification rate has increased after the enhancement to reach for instance 97% for the inflammation images with 8.5% improvement. We notice also for both polyps and inflammation cases that the classification rate improves as a function of the number of training dataset images. This is, actually, the case when the training is performed separately for polyps and inflammation datasets to improve the descriptors computation for each abnormality. However, when increasing the number of images on a training dataset containing both abnormalities frames, the classification model accuracy decrease, which affects accordingly the classification rate. This improvement in terms of WCE abnormalities classification demonstrates the usefulness of increasing the number of detected features points after the enhancement and comment on their spatial distribution (i.e. whether they are compact or well distributed to allow a better 3D liver model reconstruction and/or abnormalities classification), which proves the usefulness and the efficiency of our proposed enhancement technique.

Detected points	**Without enhancement**	**CLAHE**	**NRCIR**	**EBCE**	**Proposed enhancement**
All points	84	3058	3189	3617	3693
Stereo-matching	14	160	250	216	201

Table 4.5: Matching points detection/matching results for 3D-END dataset after the enhancement

Detected points	**Without enhancement**	**CLAHE**	**NRCIR**	**EBCE**	**Proposed enhancement**
All points	217	2027	2445	2494	2918
Stereo-matching	19	60	128	52	86

Table 4.6: Matching points detection/matching results for 2D-WCE dataset after the enhancement

Datasets	**Number of Images**	**Classification rate**		**Improvement**
		pre-enhancement	**post-enhancement**	
Normal VS inflammation	200	91.3%	98%	+ 6.7%
Normal VS polyps		98.25%	99%	+ 0.8%
Normal VS inflammation	300	88.5%	97%	+8.5%
Normal VS polyps		89%	95%	+6%

Table 4.7: WCE image classification results after a pre-processing step using our proposed method

4.4 Conclusion

In this chapter we have studied the image contrast enhancement issue from two different angles. In the first part, we constructed a dataset containing reference images, their enhanced version using conventional CE techniques, corresponding objective scores computed using CE assessment metrics, and associated subjective test rankings. The dataset is expected to be a useful benchmark to evaluate existing and newly proposed CE evaluation metrics. This work is also the first step related to one of our research plans to design a contrast enhancement evaluation metric for medical and endoscopic images, which is not studied yet in literature to the best of our knowledge especially for the stereo case. The second part of the chapter was devoted to introduce an enhancement technique for both endoscopic and WCE images. The obtained results in terms of stereo-matching and abnormalities classification for WCE frames, together with the visual post-processing images quality demonstrate the efficiency of the proposed enhancement method. However, despite the good performance with 3D images, the proposed enhancement technique is not well adapted for stereo endoscopic content since it does not take into consideration its intrinsic properties namely the depth data. Therefore the next chapter presents an adaptive edge based contrast enhancement method for stereo images combining both 2D images properties (i.e. local image activity and edginess information) and stereo images features (i.e. depth information and the binocular just noticeable difference (BJND) visibility map).

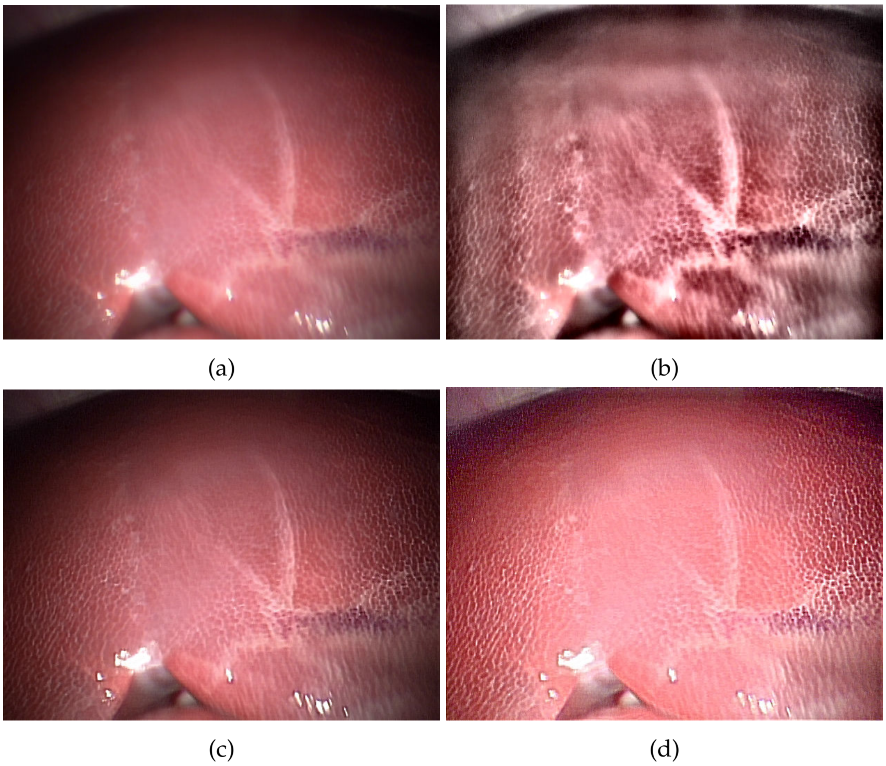
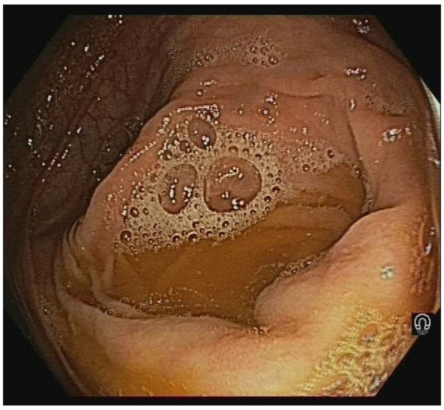
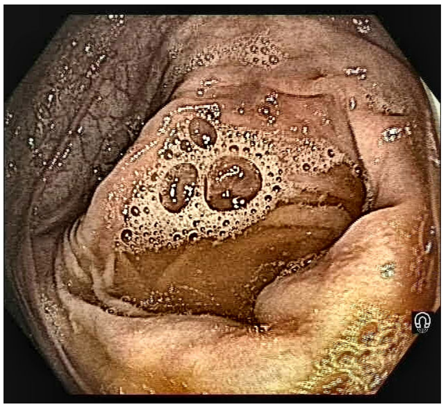


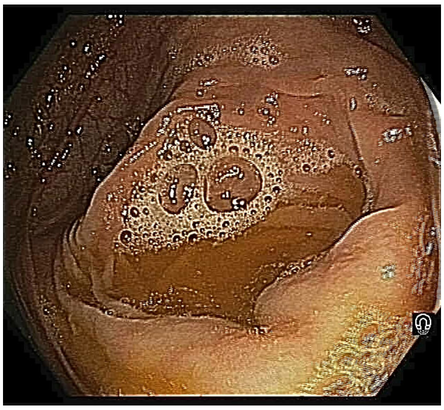
Figure 4.6: Enhancement results of a sample left view (a) from the 3D-END dataset using (b) CLAHE, (c) EBCE, and (d) the proposed method



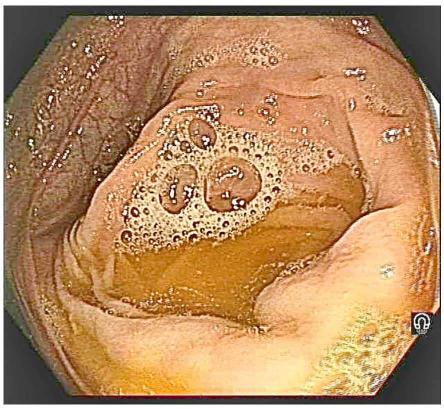
(a)



(b)

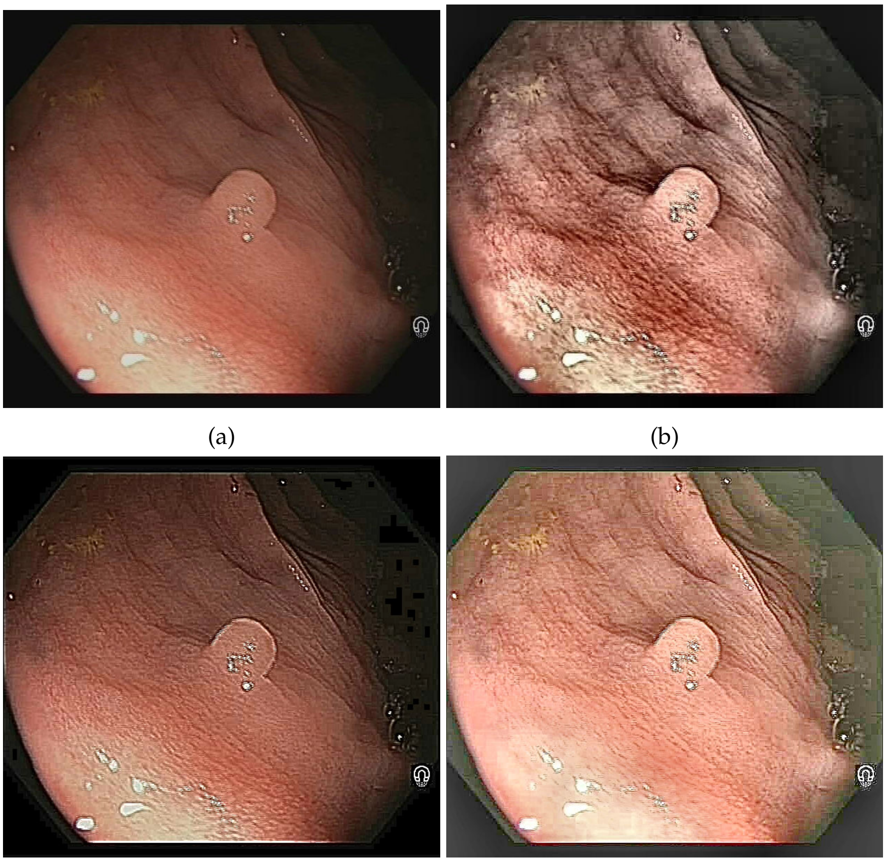


(c)



(d)

Figure 4.7: Enhancement results of a sample left view (a) from the 2D-WCE dataset using (b) CLAHE, (c) EBCE, and (d) the proposed method



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

Figure 4.8: Enhancement results of a sample left view (a) from the 2D-WCE dataset using (b) CLAHE, (c) EBCE, and (d) the proposed method

Chapter 5

Adaptive enhancement for 3D endoscopic images combining depth data and the BJND model

Vision is the art of seeing what is invisible to others.

JONATHAN SWIFT

5.1 Introduction

One of the main goals for endoscopic image processing is to provide the surgeons and the medical staff with improved visual quality images to facilitate the diagnosis and the treatment of patients. Image contrast enhancement and sharpening are two important features that affects significantly the perceived quality. In our endoscopic stereo imaging context, enhancing these properties is of great usefulness for two reasons:

- According the surgeons in *The Intervention Center* at Oslo University Hospital, having clear tissues structures, veins and vessels is very important for some operations. Particularly, in liver-resection surgeries, blood vessels should be visually distinguishable for the surgeons as they should be inhibited before performing the resection in order to avoid a bleeding, which can complicate the procedure and constraint the patient safety. Enhancing these details require processing the local contrast by improving the regular and micro edges that represent the tissue details, veins, blood vessels. Among the enhancement techniques that can achieve this processing, methods based on edge detection filtering [17, 162] seems to be promising since they address directly edginess information and local contrast.
- Improving the sharpness of stereo image views can also increase the depth perception according to some studies [188, 291]. Therefore, a

sharpness enhancement process can provide the surgeons with a better depth feeling/cues, which can facilitate their tasks by having an improved 3D visual feedback about the relative positions of instruments and assist their motion within the endoscopic site.

Processing, however, independently the left and right images using conventional 2D enhancement techniques don't yield often satisfactory results as it does not account for the special characteristics of stereoscopic images, namely interview correlation and depth information [87]. Therefore we propose in this chapter an adaptive contrast enhancement method for stereo endoscopic images combining depth information and BJND visibility thresholds. To perform the contrast enhancement, we extend the edge based method method of [17] by incorporating edginess information, depth data and the local image activity. The latter feature allow to adapt the contrast enhancement locally in each region according to its characteristics (homogeneous or boundary region). The BJND is then used to control the overall inter-view enhancement and avoid any noticeable difference that can trigger distortions, eyestrain or visual fatigue [108, 128].

The remainder of this chapter is structured as follows. In Section 5.2 we describe the algorithm for contrast enhancement based on local edge detection. Section 5.3 overviews the evolution of just noticeable difference (JND) concept from 2D to 3D content, and focus on the binocular JND model (BJND). We introduce our proposed adaptive method based on depth information in Section 5.4. Experimental results that prove the effectiveness of our technique are presented and discussed in Section 5.5. Finally conclusions are drawn in Section 5.6.

5.2 Edge-based contrast enhancement (EBCE)

This section provides an overview of the contrast enhancement technique based on local edge detection [17], which is a well-known spatial domain enhancement method for 2D images.

5.2.1 Contrast function

The edge based enhancement technique proposed in [17] accounts for special perceptual features of the HVS, namely for contour detection, by combining *Gordon*'s local contrast enhancement algorithm [82] and the theory of contour detection [200]. Given a pixel P at spatial coordinates (i, j) and its gray-level intensity $I_{i,j}$, the local contrast is defined as follows:

$$C_{i,j} = \frac{\mid I_{i,j} - \overline{E_{i,j}} \mid}{I_{i,j} + \overline{E_{i,j}}} \tag{5.1}$$

where $\overline{E_{i,j}}$ designates an estimate of the mean edge gray level computed as the average value of the gray-level pixel intensity values weighted by the edginess information $\delta_{m,n}$ within a $(2n+1) \times (2n+1)$ window centered at (i, j) and computed as follows:

$$\overline{E_{i,j}} = \frac{\sum_{(m,n)\in w_{i,j}} I_{m,n} \cdot \Phi(\delta_{m,n})}{\sum_{(m,n)\in w_{i,j}} \Phi(\delta_{m,n})} \tag{5.2}$$

where Φ is a monotonic increasing function and $\delta_{m,n}$ represents the edge value. The latter can be estimated using either the gradient or any *Sobel* operator.

5.2.2 Contrast enhancement

The improved contrast $C'_{i,j}$ can be generated by simply applying a function f to the local contrast $C_{i,j}$, satisfying the following conditions:

$$\begin{cases} f \ : \ [0,1] \rightarrow [0,1] \\ C_{i,j} \longmapsto f(C_{i,j}) = C'_{i,j} \geq C_{i,j} \end{cases} \tag{5.3}$$

The output intensity is computed as follows:

$$I'_{i,j} = \begin{cases} E_{i,j} \cdot \frac{1-C'_{i,j}}{1+C'_{i,j}} & if \ I_{i,j} \leq E_{i,j} \\ E_{i,j} \cdot \frac{1+C'_{i,j}}{1-C'_{i,j}} & otherwise \end{cases} \tag{5.4}$$

In [17], *Beghdadi et al.* demonstrated the efficiency and noise-robustness of this low complexity algorithm in sharpening the edges and the micro-edges of 2D images and discriminating objects according to their boundaries. In [162], the authors report that the technique improves as well the gray-levels distribution. The method seems to be very useful to enhance the local contrast of endoscopic images which contain very find details, edges, and micro-edges corresponding to veins, blood vessels, tissue textures, or lesions. Therefore we considered the contrast enhancement technique based on local edge detection as an important step in our proposed enhancement algorithm. Such improvement can facilitates the detection and extraction of relevant information such as feature points, which are crucial in performing a 3D organ reconstruction for the surgery navigation/resection planning.

5.3 2D / 3D - Just noticeable difference

The deep knowledge acquired about the monocular perception together with the technical challenges related to 2D images (e.g. compression, transmission, quality assessment) has led to the development of just noticeable difference models, based mainly on contrast and luminance masking effects. A JND profile indicates a visibility threshold beyond which an observer can detect visible difference in pixel values. The success of stereoscopic technologies has, however, prompted several questions linked to the extension of conventional 2D-JND models to 3D images, which require further understanding about the binocular vision and the important HVS features presented in Section 2.2.

Unlike 2D-JND, the literature dealing with 3D-JND is limited due to the recent focus on this problem beside the related challenges. Studying the sensitivity of the HVS to depth variations within the observed scene, *De Silva et al.* [54, 53] derived a mathematical model defining and explaining the JND in depth (JNDD). A JNDD visibility threshold describes the number of depth levels that can trigger a percieved depth change on a 3D display. In the same vein, the joint just noticeable difference (JJND) [174] is based on the idea that the perception of objects differs according to their depth level in the observed scene. In contrast with the JNDD, the JJND is derived using the nonlinear additively masking model [327], which is a 2D-JND accounting for texture masking and luminance adaptation.

Stereoscopic images are characterized with much content redundancy between the left and right images since they are captured with slightly different perspective. Thus, a distortion in one view can be might be masked/compensated by the other view due to the so called binocular combination/rivalry (see Sections 2.2.1.1 and 2.2.1.2). Inspired by these HVS features, the BJND model [333] measures the minimal noise/distortion in one stereoscopic view evoking noticeable perceptual difference when combined with the second view during the binocular vision process. Based on psychophysical experiments, the authors investigated the visual sensitivity to contrast masking effect, the binocular combination of noise and the luminance masking effect for stereo images.

Unlike the previously presented 3D-JND models measuring the visibility thresholds of 3D images, the stereo just noticeable difference (SJND) [248, 249] focus on stereoscopic videos JND taking into consideration four main parameters: binocular masking, sensitivity of luminance contrast, and both spatial and temporal redundancies/masking. In a multi-view video plus depth (MVD), hybrid just noticeable difference (HJND) model [334] is based on depth contrast and depth intensity which have a direct impact on the perceived depth.

5.3.1 Overview of the BJND model

In this section, we give an overview of the BJND model and the derivation of its formula as described in [333]. Given the left and the right images, the BJND map of the left view, (i.e., $BJND_l$) is defined as follows:

$$\begin{aligned} BJND_l(i,j,d) &= f(bg_r(i+d,j), eh_r(i+d,j), na_r(i+d,j)) \\ &= A_C(bg_r(i+d,j), eh_r(i+d,j)) \\ &\quad \times \left(1 - \left(\tfrac{na_r(i+d,j)}{A_C(bg_r(i+d,j), eh_r(i+d,j))}\right)^{\lambda}\right)^{\frac{1}{\lambda}} \end{aligned} \tag{5.5}$$

where *i* and *j* refers to the spatial pixel coordinates, *d* is the disparity value corresponding to the point (i, j) and na is the noise amplitude $0 \leq na_r \leq A_C$. The parameter λ controls the impact of the right-view noise and it is set experimentally to 1.25. We can notice that the BJND left is dependent on the background luminance intensity (*bg*), the edge high (*eh*) and the noise amplitude (*na*) of the right image. The inter-view pixel correspondence data is crucial for processing stereo content and generating the BNJD map. This data can be provided by the ground-truth disparity information, which is often not available with real time constraints such in our 3D endoscopic application domain. Therefore, performing a correspondence matching step (i.e., stereo matching) is important to obtain the disparity map.

Stereo matching has become a very active research field in the last decade because of the difficulty and the complexity of this task for both image processing and computer vision. Since the aim of our work is not to study nor address the stereo matching problematic, we adopted the recently proposed disparity map estimation algorithm of [167], which is reported to yield good performance in terms of accuracy and robustness, rating it among the best techniques according to the *Middleburry* website [266]. The choice of this stereo-matching algorithm is motivated by three main reasons related to its performance: i) the technique takes into account depth discontinuities and occlusions which occur in our endoscopic context mainly due to the surgical instruments and the surgeon's tasks, ii) it operates well on images with varying exposures and illumination which is the case for our endoscopic images due to the dynamic illumination conditions, and iii) running this stereo-matching technique includes a low computational cost and minimal processing time allowing the real-time constraint. A more detailed review and taxonomy of stereo-matching techniques is given by [267] and a ranking of the top performing algorithms in terms of complexity and accuracy is provided by the *Middleburry* website [266]. Note that if the right view is noise-free, which is the case for our endoscopic images, the $BJND_l$ is reduced to the expression of A_C, which is defined by:

$$A_C(bg, eh) = A_{C,limit}(bg_r(i+d,j)) + K(bg_r(i+d,j)) \cdot eh_r(i+d,j) \tag{5.6}$$

The background luminance component (bg) can be obtained by computing the average of a 5×5 sliding window centered in the spatial active-pixel location (i, j), and the edge height (eh) is given by:

$$eh(i,j) = \sqrt{E_H^2(i,j) + E_V^2(i,j)} \tag{5.7}$$

where

$$\begin{aligned} E_k(i,j) &= \tfrac{1}{24}\sum_{h=1}^{5}\sum_{v=1}^{5} l(i-3+h, j-3+h)\cdot G_k(h,v) \\ k &= H, V \end{aligned} \tag{5.8}$$

$$G_H = \begin{pmatrix} -1 & -2 & 0 & 2 & 1 \\ -2 & -3 & 0 & 3 & 2 \\ -3 & -5 & 0 & 5 & 3 \\ -2 & -3 & 0 & 3 & 2 \\ -1 & -2 & 0 & 2 & 1 \end{pmatrix}$$

$$G_V = \begin{pmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 3 & 5 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -3 & -5 & -3 & -2 \\ -1 & -2 & -3 & -2 & -1 \end{pmatrix}$$

where $l(i,j)$ refers to the luminance value related to pixel (i,j). The pass-band filtering masks G_H and G_V allow both the signal smoothing and edges detection within the processed window, similarly to extended Sobel operators. The terms $A_{C,limit}(bg)$ and $K(bg)$, which have been determined empirically, are defined as follows:

$$A_{C,limit}(bg) = \begin{cases} 0,0027\cdot(bg^2 - 96\cdot bg) + 8 & if\, 0 \le bg \le 48 \\ 0,0001\cdot(bg^2 - 32\cdot bg) + 1.7 & if\, 48 \le bg \le 255 \end{cases} \tag{5.9}$$

$$K(bg) = -10^{-6}\cdot(0.7\cdot bg^2 + 32\cdot bg) + 0.07 \tag{5.10}$$

The BJND of the right stereo view (i.e., $BJND_r$) can be generated similarly by substituting l for r and estimating the disparity map corresponding to the right image.

5.4 Proposed adaptive edge-based contrast enhancement method (AEBCE)

The proposed technique aims to exploit both monocular cues (i.e., contrast, edges) and binocular cues (i.e., disparity map, BJND) in enhancing the contrast of stereo endoscopic images and exhibiting the inner vessels/organs boundaries. The motivation behind using the edge-based method can be explained by two main reasons. First, the local contrast depends strongly

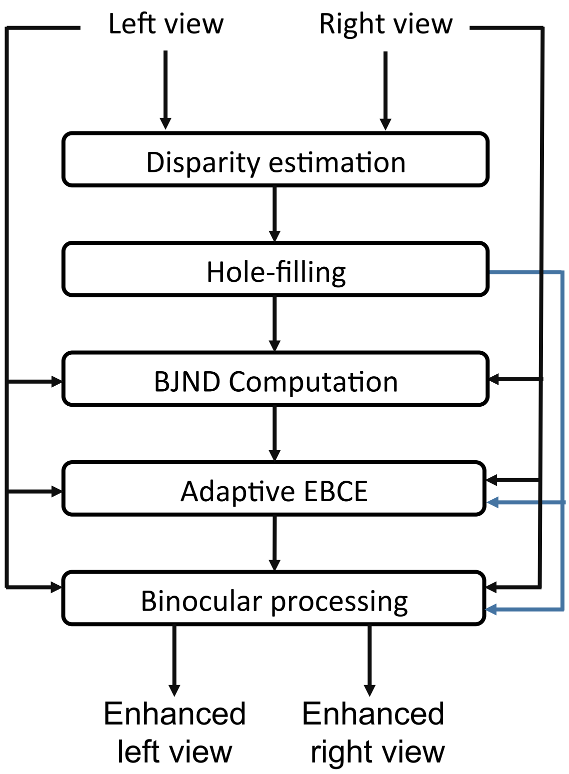


Figure 5.1: General flow-chart of transform domain image enhancement approach

on the contour sharpness which is related to the mean-edge value [16]. Second, the edge height (eh) of the BJND contrast masking component is also based on the gradient magnitude, which means that both approaches rely on the edginess information. Adapting the local edge detection based algorithm and combining it with depth and binocular information should yield more depth feeling for the surgeons and can facilitate the extraction of feature points. The contrast is adjusted locally for each object of the endoscopic scene based on its depth information available via the computed disparity map. Our proposed method is composed of 5 steps as illustrated by Figure 5.1. The following paragraphs will detail each step.

5.4.1 Disparity estimation

The first step consists in estimating the stereo correspondences of the stereoscopic views. In order to estimate the disparity map, we used the stereo matching algorithm proposed by [167], which is based on the adaptive random walk with restart (ARWR) method. We recall that the choice of this technique is justified by three main advantages that are important for this endoscopic image processing context: i) it takes into consideration occlusions and depth discontinuities which may occur due to surgical instruments handling during the surgeon's tasks, ii) it operates well on images with varying illumination and exposures which is the case for our endoscopic environment including dynamic illumination conditions, and iii) its implementation in a real-time environment revealed minimal processing time requirements with low computational costs. In other words, this stereo matching method allows a relatively accurate estimation of disparity maps with low computational complexity.

5.4.2 Hole-filling

One of the major concerns related to stereo matching is the occluded regions of the scene, which are represented by holes (i.e., pixel disparity value equal to zero) in the output estimated disparity maps. Although this problem is addressed in the used disparity estimation technique, we apply a hole-filling algorithm on the generated disparity maps to ensure that each pixel is assigned to a valid disparity value.

Many approaches have been proposed to fill disparity holes such as the Gaussian filter based method used in stereoscopic images generation via depth maps rendering [331, 332], the hole-filling based memory controller for multiview 3D videos [69], or the stereoscopic inpainting [315]. These methods can, however, involve disparity information loss and blur in the estimated disparity map specially if the hole-size is large. In our work, due the high performance of the adopted stereo-matching technique [167] tackling the occlusions problem, the generated disparity maps do not present large hole regions. To fill the very few resulting disparity holes, which size do not exceed 3 pixels, we replace each pixel hole by the mean Gaussian filtered disparity values of its surrounding pixels in a window sized 15×15.

5.4.3 BJND computation

Figure 5.2 shows the $BJND_{left}$ profiles (5.2(c),5.2(d)) of two endoscopic images (5.2(a), 5.2(b)) respectively. The BJND thresholds are mapped linearly to $[0, 255]$ for better visualization. The corresponding histograms (5.2(e),5.2(f)) show that the BJND values are usually comprised between 4 and 14, which gives an idea about the minimum and maximum visually non noticeable inter-view difference levels. We can notice that the BJND is dependent on the local background information that affects the BJND sensitivity to edges as the luminance increases. Furthermore, in regions not containing specular reflexions (i.e. where the pixels luminance intensities are not saturated resulting a loss of edges and background data), the BJND is proportional to the edginess information of the corresponding area of the other view, which is correlated with the noise-free case represented by Equation 5.6.

5.4.4 Adaptive EBCE

The adaptability of the proposed method is related to two different 2D and 3D aspects:

- Adapting the enhancement degree according the depth level of the processed object, i.e. Depth (3D) adaptation.
- Adapting the edge based contrast enhancement to the local image activity in order to avoid any halo effect or over-enhancement, i.e. spatial/intra-view (2D) adaptation.

In [291], the authors performed a perceptual subjective experiment to study the relation between depth perception and the perceived sharpness/blur in a stereoscopic imaging setup. The main aim of the experiment was to determine whether the perceived sharpness/blur varies as a function of the disparity. The obtained results revealed that the sensitivity to blur/sharpness depends significantly on the depth level of the perceived point. In other words, closer objects to the observer tend to appear shaper than the ones further in depth. This is actually correlated with the HVS visual attention feature discussed in Chapter 2 by which objects fixated using the vergence/accommodation of the eyes are projected in the *fovea*, which is a special spot of the retina allowing the sharpest vision due the high concentration of cone cells. Inspired by these experimental results, the following step of our proposed method consists in adapting the local contrast enhancement of each point according to its depth level in the endoscopic scene. Indeed, objects having high depth information are more highlighted during the contrast processing. Therefore, we integrate the estimated depth information (i.e. disparity values) in the computation of the improved contrast $C'_{i,j}$ to make the latter proportional to the depth level. The enhanced contrast expression is, thus, defined as follows:

$$C'_{i,j} = \tanh\left(\left(\frac{d_{i,j}}{d_{max}} + \lambda\right) \times C_{i,j}\right) \tag{5.11}$$

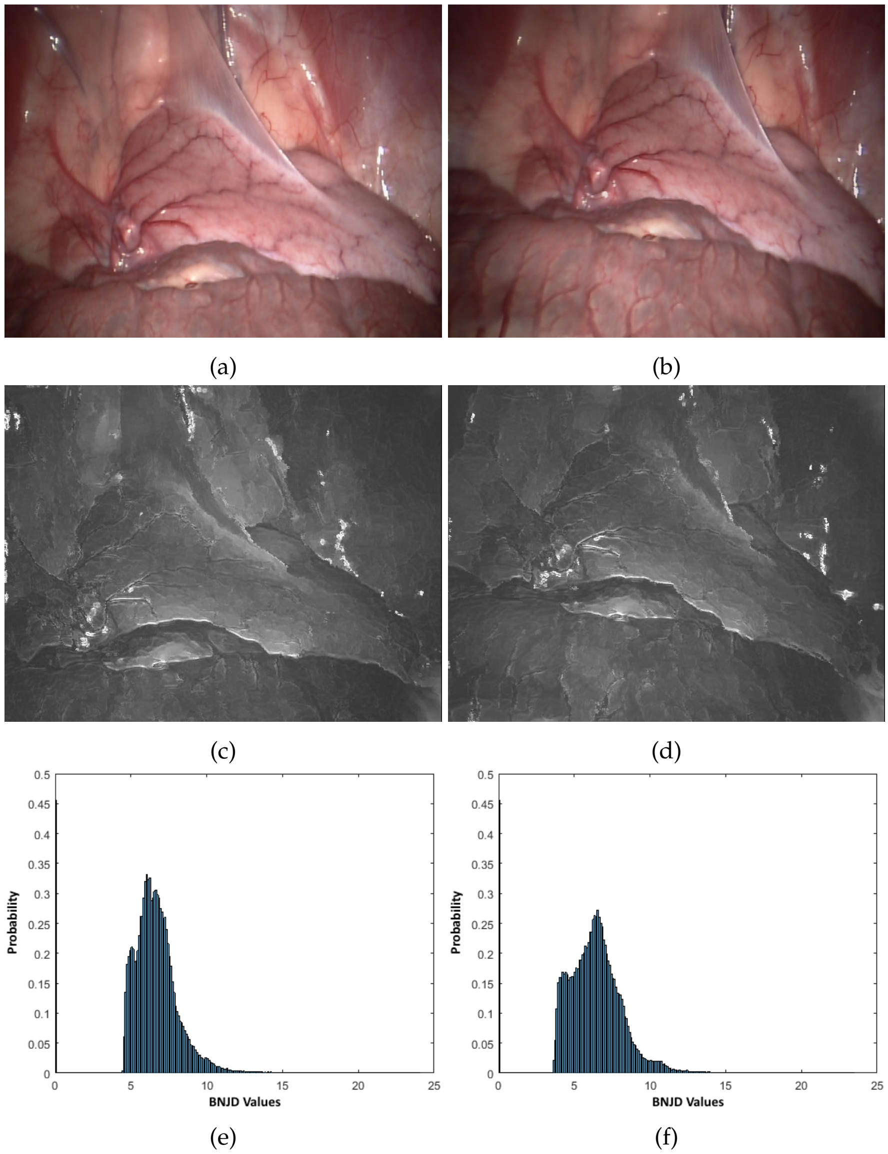


Figure 5.2: BJND profiles. Two original endoscopic left views (a) and (b). The corresponding scaled BJND maps (c,d), and histograms (e,f), respectively.

where $d_{i,j}$ refers to the disparity of the current pixel at coordinates (i, j), d_{max} is the closest perceived depth level indicated by the maximum disparity value and $\lambda \in [0, 1]$ controls the contrast increase. In order to take into consideration the local image activity and avoid any halo effect or overshooting that might occur when applying the classical edge-based contrast enhancement [17], we propose an adaptive edge-based contrast processing to compute the final intensity $I'_{i,j}$. The resulting pixel intensity $I'_{i,j}$ of the Equation 5.4 is replaced by the following expression:

$$I^{AEBCE}_{i,j} = (1 - \varphi_{i,j}) \times I'_{i,j} + \varphi_{i,j} \times I_{i,j} \tag{5.12}$$

where $\varphi_{i,j}$ indicates the local image activity and it is computed using the gradient $\nabla_{i,j}$ as follows:

$$\varphi_{i,j} = \frac{1}{1 + \nabla_{i,j}} \tag{5.13}$$

Note that when the pixel region is homogeneous, the gradient $\nabla_{i,j}$ tends to zero and we keep the original pixel intensity value $I_{i,j}$ since $\varphi_{i,j} \simeq 1$. However, when the gradient is strong, our adaptive approach gives more weight to the enhanced new intensity value $I'_{i,j}$.

5.4.5 Binocular inter-view processing

Taking into account the binocular combination property of the HVS, the following step consists in perceptually controlling the stereo enhancement using the binocular visibility thresholds provided by the BJND maps. Each inter-view pixel difference, defined as follows:

$$IVDiff_{i,j} = \mid I'_{i,j,left} - I'_{i,j,right} \mid, \tag{5.14}$$

is compared to the corresponding $BJND_{i,j}$ visibility threshold. If the interview difference does not exceed its corresponding visibility BJND threshold (i.e. $IVDiff_{i,j} \leq BJND_{i,j}$), that implies that there is no detectable visible difference for the observer and we preserve the enhanced intensity values. In the other case (i.e., $IVDiff_{i,j} > BJND_{i,j}$), we define what we will call "*visual conflict*" as the inter-view difference exceeding the binocular visibility threshold, defined by the following expression:

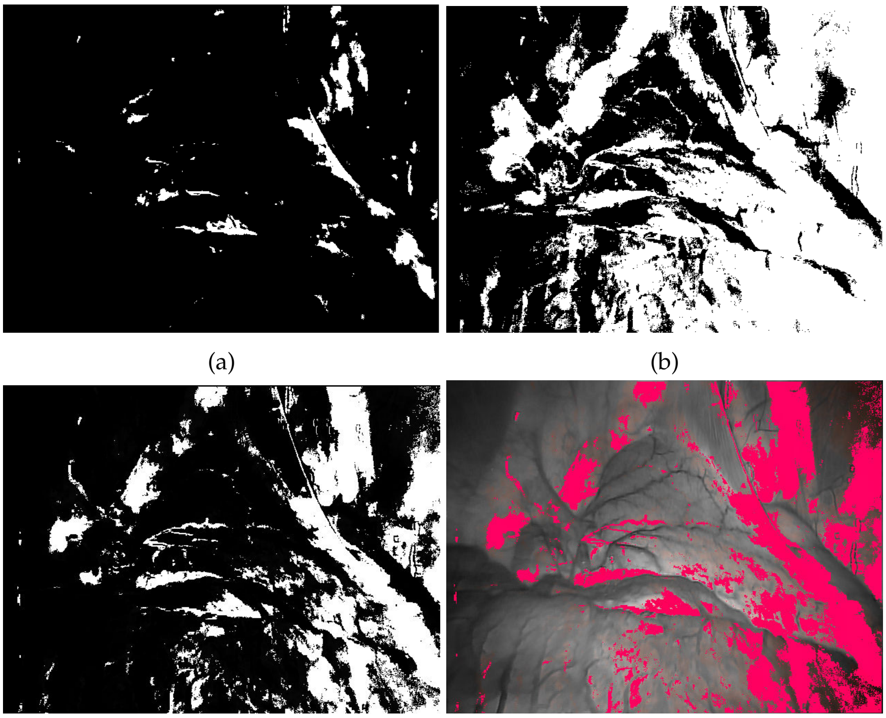
$$VC_{i,j} = IVDiff_{i,j} - BJND_{i,j} \tag{5.15}$$

Figure 5.3 shows binary maps illustrating the inter-view difference of an endoscopic stereo image before (Figure 5.3(a)) and after our adaptive edge-based contrast enhancement (Figure 5.3(b)). Figure 5.3(c) highlights the corresponding noticeable visual conflict and Figure 5.3(d) shows the visual-conflict binary map superimposed on the left gray-scale component. One might think that the inter-view difference should be negligible before the enhancement. This depends, however, on the captured scene dynamic and the video acquisition settings of the left and right stereo cameras which can yield different illuminations creating interview differences. Such case occurs in the endoscopic environment due the dynamic illumination conditions created by the endoscope light source motion and the moist tissues, surgical smoke and the instruments which might reflect the endoscope light differently to the left and the right cameras. Another non-negligible parameter is the accuracy of stereo matching algorithm used to estimate the disparity map. Indeed, if the disparity values are not accurate, then each pixel will not match with its exact correspondent in the other stereo view during the enhancement process. Unless it is a textures homogeneous region, this implies that the intensity values of the stereo pixels, supposed to be a match, are not the same. Thus, the interview difference between the concerned miss-matching pixels will be will be largely different from zero.

We can notice that the inter-view difference increases after the enhancement process and that the noticeable difference lies especially on the enhanced edges and, to a lesser extent, some homogeneous regions. Moreover, we can notice that some noticeable boundaries are shifted from the corresponding scene objects. This can be explained by the fact that the visual conflict component is generated from a binocular processing that depends on both left and right views, which makes it more or less equivalent to a *cyclopean* image (see Section 2.2.1.1). Thus, it cannot match exactly none of the stereo views objects boundaries. Once computed, we remove the visual conflict $VC_{i,j}$ from the left or the right views to suppress any noticeable perceptual difference without affecting the luminance distribution. The new intensity values are obtained as follows:

$$\begin{cases} I^{AEBCE}_{i,j,left} = I^{AEBCE}_{i,j,left} - VC_{i,j} & if\ I^{AEBCE}_{i,j,left} \geq bg_{i,j,left} \\ I^{AEBCE}_{i,j,right} = I^{AEBCE}_{i,j,right} - VC_{i,j} & if\ I^{AEBCE}_{i,j,right} \geq bg_{i,j,right} \end{cases} \tag{5.16}$$

where $bg_{i,j}$ is the average background luminance, computed the same way as for of the BJND model. When performing this interview processing, one relevant question might be raised: once detected, the visual conflict should be removed from the left or the right views ? A possible simple choice would be to remove the visual conflict randomly from either the left or the right images. In our work, however, we decided to make the interview processing adaptive and dependent on the local information in



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

Figure 5.3: Inter-view difference and binocular visual-conflict. (a) Inter-view difference before applying the AEBCE, (b) Inter-view difference after applying the AEBCE, (c) the corresponding noticeable visual stereo conflict, (d) visual conflict binary map superimposed on the left gray-scale image.

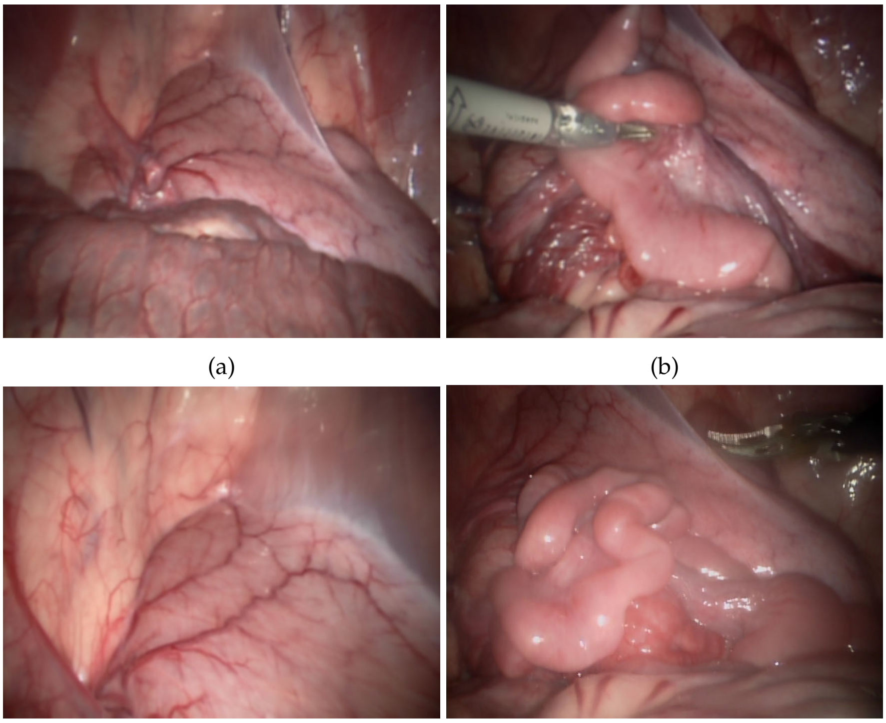
each view. Indeed, the visual conflict is removed from the pixel view which presents different intensity value than the surrounding the background luminance $bg_{i,j}$. In practice, the difference is detected if the intensity value of the processed pixel exceeds the average background luminance (i.e. $if\ I^{AEBCE}_{i,j,view} \geq bg_{i,j,view}; \forall\ view \in \{left, right\}$).

5.5 Experimental setup, results and discussions

The aim of the experiment is to emphasize the impact of the proposed adaptive depth-based contrast enhancement and the binocular enhancement control process. The subjective test is performed considering the double-stimulus continuous quality scale method (DSCQS), recommended in ITU-R BT.500 [117].

We consider the stereo enhancement method based on local edge detection and depth information [87] as a baseline for comparison.

10 endoscopic stereo images from *Hamlyn Centre* laparoscopic /endoscopic dataset [221] having the same resolution (480×640) are used to conduct the experiment. The images are first converted to Lab color-space where only the luminance component L is processed then displayed on an 23-in LG FLATRON W2363D monitor with NVIDIA wireless polarized glasses, after being reconverted to RGB color-space. Figure 5.4 illustrates some sample endoscopic views from the dataset. We can notice the fine details and the dark areas reaching up to 40 of the special image resolution in some cases due to the dynamic illumination condition. 17 expert observers, mostly males (11 males and 6 females) having an average age of 30 and with normal or corrected-to-normal visual acuity were invited to evaluate the subjective quality of the stereo endoscopic images. Their expertise is namely related to image/video processing and medical imaging quality.



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

Figure 5.4: Sample views of *Hamlyn Centre* laparoscopic /endoscopic dataset

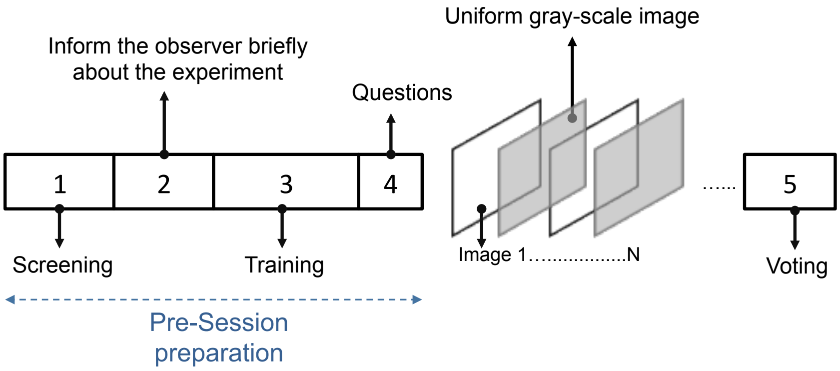


Figure 5.5: Protocol of the subjective test for stereo endoscopic image assessment.

Prior to each test session, each observer is screened and briefly informed about the experiment. A training session is then performed based on "dummy presentations" to stabilize the observer's opinion. Once the training is finished, the observer is allowed to ask questions about the evaluation/ranking process or the images content/quality. Each original endoscopic stereo image is displayed randomly with two enhanced versions using our proposed AEBCE method and the technique proposed in [87], separated by a uniform grayscale image. The observer is then invited to evaluate the visual quality by attributing a score to each stereo image using the grading scale illustrated by Figure 5.6. The subjective test protocol is summarized in Figure 5.5.

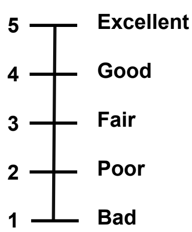


Figure 5.6: Grading scale of the subjective test.

Figure 5.7 illustrates the enhancement of three endoscopic left views (5.7(a),5.7(b)) using the proposed method (5.7(c),5.7(d)) and the technique proposed in [87] (5.7(e),5.7(f)). We can notice that the contrast of the images enhanced using our adaptive depth-based technique is well adjusted without introducing any halo effect and the tissues/veins edges are more exhibited. The images processed using the method of [87] present, however, an over-enhancement illustrated by a graying effect in the homogeneous regions and a halo effect visible specially surrounding the specular reflection regions. This can be explained by their used enhancement function (see Equation 5.17) which over-amplify the contrast and does not take into account the local image activity, unlike our proposed adaptive enhancement method. Figure 5.8 highlights two different zoomed regions of an endoscopic stereo view enhanced using our AEBCE technique (Fig. 5.8(b), 5.8(d), 5.8(f)) and the method of *Hachicha et al.* [87] (Fig. 5.8(a), 5.8(c), 5.8(e)). The images are presented in grayscale for better visualization of the over-enhancement. We can notice that zooming the ROI reveals clearly the artifacts and the distortions of Figures 5.8(a), 5.8(c), and 5.8(e). Indeed, although the edges are relatively sharper, the regions enhanced using [87] present clearly an over enhancement illustrated by halo artifacts and graying-out.

$$C'_{i,j} = C_{i,j}^{1-\frac{d_{i,j}}{d_{\max}}(1-\lambda)} \tag{5.17}$$

Although such artifacts might not be very noticeable in separate 2D views, the binocular combination of two views will exhibit even more all common image features withing each view, including the artifacts resulting from the over-enhancement. Thus, it is important to assess the overall quality of stereo images based on both left and right views. This is one of the reasons that can explain why assessing 3D images quality based on independent assessment of the left and right views using conventional 2D metrics fails often giving faithful evaluation to the HVS perception of stereo images. To address the 3D image quality assessment problem efficiently, many studies have been devoted to the further understanding the HVS stereo perception and model its binocular features. Among these studies, an interesting stereo metric proposed in [18] consists in modeling the binocular energy produced by the simple and complex cells of the visual cortex (see Chapter 2) using multi-resolution mathematical tool, namely the bandelet transform and the complex wavelet transform (CWT).

Figure 5.9 shows the results of the subjective test described above. Figure 5.9(b) illustrates the standard box-plot of the obtained scores related to the three sets of images where the black bars indicate the corresponding maximum (top bar) and minimum (bottom bar) score and the red line indicates the median. We can notice that the proposed adaptive depth-based method outperforms the technique proposed by [87]. The proposed method produces better stereo image quality then both the original images and the stereo pairs enhanced by [87], with a more depth feeling and comfortable 3D visualization for the observers. We performed as well a Paired t-test on the experimental data to compute the statistical difference between the different groups of images regarding the visual quality (Figure 5.9(a)). The paired t-test serves to compare the mean score before and after the enhancement

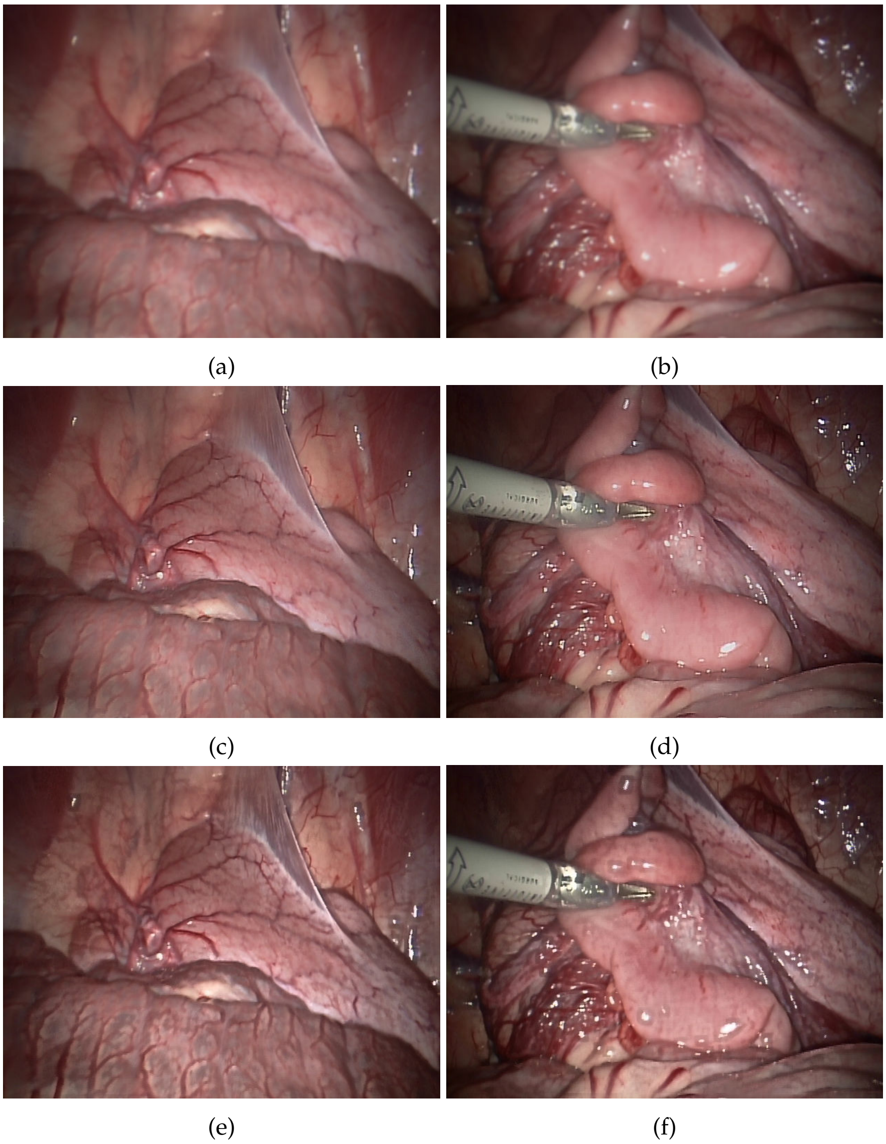
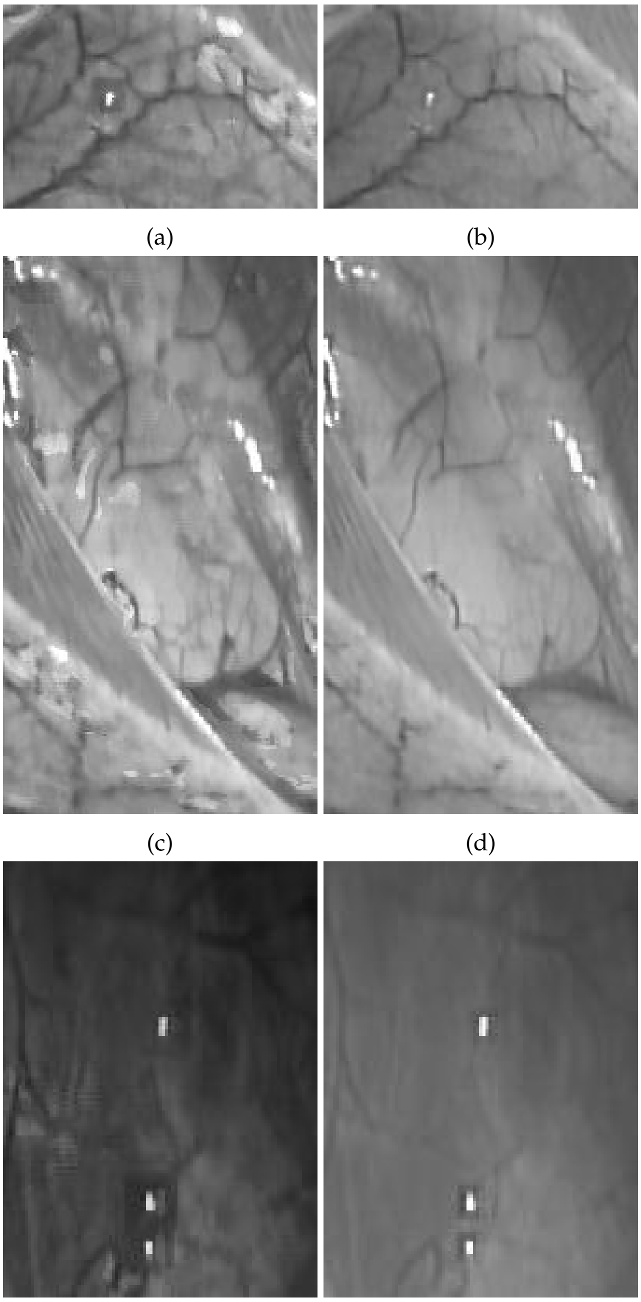


Figure 5.7: Enhancement of two endoscopic stereo views (a,b) using the proposed method (c,d) and the technique proposed in [87] (e,f).

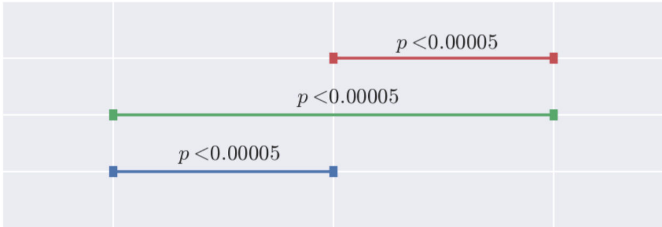


(a) (b)

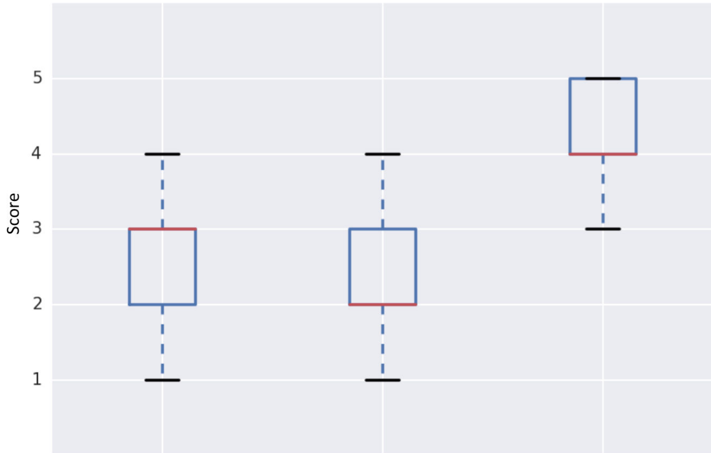
(c) (d)

(e) (f)

Figure 5.8: Zoomed ROI of a stereo endoscopic view enhanced using the method of [87] (left column) and our proposed AEBCE technique (right column)



(a)



(b)

Figure 5.9: Subjective assessment results grouped by method. The box-plot scores from the left to the right correspond respectively to: original images, enhancement method of [87], and our proposed AEBCE method (b); and the corresponding Paired T-test results (a).

using both the conventional [87] and the proposed method. This process serves to indicate whether the enhancement process affects really the perceived quality, i.e. whether there is a noticeable difference in the observer's scores after applying each enhancement technique. The obtained significance level ($p < 0.00005$) indicates a less than 0.005% risk of concluding that a visible image quality difference exists when there is no actual perceived difference, which proves statistically the effectiveness of the proposed enhancement technique.

5.6 Conclusion

In this chapter, we introduced a novel stereo image contrast enhancement method which combines the sensibility of the HVS to edginess information and the relative positions of objects in the 3D endoscopic scene. The experiment carried out on a widely used stereo image dataset show the effectiveness of the proposed method in using the depth information in stereo enhancement process. Our method produces pleasant 3D scenes, increases the perception of depth and mitigate eye strain. In this work we focused mainly on edge and depth information. The introduction of visual saliency, local features and geometrical information in the enhancement process can ameliorate further the obtained results. Furthermore, in spite of its good performance, the proposed AEBCE does not allow to reduce the dark regions or enhance specific images features while increasing the local image contrast. This processing issue is further investigated and addressed in the next chapter using a proposed wavelet based enhancement schemes.

Chapter 6

Enhancement methods for stereo endoscopic images using new joint wavelet-based processing

> There are things known and there are things unknown, and in between are the doors of perception.
>
> ALDOUS HUXLEY

6.1 Introduction

As explained in Chapter 3, endoscopic image enhancement has two main aims. The first possible purpose is to improve the perceived quality for doctors and surgeons when it is crucial to have a clear visual feedback related to some endoscopic details. For instance, the presence of lesions, polyps or blood vessels to be blocked before performing a resection in order to avoid a bleeding, which can affect the patient safety. The second possible purpose is to improve the input for subsequent post processing tasks such as feature points extraction for 3D organ reconstruction and registration. The latter tasks are important for determining the intra-operative morphology of the surgical domain, which is perquisite to register the patient specific data and establish the navigation plan.

Due to the complex structure of endoscopic images containing very fine details, it is more convenient to perform the enhancement in the wavelet domain because of its efficient spectral separation and multi-resolution characteristics discussed in Section 3.6.2.1, which are not available in the spatial domain. On the other hand, processing independently the left and right views of the stereo endoscopic image increases the inter-view difference, which adversely affect the accuracy of the stereo matching process and impact unpredictably the overall perceived resulting stereo image quality. To the best of our knowledge, no research work has been devoted to study the enhancement of stereo endoscopic images in the transform domain taking into consideration the interview redundancies/correlation, some intrinsic characteristics of 3D images namely depth data, and the multi-scale representation which offer many advantages compared to the spatial one. Therefore, we propose in this chapter a novel joint enhancement technique for stereo endoscopic images relying on a cross-view processing and depth information for both the wavelet decomposition and the enhancement steps to exploit the inter-view redundancies, together with perceptual HVS properties related to the contrast sensitivity and the binocular combination/rivalry features.

The reminder of this chapter is organized as follows. Section 6.2 presents the joint wavelet decomposition which accounts for the inter-view redundancies. In Section 6.3, we introduce our joint enhancement techniques for stereo endoscopic images. Section 6.4 presents and discusses the experimental results and, finally, Section 6.5 concludes the chapter.

6.2 Joint wavelet decomposition

Wavelet transformation has drawn significant attention in image processing and analysis for many application domains such as super resolution [56], denoising [32], enhancement [155] and pattern/shape recognition [301] due to their intrinsic properties: multiscale representation and good space-frequency localization. These properties are of great importance when processing complex endoscopic images since it allows to target specific image components and control the level of details to be processed. For this reason, we generate two wavelets representations of the left and right endoscopic images. Recall that stereo endoscopes are equipped with two cameras capturing the same scene with slightly different angles. This implies that the right and left views of each stereo image contain considerable redundant information. Despite its simplicity, decomposing independently the left and right images using conventional 2D wavelet transforms implies not taking into account the inter-view dependencies and the intrinsic properties of 3D images, namely depth data. Therefore, we resort in our work to a joint stereo wavelet decomposition using the Vector Lifting Scheme (VLS) concept [129] in order to exploit both the inter-view redundancies/correlation and depth information. To perform the cross-view processing leading to the VLS decomposition, it is thus first necessary to compute the disparity map which provides inter-view pixel correspondences information. This data is generated using a disparity estimation (DE) technique to be applied before starting the joint wavelet decomposition. Figure 6.1 illustrates the VLS analysis block diagram.

Considering a resolution scale j, for each line m and column n of the left $I_j^{(l)}$ and right $I_j^{(r)}$ input views, a Predict-Update-Predict (PUP) workflow

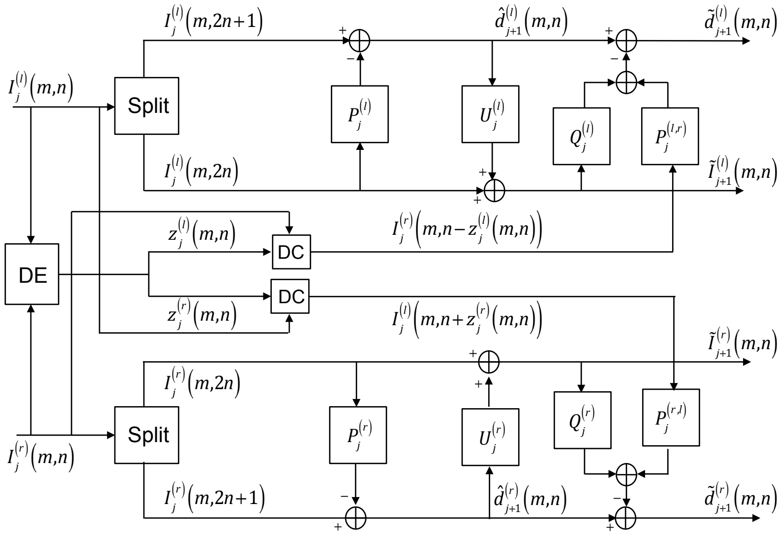


Figure 6.1: Block-diagram of the VLS decomposition

structure is performed. More precisely, dealing with the right view, a first prediction process is applied to generate an intermediate signal $\hat{d}_{j+1}^{(r)}$ of the higher resolution level $(j+1)$ as follows:

$$\hat{d}_{j+1}^{(r)}(m,n) = I_j^{(r)}(m,2n+1) - \sum_{k \in \mathcal{P}_j^{(r)}} p_{j,k}^{(r)} I_j^{(r)}(m,2n-2k), \tag{6.1}$$

where the set $\mathcal{P}_j^{(r)}$ and $p_{j,k}^{(r)}$ denote respectively the predictor support and the weights related to the odd samples of the right view $I_j^{(r)}(m,2n+1)$. The following step consists in applying an update process in order to calculate the approximation coefficients $I_{j+1}^{(r)}$ as follows:

$$\widetilde{I}_{j+1}^{(r)}(m,n) = I_j^{(r)}(m,2n) + \sum_{k \in \mathcal{U}_j^{(r)}} u_{j,k}^{(r)} \hat{d}_{j+1}^{(r)}(m,n-k), \tag{6.2}$$

where $\mathcal{U}_j^{(r)}$ and $u_{j,k}^{(r)}$ designate respectively the support and the weights of the update operator. In the last step, the final detail coefficients $\widetilde{d}_{j+1}^{(r)}$ are generated using a second hybrid step:

$$\widetilde{d}^{(r)}_{j+1}(m,n) = \hat{d}^{(r)}_{j+1}(m,n) - \Big(\sum_{k \in \mathcal{Q}^{(r)}_j} q^{(r)}_{j,k} \widetilde{I}^{(r)}_{j+1}(m, n-k)$$
$$+ \sum_{k \in \mathcal{P}^{(r,l)}_j} p^{(r,l)}_{j,k} I^{(l)}_j (m, 2n+1+z^{(r,l)}_j(m, 2n+1) - k) \Big), \tag{6.3}$$

where $q^{(r)}_{j,k}$ (resp. $p^{(r,l)}_{j,k}$) and $\mathcal{Q}^{(r)}_j$ (resp. $\mathcal{P}^{(r,l)}_j$) refer to the weights and support of the second intra (resp. inter)-image predictor related to the right image, and $z^{(r,l)}_j$ is generated by down-sampling and dividing by 2^j the input estimated disparity map $z^{(r,l)}$. The reason of the latter operation is that each decomposition level j divides the image spatial resolution by 2^j:

$$z^{(r,l)}_j(m,n) = \frac{1}{2^j} z^{(r,l)}(2^j m, 2^j n). \tag{6.4}$$

Note that the indexes order in the disparity subscript (r,l) serves to distinguish the reference view of the disparity estimation process. Indeed, the $z^{(r,l)}$ notation implies that the estimated disparity values allow to find stereo correspondences in the left view of each pixel in the right image, and vice versa for $z^{(l,r)}$. We can notice from both Figure 6.1 and Equation 6.3 that the disparity compensation (DC) step is based on $z^{(r,l)}$ to calculate $I^{(l)}_j(m, n + z^{(r,l)}_j(m,n))$, which is useful for performing the second hybrid prediction process.

The spatial supports of the prediction and update filters, which are used for the image decomposition, are set respectively to $\mathcal{P}^{(r)}_j = \{-1, 0\}$ and $\mathcal{U}^{(r)}_j = \{0, 1\}$. Similar to the conventional 5/3 lifting scheme (LS) [293], the corresponding weights are : $p^{(r)}_{j,-1} = p^{(r)}_{j,0} = \frac{1}{2}$ and $u^{(r)}_{j,0} = u^{(r)}_{j,1} = \frac{1}{4}$. As for the second hybrid prediction step, it is applied using the spatial supports $\mathcal{Q}^{(r)}_j = \{-1, 0\}$ and $\mathcal{P}^{(r,l)}_j = \{-1, 0, 1\}$. A variance minimization is performed on the detail coefficients $\widetilde{d}^{(r)}_{j+1}$ in order to optimize $q^{(r)}_{j,k}$ and $p^{(r,l)}_{j,k}$ [130, 77, 84].

After setting the right image decomposition strategy, a similar Prediction-Update-Prediction based decomposition is also performed on each line of the left view. Applying these steps along the lines and columns of each view produce one approximation subband and three detail components with three different orientations namely, horizontal, vertical and both diagonals. Similarly, performing a stereo images multi-resolution decomposition consists in repeating the same process on the left and right approximation subbands $I^{(l)}_{j+1}$ and $I^{(r)}_{j+1}$.

We can notice from Equations 6.1 and 6.3 that the odd samples prediction is based on both the surrounding pixels within the same image and corresponding neighboring samples in the other view provided by the disparity compensation (DC) steps. The latter feature is actually considered as the key property of VLS compared to the classical LS scheme since it allows exploiting both inter and intra-view redundancies when decomposing stereo images.

It is worth mentioning that the VLS was originally proposed for stereo image coding [129], which imposes a significant asymmetric decomposition constraint. Indeed for this context, the decoding process restricts the joint inter-view wavelet decomposition to only one image while a classical intra prediction-based LS is applied on the other view. Since such decoding constraint is not valid in the context of stereo endoscopic image enhancement, the joint wavelet decomposition developed and implemented in this work is applied on both the left and right images to exploit efficiently the intra and inter-view redundancies.

6.3 Proposed cross-view enhancement techniques

6.3.1 Adaptive inter-views processing technique

Given stereo endoscopic image, a disparity map is first estimated and the 3D wavelet decomposition is performed on left and right views over J resolution levels. Thus, we obtain one approximation subband at the coarsest scale $I_J^{(v)}$ and $3J$ detail subbands $\left(I_j^{(o,v)}\right)_{\substack{o\in\{HL,LH,HH\}\\ j\in\{1,\dots,J\}}}$ for each view $v \in \{l, r\}$.
Then, a transform-based mapping function, similar to those given in Section 3.6.2.1, is applied to the generated approximation and details subbands yielding intermediately intra-view enhanced wavelet coefficients of both images. The latter, called also pre-enhanced wavelet subbands, are obtained as follows:

$$
\begin{aligned}
&\forall\ v \in \{l, r\}, \quad \forall\ o \in \{HL, LH, HH\}, \\
&\begin{cases}
\hat{I}_J^{(v)}(m,n) = \kappa_J^{(v)}(m,n).I_J^{(v)}(m,n) \\
\hat{I}_j^{(o,v)}(m,n) = \kappa_J^{(v)}(m,n).\kappa_j^{(o,v)}(m,n).\lambda_j^{(o,v)}(m,n).I_j^{(o,v)}(m,n)
\end{cases}
\end{aligned}
\tag{6.5}
$$

where $\kappa_J^{(v)}$, $\kappa_j^{(o,v)}$, and $\lambda_j^{(o,v)}$ are weighting terms similar to those used in Eqs. (3.3) and (3.11).
Finally, inspired by the binocular rivalry/combination features of the HVS [127, 63], the pre-enhanced left and right images are combined using some adaptive weights to generate the final enhanced stereo images. The stereo combination is inspired by the linear formulation proposed by Levelt [170] to model the binocular rivalry when perceiving *cyclopean* images as follows:

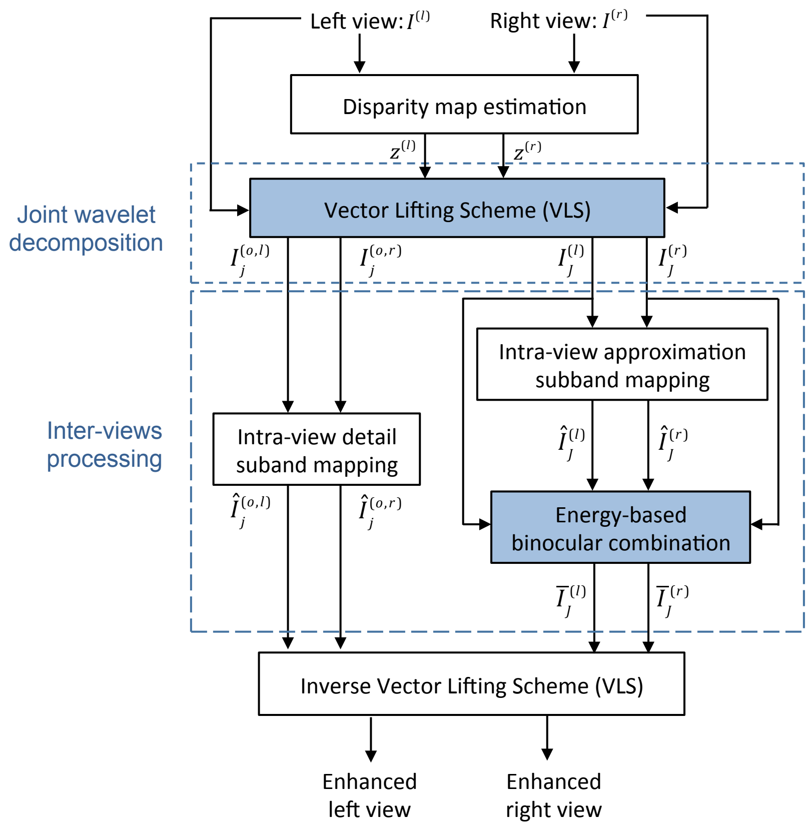


Figure 6.2: The proposed adaptive inter-view processing based technique.

$$I^{(\mathrm{C})} = w^{(l)}.I^{(l)} + w^{(r)}.I^{(r)} \tag{6.6}$$

where $I^{(l)}$ and $I^{(r)}$ designate the left and right monocular visual stimuli perceived respectively by the left and right eyes, $w^{(l)}$ and $w^{(r)}$ denote the corresponding respective weights ($w^{(l)} + w^{(r)} = 1$), and $I^{(\mathrm{C})}$ refers to the cyclopean component constructed by the visual cortex. The study of Levelt revealed that the rivalry phenomena depends on the relative stimulus strength to each view and a weighting model is proposed to set the left and right weights according to their corresponding stimulus strength.

Since the approximation coefficients correspond to a compact information of the original image at a coarsest scale, we propose in our first approach to combine *only* the approximation subbands of both views while keeping unchanged the other intra-view pre-enhanced detail subbands. More precisely, the new enhanced approximation subbands of the right image $\bar{I}_J^{(r)}$ and the left one $\bar{I}_J^{(l)}$ are given by:

$$\begin{cases} \bar{I}_J^{(r)}(m,n) = w_J^{(r)}(m,n).\hat{I}_J^{(r)}(m,n) + w_J^{(l)}(m,n+z_J^{(r)}).\hat{I}_J^{(l)}(m,n+z_J^{(r)}) \\ \bar{I}_J^{(l)}(m,n) = w_J^{(l)}(m,n).\hat{I}_J^{(l)}(m,n) + w_J^{(r)}(m,n-z_J^{(l)}).\hat{I}_J^{(r)}(m,n-z_J^{(l)}) \end{cases} \tag{6.7}$$

where $w_J^{(r)}$ and $w_J^{(l)}$ are the weighting coefficients of the corresponding right and left approximation subbands. One simple way to define these weights consists in taking an average combination of both views (i.e $w_J^{(r)} = w_J^{(l)} = \frac{1}{2}$). However, this choice is not efficient since it is not correlated with the binocular perception features namely, the binocular combination/rivalry, especially when the two views have different characteristics. For this reason, we propose to resort to a local adaptive weighting combination model allowing a stereo perception of the world that simulate the binocular processing performed by the HVS. Indeed, according to Levelt's study [170] and others investigations on binocular vision/rivalry [252], the linear weighting model given by Eq. (6.7) is equivalent to that used for simulating the *cyclopean* image (see stereo-vision Chapter 2), which is formed within the observer's visual cortex when two retinal views are captured by the left and right eyes [127, 63]. It is important to note that the use of cyclopean image model has already been investigated to assess the quality of stereo images [34, 182]. However, to the best of our knowledge, this chapter presents the first research work which exploits such model for the enhancement of stereo images.

To set the stereo weights in a way simulating the stimulus strength as suggested by Levelt's study, we used the gain control theory model [63], which is one of the latest successful schemes modeling the cyclopean perception. Since our approach operates in the wavelet transform domain, we have employed an energy weighted summation model to synthesize the final enhanced approximation subbands $\bar{I}_J^{(r)}$ and $\bar{I}_J^{(l)}$. Therefore, the weighting terms $w_J^{(r)}$ and $w_J^{(l)}$, used in Equation (6.7), are computed as follows:

$$\begin{aligned} &\forall \ v \in \{l,r\}, \quad \forall \ (m,n) \in \Omega_J, \\ &w_J^{(v)} = \frac{E_J^{(v)}}{E^{(l)} + E^{(r)}}, \end{aligned} \tag{6.8}$$

where Ω_J represents the coordinates set of the approximation coefficients I_J, and $E_J^{(v)}$ and $E^{(v)}$ are respectively the energy of the approximation subband

and all the details subbands which are given by:

$$\begin{aligned}
&\forall\ v \in \{l, r\},\\
&E_J^{(v)} = \sum_{(m,n)\in\Omega_J} |I_J^{(v)}(m,n)|^2\\
&E^{(v)} = E_J^{(v)} + \sum_{j=1}^{J} \sum_{o\in\{HL,LH,HH\}} \sum_{(m,n)\in\Omega_j} |I_j^{(o,v)}(m,n)|^2
\end{aligned} \tag{6.9}$$

x with Ω_j designates the coordinates set of the detail subband coefficients $I_j^{(o,v)}$. The stereo weights are therefore set according to the corresponding subbands energy which simulates the *stimulus strength* of Levelt's model [170]. Furthermore, we can notice that the normalization of Equation (6.8) taking into consideration both left and right views energies ensures the binocular rivalry/combination features of the HVS. Indeed, an increasing energy of one view (i.e. increasing the strength of either left or right stimulus) reduces the contribution of the other view in the overall perceived cyclopean component. A flowchart summarizing the proposed enhancement method is shown in Figure 6.2.

6.3.2 Fully adaptive inter-views processing technique

As mentioned in the previous section, the joint processing of the left and right images has been only performed on the approximation subbands of both views while their detail subbands have been separately enhanced in intra-view using Equations 3.10 and 3.11. In the second approach, we propose to add as well an efficient joint processing technique of the detail subbands yielding a fully adaptive method. The block diagram of this method is illustrated in Figure 6.3.

More precisely, similarly to the combination model used with the approximation subbands, the new enhanced detail subbands of the right image $\bar{I}_j^{(o,r)}$ and the left one $\bar{I}_j^{(o,l)}$ will be obtained as follows:

$$\forall\ o \in \{HL, LH, HH\}, \quad \forall\ j \in \{1, \ldots, J\},$$
$$\begin{cases} \bar{I}_j^{(o,r)}(m,n) = w_j^{(o,r)}(m,n).\hat{I}_j^{(o,r)}(m,n) + w_j^{(o,l)}(m,n+z_j^{(r)}).\hat{I}_j^{(o,l)}(m,n+z_j^{(r)}) \\ \bar{I}_j^{(o,l)}(m,n) = w_j^{(o,l)}(m,n).\hat{I}_j^{(o,l)}(m,n) + w_j^{(o,r)}(m,n-z_j^{(l)}).\hat{I}_j^{(o,r)}(m,n-z_j^{(l)}) \end{cases} \tag{6.10}$$

where $w_j^{(o,r)}$ and $w_j^{(o,l)}$ are the weights used for the right and left detail subbands.

To set the weights according the stimulus strength as suggested by Levelt's study [170], and since the gain-control theory is contrast sensitive [63], we

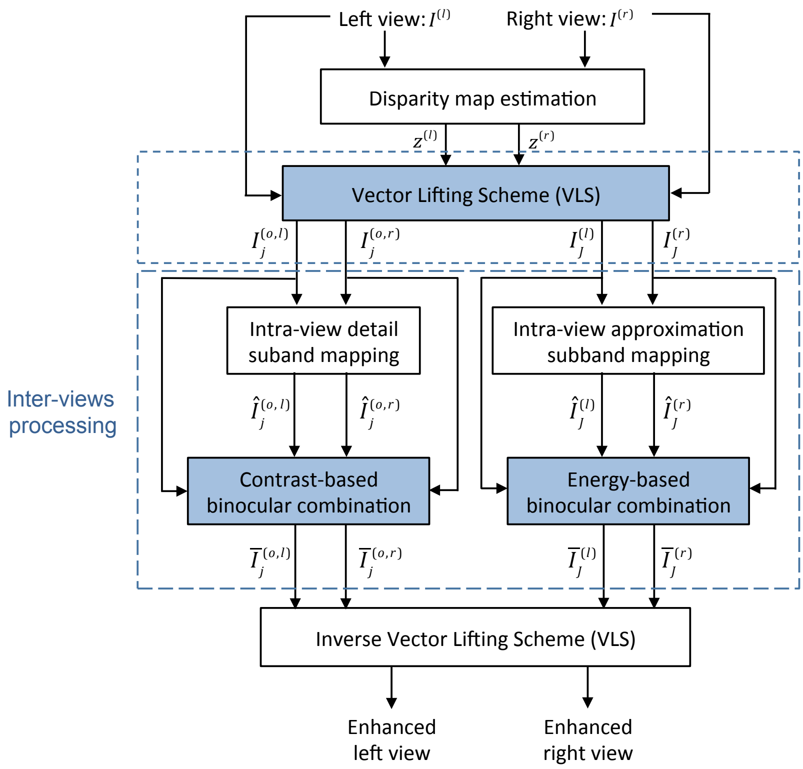


Figure 6.3: The proposed fully adaptive inter-views processing based technique.

propose to determine the details subbands weights based on a contrast measure of the left and right processed images quantifying the perceived stimulus strength in terms of contrast. Since our enhancement is performed in the wavelet domain, we adopted a contrast measure inspired of Peli's local bandlimited contrast [240] which is further extended later to process medical images in the wavelet transform domain [300, 298]. We propose, therefore, a local contrast weighted summation model to synthesize the final enhanced detail subbands $\bar{I}_j^{(o,r)}$ and $\bar{I}_j^{(o,l)}$ taking into consideration the rivalry and contrast sensitivity of the HVS. Thus, the resulting weighting terms are computed as follows:

$$\forall\ o \in \{HL, LH, HH\}, \quad \forall\ j \in \{1, \dots, J\},$$
$$\begin{cases} w_j^{(o,r)}(m,n) = \frac{C_j^{(o,r)}(m,n)}{C_j^{(o,r)}(m,n)+C_j^{(o,l)}(m,n+z_j^{(l)})} \\ w_j^{(o,l)}(m,n) = \frac{C_j^{(o,l)}(m,n)}{C_j^{(o,l)}(m,n)+C_j^{(o,r)}(m,n-z_j^{(r)})} \end{cases} \tag{6.11}$$

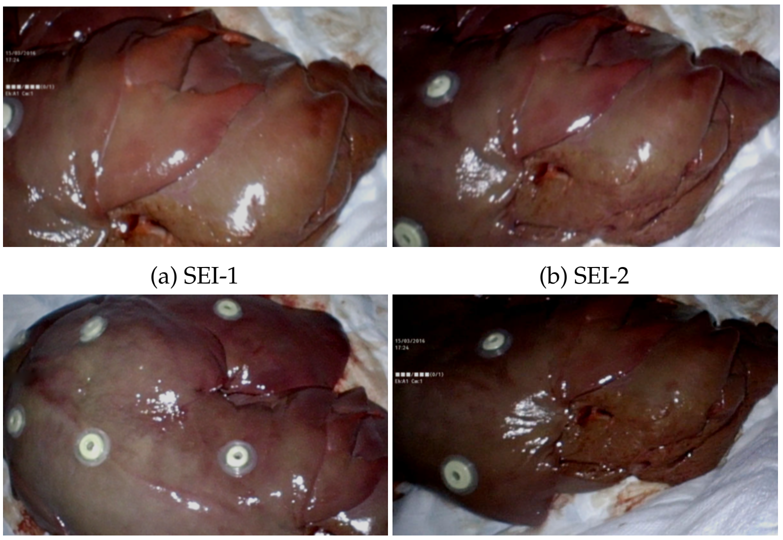
where $C_j^{(o,v)}$ denotes a contrast measure in the wavelet transform domain reflecting the HVS contrast sensitivity to details of the scene [240, 299, 300, 298]. Indeed, the local bandlimited contrast consists in attributing a contrast value to each pixel of the image for each frequency band according to the ratio between high-frequency and low-frequency information, expressed as follows:

$$C_j^{(o,v)}(m,n) = \frac{I_j^{(o,v)(m,n)}}{I_J^{(v)}(\lceil \frac{m}{2^{J-j}} \rceil, \lceil \frac{n}{2^{J-j}} \rceil)} \tag{6.12}$$

with $\lceil \cdots \rceil$ is the rounding up operator. Note that this operator is applied to the coordinates of the approximation coefficients since the size of $I_J^{(v)}$ is equal to that of $I_j^{(o,v)}$ divided by $(2^{J-j} \times 2^{J-j})$.
It is worth pointing out that, contrary to the energy-based weighted model previously presented in section 6.3.1 with the adaptive inter-views processing technique in which a similar weight is applied to all the approximation coefficients as shown in Equation (6.8), the contrast-based weights model use a local measure adapted to each pixel of the stereo image according to its local bandlimited contrast value and taking into consideration the binocular rivalry by the normalization of Equation (6.11) which adjust the contribution of each view according to its stimulus strength.

6.4 Experimental results

The simulations are carried out on two different stereo endoscopic images (SEI) datasets. The first one is composed of 10 stereo pairs, of size 1080×1728, extracted from five stereo video sequences of a pig liver acquired using different angles and zooming scales in *The Intervention Center (IVS)* at *Oslo University Hospital*. The acquisition is performed via a stereo laparoscope used in the laparoscopic surgical routine work by IVS surgeons. The choice of the pig can be explained by the fact that its liver has similar structure and tissue textures to the human liver. The second one is composed of 10 stereo images, of size 480×640, taken from different videos of *Hamlyn Centre* laparoscopic/endoscopic dataset [221]. Some SEI selected from the first and second datasets are shown in Figure 6.4 and Figure 6.5, respectively.

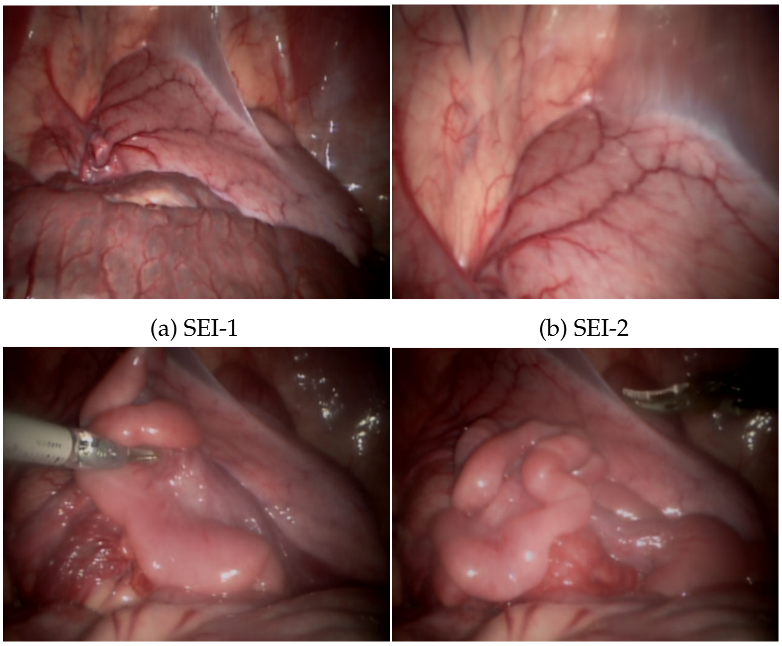


(a) SEI-1 (b) SEI-2

(c) SEI-3 (d) SEI-4

Figure 6.4: Sample right views taken from the *IVS* dataset.

In order to assess the contrast enhancement, the proposed and baseline comparison techniques are evaluated using no-reference metrics based on local and global image properties such as the edginess information, the image intensity and the local contrast. More precisely, due to the lack of stereo contrast enhancement evaluation metrics, the following 2D measures are used: Region Contrast (RC) [298], Edge Content (EC) [265], Absolute Measure of Enhancement (AME) [1], Second Derivative like Measure of Enhancement (SDME) [235], the Image Enhancement Metric (IEM) [121], Visual Information Fidelity (VIF) [276] and Mean Brightness (MB) [42]. The latter one (MB) is an indicator about the image intensity level that shows whether the image is too bright or too dark. A good MB should be close to the mid-intensity range defined as the mean of the maximum and minimum intensity values. AME combines the *Weber*'s contrast law and *Michelson*'s contrast law modulation to simulate the HVS appreciation of the image contrast enhancement. The final score is estimated based on an average value of the relation between the sum and spread of luminance intensities within each bloc. In the same way, the SDME incorporates the center pixel within each block in the computation of the visibility operator to reduce the sensitivity to noise. Both RC and EC metrics are based on the local image intensity variations computed either by Sobel operator or any other gradient-based filter depending on the level of edges to be detected and the sensitivity to noise. Inspired by the idea that variations in contrast/sharpness re-



(a) SEI-1 (b) SEI-2

(c) SEI-3 (d) SEI-4

Figure 6.5: Sample right views taken from the *Hamlyn Centre* dataset.

flect the changes between a given pixel and its neighbors, the IEM consists in comparing the absolute intra-bloc differences between the initial and the enhanced image. Inspired by the HVS, the VIF predicts the perceived quality of contrast enhanced images using an image information measure that quantifies an estimation of the information extracted by the brain when observing an image. Compared with other image quality assessment metrics, VIF is reported to be consistent and correlated with the subjective evaluation of the HVS [251, 312, 276]. Instead of using the multi-scale pixel domain implementation of the VIF, we used the available wavelet-based version provided by the authors for two main reasons. Firstly, because the latter is reported to yield better performance than the spatial domain implementation [276], and secondly since our proposed methods are processed as well in the wavelet domain. Note that these metrics are evaluated on the left and right enhanced views, and their average values correspond to the final objectives scores. Higher scores of RC, EC, IEM and VIF, and lower scores of AME and SDME, show good performance.

Since the aim of this work is not to study or improve the baseline com-

parison enhancement methods mentioned above, their implementation and test was performed using the reported default parameters. As for the settings of the proposed techniques, namely the variables of Equation 3.4, one simple idea consists in fixing a parameters vector values for all images. This solution is, however, not efficient since each image has a different spectral distribution and structure. Therefore, an automatic parameterization of the mapping function similar to the one in [42] is deployed in order to adapt the mapping parameters according to statistical distribution of each processed image. Inspired by the histogram equalization principle, the parameters estimation is based on the assumption that a proper statistical model should yield a uniform spectral distribution of the enhanced image.
Thus, under-exposed textures and tissues due to the dynamic endoscopic illumination conditions are highlighted giving a balanced distribution. This process is relevant for endoscopic images since dark regions can reach up to 40% of the spatial image resolutions in some cases. The estimation of the mapping parameters is performed by solving the following optimization problem:

$$\tilde{\Phi} = \text{argmin}_{\Phi} \sum_{k=1}^{n} \left| \frac{1}{n} - p_i \left(\frac{i}{n} f \left(\frac{i}{n}; \Phi \right) \right) \right|^2 \tag{6.13}$$

where $\Phi = \{\alpha_1, \alpha_2, \sigma_1, \sigma_2, \beta\}$ is a vector of the mapping function parameters of Equation 3.4) to be estimated dynamically and p_i is a probability density function with *i-th* intensity among n levels. A simulated annealing approach is used to solve the problem by reducing the probability of obtaining local minima. In order to optimize the results in terms of image sharpness and local image features (i.e. for features detection for instance), four annealing intervals are tested with different lower and upper bands. The process is controlled using the EC and AME metrics by selecting the output having the highest edge content score with the lowest AME value. Another reason for this control is to avoid any over-enhancement when processing the image in multi-scale decomposition. Indeed, using the optimization interval yielding the approximation with the highest edge content and lowest AME score in each level reduces the probability of having an over-enhanced output view (i.e. excessive brightness with bad local contrast). This can be explained by the simple fact that increasing or decreasing drastically the image brightness can reduce the image contrast, sharpness and saturation.

The performance of the different enhancement methods for a sample SEI as well as the average scores on all images are given in Tables 6.1-6.4 and Tables 6.5-6.8, for the *IVS* and *Hamlyn Centre* datasets respectively. For each dataset, the proposed methods are compared separately in two different tables with 2D and 3D enhancement techniques. Note that the first line of each table, denoted by "Original", indicate the initial scores computed on the input data before the enhancement step. Moreover, the two best enhancement methods have been highlighted in bold. From these tables, it can be firstly noticed that the endoscopic image enhancement methods [271, 8] as well as the 3D enhancement approach [291] often outperform the other state-of-the-art methods. Moreover, our first proposed approach "VLS-AIVP" relying on the stereo approximation subbands processing leads to better results, especially for the IVS liver dataset which is characterized by large homogeneous liver structures and more drastic dynamic illumination conditions. Finally, thanks to the fully adaptive inter-view processing technique "VLS-FAIVP", a significant gain is achieved for both dataset images in terms of image brightness, contrast and edginess information. The gain is more significant for the *Hamlyn* dataset SEI in terms of sharpness and edge information (RC and EC metrics) since these images contain much more structure details with regular/micro edges representing fine veins and vessels. Indeed, the stereo high-frequency subbands processing of the proposed "VLS-FAIVP" technique enhances the images details and local contrast, which exhibits the tissues structures and fine vessels.

In addition to this quantitative evaluation, the perceived quality of the processed images demonstrate the efficiency of the proposed methods. Figures **??** and **??** illustrate the original and enhanced right views for two SEI taken from IVS and *Hamlyn* datasets, respectively. It can be observed that the 2D endoscopic image enhancement methods [271, 8] result in an over-enhancement illustrated by both over brightened and darkened regions despite better exposure of the edges in some other areas. Furthermore, the 3D endoscopic image enhancement methods [88, 270] yield better contrasted views without efficiently enhancing the small image details. Using our approach "VLS-FAIVP", the obtained images show better brightness and contrast while improving and emphasizing the local image details such as the liver edges of the IVS dataset and the fine vessels of the *Hamlyn* dataset.

Furthermore, in order to better show the benefits of the different blocks related to the proposed approaches, we will focus now on the following comparisons. First, our methods will be also evaluated by modifying the wavelet decomposition step. Indeed, instead of using a 3D wavelet transform based on VLS, a standard 2D 5/3 Lifting Scheme (LS) is used by applying it separately to each view of the stereo pairs. These methods are denoted by "LS-AIVP" and "LS-FAIVP". Moreover, the results obtained with the wavelet-based intra-view processing technique [42] allow us to show the interest of exploiting both the depth and binocular stereo information by resorting to an inter-view processing. Tables 6.9 and 6.10 provide the performance of these latter methods. Thus, it can be seen that our approaches outperform significantly the method developed by Cho *et al.* [42] which confirms the benefits of the proposed inter-views processing techniques compared to a simple intra-view processing. Further improvements are also achieved by using a 3D wavelet decomposition (VLS) compared to a conventional 2D LS-based one. Moreover, Figs. 6.6 and 6.7 illustrate the enhanced images obtained by our approaches "VLS-AIVP" and "VLS-FAIVP" as well as that of Cho *et al.* [42]. Compared to the latter, it can be noticed that the "VLS-AIVP" leads to better image quality in terms of local contrast and image brightness. Indeed, the dark regions are reduced and the liver structures and edges are sharper and clearer. The images are further enhanced by resorting to an inter-view processing technique fully adapted to the local characteristics and activity of the stereo endoscopic images.

Finally, we have tested the efficiency of the multi-scale processing for both of the proposed methods. The results illustrated in Table 6.12 and Table 6.11 shows that the obtained scores improve generally but with different ratios between the two datasets. Indeed, the improvements are more significant for *Hamlyn Centre* dataset than the *IVS* images, especially using two decomposition levels. The explanation of these results is that *Hamlyn Centre* dataset images have smaller spatial resolution (the half approximately) and contain much more details in the High-frequency subbands (fine veins, tissues, vessels, etc.) compared with the *IVS* textureless pig liver images. Therefore, all the obtained results confirm the efficiency of the proposed enhancement methods in improving the quality of stereo endoscopic images.

Method/Measure	**MB**	**AME**	**SDME**	**EC**	**RC**	**IEM**	**VIF**
Original	122.87	60.43	134.11	65.49	5.97	-	1
VLS-AIVP	134.44	**50.35**	**105.87**	**91.69**	**33.10**	**2.52**	**1.31**
VLS-FAIVP	**128.54**	**48.14**	**98.40**	**110.33**	**46.27**	**3.03**	**1.52**
Cho *et al.* [42]	165.42	61.26	116.82	77.63	7.89	1.26	1.11
Selka *et al.* [271]	98.87	51.94	118.11	81.27	5.58	1.25	1.24
Attar *et al.* [8]	122.42	56.84	137.97	88.30	12.39	1.27	1.17

Table 6.1: Performance comparison of the proposed methods with 2D enhancement techniques for a SEI of the *IVS* dataset.

Method/Measure	**MB**	**AME**	**SDME**	**EC**	**RC**	**IEM**	**VIF**
VLS-AIVP	**121.70**	**40.33**	**91.13**	**81.19**	**20.41**	**1.97**	**1.39**
VLS-FAIVP	**123.73**	**38.59**	**88.60**	**96.53**	**23.81**	**2.19**	**1.47**
Cho *et al.* [42]	108.23	48.42	112.95	52.17	5.51	0.89	1.14
Selka *et al.* [271]	78.37	50.41	92.46	64.35	4.74	1.11	1.33
Attar *et al.* [8]	102.02	47.71	108.87	70.24	10.05	1.13	1.26

Table 6.2: Average scores of the proposed methods compared with 2D enhancement techniques for the *IVS* dataset images.

Method/Measure	MB	AME	SDME	EC	RC	IEM	VIF
Original	122.87	60.43	134.11	65.49	5.97	-	1
VLS-AIVP	134.44	**50.35**	**105.87**	**91.69**	**33.10**	**2.52**	**1.31**
VLS-FAIVP	**128.54**	**48.14**	**98.40**	**110.33**	**46.27**	**3.03**	**1.52**
Hai *et al.* [88]	115.11	54.04	118.86	80.58	8.47	1.29	1.17
Sdiri *et al.* [270]	122.86	53.78	120.20	86.97	11.00	1.52	1.19
Subedar and Karam [291]	**123.10**	50.79	118.18	88.33	25.31	2.01	1.16
Hachicha *et al.* [87]	122.31	56.66	122.71	79.06	7.74	1.26	1.14

Table 6.3: Performance comparison of the proposed methods with stereo enhancement techniques for a SEI of the *IVS* dataset.

Method/Measure	MB	AME	SDME	EC	RC	IEM	VIF
VLS-AIVP	**121.70**	**40.33**	**91.13**	**81.19**	**20.41**	**1.97**	**1.39**
VLS-FAIVP	**123.73**	**38.59**	**88.60**	**96.53**	**23.81**	**2.19**	**1.47**
Hai *et al.* [88]	87.26	44.60	98.15	55.64	5.87	0.98	1.15
Sdiri *et al.* [270]	113.04	41.72	99.59	79.67	10.54	1.57	1.17
Subedar and Karam [291]	90.63	49.19	96.44	67.08	17.93	1.80	1.14
Hachicha *et al.* [87]	90.01	46.28	103.34	52.89	5.46	0.92	1.13

Table 6.4: Average scores of the proposed methods compared with stereo enhancement techniques for the *IVS* dataset images.

Method/Measure	MB	AME	SDME	EC	RC	IEM	VIF
Original	100.56	57.22	137.83	64.95	42.75	-	1
VLS-AIVP	**123.08**	**50.35**	**101.54**	**99.98**	**92.20**	**1.98**	**1.19**
VLS-FAIVP	**130.54**	**48.14**	**96.04**	**119.25**	**126.27**	**2.65**	**1.28**
Cho *et al.* [42]	150.58	59.90	112.38	84.62	67.65	1.44	1.09
Selka *et al.* [271]	98.41	54.04	107.12	70.63	40.01	1.03	1.04
Attar *et al.* [8]	117.58	50.41	147.67	88.77	45.07	1.05	1.06

Table 6.5: Performance comparison of the proposed methods with 2D enhancement techniques for a SEI of the *Hamlyn Centre* dataset.

Method/Measure	MB	AME	SDME	EC	RC	IEM	VIF
VLS-AIVP	126.71	**42.78**	**99.40**	99.31	**66.36**	**1.87**	**1.32**
VLS-FAIVP	**128.35**	**40.44**	**93.22**	**119.81**	**87.03**	**2.59**	**1.48**
Cho *et al.* [42]	130.41	60.20	116.20	71.94	29.83	0.88	1.09
Selka *et al.* [271]	93.53	45.32	122.59	84.56	22.53	1.02	1.07
Attar *et al.* [8]	**127.54**	47.01	135.32	**110.99**	45.46	1.21	1.13

Table 6.6: Average scores of the proposed methods compared with 2D enhancement methods for the *Hamlyn Centre* dataset images.

Method/Measure	MB	AME	SDME	EC	RC	IEM	VIF
Original	100.56	57.22	137.83	64.95	42.75	-	1
VLS-AIVP	**123.08**	**50.35**	**101.54**	**99.98**	**92.20**	**1.98**	**1.19**
VLS-FAIVP	**130.54**	**48.14**	**96.04**	**119.25**	**126.27**	**2.65**	**1.28**
Hai *et al.* [88]	103.10	53.63	127.02	74.64	42.71	1.19	1.09
Sdiri *et al.* [270]	100.47	53.26	129.76	82.20	68.12	1.45	1.11
Subedar and Karam [291]	100.66	52.18	110.50	77.05	64.66	1.73	1.13
Hachicha *et al.* [87]	100.53	56.91	136.91	65.96	43.23	1.02	1.03

Table 6.7: Performance comparison of the proposed methods with stereo enhancement techniques for a SEI of the *Hamlyn Centre* dataset.

Method/Measure	MB	AME	SDME	EC	RC	IEM	VIF
VLS-AIVP	126.71	**42.78**	**99.40**	99.31	**66.36**	**1.87**	**1.32**
VLS-FAIVP	**128.35**	**40.44**	**93.22**	**119.81**	**87.03**	**2.59**	**1.48**
Hai *et al.* [88]	105.34	50.77	119.66	85.15	39.97	1.24	1.11
Sdiri *et al.* [270]	104.33	50.46	120.57	90.38	51.44	1.35	1.14
Subedar and Karam [291]	104.43	48,12	112.72	90.91	53.62	1.24	1.13
Hachicha *et al.* [87]	104.31	54.61	128.39	73.21	36.88	1.04	1.10

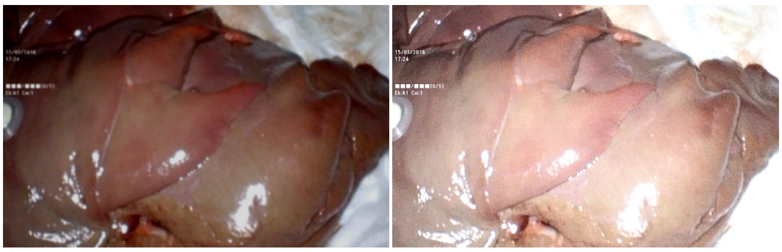
Table 6.8: Average scores of the proposed methods compared with stereo enhancement methods for the *Hamlyn Centre* dataset images.

Method/Measure	MB	AME	SDME	EC	RC	IEM	VIF	
Original	122.87	60.43	134.11	65.49	5.97	-	1	**SEI-1**
Cho *et al.* [42]	165.42	61.26	116.82	77.63	7.89	1.26	1.11	
LS-AIVP	148.12	57.26	109.00	83.86	21.45	1.67	1.27	
VLS-AIVP	134.44	50.35	105.87	91.69	33.10	2.52	1.31	
LS-FAIVP	126.94	52.25	106.33	94.13	34.64	2.56	1.47	
VLS-FAIVP	128.54	48.14	98.40	110.33	46.27	3.03	1.52	
Original	97.68	53.87	126.75	66.80	5.31	-	1	**SEI-2**
Cho *et al.* [42]	130.62	52.54	108.79	83.92	7.98	1.35	1.11	
LS-AIVP	103.25	50.54	102.88	77.58	10.44	1.47	1.16	
VLS-AIVP	120.59	43.46	96.67	98.44	16.38	1.70	1.19	
LS-FAIVP	122.10	50.19	107.49	94.53	22.42	1.90	1.29	
VLS-FAIVP	126.12	44.35	91.73	104.34	28.48	2.68	1.33	

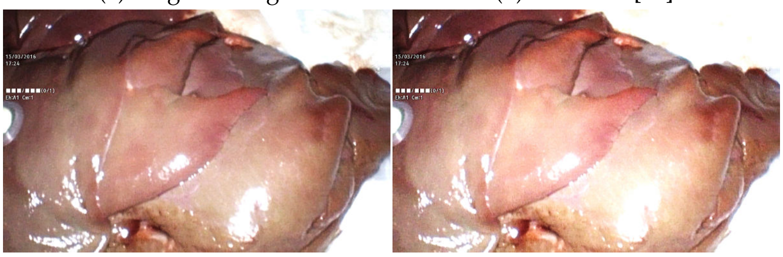
Table 6.9: Performance of LS and VLS-based enhancement methods for two SEI of the *IVS* dataset.

Method/Measure	MB	AME	SDME	EC	RC	IEM	VIF	
Original	100.56	57.22	137.83	64.95	42.75	-	1	**SEI-1**
Cho *et al.* [42]	150.58	59.90	112.38	84.62	67.65	1.44	1.09	
LS-AIVP	121.51	52.45	104.83	96.11	86.19	1.81	1.16	
VLS-AIVP	123.08	50.95	101.54	99.98	92.20	1.98	1.19	
LS-FAIVP	139.55	51.58	102.76	102.90	110.50	2.21	1.24	
VLS-FAIVP	**130.54**	**48.14**	**96.04**	**119.25**	**126.27**	**2.65**	1.28	
Original	93.84	48.66	139.32	55.98	28.04	-	1	**SEI-2**
Cho *et al.* [42]	147.06	59.08	111.94	82.50	47.96	1.63	1.11	
LS-AIVP	116.14	52.97	104.45	85.53	47.04	1.92	1.13	
VLS-AIVP	117.23	48.02	102.78	87.08	49.44	2.02	1.17	
LS-FAIVP	114.75	49.67	105.39	93.24	52.83,	2.40	1.20	
VLS-FAIVP	128.12	47.35	97.53	102.18	58.11	2.78	1.23	

Table 6.10: Performance of LS and VLS-based enhancement methods for two SEI of the *Hamlyn Centre* dataset.

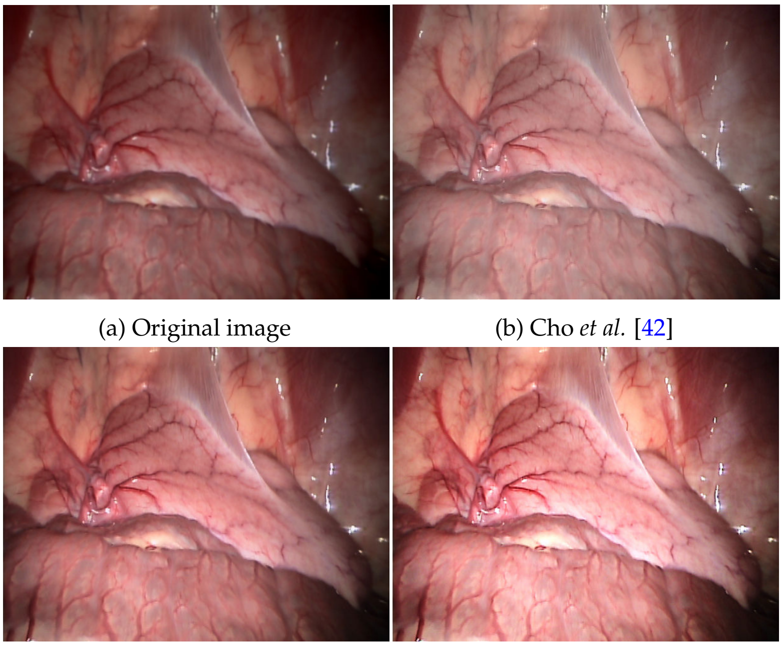


(a) Original image (b) Cho *et al.* [42]



(c) VLS-AIVP (d) VLS-FAIVP

Figure 6.6: The enhanced right images for the SEI-1 of *IVS* dataset using the two proposed inter-views processing techniques and the intra-view processing one [42].



(a) Original image (b) Cho *et al.* [42]

(c) VLS-AIVP (d) VLS-FAIVP

Figure 6.7: The enhanced right images for the SEI-1 of *Hamlyn Centre* dataset using the two proposed inter-views processing techniques and the intra-view processing one [42].

Method/Measure	MB	AME	SDME	EC	RC	IEM	
Original	122.87	60.43	134.11	65.49	5.97	-	**IMG1**
VLS-AIVP (J=1)	137.96	55.20	110.33	69.69	23.92	1.93	
VLS-AIVP (J=2)	134.44	50.35	105.87	91.69	33.10	2.52	
VLS-FAIVP (J=1)	132.58	54.34	103.04	78.15	31.48	2.40	
VLS-FAIVP (J=2)	128.54	48.14	98.40	110.33	46.27	3.03	
Original	97.68	53.87	126.75	66.80	5.31	-	**IMG2**
VLS-AIVP (J=1)	116.61	49.20	102.43	78.14	14.00	1.43	
VLS-AIVP (J=2)	120.59	43.46	96.67	98.44	16.38	1.70	
VLS-FAIVP (J=1)	110.76	45.52	95.41	86.41	17.81	2.53	
VLS-FAIVP (J=2)	126.12	44.35	91.73	104.34	28.48	2.68	

Table 6.11: Multiscale performance of the proposed methods on two sample SEI from*IVS*dataset.

Method/Measure	MB	AME	SDME	EC	RC	IEM	
Original	100.56	57.22	137.83	64.95	42.75	-	**IMG1**
VLS-AIVP (J=1)	126.16	54.33	105.01	80.74	86.41	1.78	
VLS-AIVP (J=2)	123.08	50.95	101.54	99.98	92.20	1.98	
VLS-FAIVP (J=1)	126.17	52.06	101.30	85.66	110.97	2.24	
VLS-FAIVP (J=2)	130.54	48.14	96.04	119.25	126.27	2.65	
Original	93.84	48.66	139.32	55.98	28.04	-	**IMG2**
VLS-AIVP (J=1)	120.79	48.37	105.94	73.78	46.62	1.83	
VLS-AIVP (J=2)	117.23	48.02	102.78	87.08	49.44	2.02	
VLS-FAIVP (J=1)	120.81	53.20	101.68	77.55	54.01	2.39	
VLS-FAIVP (J=2)	128.12	47.35	97.53	102.18	58.11	2.78	

Table 6.12: Multiscale performance of the proposed methods on two sample SEI from *Hamlyn Centre* dataset.

6.5 Conclusion

In this chapter, we have focused on the enhancement of stereo endoscopic images. For this purpose, two efficient methods have been designed in the wavelet transform domain. More precisely, a joint 3D wavelet decomposition is developed to generate the wavelet subbands of the left and right views, and two inter-views processing techniques are proposed to exploit the binocular vision properties via a cross-view processing. Experimental results show the efficiency of the proposed techniques compared to the state-of-the-art ones and demonstrate the usefulness of the interview processing for both the stereo wavelet decomposition and the enhancement taking into consideration the binocular combination of the HVS and exploiting interview redundancies.

Chapter 7

Conclusion

> The end of a work should always suggest the beginning.
>
> JOSEPH JOUBERT

7.1 Summary of contributions

Endoscopic image enhancement has become a very active research field thanks to the success of minimally invasive interventions and novel technological treatment and diagnosis tools such as stereoscopic laparoscopes and wireless capsule endoscopy. In spite of the important advances achieved in terms of image processing and enhancement, only a few techniques have been proposed or adapted to stereo endoscopic images. This can be explained by the specificities of the stereo endoscopic video acquisition process; the endoscopic domain characteristics including moist tissues with high reflectance and dynamic illumination conditions yielding large dark areas with specular reflections; and surgical tasks artifacts including blur due to patient's and surgeon's motion and surgical smoke.

We recall as well that we distinguish two main aims for endoscopic image enhancement based on both our literature study and our clinical research context. The first aim is to improve the perceived endoscopic image quality by exhibiting specific tissue details and abnormalities in order to improve the diagnosis accuracy and reliability. The intra-operative visual feedback quality is also of great importance for surgeons since it can affect the efficiency of their tasks. Indeed, the intra-operative endoscopic videos should highlight some significant details such as tissues texture, organs boundaries and blood vessels. For instance, liver blood vessels should be blocked before performing a resection in order to avoid causing bleeding, which can compromise the patient safety. If not properly controlled, the bleeding may oblige the doctors to convert to an open surgery, in which case the patient gets the worst of both MIS and OS: increasing the operating time without any post-operative advantage. The second possible goal of endoscopic image enhancement is to improve the outcomes of subsequent post-processing tasks such as features extraction for 3D organ reconstruction and registration [238]. For some laparoscopic surgeries, namely for liver resection operations at the *Intervention Centre* of *Oslo University Hospital*, these tasks are prerequisite to register the 3D patient-specific data, establish resection and navigation plans, and provide the surgeon with an efficient control of robotic-assisted surgical systems.

Since this thesis deals with 3D images, understanding the human visual system and its binocular features was prominent to comprehend the factors that may affect the perceived quality of stereo images and the different 3D capture/display systems and their parameters. Therefore, Chapter 2 reviewed the idea of depth perception with the fundamentals of human binocular vision and emphasized the relationship between left and right views of stereo images from a geometrical point of view. In addition to that, the chapter outlined the most important methods for disparity information estimation, which is crucial to exploit the cross-view redundancies, and reviewed the relationship between depth and disparity maps. To identify endoscopic imaging issues, Chapter 3 presented the principal challenges and problems related to endoscopic images together with the main artifacts and noise that can occur. The chapter investigated and categorized as well the different existent technologies and methods for endoscopic image enhancement.

The contributions presented in this dissertation are all linked to at least one or both enhancement purposes recalled above (i.e. improve visual quality or post processing tasks outcomes). In Chapter 4, we presented an enhancement method for 2D endoscopic images and wireless capsule endoscopy improving both local and global contrast. Beside targeting the previously mentioned enhancement goals above, the proposed method aims to exhibit inner subtle structures and tissues details, which improves the features detection process and the automatic classification of GI tract abnormalities for cancer detection. It has been proven through this work that the enhanced images using the proposed method improved the classification rate by up to 6% and 8.5% for inflammation and polyps detection respectively on the bases of manually labeled frames by expert physicians and doctors according to each possible abnormality. The performance of the processing was evaluated as well in terms of matched detected features since such parameter can impact significantly the 3D organ surface reconstruction from stereo endoscopic images [238]. The obtained results demonstrate that our enhancement increased the number of stereo matched features and allowed detecting feature points in regions with poor contrast and low brightness. The visual quality of the processed WCE frames and endoscopic images shows that our processing adjusts well the brightness and improves the contrast of original images, without any over-enhancement while preserving the image naturalness property in terms of color and illumination distribution, which is of great importance for endoscopic images since it can affect the diagnosis.

In a stereo imaging context, Chapter 5 introduced an adaptive enhancement technique for stereo endoscopic images combining depth and edginess information. Inspired by a perceptual study [291, 174] stipulating that the HVS attributes more attention and is more sensitive to closer objects of the scene due to its projection in the fovea, the contrast enhancement degree of of the proposed technique is adjusted locally to each object according to its depth level. The other aspect of adaptability related to the proposed method is that the enhancement adjustment according to the local image activity (i.e. whether the processed region is homogeneous or presents edges/textures) in order to avoid any over-enhancement, halo effect, or noise amplification. The following step consists in controlling the intra-view processing using the binocular just noticeable difference model to suppress any inter-view difference exceeding the BJND visibility threshold, which may result in visible artifacts or visual fatigue. In order to assess the perceived quality of the enhanced images, we performed subjective tests in a controlled environment including both expert and non-expert observers who were invited to evaluate each displayed image processed using either our proposed technique or some state-of-the-art methods and assign a score according to a standard grading quality scale. The obtained subjective scores demonstrated the effectiveness of the proposed adaptive edge-based stereo enhancement method in generating more contrasted images with sharper edges and micro-edges, without resulting in any over-enhancement or halo effect. The statistical analysis performed on the resulting experimental data proves statistically the effectiveness of the proposed enhancement technique with less than 0.005% risk of concluding that a visible image quality difference exists when there is no actual perceived difference.

Notice that all the previously presented enhancement methods were performed in the spatial domain, which offered the advantage of a low computational complexity allowing the real-time execution of our proposed techniques. The spatial domain does not allow, however, to target the enhancement of specific image components or to perform multiscale processing with low computational cost. Therefore, with this in mind, we resorted in another recent work to the wavelet domain for its intrinsic properties including multiscale representation and good space-frequency localization. Thus, Chapter 6 presented a novel enhancement algorithm for stereo endoscopic images using a joint wavelet decomposition to exploit the inter-view redundancies and depth data. The proposed enhancement consists in mapping the approximation subband to adjust the image illumination and processing accordingly the corresponding detail subbands coefficients to improve the local image contrast together with shrinking the trivial signals that represent noise. The mapping parameters are computed adaptively for each image using an annealing optimization process that accounts on its statistical histogram data to ensure a uniform spectral distribution. An inter-view processing is then performed on each intra-view enhanced image to generate the final enhanced version using a weighted combination model inspired from the rivalry/combination features of the HVS and exploiting depth information and interview redundancies. The stereo weighting models are computed in a way that determines adaptively the weight of each left and right view pixel according to the strength of its related stimuli, which is simulated using the subbands energy [63] and local bandlimited contrast measures in the wavelet domain [240, 298, 300]. To evaluate the performance of the proposed algorithms, the images enhanced using our methods and other baseline comparison techniques, for both 2D and 3D endoscopic and regular images, are assessed using some conventional metrics based on local and global image features. The obtained objective scores demonstrated the efficiency of the proposed techniques compared to the 2D endoscopic and stereo enhancement methods in both the spatial and transform domains. The visual quality of the enhanced images shows that our technique adjust well the endoscopic images brightness by reducing the dark regions and adjusting the contrast in saturated areas. Furthermore, the details subbands processing improved local contrast and sharpness of the images by exhibiting liver tissue textures, vessels and boundaries including edges and micro-edges, without yielding any over-enhancement or halo effects.

In the context of the evaluation of contrast enhancement techniques, we noticed the lack of a benchmark allowing the assessment of different CE evaluation metrics, which can be very useful especially with the rapidly increasing number of CE methods and CEE metrics for various application domains. Additionally, the few studies addressing this topic in the literature focus on evaluating image quality assessment metrics designed mainly to distorted or degraded images, without any attention to contrast enhanced images. To tackle this problem, we proposed in the first part of Chapter 4 a database containing original images with various color distributions, textures, contrast variations, and different realistic contrast enhancement artifacts; the associated contrast enhanced versions using representatives of the most common CE techniques in the literature; the corresponding subjective evaluation scores; objective assessment measures; and statistical correlation analysis data. The proposed dataset, which is publicity available online[1], allows a comprehensive analysis of the performance of different state-of-the-art contrast enhancement evaluation metrics in terms of correlation with the HVS appreciation. The latter is provided by the scores resulting from the subjective evaluation experiment performed in a proper controlled test environment for CEE. The database can be useful as well to evaluate both newly proposed contrast enhancement techniques and CEE metrics.

[1]MENDELEY data plateform: Contrast Enhancement Evaluation Database (CEED2016)

7.2 Future research

In the course of the research activities carried out as a part of this thesis, a number of possible directions for further studies related to endoscopic image enhancement have been identified for different parts of this dissertation.

While both spatial and wavelet domains have been used in the proposed enhancement techniques, we believe that other transform domains should be considered and investigated for the particular characteristics of endoscopic images including unbalanced brightness due to dynamic illumination conditions in the endoscopic environment and the fine tissues textures, micro-edges, and subtle details (e.g. blood vessels, cancer tumor tissues, and thin nerves) that might be of great importance for the early diagnosis of diseases or abnormalities classification. We think particularly of the pyramidal decomposition which allows efficient multiscale decomposition and smooth processing of large structures [64]. Furthermore, unlike the classical wavelets which are sensitive to three directions (i.e. horizontal, vertical and diagonal), geometric wavelets such as curvelets [30, 29], contourlet [65], and steerable filters [74] allow the capture of very fine details and salient features in many directions which is very useful for endoscopic images. In the case of inter-view processing of stereo images, saliency models should be investigated based on the most relevant features for doctors and surgeons. In the same way, stereo contrast enhancement evaluation metrics should be investigated, namely for 3D endoscopic and medical images in which contrast is a very important feature that can affect diagnosis and doctors/surgeons work.

In another perspective, the rapid rate of growth of voluminous and complex biomedical data has paved the way to an analysis/processing trend that aims to extract relevant valuable information from this "big data" : machine learning (ML). Particularly, deep learning (DL) which is a branch of ML, has recently attracted the attention of researchers for biomedical data processing due to its higher level of abstraction based on distributed and parallel computing and sophisticated algorithms to converge to "data-driven" features. For instance Google has launched recently *DeepMind Health*[2] [281] technology to facilitate the transition of patients from diagnostic test to treatment steps as accurately and quickly as possible. In biomedical imaging, although that DL has been used mainly for segmentation [242, 218] or abnormalities detection and classification such as proposed in Chapter 4 for WCE anomaly classification or recently in [263, 284, 66, 6], we believe that neural networks, namely convolution neural networks (CNNs) can be used to estimate proper mapping parameters for endoscopic image enhancement or to design a contrast enhancement evaluation metric for 2D and especially 3D endoscopic/medical images. The choice of CNNs can be explained by

[2] https://www.deepmind.com/health

its adaptability for 2D images and its design inspired by the powerful processing of the visual cortex to simulate two main features: local connectivity, and invariance to local transition and location. One of the main key points related to DL technologies is that the training benefits and relies on large datasets. Developing large medical and endoscopic image datasets is, however, a challenging task for two main reasons: the expensive medical expertise required for good annotation, and the ethical issues related to the processing or sharing of private patients' data.

Last but not least, with the increasing optical and technological evolution of capture and visualization systems especially in terms of resolution and the merging trend of using ultra high-definition 4K screens in the operating theatre and endoscopic towers [326], we believe that the need for contrast enhancement methods for endoscopic images will be reduced because of the very fine structures and tissues HD visualization and recognition provided by such devices. Indeed, 4K technologies are expected to open up new intricate surgeries such as the anastomoses of blood vessels and thin nerves, and provide more confidence for different surgical resection procedures of cancer tissues [326].

Bibliography

[1] AGAIAN, S. S., SILVER, B., AND PANETTA, K. A. Transform coefficient histogram-based image enhancement algorithms using contrast entropy. *Transactions on Image Processing 16*, 3 (2007), 741–758. 123

[2] AGHA, R., AND MUIR, G. Does laparoscopic surgery spell the end of the open surgeon? *Journal of the Royal Society of Medicine 96*, 11 (2003), 544–546. 2

[3] AGRAWAL, A., RASKAR, R., NAYAR, S. K., AND LI, Y. Removing photography artifacts using gradient projection and flash-exposure sampling. *ACM Transactions on Graphics 24*, 3 (2005), 828–835. 40

[4] ALVAREZ, L., DERICHE, R., SANCHEZ, J., AND WEICKERT, J. Dense disparity map estimation respecting image discontinuities: A pde and scale-space based approach. *Journal of Visual Communication and Image Representation 13*, 1 (2002), 3–21. 32

[5] ANGELOPOULOU, E. Specular highlight detection based on the fresnel reflection coefficient. In *International Conference on Computer Vision* (2007), IEEE, pp. 1–8. 42

[6] ANTHIMOPOULOS, M., CHRISTODOULIDIS, S., EBNER, L., CHRISTE, A., AND MOUGIAKAKOU, S. Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. *IEEE transactions on medical imaging 35*, 5 (2016), 1207–1216. 137

[7] ARTUSI, A., BANTERLE, F., AND CHETVERIKOV, D. A survey of specularity removal methods. In *Computer Graphics Forum* (2011), vol. 30, Wiley Online Library, pp. 2208–2230. 41

[8] ATTAR, A., XIE, X., ZHANG, C., WANG, Z., AND YUE, S. Wireless micro-ball endoscopic image enhancement using histogram information. In *International Conference of the Engineering in Medicine and Biology Society* (2014), IEEE, pp. 3337–3340. 63, 64, 126, 127, 128

[9] BAJCSY, R., LEE, S. W., AND LEONARDIS, A. Detection of diffuse and specular interface reflections and inter-reflections by color image

segmentation. *International Journal of Computer Vision 17*, 3 (1996), 241–272. 40

[10] BAKER, D. H., AND GRAF, E. W. On the relation between dichoptic masking and binocular rivalry. *Vision research 49*, 4 (2009), 451–459. 13

[11] BARBOSA, D. J., RAMOS, J., CORREIA, J. H., AND LIMA, C. S. Automatic detection of small bowel tumors in capsule endoscopy based on color curvelet covariance statistical texture descriptors. In *International Conference of the Engineering in Medicine and Biology Society* (2009), IEEE, pp. 6683–6686. 76

[12] BARENDREGT, M., HARVEY, B. M., ROKERS, B., AND DUMOULIN, S. O. Transformation from a retinal to a cyclopean representation in human visual cortex. *Current Biology 25*, 15 (2015), 1982–1987. 13

[13] BARLOW, H. Temporal and spatial summation in human vision at different background intensities. *The Journal of physiology 141*, 2 (1958), 337–350. 67

[14] BAYLOR, D., LAMB, T., AND YAU, K.-W. Responses of retinal rods to single photons. *The Journal of physiology 288*, 1 (1979), 613–634. 11

[15] BECKER, V., VERCAUTEREN, T., VON WEYHERN, C. H., PRINZ, C., SCHMID, R. M., AND MEINING, A. High-resolution miniprobe-based confocal microscopy in combination with video mosaicing (with video). *Gastrointestinal endoscopy 66*, 5 (2007), 1001–1007. 48

[16] BEGHDADI, A., LARABI, M.-C., BOUZERDOUM, A., AND IFTEKHARUDDIN, K. M. A survey of perceptual image processing methods. *Signal Processing: Image Communication 28*, 8 (2013), 811–831. 70, 99

[17] BEGHDADI, A., AND LE NEGRATE, A. Contrast enhancement technique based on local detection of edges. *Computer Vision, Graphics, and Image Processing 46*, 2 (1989), 162–174. 53, 54, 65, 70, 79, 83, 85, 93, 94, 95, 103

[18] BENSALMA, R., AND LARABI, M.-C. Towards a perceptual quality metric for color stereo images. In *International Conference on Image Processing* (2010), IEEE, pp. 4037–4040. 108

[19] BHAT, Y. M., DAYYEH, B. K. A., CHAUHAN, S. S., GOTTLIEB, K. T., HWANG, J. H., KOMANDURI, S., KONDA, V., LO, S. K., MANFREDI, M. A., MAPLE, J. T., ET AL. High-definition and high-magnification endoscopes. *Gastrointestinal endoscopy 80*, 6 (2014), 919–927. 43, 44

[20] BHAYANI, S. B., AND ANDRIOLE, G. L. Three-dimensional (3d) vision: does it improve laparoscopic skills? an assessment of a 3d head-mounted visualization system. *Reviews in urology 7*, 4 (2005), 211. 2

[21] BIGONY, L. Risks associated with exposure to surgical smoke plume: a review of the literature. *AORN journal 86*, 6 (2007), 1013–1024. 42

[22] BOCK, H.-H. Clustering methods: a history of k-means algorithms. *Selected contributions in data analysis and classification* (2007), 161–172. 87

[23] BOUARFA, L., AKMAN, O., SCHNEIDER, A., JONKER, P. P., AND DANKELMAN, J. In-vivo real-time tracking of surgical instruments in endoscopic video. *Minimally Invasive Therapy & Allied Technologies 21*, 3 (2012), 129–134. 42

[24] BOYKOV, Y., VEKSLER, O., AND ZABIH, R. Fast approximate energy minimization via graph cuts. *Transactions on pattern analysis and machine intelligence 23*, 11 (2001), 1222–1239. 32

[25] BRUCE, V., GREEN, P. R., AND GEORGESON, M. A. *Visual perception: Physiology, psychology, & ecology*. Psychology Press, 2003. 15

[26] BRYANT, R., LAURENT, A., TAYAR, C., VAN NHIEU, J. T., LUCIANI, A., AND CHERQUI, D. Liver resection for hepatocellular carcinoma. *Surgical oncology clinics of North America 17*, 3 (2008), 607–633. 4

[27] CALDERBANK, A., DAUBECHIES, I., SWELDENS, W., AND YEO, B.-L. Wavelet transforms that map integers to integers. *Applied and computational harmonic analysis 5*, 3 (1998), 332–369. 56

[28] CAMPBELL, F., AND GREEN, D. Monocular versus binocular visual acuity. *Nature 208*, 5006 (1965), 191–192. 13

[29] CANDES, E., DEMANET, L., DONOHO, D., AND YING, L. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation 5*, 3 (2006), 861–899. 137

[30] CANDES, E. J., AND DONOHO, D. L. Curvelets: A surprisingly effective nonadaptive representation for objects with edges. Tech. rep., Stanford Univ Ca Dept of Statistics, 2000. 137

[31] CARBAJO-RODRÍGUEZ, H., AGUAYO-ALBASINI, J. L., SORIA-ALEDO, V., AND GARCÍA-LÓPEZ, C. Surgical smoke: risks and preventive measures. *Cirugía Española (English Edition) 85*, 5 (2009), 274–279. 42

[32] CHANG, S. G., YU, B., AND VETTERLI, M. Adaptive wavelet thresholding for image denoising and compression. *Transactions on image processing 9*, 9 (2000), 1532–1546. 56, 114

[33] CHAUHAN, S. S., DAYYEH, B. K. A., BHAT, Y. M., GOTTLIEB, K. T., HWANG, J. H., KOMANDURI, S., KONDA, V., LO, S. K., MANFREDI, M. A., MAPLE, J. T., ET AL. Confocal laser endomicroscopy. *Gastrointestinal endoscopy 80*, 6 (2014), 928–938. 38

[34] CHEN, M.-J., SU, C.-C., KWON, D.-K., CORMACK, L. K., AND BOVIK, A. C. Full-reference quality assessment of stereopairs accounting for rivalry. *Signal Processing: Image Communication 28*, 9 (2013), 1143–1155. 35, 119

[35] CHEN, S., AND BEGHDADI, A. Natural rendering of color image based on retinex. In *International Conference on Image Processing* (2009), IEEE, pp. 1813–1816. 78, 79, 81, 83, 85

[36] CHEN, S., AND BEGHDADI, A. Natural enhancement of color image. *EURASIP Journal on Image and Video Processing 2010*, 1 (2010), 175203. 79

[37] CHEN, S.-D., AND RAMLI, A. R. Minimum mean brightness error bi-histogram equalization in contrast enhancement. *Transactions on Consumer Electronics 49*, 4 (2003), 1310–1319. 70

[38] CHEN, T., GOESELE, M., AND SEIDEL, H.-P. Mesostructure from specularity. In *Conference on Computer Vision and Pattern Recognition* (2006), vol. 2, IEEE, pp. 1825–1832. 41

[39] CHENG, H., MIN, R., AND ZHANG, M. Automatic wavelet base selection and its application to contrast enhancement. *Signal Processing 90*, 4 (2010), 1279–1289. 57, 60

[40] CHO, D., AND BUI, T. D. Multivariate statistical modeling for image denoising using wavelet transforms. *Signal Processing: Image Communication 20*, 1 (2005), 77–89. 60

[41] CHO, D., AND BUI, T. D. Image contrast enhancement in compressed wavelet domain. In *International Conference on Image Processing* (2011), IEEE, pp. 3421–3424. 57

[42] CHO, D., AND BUI, T. D. Fast image enhancement in compressed wavelet domain. *Signal Processing 98* (2014), 295–307. 23, 57, 58, 60, 123, 125, 126, 127, 128, 129, 130, 131

[43] CHOI, K.-S., AND JUNG, H.-Y. Confocal laser endomicroscopy and molecular imaging in barrett esophagus and stomach. *Clinical endoscopy 47*, 1 (2014), 23. 48

[44] CLERC, M., AND MALLAT, S. The texture gradient equation for recovering shape from texture. *Transactions on Pattern Analysis and Machine Intelligence 24*, 4 (2002), 536–549. 25

[45] CONNOR, C. E., BRINCAT, S. L., AND PASUPATHY, A. Transformation of shape information in the ventral pathway. *Current opinion in neurobiology 17*, 2 (2007), 140–147. 12

[46] CORNSWEET, T. N. *Visual Perception* (1970). 53

[47] CURVERS, W. L., SINGH, R., SONG, L. W.-K., WOLFSEN, H. C., RAGUNATH, K., WANG, K., WALLACE, M. B., FOCKENS, P., AND BERGMAN, J. Endoscopic tri-modal imaging for detection of early neoplasia in barrett's oesophagus: a multi-centre feasibility study using high-resolution endoscopy, autofluorescence imaging and narrow band imaging incorporated in one endoscopy system. *Gut 57*, 2 (2008), 167–172. 48

[48] CUTTING, J., AND VISHTON, P. Perceiving layout and knowing distances : the integration, relative potency and contextual use of different information about depth. In *William Epstein and Rogers Sheena*, P. of space and motion, Eds. Academic press, 1995, ch. 3, pp. 71–110. 25, 17, 19

[49] CYGANEK, B., AND SIEBERT, J. P. *An introduction to 3D computer vision techniques and algorithms*. John Wiley & Sons, 2011. 30

[50] DALY, S. The visible differences predictor: an algorithm for the assessment of image fidelity. *Digital images and human vision 4* (1993), 124–125. 79

[51] DE, T. A simple programmable s-function for digital image processing. In *International Technical Conference of IEEE Region 10* (1989), IEEE, pp. 573–576. 59

[52] DE RIDDER, H. Naturalness and image quality: saturation and lightness variation in color images of natural scenes. *Journal of imaging science and technology 40*, 6 (1996), 487–493. 78

[53] DE SILVA, D. V. S., EKMEKCIOGLU, E., FERNANDO, W. A. C., AND WORRALL, S. T. Display dependent preprocessing of depth maps based on just noticeable depth difference modeling. *Journal of Selected Topics in Signal Processing 5*, 2 (2011), 335–351. 96

[54] DE SILVA, D. V. S., FERNANDO, W. A. C., WORRALL, S. T., YASAKETHU, S., AND KONDOZ, A. M. Just noticeable difference in depth model for stereoscopic 3d displays. In *Multimedia and Expo*

(ICME), 2010 IEEE International Conference on (2010), IEEE, pp. 1219–1224. 96

[55] DEEBA, F., MOHAMMED, S. K., BUI, F. M., AND WAHID, K. A. Unsupervised abnormality detection using saliency and retinex based color enhancement. In *International Conference of the Engineering in Medicine and Biology Society* (2016), IEEE, pp. 3871–3874. 62, 64

[56] DEMIREL, H., AND ANBARJAFARI, G. Image resolution enhancement by using discrete and stationary wavelet decomposition. *Transactions on Image Processing 20*, 5 (2011), 1458–1460. 114

[57] DEMIREL, H., OZCINAR, C., AND ANBARJAFARI, G. Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition. *IEEE Geoscience and remote sensing letters 7*, 2 (2010), 333–337. 58

[58] DENTINO, M., MCCOOL, J., AND WIDROW, B. Adaptive filtering in the frequency domain. *Proceedings of the IEEE 66*, 12 (1978), 1658–1659. 65

[59] DERRINGTON, A., AND LENNIE, P. Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *The Journal of Physiology 357*, 1 (1984), 219–240. 67

[60] DEVISME, C. *Etude de l'influence des disparités horizontales et verticales sur la perception de la profondeur en champ visuel périphérique*. Ph.D. thesis, Université Pierre et Marie Curie-Paris VI, 2008. 16

[61] DEVORE, R. A., JAWERTH, B., AND LUCIER, B. J. Image compression through wavelet transform coding. *IEEE Transactions on information theory 38*, 2 (1992), 719–746. 56

[62] DING, J., AND SPERLING, G. A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America 103*, 4 (2006), 1141–1146. 13

[63] DING, J., AND SPERLING, G. A gain-control theory of binocular combination. *Proceedings of the National Academy of Sciences of the United States of America 103*, 4 (2006), 1141–1146. 117, 119, 120, 136

[64] DIPPEL, S., STAHL, M., WIEMKER, R., AND BLAFFERT, T. Multiscale contrast enhancement for radiographies: Laplacian pyramid versus fast wavelet transform. *Transactions on Medical Imaging 21*, 4 (2002), 343–353. 137

[65] Do, M. N., and Vetterli, M. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on image processing 14*, 12 (2005), 2091–2106. 137

[66] Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D., Mok, V. C., Shi, L., and Heng, P.-A. Automatic detection of cerebral microbleeds from mr images via 3d convolutional neural networks. *transactions on medical imaging 35*, 5 (2016), 1182–1195. 137

[67] Drago, F., Myszkowski, K., Annen, T., and Chiba, N. Adaptive logarithmic mapping for displaying high contrast scenes. In *Computer Graphics Forum* (2003), vol. 22, Wiley Online Library, pp. 419–426. 52

[68] Ebner, M. *Color constancy*, vol. 6. John Wiley & Sons, 2007. 54, 62

[69] Fan, Y.-C., Chiou, J.-C., and Jiang, Y.-H. Hole-filling based memory controller of disparity modification system for multiview three-dimensional video. *Transactions on Magnetics 47*, 3 (2011), 679–682. 100

[70] Felleman, D. J., and Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991) 1*, 1 (1991), 1–47. 12

[71] Felzenszwalb, P. F., and Huttenlocher, D. P. Efficient belief propagation for early vision. *International journal of computer vision 70*, 1 (2006), 41–54. 32

[72] Fercher, A. F. Optical coherence tomography. *Journal of Biomedical Optics 1*, 2 (1996), 157–173. 39

[73] Feris, R., Raskar, R., Tan, K.-H., and Turk, M. Specular reflection reduction with multi-flash imaging. In *Brazilian Symposium on Computer Graphics and Image Processing* (2004), IEEE, pp. 316–321. 40

[74] Freeman, W. T., Adelson, E. H., et al. The design and use of steerable filters. *IEEE Transactions on Pattern analysis and machine intelligence 13*, 9 (1991), 891–906. 137

[75] Funt, B., Ciurea, F., and McCann, J. Retinex in matlab™. *Journal of electronic imaging 13*, 1 (2004), 48–57. 78

[76] Fusiello, A., Trucco, E., and Verri, A. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications 12*, 1 (2000), 16–22. 29

[77] Gerek, O. N., and Çetin, A. E. A 2-d orientation-adaptive prediction filter in lifting structures for image coding. *Transactions on Image Processing 15*, 1 (2006), 106–111. 116

[78] G.H BALLANTYNE, F. M. The da vinci telerobotic surgical system : the virtual operative field and telepresence surgery. *Surgical Clinics of North America 6*, 83 (2003), 1293–1304,. 2

[79] GIBSON, J. J. The perception of the visual world. 25

[80] GOETZ, M. Confocal laser endomicroscopy: Applications in clinical and translational science—a comprehensive review. *ISRN Pathology 2012* (2012). 49

[81] GONZALEZ, R. C., AND WOODS, R. E. Filtering in the frequency domain. In *Digital image processing*. Pearson education, 2008, ch. 4, pp. 199–310. 65

[82] GORDON, R., AND RANGAYYAN, R. M. Feature enhancement of film mammograms using fixed and adaptive neighborhoods. *Applied optics 23*, 4 (1984), 560–564. 53, 94

[83] GOSTOUT, C. J., AND JACQUES, S. L. Infrared video imaging of subsurface vessels: a feasibility study for the endoscopic management of gastrointestinal bleeding. *Gastrointestinal endoscopy 41*, 3 (1995), 218–224. 48

[84] GOUZE, A., ANTONINI, M., BARLAUD, M., AND MACQ, B. Design of signal-adapted multidimensional lifting scheme for lossy coding. *IEEE Transactions on Image Processing 13*, 12 (2004), 1589–1603. 116

[85] GREENBERG, J. A. Viking 3dhd vision system. *Reviews in Obstetrics and Gynecology 5*, 1 (2012), 62. 2

[86] GUO, J.-J., SHEN, D.-F., LIN, G.-S., HUANG, J.-C., LIU, K.-C., AND LIE, W.-N. A specular reflection suppression method for endoscopic images. In *International Conference on Multimedia Big Data* (2016), IEEE, pp. 125–128. 40

[87] HACHICHA, W., BEGHDADI, A., AND CHEIKH, F. A. Combining depth information and local edge detection for stereo image enhancement. In *European Signal Processing Conference* (2012), IEEE, pp. 250–254. 22, 61, 65, 94, 106, 107, 108, 109, 110, 111, 128, 129

[88] HAI, Y., LI, L., AND GU, J. Image enhancement based on contrast limited adaptive histogram equalization for 3d images of stereoscopic endoscopy. In *International Conference on Information and Automation* (2015), IEEE, pp. 668–672. 63, 64, 85, 126, 128, 129

[89] HANSEN, C., ZIDOWITZ, S., PREIM, B., STAVROU, G., OLDHAFER, K. J., AND HAHN, H. K. Impact of model-based risk analysis for liver surgery planning. *International journal of computer assisted radiology and surgery 9*, 3 (2014), 473–480. 4

[90] HARINGSMA, J., AND TYTGAT, G. Fluorescence and autofluorescence. *Best Practice & Research Clinical Gastroenterology 13*, 1 (1999), 1–10. 47

[91] HARINGSMA, J., TYTGAT, G. N., YANO, H., IISHI, H., TATSUTA, M., OGIHARA, T., WATANABE, H., SATO, N., MARCON, N., WILSON, B. C., ET AL. Autofluorescence endoscopy: feasibility of detection of gi neoplasms unapparent to white light endoscopy with an evolving technology. *Gastrointestinal endoscopy 53*, 6 (2001), 642–650. 47

[92] HARTLEY, R., AND GUPTA, R. Computing matched-epipolar projections. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on* (1993), IEEE, pp. 549–555. 29

[93] HAXBY, J. V., GRADY, C. L., HORWITZ, B., UNGERLEIDER, L. G., MISHKIN, M., CARSON, R. E., HERSCOVITCH, P., SCHAPIRO, M. B., AND RAPOPORT, S. I. Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences 88*, 5 (1991), 1621–1625. 12

[94] HAZEWINKEL, Y., AND DEKKER, E. Colonoscopy: basic principles and novel techniques. *Nature Reviews Gastroenterology and Hepatology 8*, 10 (2011), 554–564. 44

[95] HECHT, S., SHLAER, S., AND PIRENNE, M. H. Energy, quanta, and vision. *The Journal of general physiology 25*, 6 (1942), 819–840. 11

[96] HENRY, M. M., AND THOMPSON, J. N. *Clinical Surgery*. Elsevier Health Sciences, 2012. 2

[97] HERSHBERGER, W. Attached-shadow orientation perceived as depth by chickens reared in an environment illuminated from below. *Journal of comparative and physiological psychology 73*, 3 (1970), 407. 21

[98] HOFFMAN, A., GOETZ, M., VIETH, M., GALLE, P., NEURATH, M., AND KIESSLICH, R. Confocal laser endomicroscopy: technical status and current indications. *Endoscopy 38*, 12 (2006), 1275–1283. 48

[99] HOFFMAN, A., SAR, F., GOETZ, M., TRESCH, A., MUDTER, J., BIESTERFELD, S., GALLE, P., NEURATH, M., AND KIESSLICH, R. High definition colonoscopy combined with i-scan is superior in the detection of colorectal neoplasias compared with standard video colonoscopy: a prospective randomized controlled trial. *Endoscopy 42*, 10 (2010), 827–833. 38

[100] HOWARD, I., AND ROGERS, B. Perceiving in depth, volume 2: Stereoscopic vision, no. 29, 2012. 13

[101] HOWARD, I. P., AND ROGERS, B. J. *Binocular vision and stereopsis*. Oxford University Press, USA, 1995. 13

[102] HUANG, C.-B., ZHOU, J., LU, Z.-L., FENG, L., AND ZHOU, Y. Binocular combination in anisometropic amblyopia. *Journal of vision 9*, 3 (2009), 17–17. 13

[103] HUANG, D., SWANSON, E. A., LIN, C. P., SCHUMAN, J. S., STINSON, W. G., CHANG, W., HEE, M. R., FLOTTE, T., GREGORY, K., PULIAFITO, C. A., ET AL. Optical coherence tomography. *Science (New York, NY) 254*, 5035 (1991), 1178. 39

[104] HUANG, S.-C., CHENG, F.-C., AND CHIU, Y.-S. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Transactions on Image Processing 22*, 3 (2013), 1032–1041. 53

[105] HUANG, Y.-P., HASSANFIROOZI, A., SHIEH, H.-P. D., AND JAVIDI, B. Specular reflection reduction for endoscopic imaging based on liquid crystal lens. In *3D Image Acquisition and Display: Technology, Perception and Applications* (2016), Optical Society of America, pp. TW2A–1. 41

[106] HUBEL, D. H., AND WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology 195*, 1 (1968), 215–243. 13

[107] HUBENS, G., COVELIERS, H., BALLIU, L., RUPPERT, M., AND VANEERDEWEG, W. A performance study comparing manual and robotically assisted laparoscopic surgery using the da vinci system. *Surgical Endoscopy and other interventional techniques 17*, 10 (2003), 1595–1599. 2

[108] IATSUN, I., LARABI, M.-C., AND FERNANDEZ-MALOIGNE, C. Investigation of visual fatigue/discomfort generated by s3d video using eye-tracking data. *IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics* (2013), 864803–864803. 94

[109] IBRAHIM, H., AND KONG, N. S. P. Brightness preserving dynamic histogram equalization for image contrast enhancement. *Transactions on Consumer Electronics 53*, 4 (2007). 55

[110] IMTIAZ, M. S., KHAN, T. H., AND WAHID, K. New color image enhancement method for endoscopic images. In *International Conference on Advances in Electrical Engineering* (2013), IEEE, pp. 263–266. 62, 64

[111] IMTIAZ, M. S., AND WAHID, K. A color reproduction method with image enhancement for endoscopic images. In *Middle East Conference on Biomedical Engineering* (2014), IEEE, pp. 135–138. 62, 64

[112] IMTIAZ, M. S., AND WAHID, K. A. Image enhancement and space-variant color reproduction method for endoscopic images using adaptive sigmoid function. In *International Conference of the Engineering in Medicine and Biology Society* (2014), IEEE, pp. 3905–3908. 62, 64

[113] INOUE, H., SASAJIMA, K., KAGA, M., SUGAYA, S., SATO, Y., WADA, Y., INUI, M., SATODATE, H., KUDO, S.-E., KIMURA, S., ET AL. Endoscopic in vivo evaluation of tissue atypia in the esophagus using a newly designed integrated endocytoscope: a pilot trial. *Endoscopy 38*, 09 (2006), 891–895. 38

[114] ISEKI, K., TATSUTA, M., IISHI, H., SAKAI, N., YANO, H., AND ISHIGURO, S. Effectiveness of the near-infrared electronic endoscope for diagnosis of the depth of involvement of gastric cancers. *Gastrointestinal endoscopy 52*, 6 (2000), 755–762. 48

[115] ISHIHARA, R. Infrared endoscopy in the diagnosis and treatment of early gastric cancer. *Endoscopy 42*, 08 (2010), 672–676. 48

[116] ITO, S., MUGURUMA, N., KIMURA, T., YANO, H., IMOTO, Y., OKAMOTO, K., KAJI, M., SANO, S., AND NAGAO, Y. Principle and clinical usefulness of the infrared fluorescence endoscopy. *The Journal of Medical Investigation 53*, 1, 2 (2006), 1–8. 48

[117] ITU GENEVA SWITZERLAND. *Methodology for the Subjective Assessment of the Quality of Television Pictures , Recommendation ITU-R BT.500-11*, August 2012. 105

[118] JAIN, A. K. *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989. 51, 52

[119] JAMESON, D., AND HURVICH, L. M. Theory of brightness and color contrast in human vision. *Vision research 4*, 1 (1964), 135–154. 67

[120] JANG, J.-Y. The past, present, and future of image-enhanced endoscopy. *Clinical endoscopy 48*, 6 (2015), 466–475. 38

[121] JAYA, V., AND GOPIKAKUMARI, R. IEM: a new image enhancement metric for contrast and sharpness measurements. *International Journal of Computer Applications 79*, 9 (2013). 70, 123

[122] JEMAL, A., BRAY, F., CENTER, M. M., FERLAY, J., WARD, E., AND FORMAN, D. Global cancer statistics. *CA: a cancer journal for clinicians 61*, 2 (2011), 69–90. 4

[123] JI, T.-L., SUNDARESHAN, M. K., AND ROEHRIG, H. Adaptive image contrast enhancement based on human visual properties. *IEEE transactions on medical imaging 13*, 4 (1994), 573–586. 53, 65

[124] JOBSON, D. J., RAHMAN, Z.-U., AND WOODELL, G. A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Transactions on Image processing 6*, 7 (1997), 965–976. 53, 62, 65, 70, 81

[125] JOBSON, D. J., RAHMAN, Z.-U., AND WOODELL, G. A. Properties and performance of a center/surround retinex. *transactions on image processing 6*, 3 (1997), 451–462. 53, 65

[126] JULESZ, B. Foundations of cyclopean perception. 13

[127] JULESZ, B. *Foundations of cyclopean perception*. Chicago: University of Chicago Press, 1971. 117, 119

[128] JUNG, S.-W., JEONG, J.-Y., AND KO, S.-J. Sharpness enhancement of stereo images using binocular just-noticeable difference. *IEEE Transactions on Image Processing 21*, 3 (2012), 1191–1199. 62, 65, 94

[129] KAANICHE, M., BENAZZA-BENYAHIA, A., PESQUET-POPESCU, B., AND PESQUET, J.-C. Vector lifting schemes for stereo image coding. *Transactions on Image Processing 18*, 11 (2009), 2463–2475. 114, 117

[130] KAANICHE, M., PESQUET-POPESCU, B., BENAZZA-BENYAHIA, A., AND PESQUET, J.-C. Adaptive lifting scheme with sparse criteria for image coding. *EURASIP Journal on Advances in Signal Processing 2012*, 1 (2012), 10. 116

[131] KAISE, M., OHKURA, Y., IIZUKA, T., KIMURA, R., NOMURA, K., KURIBAYASHI, Y., YAMADA, A., YAMASHITA, S., FURUHATA, T., KIKUCHI, D., ET AL. Endocytoscopy is a promising modality with high diagnostic accuracy for gastric cancer. *Endoscopy 47*, 01 (2015), 19–25. 50

[132] KARA, M., AND BERGMAN, J. Autofluorescence imaging and narrowband imaging for the detection of early neoplasia in patients with barrett's esophagus. *Endoscopy 38*, 06 (2006), 627–631. 48

[133] KATO, M., KAISE, M., YONEZAWA, J., GODA, K., TOYOIZUMI, H., YOSHIMURA, N., YOSHIDA, Y., KAWAMURA, M., AND TAJIRI, H. Trimodal imaging endoscopy may improve diagnostic accuracy of early gastric neoplasia: a feasibility study. *Gastrointestinal endoscopy 70*, 5 (2009), 899–906. 48

[134] KATO, M., KAISE, M., YONEZAWA, J., YOSHIDA, Y., AND TAJIRI, H. Autofluorescence endoscopy versus conventional white light endoscopy for the detection of superficial gastric neoplasia: a prospective comparative study. *Endoscopy 39*, 11 (2007), 937–941. 47

[135] KAUFMAN, L. *The American Journal of Psychology 85*, 1 (1972), 141–152. [link]. 53

[136] KAUR, M., KAUR, J., AND KAUR, J. Survey of contrast enhancement techniques based on histogram equalization. *International Journal of Advanced Computer Science and Applications 2*, 7 (2011), 137–141. 54

[137] KELLMAN, P. J., AND SPELKE, E. S. Perception of partly occluded objects in infancy. *Cognitive psychology 15*, 4 (1983), 483–524. 21

[138] KIESSLICH, R., BURG, J., VIETH, M., GNAENDIGER, J., ENDERS, M., DELANEY, P., POLGLASE, A., MCLAREN, W., JANELL, D., THOMAS, S., ET AL. Confocal laser endoscopy for diagnosing intraepithelial neoplasias and colorectal cancer in vivo. *Gastroenterology 127*, 3 (2004), 706–713. 48

[139] KIESSLICH, R., GOETZ, M., AND NEURATH, M. F. Confocal laser endomicroscopy for gastrointestinal diseases. *Gastrointestinal endoscopy clinics of North America 18*, 3 (2008), 451–466. 48

[140] KIESSLICH, R., GOETZ, M., VIETH, M., GALLE, P. R., AND NEURATH, M. F. Confocal laser endomicroscopy. *Gastrointestinal endoscopy clinics of North America 15*, 4 (2005), 715–731. 38

[141] KIESSLICH, R., AND JUNG, M. Magnification endoscopy: does it improve mucosal surface analysis for the diagnosis of gastrointestinal neoplasias? *Endoscopy 34*, 10 (2002), 819–822. 44

[142] KIHARA, K., FUJII, Y., MASUDA, H., SAITO, K., KOGA, F., MATSUOKA, Y., NUMAO, N., AND KOJIMA, K. New three-dimensional head-mounted display system, tmdu-s-3d system, for minimally invasive surgery application: Procedures for gasless single-port radical nephrectomy. *International Journal of Urology 19*, 9 (2012), 886–889. 2

[143] KIM, J. C., LEE, K. M., CHOI, B. T., AND LEE, S. U. A dense stereo matching using two-pass dynamic programming with generalized ground control points. In *Computer Society Conference on Computer Vision and Pattern Recognition* (2005), vol. 2, IEEE, pp. 1075–1082. 32

[144] KIMMEL, R., ELAD, M., SHAKED, D., KESHET, R., AND SOBEL, I. A variational framework for retinex. *International Journal of computer vision 52*, 1 (2003), 7–23. 53, 65

[145] KIMURA, T., MUGURUMA, N., ITO, S., OKAMURA, S., IMOTO, Y., MIYAMOTO, H., KAJI, M., AND KUDO, E. Infrared fluorescence endoscopy for the diagnosis of superficial gastric tumors. *Gastrointestinal endoscopy 66*, 1 (2007), 37–43. 48

[146] KLAUS, A., SORMANN, M., AND KARNER, K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *International Conference on Pattern Recognition* (2006), vol. 3, IEEE, pp. 15–18. 32

[147] KLETTE, R., KOSCHAN, A., AND SCHLUNS, K. Three-dimensional data from images. *Springer-Verlag* (1998). 30

[148] KOLMOGOROV, V., AND ZABIH, R. Computing visual correspondence with occlusions using graph cuts. In *International Conference on Computer Vision* (2001), vol. 2, IEEE, pp. 508–515. 32

[149] KONG, X., AND GOUTSIAS, J. A study of pyramidal techniques for image representation and compression. *Journal of Visual Communication and Image Representation 5*, 2 (1994), 190–203. 55

[150] KOSOV, S., THORMÄHLEN, T., AND SEIDEL, H.-P. Accurate real-time disparity estimation with variational methods. *Advances in Visual Computing* (2009), 796–807. 32

[151] KRYSTALLIS, C., KOULAOUZIDIS, A., DOUGLAS, S., AND PLEVRIS, J. N. Chromoendoscopy in small bowel capsule endoscopy: Blue mode or fuji intelligent colour enhancement? *Digestive and Liver Disease 43*, 12 (2011), 953–957. 38

[152] KUANG, J., YAMAGUCHI, H., JOHNSON, G. M., AND FAIRCHILD, M. D. Testing hdr image rendering algorithms. In *Color and Imaging Conference* (2004), vol. 2004, Society for Imaging Science and Technology, pp. 315–320. 79

[153] KUZNETSOV, K., LAMBERT, R., AND REY, J.-F. Narrow-band imaging: potential and limitations. *Endoscopy 38*, 01 (2006), 76–81. 38, 45

[154] KWON, R. S., SONG, L.-M. W. K., ADLER, D. G., CONWAY, J. D., DIEHL, D. L., FARRAYE, F. A., KANTSEVOY, S. V., KAUL, V., KETHU, S. R., MAMULA, P., ET AL. Endocytoscopy. *Gastrointestinal endoscopy 70*, 4 (2009), 610–613. 49

[155] LAINE, A., FAN, J., AND YANG, W. Wavelets for contrast enhancement of digital mammography. *Engineering in Medicine and Biology Magazine 14*, 5 (1995), 536–550. 57, 114

[156] LAMADÉ, W., GLOMBITZA, G., FISCHER, L., CHIU, P., CÁRDENAS SR, C. E., THORN, M., MEINZER, H.-P., GRENACHER, L., BAUER, H., LEHNERT, T., ET AL. The impact of 3-dimensional reconstructions on operation planning in liver surgery. *Archives of surgery 135*, 11 (2000), 1256–1261. 4

[157] LAMATA, P., LAMATA, F., SOJAR, V., MAKOWSKI, P., MASSOPTIER, L., CASCIARO, S., ALI, W., STÜDELI, T., DECLERCK, J., ELLE, O. J., ET AL. Use of the resection map system as guidance during hepatectomy. *Surgical endoscopy 24*, 9 (2010), 2327–2337. 4

[158] LAMBOOIJ, M., FORTUIN, M., HEYNDERICKX, I., AND IJSSELSTEIJN, W. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology 53*, 3 (2009), 30201–1. 62

[159] LAND, E. H. Recent advances in retinex theory. *Vision research 26*, 1 (1986), 7–21. 81

[160] LAND, E. H., AND MCCANN, J. J. Lightness and retinex theory. *Josa 61*, 1 (1971), 1–11. 53, 56, 61, 65

[161] LANG, H., RADTKE, A., HINDENNACH, M., SCHROEDER, T., FRÜHAUF, N. R., MALAGÓ, M., BOURQUAIN, H., PEITGEN, H.-O., OLDHAFER, K. J., AND BROELSCH, C. E. Impact of virtual tumor resection and computer-assisted risk analysis on operation planning and intraoperative strategy in major hepatic resection. *Archives of surgery 140*, 7 (2005), 629–638. 4

[162] LE NÉGRATE, A., BEGHDADI, A., AND DUPOISOT, H. An image enhancement technique and its evaluation through bimodality analysis. *Graphical Models and Image Processing (CVGIP) 54*, 1 (1992), 13–22. 93, 95

[163] LEE, B. B., AND SUN, H. Contrast sensitivity and retinal ganglion cell responses in the primate. *Psychology & Neuroscience 4*, 1 (2011), 11. 67

[164] LEE, D. N., AND KALMUS, H. The optic flow field: the foundation of vision. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* (1980), 169–179. 19

[165] LEE, J., OH, J., SHAH, S. K., YUAN, X., AND TANG, S. J. Automatic classification of digestive organs in wireless capsule endoscopy videos. In *Proceedings of the ACM symposium on Applied computing* (2007), ACM, pp. 1041–1045. 76

[166] LEE, S. An efficient content-based image enhancement in the compressed domain using retinex theory. *IEEE Transactions on Circuits and Systems for Video Technology 17*, 2 (2007), 199–213. 56, 59

[167] LEE, S., LEE, J. H., LIM, J., AND SUH, I. H. Robust stereo matching using adaptive random walk with restart algorithm. *Image and Vision Computing 37* (2015), 1–11. 97, 100

[168] LEFCOUT, A. M., AND KIM, M. S. Technique for normalizing intensity histograms of images when the approximate size of the target is known: Detection of feces on apples using fluorescence imaging. *Computers and electronics in agriculture 50*, 2 (2006), 135–147. 84

[169] LEGGE, G. E., AND RUBIN, G. S. Binocular interactions in suprathreshold contrast perception. *Attention, Perception, & Psychophysics 30*, 1 (1981), 49–61. 13

[170] LEVELT, W. J. *On binocular rivalry*. Ph.D. thesis, Van Gorcum Assen, 1965. 117, 119, 120

[171] LI, B., AND MENG, M. Q.-H. Wireless capsule endoscopy images enhancement using contrast driven forward and backward anisotropic diffusion. In *International Conference on Image Processing* (2007), vol. 2, IEEE, pp. II–437. 64

[172] LI, B., AND MENG, M. Q.-H. Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. *Computers in biology and medicine 39*, 2 (2009), 141–147. 76

[173] LI, L., ZOU, Y., AND LI, Y. Wireless capsule endoscopy images enhancement based on adaptive anisotropic diffusion. In *International Conference on Signal and Information Processing* (2013), IEEE, pp. 273–277. 64

[174] LI, X., WANG, Y., ZHAO, D., JIANG, T., AND ZHANG, N. Joint just noticeable difference model based on depth perception for stereoscopic images. In *Visual Communications and Image Processing (VCIP)* (2011), IEEE, pp. 1–4. 96, 135

[175] LI, X.-W., AND KANG, Y.-X. A novel medical image enhancement method based on wavelet multi-resolution analysis. In *International Conference on Biomedical Engineering and Informatics* (2015), IEEE, pp. 727–731. 57

[176] LIM, S.-N., MITTAL, A., DAVIS, L. S., AND PARAGIOS, N. Uncalibrated stereo rectification for automatic 3d surveillance. In *Image Processing, 2004. ICIP'04. 2004 International Conference on* (2004), vol. 2, IEEE, pp. 1357–1360. 29

[177] LIM, Y., CHAE, E., LEE, E., KANG, W., AND PAIK, J. Image enhancement using similarity-based color matching for high-quality stereoscopic 3D image acquisition. In *Technical Conference & Exhibition* (2012), SMPTE, pp. 2–12. 61

[178] LIN, C., AND HUANG, C. Cnn-based retinex technology. In *European Conference on Circuit Theory and Design* (2003), vol. 2, pp. 69–72. 53, 65

[179] LIN, G.-S., CHEN, C.-Y., KUO, C.-T., LIE, W.-N., AND LIU, K.-C. Adaptive support-window approximation to bilateral filtering. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (2012), IEEE, pp. 2013–2016. 62

[180] LIN, J., AND QU, L. Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis. *Journal of sound and vibration 234*, 1 (2000), 135–148. 56

[181] LIN, Y.-H., AND WU, J.-L. Quality assessment of stereoscopic 3d image compression by binocular integration behaviors. *Transactions on Image Processing 23*, 4 (2014), 1527–1542. 35

[182] LIN, Y.-H., AND WU, J.-L. Quality assessment of stereoscopic 3d image compression by binocular integration behaviors. *IEEE Transactions on Image Processing 23*, 4 (2014), 1527–1542. 119

[183] LIU, J. G., AND MASON, P. J. *Image processing and GIS for remote sensing: Techniques and applications*. John Wiley & Sons, 2016. 84

[184] LIU, L., JIA, Z., YANG, J., AND KASABOV, N. A remote sensing image enhancement method using mean filter and unsharp masking in non-subsampled contourlet transform domain. *Transactions of the Institute of Measurement and Control 39*, 2 (2017), 183–193. 65

[185] LIU, X., PEDERSEN, M., AND HARDEBERG, J. Y. Cid: Iq–a new image quality database. In *International Conference on Image and Signal Processing* (2014), Springer, pp. 193–202. 68, 69, 72

[186] LOGOTHETIS, N. K., AND SHEINBERG, D. L. Visual object recognition. *Annual review of neuroscience 19*, 1 (1996), 577–621. 12

[187] LOWE, D. G. Object recognition from local scale-invariant features. In *Proc. International conference on Computer vision* (1999), vol. 2, IEEE, pp. 1150–1157. 86

[188] LUFT, T., COLDITZ, C., AND DEUSSEN, O. Image enhancement by unsharp masking the depth buffer. vol. 25, ACM Transactions on Graphics. 93

[189] LUONG, Q.-T., DERICHE, R., FAUGERAS, O., AND PAPADOPOULO, T. *On determining the fundamental matrix: Analysis of different methods and experimental results*. Ph.D. thesis, INRIA, 1993. 28

[190] LUONG, Q.-T., AND FAUGERAS, O. D. The fundamental matrix: Theory, algorithms, and stability analysis. *International journal of computer vision 17*, 1 (1996), 43–75. 28

[191] MA, W.-C., HAWKINS, T., PEERS, P., CHABERT, C.-F., WEISS, M., AND DEBEVEC, P. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Eurographics conference on Rendering Techniques* (2007), Eurographics Association, pp. 183–194. 41

[192] MA, X., CHEN, H., AND ZHAO, Y. Stereo image coding method using stereo matching with difference-based adaptive searching windows. In *International Workshop on Imaging Systems and Techniques* (2009), IEEE, pp. 373–376. 31

[193] MACQUEEN, J., ET AL. Some methods for classification and analysis of multivariate observations. In *Proc. Berkeley symposium on mathematical statistics and probability* (1967), vol. 1, Oakland, CA, USA., pp. 281–297. 87

[194] MAEHARA, G., AND GORYO, K. Binocular, monocular and dichoptic pattern masking. *Optical Review 12*, 2 (2005), 76–82. 13

[195] MAINI, R., AND AGGARWAL, H. A comprehensive review of image enhancement techniques. *arXiv preprint arXiv:1003.4053* (2010). 52

[196] MALLICK, S. P., ZICKLER, T., BELHUMEUR, P. N., AND KRIEGMAN, D. J. Specularity removal in images and videos: A pde approach. In *European Conference on Computer Vision* (2006), Springer, pp. 550–563. 41

[197] MAMASSIAN, P., KNILL, D. C., AND KERSTEN, D. The perception of cast shadows. *Trends in cognitive sciences 2*, 8 (1998), 288–295. 21

[198] MAMONOV, A. V., FIGUEIREDO, I. N., FIGUEIREDO, P. N., AND TSAI, Y.-H. R. Automated polyp detection in colon capsule endoscopy. *IEEE transactions on medical imaging 33*, 7 (2014), 1488–1502. 78

[199] MANIVANNAN, S., WANG, R., TRUCCO, E., AND HOOD, A. Automatic normal-abnormal video frame classification for colonoscopy. In *International Symposium on Biomedical Imaging* (2013), IEEE, pp. 644–647. 76

[200] MARR, D., AND HILDRETH, E. Theory of edge detection. *Proceedings of the Royal Society of London Biological Sciences 207*, 1167 (1980), 187–217. 53, 94

[201] MARR, D., AND POGGIO, T. A computational theory of human stereo vision. *Proceedings of the Royal Society of London B: Biological Sciences 204*, 1156 (1979), 301–328. 13, 17

[202] MASSOPUST, L. C. Infrared photographic study of the superficial veins of the thorax in relation to breast tumors; a preliminary report. *Surgery, gynecology & obstetrics 86*, 1 (1948), 54–58. 48

[203] MATAKI, N., NAGAO, S., KAWAGUCHI, A., MATSUZAKI, K., MIYAZAKI, J., KITAGAWA, Y., NAKAJIMA, H., TSUZUKI, Y., ITOH, K., NIWA, H., ET AL. Clinical usefulness of a new infrared videoendoscope system for diagnosis of early stage gastric cancer. *Gastrointestinal endoscopy 57*, 3 (2003), 336–342. 48

[204] MATKOVIC, K., NEUMANN, L., NEUMANN, A., PSIK, T., AND PURGATHOFER, W. Global contrast factor-a new approach to image contrast. *Computational Aesthetics 2005* (2005), 159–168. 72

[205] MAUNSELL, J. H. The brain's visual world: representation of visual targets in cerebral cortex. *Science 270*, 5237 (1995), 764. 12

[206] MCCANN, J. J. Capturing a black cat in shade: past and present of retinex color appearance models. *Journal of Electronic Imaging 13*, 1 (2004), 36–47. 81

[207] MEESE, T. S., GEORGESON, M. A., AND BAKER, D. H. Binocular contrast vision at and above threshold. *Journal of vision 6*, 11 (2006), 7–7. 13

[208] MEMON, N. D., AND SAYOOD, K. Lossless image compression: A comparative study. In *Proc. SPIE* (1995), vol. 2418, p. 20. 35

[209] MESLOUHI, O., KARDOUCHI, M., ALLALI, H., GADI, T., AND BENKADDOUR, Y. Automatic detection and inpainting of specular reflections for colposcopic images. *Open Computer Science 1*, 3 (2011), 341–354. 40

[210] MEYLAN, L. Tone mapping for high dynamic range images. 79

[211] MEYLAN, L., AND SÜSSTRUNK, S. Bio-inspired image enhancement for natural color images. In *Proc. IST/SPIE Electronic Imaging: Human Vision and Electronic Imaging* (2004), vol. 5292, pp. 46–56. 70, 78, 79, 81

[212] MICHELSON, A. *Studies in Optics*. The Univ. of Chicago Science Series. University of Chicago Press, 1927. 68

[213] MILED, W., AND PESQUET, J.-C. Disparity map estimation using a total variation bound. In *Canadian Conference on Computer and Robot Vision* (2006), IEEE, pp. 48–48. 32

[214] MILED, W., PESQUET, J.-C., AND PARENT, M. A convex optimization approach for depth estimation under illumination variation. *IEEE Transactions on Image Processing 18*, 4 (2009), 813–830. 32

[215] MISIAKOS, E. P., KARIDIS, N. P., AND KOURAKLIS, G. Current treatment for colorectal liver metastases. *World journal of gastroenterology: WJG 17*, 36 (2011), 4067. 4

[216] MITRA, S. K., AND YU, T.-H. Transform amplitude sharpening: A new method of image enhancement. *Computer vision, graphics, and image processing 40*, 2 (1987), 205–218. 59

[217] MIYAKE, Y., KOUZU, T., TAKEUCHI, S., YAMATAKA, S., NAKAGUCHI, T., AND TSUMURA, N. Development of new electronic endoscopes using the spectral images of an internal organ. In *Color and Imaging Conference* (2005), vol. 2005, Society for Imaging Science and Technology, pp. 261–263. 46

[218] MOESKOPS, P., VIERGEVER, M. A., MENDRIK, A. M., DE VRIES, L. S., BENDERS, M. J., AND IŠGUM, I. Automatic segmentation of mr brain images with a convolutional neural network. *IEEE transactions on medical imaging 35*, 5 (2016), 1252–1261. 137

[219] MORADI, F., AND HEEGER, D. J. Inter-ocular contrast normalization in human visual cortex. *Journal of Vision 9*, 3 (2009), 13–13. 13

[220] MORADI, M., FALAHATI, A., SHAHBAHRAMI, A., AND ZARE-HASSANPOUR, R. Improving visual quality in wireless capsule endoscopy images with contrast-limited adaptive histogram equalization. In *International Conference on Pattern Recognition and Image Analysis* (2015), IEEE, pp. 1–5. 64, 85

[221] MOUNTNEY, P., STOYANOV, D., AND YANG, G.-Z. Three-dimensional tissue deformation recovery and tracking: Introducing techniques based on laparoscopic or endoscopic images. *Signal Processing Magazine 27*, 4 (2010), 14–24. 85, 106, 122

[222] MOVSHON, J. A., THOMPSON, I., AND TOLHURST, D. Receptive field organization of complex cells in the cat's striate cortex. *The Journal of physiology 283*, 1 (1978), 79–99. 12

[223] MOVSHON, J. A., THOMPSON, I. D., AND TOLHURST, D. J. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *The Journal of physiology 283*, 1 (1978), 53–77. 12

[224] MUKHERJEE, J., AND MITRA, S. K. Enhancement of color images by scaling the dct coefficients. *Transactions on Image Processing 17*, 10 (2008), 1783–1794. 56, 70

[225] MUKHOPADHYAY, S., AND CHANDA, B. A multiscale morphological approach to local contrast enhancement. *Signal Processing 80*, 4 (2000), 685–696. 70

[226] MUTO, M., HORIMATSU, T., EZOE, Y., HORI, K., YUKAWA, Y., MORITA, S., MIYAMOTO, S., AND CHIBA, T. Narrow-band imaging of the gastrointestinal tract. *Journal of gastroenterology 44*, 1 (2009), 13–25. 45

[227] MYLONAKI, M., FRITSCHER-RAVENS, A., AND SWAIN, P. Wireless capsule endoscopy: a comparison with push enteroscopy in patients with gastroscopy and colonoscopy negative gastrointestinal bleeding. *Gut 52*, 8 (2003), 1122–1126. 76

[228] MÜLLER, C. Innovation in endoscopy: technological trends in visualisation. `http://www.medicaldevice-developments.com`, November 2016. 43, 44, 50

[229] NASSI, J. J., AND CALLAWAY, E. M. Parallel processing strategies of the primate visual system. *Nature reviews. Neuroscience 10*, 5 (2009), 360. 12

[230] NELSON, D., BLOCK, K., BOSCO, J., BURDICK, J., CURTIS, W., FAIGEL, D., GREENWALD, D., KELSEY, P., RAJAN, E., SLIVKA, A., ET AL. Technology status evaluation report-high resolution and high-magnification endoscopy. *Gastrointestinal Endoscopy 52*, 6 (2000), 864–866. 44

[231] OHTA, H., KOHGO, Y., GOTO, Y., TAKAHASI, Y., MOGI, Y., WATANABE, N., AND NIITSU, Y. The near-infrared electronic endoscope for diagnosis of esophageal varices. *Gastrointestinal endoscopy 38*, 3 (1992), 330–335. 48

[232] OHTA, Y., AND KANADE, T. Stereo by intra-and inter-scanline search using dynamic programming. *Transactions on pattern analysis and machine intelligence*, 2 (1985), 139–154. 32

[233] OKUHATA, H., NAKAMURA, H., HARA, S., TSUTSUI, H., AND ONOYE, T. Application of the real-time retinex image enhancement for endoscopic images. In *International Conference of the Engineering in Medicine and Biology Society* (2013), IEEE, pp. 3407–3410. 62, 64

[234] PALMER, S. E. *Vision science: Photons to phenomenology*. MIT press, 1999. 25, 17, 19

[235] PANETTA, K., GAO, C., AND AGAIAN, S. No reference color image contrast and quality measures. *Transactions on Consumer Electronics 59*, 3 (2013), 643–651. 70, 123

[236] PANETTA, K., ZHOU, Y., AGAIAN, S., AND JIA, H. Nonlinear unsharp masking for mammogram enhancement. *Transactions on Information Technology in Biomedicine 15*, 6 (2011), 918–928. 54

[237] PAPADIMITRIOU, D. V., AND DENNIS, T. J. Epipolar line estimation and rectification for stereo image pairs. *IEEE transactions on image processing 5*, 4 (1996), 672–676. 29

[238] PARCHAMI, M., AND MARIOTTINI, G.-L. A comparative study on 3-d stereo reconstruction from endoscopic images. In *Proceedings of International Conference on Pervasive Technologies Related to Assistive Environments* (2014), ACM, p. 25. 37, 134

[239] PAWLIK, T. M., AND VAUTHEY, J.-N. Surgical margins during hepatic surgery for colorectal liver metastases: complete resection not millimeters defines outcome. *Annals of surgical oncology 15*, 3 (2008), 677–679. 4

[240] PELI, E. Contrast in complex images. *JOSA A 7*, 10 (1990), 2032–2040. 68, 121, 122, 136

[241] PELLI, D. G., AND BEX, P. Measuring contrast sensitivity. *Vision research 90* (2013), 10–14. 68

[242] PEREIRA, S., PINTO, A., ALVES, V., AND SILVA, C. A. Brain tumor segmentation using convolutional neural networks in mri images. *transactions on medical imaging 35*, 5 (2016), 1240–1251. 137

[243] PIRENNE, M. H. Binocular and uniocular threshold of vision. *Nature* (1943). 13

[244] PITAS, I. *Digital image processing algorithms and applications*. John Wiley & Sons, 2000. 51, 52

[245] PIZER, S. M., AMBURN, E. P., AUSTIN, J. D., CROMARTIE, R., GESELOWITZ, A., GREER, T., TER HAAR ROMENY, B., ZIMMERMAN, J. B., AND ZUIDERVELD, K. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing 39*, 3 (1987), 355–368. 55

[246] PLATANIOTIS, K., AND VENETSANOPOULOS, A. N. *Color image processing and applications*. Springer Science & Business Media, 2013. 53

[247] POLESEL, A., RAMPONI, G., AND MATHEWS, V. J. Image enhancement via adaptive unsharp masking. *transactions on image processing 9*, 3 (2000), 505–510. 54, 62

[248] QI, F., JIANG, T., FAN, X., MA, S., AND ZHAO, D. Stereoscopic video quality assessment based on stereo just-noticeable difference model. In *Image Processing (ICIP), 2013 20th IEEE International Conference on* (2013), IEEE, pp. 34–38. 96

[249] QI, F., ZHAO, D., FAN, X., AND JIANG, T. Stereoscopic video quality assessment based on visual attention and just-noticeable difference models. *Signal, Image and Video Processing 10*, 4 (2016), 737–744. 96

[250] QIU, G. Indexing chromatic and achromatic patterns for content-based colour image retrieval. *Pattern Recognition 35*, 8 (2002), 1675–1686. 86

[251] QURESHI, M. A., BEGHDADI, A., AND DERICHE, M. Towards the design of a consistent image contrast enhancement evaluation measure. *Signal Processing: Image Communication 58* (2017), 212–227. 124

[252] R. BLAKE, Y. Y., AND WILSON, H. R. On the coexistence of stereopsis and binocular rivalry. *Elsevier, Vision Research 31*, 7-8 (1991), 1191–1203. 119

[253] RAHMAN, Z.-U., JOBSON, D. J., AND WOODELL, G. A. Retinex processing for automatic image enhancement. *Journal of Electronic Imaging 13*, 1 (2004), 100–110. 70, 78, 81

[254] RAHMAN, Z.-U., WOODELL, G. A., AND JOBSON, D. J. A comparison of the multiscale retinex with other image enhancement techniques. 54

[255] RAMARAJ, M., RAGHAVAN, S., AND KHAN, W. A. Homomorphic filtering techniques for wce image enhancement. In *International Conference on Computational Intelligence and Computing Research* (2013), IEEE, pp. 1–5. 64

[256] RAMPONI, G. A cubic unsharp masking technique for contrast enhancement. *Signal Processing 67*, 2 (1998), 211–222. 54

[257] RAMPONI, G., AND POLESEL, A. Rational unsharp masking technique. *Journal of Electronic Imaging 7*, 2 (1998), 333–338. 54

[258] RASTI, P., DANESHMAND, M., ALISINANOGLU, F., OZCINAR, C., AND ANBARJAFARI, G. Medical image illumination enhancement and sharpening by using stationary wavelet transform. In *Signal Processing and Communication Application Conference* (2016), IEEE, pp. 153–156. 58

[259] REINHARD, E., STARK, M., SHIRLEY, P., AND FERWERDA, J. Photographic tone reproduction for digital images. *ACM transactions on graphics 21*, 3 (2002), 267–276. 79

[260] RISUCCI, D. A. Visual spatial perception and surgical competence. *The American Journal of Surgery 184*, 3 (2002), 291–295. 2

[261] RIZZI, A., ALGERI, T., MEDEGHINI, G., AND MARINI, D. A proposal for contrast measure in digital images. In *Conference on colour in graphics, imaging, and vision* (2004), vol. 2004, Society for Imaging Science and Technology, pp. 187–192. 68, 69

[262] ROCK, I. *Indirect perception*. Mit Press, 1997. 9

[263] ROTH, H. R., LU, L., LIU, J., YAO, J., SEFF, A., CHERRY, K., KIM, L., AND SUMMERS, R. M. Improving computer-aided detection using convolutional neural networks and random view aggregation. *transactions on medical imaging 35*, 5 (2016), 1170–1181. 137

[264] SĂFTOIU, A., AND VILMANN, P. Autofluorescence imaging and magnification endoscopy. *World. J. Gastroenterol 17* (2011), 9–14. 45

[265] SALEEM, A., BEGHDADI, A., AND BOASHASH, B. Image fusion-based contrast enhancement. *Journal on Image and Video Processing 2012*, 1 (2012), 10. 70, 123

[266] SCHARSTEIN, D., AND SZELISKI, R. Middlebury stereo vision website. `http://vision.middlebury.edu/stereo/`. 97

[267] SCHARSTEIN, D., AND SZELISKI, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision 47*, 1-3 (2002), 7–42. 30, 97

[268] SCHLÜNS, K., AND TESCHNER, M. Fast separation of reflection components and its application in 3d shape recovery. In *Color and Imaging Conference* (1995), vol. 1995, Society for Imaging Science and Technology, pp. 48–51. 40

[269] SCHMITZ-VALCKENBERG, S., HOLZ, F. G., BIRD, A. C., AND SPAIDE, R. F. Fundus autofluorescence imaging: review and perspectives. *Retina 28*, 3 (2008), 385–409. 38

[270] SDIRI, B., BEGHDADI, A., CHEIKH, F. A., PEDERSEN, M., AND ELLE, O. J. An adaptive contrast enhancement method for stereo endoscopic images combining binocular just noticeable difference model and depth information. *Electronic Imaging 2016*, 13 (2016), 1–7. 126, 128, 129

[271] SELKA, F., NICOLAU, S. A., AGNUS, V., BESSAID, A., MARESCAUX, J., AND SOLER, L. Evaluation of endoscopic image enhancement for feature tracking: a new validation framework. In *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*. Springer, 2013, pp. 75–85. 63, 64, 126, 127, 128

[272] SELVAN, S., AND RAMAKRISHNAN, S. SVD-based modeling for image texture classification using wavelet transformation. *IEEE Transactions on Image Processing 16*, 11 (2007), 2688–2696. 58

[273] SHAFER, S. A. Using color to separate reflection components. *Color Research & Application 10*, 4 (1985), 210–218. 40

[274] SHAHRIL, R., BAHARUN, S., ISLAM, A. M., AND KOMAKI, S. Anisotropic contrast diffusion enhancement using variance for wireless capsule endoscopy images. In *International Conference on Informatics, Electronics & Vision* (2014), IEEE, pp. 1–6. 64

[275] SHEET, D., GARUD, H., SUVEER, A., MAHADEVAPPA, M., AND CHATTERJEE, J. Brightness preserving dynamic fuzzy histogram equalization. *IEEE Transactions on Consumer Electronics 56*, 4 (2010). 55

[276] SHEIKH, H. R., AND BOVIK, A. C. Image information and visual quality. *Transactions on image processing 15*, 2 (2006), 430–444. 123, 124

[277] SHEN, H.-L., AND CAI, Q.-Y. Simple and efficient method for specularity removal in an image. *Applied optics 48*, 14 (2009), 2711–2719. 41

[278] SHIMOJO, S., AND NAKAYAMA, K. Real world occlusion constraints and binocular rivalry. *Vision research 30*, 1 (1990), 69–80. 21

[279] SIDAR, I., DAVIDSON, T., KRONMAN, A., LIOR, M., AND LEVY, I. Endoscopic image enhancement using contrast limited adaptive histogram equalization (clahe) implemented in a processor, 2016. US Patent App. 15/155,814. 64, 85

[280] SIEGAL, S. *Nonparametric statistics for the behavioral sciences*. McGraw-hill, 1956. 73

[281] SILVER, D., HUANG, A., MADDISON, C. J., GUEZ, A., SIFRE, L., VAN DEN DRIESSCHE, G., SCHRITTWIESER, J., ANTONOGLOU, I., PANNEERSHELVAM, V., LANCTOT, M., ET AL. Mastering the game of go with deep neural networks and tree search. *Nature 529*, 7587 (2016), 484–489. 137

[282] SIMONE, G., PEDERSEN, M., AND HARDEBERG, J. Y. Measuring perceptual contrast in digital images. *Journal of Visual Communication and Image Representation 23*, 3 (2012), 491–506. 68

[283] SINGH, R., SATHANANTHAN, D., TAM, W., AND RUSZKIEWICZ, A. Endocytoscopy for diagnosis of gastrointestinal neoplasia: the expert's approach. *Video Journal and Encyclopedia of GI Endoscopy 1*, 1 (2013), 18–19. 49

[284] SIRINUKUNWATTANA, K., RAZA, S. E. A., TSANG, Y.-W., SNEAD, D. R., CREE, I. A., AND RAJPOOT, N. M. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *transactions on medical imaging 35*, 5 (2016), 1196–1206. 137

[285] SLATER, A., MORISON, V., SOMERS, M., MATTOCK, A., BROWN, E., AND TAYLOR, D. Newborn and older infants perception of partly occluded objects. *Infant Behavior and Development 13*, 1 (1990), 33–49. 21

[286] SLESAREVA, N., BRUHN, A., AND WEICKERT, J. Optic flow goes stereo: A variational method for estimating discontinuity-preserving dense disparity maps. *Pattern Recognition* (2005), 33–40. 32

[287] SMITH, A. T. Binocular vision: joining up the eyes. *Current Biology 25*, 15 (2015), R661–R663. 13

[288] SONG, L. M. W. K., ADLER, D. G., CONWAY, J. D., DIEHL, D. L., FARRAYE, F. A., KANTSEVOY, S. V., KWON, R., MAMULA, P., RODRIGUEZ, B., SHAH, R. J., ET AL. Narrow band imaging and multiband imaging. *Gastrointestinal endoscopy 67*, 4 (2008), 581–589. 38, 45

[289] SPEIDEL, S., DELLES, M., GUTT, C., AND DILLMANN, R. Tracking of instruments in minimally invasive surgery for surgical skill analysis. In *MIAR* (2006), Springer, pp. 148–155. 42

[290] STARCK, J.-L., MURTAGH, F., CANDES, E. J., AND DONOHO, D. L. Gray and color image contrast enhancement by the curvelet transform. *Transactions on image processing 12*, 6 (2003), 706–717. 60

[291] SUBEDAR, M. M., AND KARAM, L. J. Increased depth perception with sharpness enhancement for stereo video. In *Electronic Imaging* (2010), International Society for Optics and Photonics, pp. 75241B–75241B. 61, 65, 93, 101, 126, 128, 129, 135

[292] SUN, J., ZHENG, N.-N., AND SHUM, H.-Y. Stereo matching using belief propagation. *IEEE Transactions on pattern analysis and machine intelligence 25*, 7 (2003), 787–800. 32

[293] SWELDENS, W., ET AL. The lifting scheme: A new philosophy in biorthogonal wavelet constructions. *Wavelet Applications in Signal and Image Processing 3* (1995), 68–79. 116

[294] SZELISKI, R., AND ZABIH, R. An experimental comparison of stereo algorithms. In *International Workshop on Vision Algorithms* (1999), Springer, pp. 1–19. 30

[295] TADMOR, Y., AND TOLHURST, D. Calculating the contrasts that retinal ganglion cells and lgn neurones encounter in natural scenes. *Vision research 40*, 22 (2000), 3145–3157. 68

[296] TAN, P., QUAN, L., AND LIN, S. Separation of highlight reflections on textured surfaces. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* (2006), vol. 2, IEEE, pp. 1855–1860. 40

[297] TANAGHO, Y. S., ANDRIOLE, G. L., PARADIS, A. G., MADISON, K. M., SANDHU, G. S., VARELA, J. E., AND BENWAY, B. M. 2d versus 3d visualization: impact on laparoscopic proficiency using the fundamentals of laparoscopic surgery skill set. *Journal of Laparoendoscopic & Advanced Surgical Techniques 22*, 9 (2012), 865–870. 2

[298] TANG, J., LIU, X., AND SUN, Q. A direct image contrast enhancement algorithm in the wavelet domain for screening mammograms. *Journal of Selected Topics in Signal Processing 3*, 1 (2009), 74–80. 58, 65, 121, 122, 123, 136

[299] TANG, J., PELI, E., AND ACTON, S. Image enhancement using a contrast measure in the compressed domain. *IEEE Signal Processing Letters 10*, 10 (2003), 289–292. 58, 65, 122

[300] TANG, J., SUN, Q., AND AGYEPONG, K. An image enhancement algorithm based on a contrast measure in the wavelet domain for screening mammograms. In *International Conference on Image Processing* (2007), vol. 5, IEEE, pp. V–29. 58, 65, 121, 122, 136

[301] TANG, Y. Y. *Wavelet theory approach to pattern recognition*. World Scientific, 2009. 114

[302] TAPPEN, M. F., FREEMAN, W. T., AND ADELSON, E. H. Recovering intrinsic images from a single image. In *Advances in neural information processing systems* (2003), pp. 1367–1374. 40

[303] TCHAKAA, K., PAWARA, V. M., AND STOYANOVA, D. Chromaticity based smoke removal in endoscopic images. In *Proc. of SPIE Vol* (2017), vol. 10133, pp. 101331M–1. 42

[304] THOMPSON, B. J., RAHMAN, Z.-U., AND PARK, S. K. Multiscale retinex for improved performance in multispectral image classification., 2000. 53, 65

[305] TORR, P. H., AND MURRAY, D. W. The development and comparison of robust methods for estimating the fundamental matrix. *International journal of computer vision 24*, 3 (1997), 271–300. 28

[306] Tumblin, J., and Rushmeier, H. Tone reproduction for realistic images. *Computer graphics and Applications 13*, 6 (1993), 42–48. 79

[307] Uedo, N., Iishi, H., Tatsuta, M., Yamada, T., Ogiyama, H., Imanaka, K., Sugimoto, N., Higashino, K., Ishihara, R., Narahara, H., et al. A novel videoendoscopy system by using autofluorescence and reflectance imaging for diagnosis of esophagogastric cancers. *Gastrointestinal endoscopy 62*, 4 (2005), 521–528. 47

[308] Umeyama, S., and Godin, G. Separation of diffuse and specular components of surface reflection by use of polarization and statistical analysis of images. *Pattern Analysis and Machine Intelligence 26*, 5 (2004), 639–647. 41

[309] Vani, V., and Prashanth, K. M. Color image enhancement techniques in wireless capsule endoscopy. In *International Conference on Trends in Automation, Communications and Computing Technology* (2015), vol. 1, IEEE, pp. 1–6. 64, 85

[310] Veksler, O. Stereo matching by compact windows via minimum ratio cycle. In *International Conference on Computer Vision* (2001), vol. 1, IEEE, pp. 540–547. 31

[311] Veksler, O. Stereo correspondence by dynamic programming on a tree. In *Computer Society Conference on Computer Vision and Pattern Recognition* (2005), vol. 2, IEEE, pp. 384–390. 32

[312] Vu, C. T., Phan, T. D., Banga, P. S., and Chandler, D. M. On the quality assessment of enhanced images: A database, analysis, and strategies for augmenting existing methods. In *Southwest Symposium on Image Analysis and Interpretation* (2012), IEEE, pp. 181–184. 68, 124

[313] Wallach, H., and Floor, L. The use of size matching to demonstrate the effectiveness of accommodation and convergence as cues for distance. *Perception & Psychophysics 10*, 6 (1971), 423–428. 14

[314] Wang, K. K., Carr-Locke, D. L., Singh, S. K., Neumann, H., Bertani, H., Galmiche, J.-P., Arsenescu, R. I., Caillol, F., Chang, K. J., Chaussade, S., et al. Use of probe-based confocal laser endomicroscopy (pcle) in gastrointestinal applications. a consensus report based on clinical evidence. *United European gastroenterology journal 3*, 3 (2015), 230–254. 48

[315] Wang, L., Jin, H., Yang, R., and Gong, M. Stereoscopic inpainting: Joint color and depth completion from stereo images. In *Conference on Computer Vision and Pattern Recognition* (2008), IEEE, pp. 1–8. 100

[316] WANG, T. D., AND TRIADAFILOPOULOS, G. Autofluorescence imaging: have we finally seen the light? *Gastrointestinal Endoscopy 61*, 6 (2005), 686–688. 48

[317] WEINSTEIN, S. The perception of depth in the absence of texture-gradient. *The American journal of psychology 70*, 4 (1957), 611–615. 25

[318] WHEATSTONE, C. Contributions to the physiology of vision.–part the first. on some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical transactions of the Royal Society of London* (1838), 371–394. 13

[319] WHEATSTONE, C. *The Scientific Papers of Sir Charles Wheatstone*. Cambridge University Press, 2011. 14

[320] WINKLER, S. Analysis of public image and video databases for quality assessment. *IEEE Journal of Selected Topics in Signal Processing 6*, 6 (2012), 616–625. 72

[321] WISKOTT, L., KRÜGER, N., KUIGER, N., AND VON DER MALSBURG, C. Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence 19*, 7 (1997), 775–779. 56

[322] WONG, K. Y., DUNN, F. A., AND BERSON, D. M. Photoreceptor adaptation in intrinsically photosensitive retinal ganglion cells. *Neuron 48*, 6 (2005), 1001–1010. 12

[323] WOO, W., AND ORTEGA, A. Overlapped block disparity compensation with adaptive windows for stereo image coding. *Transactions on Circuits and Systems for Video Technology 10*, 2 (2000), 194–200. 31

[324] WU, J. C.-H., LIN, G.-S., HSU, H.-T., LIAO, Y.-P., LIU, K.-C., AND LIE, W.-N. Quality enhancement based on retinex and pseudo-hdr synthesis algorithms for endoscopic images. In *Visual Communications and Image Processing* (2013), IEEE, pp. 1–5. 62, 64

[325] YADAV, G., MAHESHWARI, S., AND AGARWAL, A. Fog removal techniques from images: A comparative review and future directions. In *Signal Propagation and Computer Technology (ICSPCT), 2014 International Conference on* (2014), IEEE, pp. 44–52. 42

[326] YAMASHITA, H., AOKI, H., TANIOKA, K., MORI, T., AND CHIBA, T. Ultra-high definition (8k uhd) endoscope: our first clinical success. *SpringerPlus 5*, 1 (2016), 1445. 138

[327] YANG, X., LING, W., LU, Z., ONG, E. P., AND YAO, S. Just noticeable distortion model and its applications in video coding. *Signal Processing: Image Communication 20*, 7 (2005), 662–680. 96

[328] YENDRIKHOVSKIJ, S. N., BLOMMAERT, F., AND DE RIDDER, H. Color reproduction and the naturalness constraint. *Color Research & Application 24*, 1 (1999), 52–67. 78

[329] YOON, K.-J., CHOI, Y., AND KWEON, I. S. Fast separation of reflection components using a specularity-invariant image representation. In *International Conference on Image Processing* (2006), IEEE, pp. 973–976. 41

[330] ZEKI, S. M., ET AL. Functional specialisation in the visual cortex of the rhesus monkey. *Nature 274*, 5670 (1978), 423–428. 13

[331] ZHANG, L., AND TAM, W. J. Stereoscopic image generation based on depth images for 3d tv. *Transactions on broadcasting 51*, 2 (2005), 191–199. 100

[332] ZHANG, L., TAM, W. J., AND WANG, D. Stereoscopic image generation based on depth images. In *International Conference on Image Processing* (2004), vol. 5, IEEE, pp. 2993–2996. 100

[333] ZHAO, Y., CHEN, Z., ZHU, C., TAN, Y.-P., AND YU, L. Binocular just-noticeable-difference model for stereoscopic images. *IEEE Signal Processing Letters 18*, 1 (2011), 19–22. 35, 62, 96

[334] ZHONG, R., HU, R., WANG, Z., AND WANG, S. 3d hybrid just noticeable distortion modeling for depth image-based rendering. *Multimedia Tools and Applications 74*, 23 (2015), 10457–10478. 96

[335] ZUIDERVELD, K. Contrast limited adaptive histogram equalization. In *Graphics gems IV* (1994), Academic Press Professional, Inc., pp. 474–485. 55, 70, 85