3-8-2021 2:00 PM

Laparoscopic Image Recovery and Stereo Matching

Wenyao Xia, *The University of Western Ontario*

Supervisor: Terry Peters M., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Medical Biophysics

© Wenyao Xia 2021

Abstract

Laparoscopic imaging can play a significant role in the minimally invasive surgical procedure. However, laparoscopic images often suffer from insufficient and irregular light sources, specular highlight surfaces, and the lack of depth information. These problems can negatively influence the surgeons during surgery, and lead to erroneous visual tracking and potential surgical risks. Thus, developing effective image-processing algorithms for laparoscopic vision recovery and stereo matching is of significant importance. Most of related algorithms are effective on nature images, but less effective on laparoscopic images.

The first purpose of this thesis is to restore low-light laparoscopic vision, where an effective image enhancement method is proposed by identifying different illumination regions and designing the enhancement criteria for desired image quality. This method can enhance the low-light region by reducing noise amplification during the enhancement process. In addition, this thesis also proposes a simplified Retinex optimization method for non-uniform illumination enhancement. By integrating the prior information of the illumination and reflectance into the optimization process, this method can significantly enhance the dark region while preserving naturalness, texture details, and image structures. Moreover, due to the replacement of the total variation term with two l_2-norm terms, the proposed algorithm has a significant computational advantage.

Second, a global optimization method for specular highlight removal from a single laparoscopic image is proposed. This method consists of a modified dichromatic reflection model and a novel diffuse chromaticity estimation technique. Due to utilizing the limited color variation of the laparoscopic image, the estimated diffuse chromaticity can approximate the true diffuse chromaticity, which allows us to effectively remove the specular highlight with texture detail preservation.

Third, a robust edge-preserving stereo matching method is proposed, based on sparse feature matching, left and right illumination equalization, and refined disparity optimization processes. The sparse feature matching and illumination equalization techniques can provide a

good disparity map initialization so that our refined disparity optimization can quickly obtain an accurate disparity map. This approach is particularly promising on surgical tool edges, smooth soft tissues, and surfaces with strong specular highlight.

Summary for Lay Audience

Laparoscopic surgery is increasingly performed as a minimally invasive procedure for many life-threatening diseases. It uses stereoscopic laparoscopes/endoscopes to intuitively visualize the organ surface in the body and manipulate various surgical tools. In principle, the data acquired are high-quality HD stereoscopic images, with the potential to provide secondary information to the surgeons, such as 3D reconstructed scenes, spectroscopic tissue analysis, and the enhancement of subtle repetitious motions. Nevertheless, all these applications assume that the images are required in an ideal environment that is free of artifacts, noises, and illumination non-uniformities. In practice, the laparoscopic images suffer from a number of problems including a high amount of specularity, insufficient illuminations, and a relatively narrow field of view. Due to these problems, computer-assisted interventions such as 3D scene reconstruction and motion magnification suffer greatly in terms of robustness and produce a large amount of erroneous results. Thus, to overcome this problem, I would like to transform the non-ideal laparoscopic images into ideal ones suitable for other computer-assisted interventions and improve their accuracies. In this thesis, I have proposed several methods for specular highlight removal and image enhancement to improve the image quality of laparoscopic/endoscopic images and demonstrated that they are beneficial to the surgeons during the surgery. Moreover, I have also proposed a fast and robust stereo-matching algorithm for laparoscopic images to provide surgeons with accurate depth information and 3D surface reconstructions.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Terry Peters for providing me this opportunity to join his world-class research lab, and for all his valuable guidance and continued encouragement. He motivated me to pursue my passion and supported my research in every way he can. His profound knowledge and suggestions in medical image guidance are both insightful and instrumental for my research. Not only Dr.Peters gave me great help academically, he also encouraged me to participate many leadership activities. Because of his support, I organized the first Deep Learning Club in Robarts Research Institute, which was a great success. Every year, Dr.Peters and Jackie Williams would also host a fantastic Chinese new year event, which was very heart warming and memorable. In addition, I would like to thank both of them for taking their precious time to proofread my research papers and helping me correct many grammatic mistakes.

I would like to thank my advisory committee members, Dr. Elvis Chen, Dr. Stephen Pautler, and Dr. Hanif Ladak, for their valuable comments and suggestions during my Ph.D graduate study. I would also like to thank Dr. John Baxter, Dr. Uditha Jayarathne, Dr. Jonathan McLeod, Dr. Jing Yuan, and Dr. Xiongbiao Luo, for their kind help in related research discussions. I would like to thank John Moore and Adam Rankin for their technical supports for many collaboration projects. Also, I would like to express my appreciation to Janette Wallace, Jackie Williams, Kitty Wong, and Johanne Langford for their time and efforts to organize many fantastic events and conferences.

Finally, I would like to thank my parents, who have unconditionally supported me even under the most stressful and difficult times. Their devotion and love are the foundation for all I have achieved today.

Contents

List of Figures

List of Tables

List of Abbreviations

OS   Open Surgery
MIS   Minimally Invasive Surgery
2D   Two Dimensional (typically (x,y) domain)
3D   Three Dimensional (typically (x,y,z) domain)
HD   High Definition
WTA   Winner-Takes-All
ASW   Adaptive Support Weights
BF   Bilateral Filter
GF   Guided Filter
IGF   Iterative Guided Filter
NIQE   Naturalness Image Quality Evaluator
IU   Illumination Uniformity
PSNR   Peak Signal-to-Noise-Ratio
SSIM   Structural Similarity Image Measurement
SSI   Structural Similarity Index
RGB   Red-Green-Blue (a type of color model)
HSV   Hue-Saturation-Value (another type of color model)
RMSD Root Mean Square Distance
GPU   Graphics Processing Unit
CVF   Cost Volume Filtering
SIFT   Scale-Invariant Feature Transform

Chapter 1

Introduction

1.1   Background and significance

Traditional surgery is commonly used to deliver therapy to the desired location in the human body. It has become a necessary treatment option for many life-threatening diseases due to being efficient and controllable procedure to treat the patients. The major challenge of surgery is to gain access and visualization to the target regions inside the human body. To achieve this, the surgeons need to remove the obstructing tissue and make large incisions to allow them to visualize and operate directly. Due to the large incisions, the traditional surgery is also named as open surgery (OS). Nevertheless, the large incisions often lead to huge undesired side-effects, which not only cause a significant amount of pain with long recovery time, but also produce other life-threatening complications such as hemorrhage, wound infection, and shock, etc. Aside from the high mortality rate, many patients would require secondary interventions to address these OS-related complications.

To reduce these unnecessary traumas and risks to the patients, minimally invasive surgery (MIS) was introduced as an alternative to the traditional surgery. In general, the minimally invasive procedure utilizes specialized tools and scopes for surgeons to gain access to the target regions and observe the surgical scene. Instead of making large incisions for surgeons to operate directly, the MIS tools and scopes are inserted into the patient body via keyhole openings (small openings) or natural orifices to allow the surgeons to perform various kind of operations

indirectly [39]. Since the incision wound is reduced, many associated complications can be significantly reduced as well. In recent years, the MIS has become popular in many surgical interventions for treating brain, heart, and abdomen and demonstrated better surgical outcomes than the traditional OS. An example of the modern MIS is shown in Fig. 1.1.

Figure 1.1: An illustration of the modern minimally invasive surgery.

1.1.1 Development of minimally invasive surgery

The implementation of minimally invasive surgery (MIS) is not straightforward. There are two major challenges to implement the MIS. Not only the surgeons need to have access to the target region to accurately perform surgery, but also they need to visualize the surgical scene in real time through the obstructed human tissue. In past two centuries, various specialized tools were developed to help the surgeons overcome the access difficulty in performing operations on the target organs via a small incision. For a detailed survey, the reader may refer to [26] and references therein.

In 1805, Phillip Boziini developed the first cystoscope, a laparoscope for the urinary bladder, and conducted an animal study. Utilizing this design, surgeons in the 19th century experimented with similar operations such as cystoscopy, proctoscopy, laryngoscopy, and esophagogastroscopy.

In 1901, George Kelling developed the first celioscopy technique on dogs, which inflates the abdominal cavity with gas and have the laparoscopy, through the abdominal wall. Later in 1910, George Kelling performed the first trial on humans, and in the meantime, Hans Christian also performed multiple animal/clinical trials and named such procedure Laparoscopy. In 1929, Heinz Kalk developed a superior laparoscope with an improved lens and forward viewing scope, and in 1933 Karl Fervers developed an adhesion technique using cautery to treat internal bleeding during the procedure. While the development of MIS has been slowly gaining recognition, due to the limitation of tools and technologies, many concerns were raised regarding the effectiveness of the MIS, and only limited laparoscopy procedures were allowed.

From 1960 to 1970, Kurt Semm improved many aspects of the laparoscope such as introducing the automated insufflator to blow gas into the abdominal cavity, a suction irrigator to inject water to clean blood, a safer electrocoagulation instrument to cauterize bleeding tissue, intracorporeal and extracorporeal knotting techniques, and an electoral morcellator to remove large tissue mass from inside of the body. In 1971, Harrith Hasson developed the Hasson trocar to withdraw fluid from the human body. In 1983, Kurt Semm performed the first MIS to remove the appendix, and Erich Muhe also performed an MIS to remove the gallbladder.

The most important advancement in laparoscopic surgery is the introduction and integration of digital imaging technology and video technology.

In early 1980, Camran Nezhat developed a video system adequate for surgery [26]. In this setup, the surgeon inserts a video camera with a light source into the patient body cavity, and can visualize the surgical site in a display monitor in front of the surgeon. In addition, he also built a multidisciplinary team to work with video laparoscopic surgeries. Gradually, people began to see the benefits, and MIS has become popular among surgeons and in 1990, MIS procedures such as appendectomy became common practice. Aside from abdominal surgeries, vascular surgery was also revolutionized by MIS during the 90s. Since 2000, robot-assisted surgical systems were introduced to MIS, where most of the surgical tools are installed on robotic arms and remotely controlled by the surgeon using a control console. In addition, the con-

ventional monoscopic camera was replaced by a stereoscopic camera, and the captured stereo images were displayed at the surgeon console via a binocular display. These improvements not only provide more accurate manipulation of surgical tools but also improved many ergonomic issues of traditional laparoscopic surgery, allowing the surgeons to perform the surgery more intuitively and comfortably.

1.1.2   Vision artifacts in laparoscopic/endoscopic imaging

Due to the obstruction of direct vision from the small keyhole opening, the effectiveness of MIS relies on optical digital cameras for indirect visualization of the surgical scene. For visualization, a digital video camera is installed on the laparoscope, along with an attached light source, to capture the surgical scenes and display them on monitors. Although this indirect visualization method can successfully avoid the line-of-sight obstruction due to the surrounding tissues from the small incision, it also has several drawbacks. In principle, the data acquired are high-quality HD stereoscopic images, with the potential to provide secondary information to the surgeons, such as 3D reconstructed scenes, spectroscopic tissue analysis, and the enhancement of subtle repetitious motions. Nevertheless, all these applications assume that the images are acquired under an ideal environment. In practice, the received laparoscopic images are often degraded by non-uniform illumination, noise, insufficient and irregular light sources, and specularity, as shown in Fig. 1.2 (a) and (b). In addition, the 2D video images captured from the digital camera do not contain depth information and Fig. 1.2(c) and (d) display the 2D stereo laparoscopic images without depth information. These drawbacks can negatively affect the surgeon's vision during surgical guidance, and prevent the surgeon from clearly observing important structures during robotic surgery. Therefore, it is important to develop image-processing algorithms to restore such degradations and provide accurate depth information.
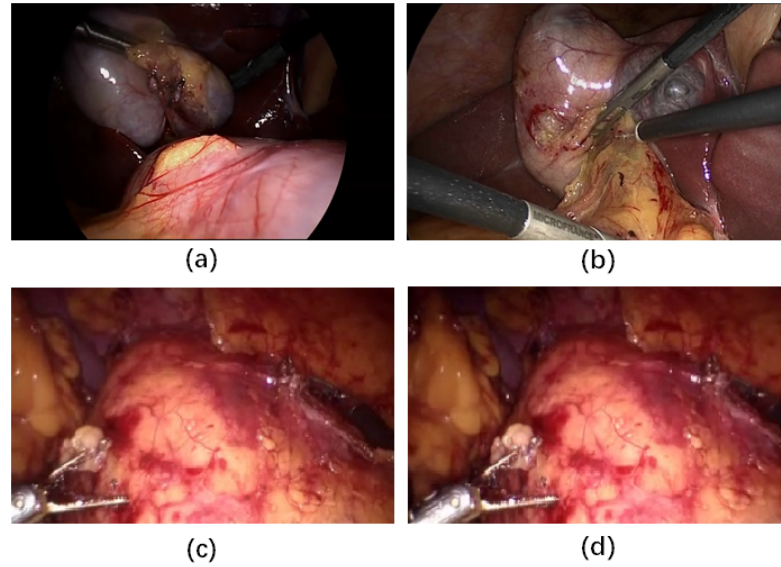
Figure 1.2: Examples of surgical laparoscopic images: (a) Low light image, (b) Specular highlight image, (c) and (d) Left and right stereoscopic image pairs, where the left pair is for "wall-eye" of parallel-eye viewing and the right pair is for cross-eye viewing.

1.2 Research status of laparoscopic imaging methods

To improve the visualization of digital images, scientists and researchers in computer vision have made significant advancements in the past decades. Many image processing algorithms were developed to intelligently remove undesired image artifacts, extract image content information, and mimic functions and features of the human visual system. There are many existing methods designed for generic image processing, such as visual tracking, image enhancement, image defogging, specular highlight removal, and stereo-matching. Nevertheless, most of the existing algorithms are designed using phantoms and natural images and there is a large difference between the images acquired from the natural scene and the images from the laparoscopic ones. For example, natural images tend to have abundant and smooth varying ambient lighting. Furthermore, the objects in the natural images have distinct colors and edges, and the images are usually captured at a greater depth, etc. In contrast, the laparoscopic images have much higher illumination variance due to the single light source and the images acquired from the laparoscopic scene are mostly monotonic. Moreover, the laparoscopic images have a much

lower depth due to the proximity of the camera. Because of these differences, directly apply-
ing conventional natural image processing algorithms on the laparoscopic images are often not
effective. Thus, there is need for incorporating laparoscopy-specific prior information in the
laparoscopic image algorithms for visualization improvements.

1.2.1   Vision enhancement

The insufficient and irregular illumination present major challenges for endoscopic surgery.
Often during surgery, the light source will be occluded by surgical tools and tissues, creat-
ing shadows and low-light regions on the operating organs, as illustrated in Fig. 1.3. These
issues unavoidably deteriorate the clear and high-quality visualization of both the organ and
its surroundings. As a result, the visibility of important organs and surgical tools at low-light
regions may be poor so that the risk of damaging critical structures is increased. Therefore, it
is necessary to develop image enhancement techniques to restore or augment such as degraded
visualization.
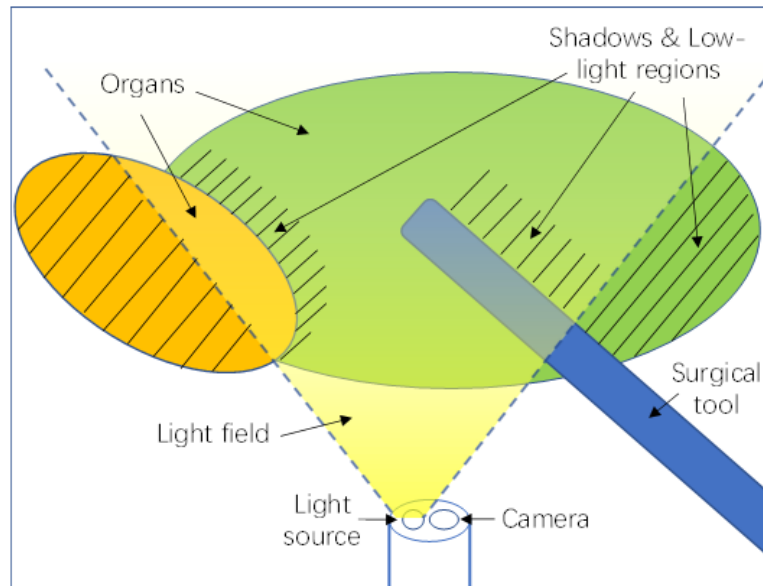
Figure 1.3: An illustration of the low-light surgical scene

Various natural image enhancement methods have been presented in the past decades.

Histogram-based methods were first proposed to adjust pixel values using the cumulative distribution function of the input image [2, 29]. Although this class of enhancement techniques is popular due to its simple and fast implementations, it often causes over-enhancement in the bright region due to the extremely narrow histogram of the low-light image. Thus, this method is not suitable for non-uniform illumination enhancement for laparoscopic images containing large illumination differences. The gamma correction-based method [23] is simple and effective in maintaining the visibility of the bright region, but its enhancement ability is insufficient for low-light regions. Tone mapping algorithms are effective on images with high dynamic range [14], however, they may also cause halo artifacts and over-enhancement, which degrade the quality of the enhanced image. Multiscale tone mapping algorithm [33] can mitigate these problems but suffers from color inconsistency, and while logarithmic processing algorithms may overcome the effect of saturated pixels in high exposure regions [42, 41], colour inconsistency still remains an important problem nevertheless. Deep learning-based methods for image enhancement have been developed in recent years. Among them, Hua [22] presented a Low-light image enhancement based on a joint generative adversarial network and image quality assessment, and Ren et al. [47] proposed a deep hybrid network for image enhancement. Although these learning-based methods were reported to be effective on a class of natural images, a large number of ground truth images must first be collected as the training set. For surgical scenarios, however, similar ground truth data are difficult to generate for medical images.

Retinex theory was developed by Edward Land [31] to model the colour constancy feature of the human visual system, such that a human observer would perceive the colour consistently under various illumination conditions. According to this theory, the human visual system perceives colours in a relative scale instead of an absolute one, and the colour perception depends on the local variation of lightness in the image region. Although the overall appearance of colour may change under different lighting, the relative colour differences between objects are constant. While this theory is used to explain the property of the human visual system, many computer vision algorithms are derived from retinex theory to enhance the local image contrast

and reveal details in the low-light image regions. Retinex-based image enhancement methods have recently received much attention due to their effectiveness in enhancing non-uniform illumination. The single-scale Retinex method [25] is the foundation of this field, while multiscale Retinex approaches [24, 34] offer various improvements to the original. By using a bright-pass filter to decompose the observed image into reflectance and illumination, this class of methods can enhance the lightness and image contrast by color restoration and histogram equalization. Recently, Wang et al. [47] proposed a multi-layer decomposition-based method for naturalness preservation. However, these methods suffer from blurred texture and edges as well as colour cast(distortion), caused by non-uniform illumination or scattering. Since estimating illumination and reflectance from a single observed image is an ill-posed problem, Retinex-based optimization methods have become popular, where different illumination regularization terms are used to establish the objective function [6]-[30]. The illumination is first estimated and then enhanced by Gamma correction to enhance the image. Illustrations of such approaches have been presented by Kimmel et al. [30], who first introduced a Retinex-based variational method, and Wang et al. [61] who proposed a Retinex-based Bayesian method to enhance images while preserving their naturalness. Zosso et al. [72] presented a Retinex-based unifying framework, and Fu et al. [11] proposed a Retinex-based optimization method with l_1 and l_2-norm gradient regularization terms to better capture the nature of the reflectance and illumination components. Parks et al. [43] presented an improved Retinex-based optimization method with l_2-norm gradient regularization term for reduced computational complexity. Guo et al. [16] proposed a Retinex-based optimization method with improved illumination map estimation by using a structure image as the prior information, while Rao et al. [45] presented a method with l_1 and l_2-norm error terms to model the reflectance and illumination terms. In summary, current Retinex-based optimization methods mainly focus on non-uniform enhancement by using the gradient regularization to estimate the illumination element.

Although the existing methods can somewhat improve the illumination of low-light images, there are two major obstacles to be overcome. First, in the low light region, the image

pixels may include noise, and most of image enhancement methods usually suffer from noise amplification in low-light region. Although the optimization-based enhancement algorithms are trying to address this problem, the algorithm convergence to the optimal solution is still an open issue. Second, the image naturalness, color, and texture details may be distorted during the enhancement process. So, it is desirable to develop an effective enhancement method to overcome these difficulties.

1.2.2 Specular highlight removal

Endoscopic/laparoscopic images often contain specular highlights on the organ surfaces, which are caused by the inherent frontal illumination as illustrated in the Fig. 1.4.
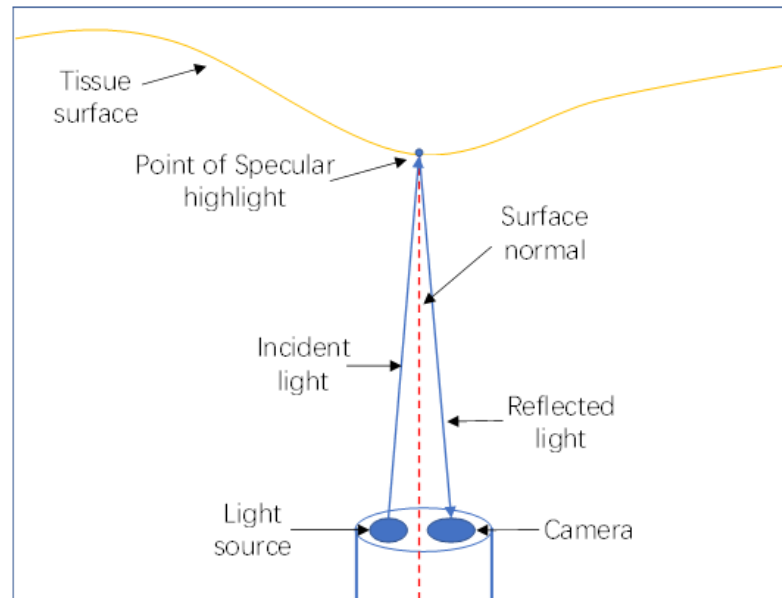
Figure 1.4: An illustration of specular highlight in endoscopic/laparoscopic procedure

The presence of strong specular highlights can both negatively affect the visual quality and degrade the subsequent tasks of computer vision algorithms, such as visual tracking and stereo reconstruction, and image segmentation [58, 40]. In addition, regions with specular highlights may contain vital information relating to the organ such as color and texture. Therefore, it is necessary to remove specular reflections while preserving the original color and texture details

of the organ surface for computer vision applications.

Various highlight removal methods have been presented in literature, and can be categorized into multi-and single-image methods. The multi-image method employs different light information from a group of images. For example, Wang et al. [60] proposed a energy minimization with respect to the local weighting coefficient for highlight removal from multiple images, based on a polarization filter, while Shah et al. [51] proposed a specular highlight removal method from image sequences by using feature corresponding points. In contrast, the single-image method separates the reflection component using a single image only. In addition to conventional color histogram analysis [28], categorically they can be divided into dichromatic reflection model-based [1]-[15],[50]-[68], inpainting-based [44]-[48], and learning-based methods [3]-[12].

The dichromatic reflection model was proposed by Shafer [50], which describes the interaction of the illumination with opaque dielectric materials. During surface reflection, the material surface immediately reflects a portion of the incident light, where the reflected light has a similar colour to the incident light. This type of reflection is often referred to as interface reflection or specular reflection. The remaining portion of the incident light will be absorbed and scattered by the surface materials, and then re-emitted at a different wavelength. The colour of this particular reflection depends on the surface material, and this reflection is named body reflection or diffuse reflection. Thus, the overall colour of the reflected light is the addition of both illumination colour and the diffuse reflection colour. Specular highlight occurs when the interface reflection greatly exceeds the diffuse reflection. The dichromatic highlight removal approach aims to separate these two types of reflections, where the diffuse reflection corresponds to the highlight-free image.

Since the introduction of the dichromatic reflection model, many model-based methods have been presented. In particular, by introducing a pseudo specular-free image Tan and Ikeuchi [56] proposed a highlight removal method without explicit color segmentation, based on diffuse pixel identification and chromaticity analysis [57]. Yang et al. [68] presented a robust

diffuse chromaticity estimation method by applying a low-pass filter. To enhance computation speed, Shen and Cai [52] approximated the chromaticity of diffuse reflection using the pseudo specular-free image. By choosing a proper chromaticity threshold, Shen and Zheng [53] proposed a intensity ratio-based highlight removal method with color clustering so that the specular components can be easily computed. By using a bilateral filter, Yang et al. [67] proposed a specular highlight removal method under appropriate bilateral filter parameters. Suo et al. [55] proposed an effective highlight removal method, by introducing adaptive color clustering so that diffuse components can be quickly recovered. By using color clustering, Ren et al. [46] proposed a highlight removal method with color-line constraint so that the specular and diffuse components are fast computed. By using adaptive principal component analysis, Li et al. [32] presented a highlight removal method. Most of the approaches outlined above were reported to be effective on natural images, but less so on medical images, since some assumptions made for natural images are not applicable in a medical context such as endoscopy. In contrast, the inpainting-based method used in medical imaging mainly includes both highlight detection and inpainting correction. The inpainting-based method provides visually pleasing and coherent images when highlight regions are very small, yet their texture details in the inpainted specular highlight region are often lost, due to neighborhood interpolation limitation. Recently, deep learning-based methods were successfully used for image specular highlight removal [12], but the learning-based method requires a good training set to be available and thus can not be generalizable for different image scenes.

In summary, developing an accurate and efficient method for removing specular highlights from a single image with texture detail preservation remains a challenging issue.

1.2.3 Stereo matching

In traditional endoscopic/laparoscopic surgery, the surgeon is manipulating the tools while looking at the surgical scene displayed in the LCD monitor. Since the video displayed on the monitor is two-dimensional, the surgeons often need to rely on their experience to accurately

and safely perform the surgery due to the depth information lack. A comparative study shows that surgeons with depth perception would greatly outperform those using traditional 2D endoscope [38]. In addition, the 3D information can also be used for tracking, augmented vision, and allowing on site registration between the surgical scene and treatment plans. With the development of minimally invasive surgery technology, binocular stereo laparoscopes have been introduced as a potential solution to address the depth problem. By developing a binocular stereo matching technique, it becomes possible to explicitly estimate the depth information and reconstruct the 3D surgical scene [10]-[20].

Stereo matching is a common technique used in computer vision to estimate the distance between the camera and imaged objects using a stereo camera. Inspired by the human visual system, the stereo matching algorithm seeks to establish correspondences between images captured by the left and right cameras as shown in Fig. 1.5. Based on these correspondences, the distances between the corresponding left and right feature points can be computed, also known as disparities. Given the disparities and camera intrinsics, such as the baseline distance (the distance between the left and right camera) and the focal depth, I can estimate the depth information using triangulation, where the ratio between the depth and baseline is equivalent to the ratio between disparity and the difference of depth and focal length. The depth map provides depth information for each pixel, where the depth value describes the distance between the object surface and the camera. When combining the depth information with the 2D pixel coordinate information, each pixel can be mapped into the 3D space. Using the camera position as origin, a 3D point cloud can be created by considering each pixel as a point in the 3D space, representing a sampled surface or environment. Each point can also be represented with additional information such as colour and intensity, allowing us to map the RGB information from the original 2D image onto the 3D point cloud. This technique can help us visualize the reconstructed 3D scene.

Many stereo matching methods for natural images have been developed in recent years, and for a detailed review on the stereo matching methods, the reader may refer to [49]. Binoc-
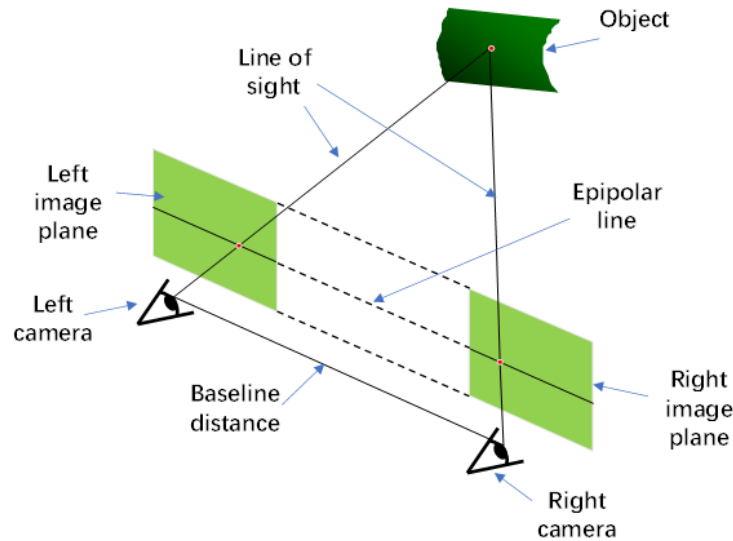
Figure 1.5: An illustration of the principle of stereo-matching

ular stereo matching methods can be divided into local methods and global methods. The global methods estimate disparity maps by minimizing an energy function with a data term and a smoothness term. The optimal results are usually obtained using either graph cut algorithms [4, 5], belief propagation algorithms [17, 54], and variational approaches. The local methods use image intensity or color information to obtain a disparity map based on cost volume computation, cost aggregation, and disparity optimization, where the latter employs a winner-takes-all (WTA) algorithm. Among these methods, Gerrits and Bekaert [13] presented a segmentation-based method to reduce computational cost. However, this method may fail if the segmentation overlaps with depth boundaries. To enhance stereo matching accuracy, Yoon and Kweon [70] introduced the adaptive support weight (ASW) method for stereo matching, where the cost aggregation step is based on filtering the cost volume with a bilateral filter (BF) [59]. In order to speed up the computation of the BF weight function, Yang [65] presented a hardware-efficient bilateral filter for stereo matching. Hosni et al. [19] applied a guided filter (GF) to the cost volume and demonstrated that the resulting algorithm outperformed most local methods in terms of both speed and accuracy [21]. To avoid the depth discontinuity caused by the kernel windows of GF-based methods, Yang et al. [64] proposed an adaptive guided

filtering method. By combining the GF and BF to filter the cost volume, Hamzah et al. [18] presented the iterative guided image filter (IGF) for better edge-preservation. To improve the performance of the ASW methods, Zhang et al. [71] subsequently proposed a cross-scale cost aggregation-based method. Yang proposed a tree-filtering method to address the kernel size problem during cost aggregation; in addition, the non-local cost aggregation using tree-filter has extremely low computational complexity [66]. Cheng et al. improved the tree-filtering method by using a horizontal tree and a vertical tree filter for cost aggregation [9]. Although these non-local methods can handle low texture regions, they show poor performance in regions containing dense textures.

To improve the algorithm's robustness towards handling textures, Wu et al. proposed a strategy to fuse ASW for stereo matching by combining a local edge-aware filter and a non-local edge-aware filter [63]. For image-guided robotic surgery, Chang et al. [7] proposed a convex optimization scheme on the cost-volume for efficient dense disparity matching, creating a smooth disparity map for the tissue surface. Nevertheless, these depth boundaries may also be blurred. In 2017, Chang et al. [8] proposed a pixel-based adaptive normalized cross-correlation as a matching cost for illumination invariant stereo matching, while in 2018, Khan et al. [27] follow the traditional local correspondence search framework with a newly proposed intensity guided cost metric for accuracy and speed improvements. In summary, local correspondence algorithms may efficiently obtain an approximate disparity map, but do not perform well for regions with weak or repetitive textures, due to the assumption that all pixels in each image patch have similar disparities. The global methods may have more accurate matching results but suffer from high computational cost due to the optimization iteration process [49].

Although current stereo matching approaches may be effective on natural image pairs, their efficacy is still limited for the laparoscopic images, since these images often suffer from weak textures, monotonic colors, varying illumination, and strong specular highlights. Such problems deteriorate the quality of the laparoscopic images and affect the performances of these stereo matching algorithms. Therefore, a robust and accurate stereo matching for laparoscopic

images remains a challenging issue.

1.3   Vision assessment metrics

To validate the effectiveness of the proposed method in this thesis, both subjective and objective assessments are used. The subjective assessment is used to directly and intuitively inspect processed images with human eyes. The objective (quantitative) evaluation is to precisely parameterize the image properties in terms of quantitative metrics. To quantitatively evaluate the proposed method, the analytic assessment metrics are usually employed by using benchmark images with publicly available ground truth. On the other side, since no ground truth in surgical endoscopic images is available, the evaluation of the enhanced and recovered images are necessarily through blind image qualitative assessments methods.

To quantitatively evaluate enhanced images, six blind image assessment metrics are used. The first image metric is the naturalness image quality evaluator (NIQE), based on statistical regularities from natural and undistorted images [37]. Since conventional NIQE is based on the statistical model trained with naturral images, it is more suitable for evaluating natural images. For our study, we re-train the statistical model based on hand-picked pristine endoscopic images that are free of visible image noise/artifacts and have relatively bright illumination. A lower NIQE value represents a higher image quality. The second image metric is the illumination uniformity (IU), which can describe the local consistency of enhanced images by computing the local intensity variations of the estimated illumination. Since the smaller variation would indicate higher local consistency, a lower IU value would represent a higher image quality in terms of illumination distribution.

The other four medical image metrics were suggested by Luo et al. [35]. They include a sharpness metric characterizing how much structural information of contours and boundaries on images, a naturalness metric describing how natural endoscopic vision appears, a contrast metric depicting the difference in luminance to describe the distinguishbility of the region

of interest, and a hybrid metric representing an overall balance of the previous metrics. For simplified comparison, we evaluate the quality of enhanced images by using the average of four blind metric values. The higher average metric values will indicate the higher image quality of the enhanced endoscopic image.

To analyze highlight removal images quantitatively, the peak signal-to-noise ratio (PSNR) and structural similarity image measurement (SSIM) are used as objective measures [62] for the benchmark images with ground truth. Next, the accuracy of surface reconstructions is compared before and after applying our highlight removal algorithm to evaluate its impact on stereo-matching. This thesis also use the benchmark endoscopic image data from the "TMI Distance dataset" with ground truth surface reconstructions [36].

To analyze disparity maps estimated by stereo matching methods quantitatively, the structural similarity measure [69] is used. Assume that a disparity map is estimated by using left and right image, the right image is then warped by using the disparity map as a deformation field. As a result, the warped right image should be similar to the left image. The quantitative objective assessment metric of the endoscopic disparity map is defined as the SSI value on $[0, 1]$ between the left image and the warped right image. The closer the SSI gets to 1, the better the disparity map is estimated. Next, the surface reconstruction error measurement [36] is employed, based on benchmark endoscopic stereo images from the "TMI Distance Dataset" and "TMI Angle Dataset" as mentioned above.

1.4   Research Challenges

Modern minimally invasive interventions rely on 2D video images captured by laparoscopic or endoscopic imaging system to perform sophisticated surgical tasks without the need for open surgery. The acquired laparoscopic or endoscopic video images are lack of depth information, and are often degraded due to surgical instrument and environment limitations. In order to improve guidance for minimally invasive surgery, this thesis aims to restore such degraded

video images in three main areas:

Vision augmentation of laparoscopic/endoscopic images

Due to the insufficient and irregular light sources, and the hostile imaging environment inside the human body, the laparoscopic/endoscopic images often suffer from local low-light region, noise, and non-uniform illumination problems. These problems can decrease the image quality and negatively affect the surgeon's, or surgical, performance, which may lead to potential surgical risks during surgery. This thesis aims to enhance the on-site laparoscopic/endoscopic image qualities in a surgical setting.

Specular highlight removal for laparoscopic/endoscopic images

Due to the proximity of the camera light source and organ surfaces, received images often suffer from specular highlight problems. The presence of specular highlight is a critical issue for both natural and medical images such as those produced by laparoscopes, which can lead to erroneous visual tracking, stereo reconstruction, and image segmentation. Existing methods have only been effective on natural images or medical images with textureless regions. This thesis aims to develop an effective specular highlight removal method for laparoscopic/endoscopic images, while preserving the original color and texture details of the organ surface.

Stereo matching for laparoscopic image pairs

Although the laparoscopic/endoscopic imaging system can capture and display the surgical scene inside the patient's body, the surgeons cannot perceive the depth information on the 2D monitor, which could affect the surgeon ability to perform the surgery safely and accurately. With the introduction of the binocular stereo laparoscope, it becomes possible to explicitly acquire accurate depth information using stereo matching techniques. This thesis aims to develop an effective stereo matching method for providing depth information of laparoscopic image pairs. It will be useful for laparoscopic/endoscopic three-dimensional (3-D) surface recon-

struction. Moreover, it is also a route to surface-based image registration, as well as allowing views of the surface different from the original stereo perspective.

1.5   Thesis Outline

This thesis is presented in integrated article format with Chapter 2-5 derived from published and submitted journal articles. I was the first author in all the supporting articles for these chapters and I was responsible for designing, developing, experimenting, and analyzing the proposed methods. I really appreciate the support and contributions from all the co-authors and this thesis would not have been possible without them.

1.5.1   Chapters 2 and 3: Vision enhancements

With intrinsic non-uniform illumination variations, the laparoscopic/endoscopic images often suffer from low visibility. This in turn may cause incorrect targeting, surgical risk, and extended operating time during laparoscopic surgery. Chapter 2 proposes a parametric approach to enhance the low-light laparoscopic images. Threshold image segmentations are used to decompose the low-light image into different regions, and each image region is enhanced differently. The enhanced image regions are combined together to obtain the final enhanced image. This method can reduce noise amplification during the enhancement process and avoid over enhancement in the bright region. This chapter quantitatively evaluates the proposed algorithm by using two blind assessment metrics: NIQE and IU. Chapter 3 proposes a simplified Retinex optimization approach with prior information fusion. A texture-free structural image is first computed to help me obtain the illumination and reflectance prior enhanced information. An objective function is proposed, consisting of a data fidelity term, prior enhanced illumination term, and prior enhanced reflectance term. These regularization terms can satisfy the convergence condition and preserve the overall naturalness and textures. By minimizing this objective function, the optimal reflectance and illumination can be obtained. As a result, this method can

efficiently enhance the non-uniform illumination images, while preserving image naturalness, details, and texture information. Moreover, the proposed algorithm has low computational complexity and satisfies the convergence condition. The performance of the proposed algorithm is quantitatively evaluated by four blind image metrics and a preference user study with experienced clinicians. In contrast, the proposed parametric approach is a non-optimization method with robust performance against noise.

1.5.2   Chapter 4: Specular highlight removal

Strong specular highlights often appears in laparoscopic images, which negatively affect the visual quality and degrade the subsequent tasks of computer vision algorithms, such as visual tracking, stereo reconstruction, and image segmentation. Chapter 4 proposes a global optimization method to remove specular highlight in laparoscopic images, based on the dichromatic reflection model. In this work, I first used the adaptive inpainting technique to correct the colour at highlight regions and then minimized a convex cost function to obtain the optimal intensity values at highlight regions. The colour correction is based on hue-saturation-value (HSV) representation instead of the traditional red-green-blue (RGB) representation. This is because, under white light, the highlight affects the saturation component of the colour, and using HSV colour representation can improve the accuracy and efficiency of the algorithm. The proposed method is quantitatively validated by using specular highlight images with highlight-free ground truth and a preference user study with experience clinicians. Moreover, we also demonstrate that our specular highlight removal can effectively enhance the accuracy of stereo matching on laparoscopic images.

1.5.3   Chapter 5: Stereo matching method

In minimally invasive surgery, stereo matching can provide depth information to surgeons so as to increase the safety and accuracy of surgical procedures. Although current stereo matching methods have been reported to work well on natural images, but for laparoscopic images,

they are still limited by illumination differences, weak texture content, specular highlights, and occlusions. Chapter 5 proposes a stereo matching method for stereo laparoscopic images, comprising a fast sparse-dense matching step, left and right illumination equalization, and refined disparity optimization. The sparse-dense matching approach uses feature point detection, feature matching, and interpolation to create an initial estimate of the disparity map. Based on the matched features, the intensity ratios are computed between the matched points to equalize the illumination between left and right images. Finally, an objective function is minimized to obtain the optimal disparity map. The initial disparity map and illumination equalization can improve the accuracy and efficiency of the optimization process. A popular blind qualitative metric, SSI, is used to evaluate the proposed method. Moreover, a laparoscopic image dataset with surface reconstruction ground truth is used to further evaluate the robust performance of the proposed method.

Bibliography

[1] Yasuhiro Akashi and Takayuki Okatani. Separation of reflection components by sparse non-negative matrix factorization. In *Asian Conference on Computer Vision*, pages 611–625. Springer, 2014.

[2] Tarik Arici, Salih Dikbas, and Yucel Altunbasak. A histogram modification framework and its application for image contrast enhancement. *IEEE Transactions on image processing*, 18(9):1921–1935, 2009.

[3] Jorge Bernal, Javier Sánchez, and Fernando Vilarino. Impact of image preprocessing methods on polyp localization in colonoscopy frames. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7350–7354. IEEE, 2013.

[4] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-

flow algorithms for energy minimization in vision. *IEEE transactions on pattern analysis and machine intelligence*, 26(9):1124–1137, 2004.

[5] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on pattern analysis and machine intelligence*, 23(11):1222–1239, 2001.

[6] Bolun Cai, Xianming Xu, Kailing Guo, Kui Jia, Bin Hu, and Dacheng Tao. A joint intrinsic-extrinsic prior model for retinex. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4000–4009, 2017.

[7] Ping-Lin Chang, Danail Stoyanov, Andrew J Davison, et al. Real-time dense stereo reconstruction using convex optimisation with a cost-volume for image-guided robotic surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 42–49. Springer, 2013.

[8] Yong-Jun Chang and Yo-Sung Ho. Pixel-based adaptive normalized cross correlation for illumination invariant stereo matching. *Electronic Imaging*, 2017(5):124–129, 2017.

[9] Feiyang Cheng, Hong Zhang, Mingui Sun, and Ding Yuan. Cross-trees, edge and super-pixel priors-based cost aggregation for stereo matching. *Pattern recognition*, 48(7):2269–2278, 2015.

[10] Frédéric Devernay, Fabien Mourgues, and Ève Coste-Manière. Towards endoscopic augmented reality for robotically assisted minimally invasive cardiac surgery. In *Proceedings International Workshop on Medical Imaging and Augmented Reality*, pages 16–20. IEEE, 2001.

[11] Xueyang Fu, Yinghao Liao, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. *IEEE Transactions on Image Processing*, 24(12):4965–4977, 2015.

[12] Isabel Funke, Sebastian Bodenstedt, Carina Riediger, Jürgen Weitz, and Stefanie Speidel. Generative adversarial networks for specular highlight removal in endoscopic images. In *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10576, page 1057604. International Society for Optics and Photonics, 2018.

[13] Mark Gerrits and Philippe Bekaert. Local stereo matching with segmentation-based outlier rejection. In *The 3rd Canadian Conference on Computer and Robot Vision (CRV'06)*, pages 66–66. IEEE, 2006.

[14] Bo Gu, Wujing Li, Minyun Zhu, and Minghui Wang. Local edge-preserving multiscale decomposition for high dynamic range image tone mapping. *IEEE Transactions on image Processing*, 22(1):70–79, 2012.

[15] Jie Guo, Zuojian Zhou, and Limin Wang. Single image highlight removal with a sparse and low-rank reflection model. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 268–283, 2018.

[16] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016.

[17] Raj Kumar Gupta and Siu-Yeung Cho. Stereo correspondence using efficient hierarchical belief propagation. *Neural Computing and Applications*, 21(7):1585–1592, 2012.

[18] Rostam Affendi Hamzah, A Fauzan Kadmin, M Saad Hamid, S Fakhar A Ghani, and Haidi Ibrahim. Improvement of stereo matching algorithm for 3d surface reconstruction. *Signal Processing: Image Communication*, 65:165–172, 2018.

[19] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, 35(6):1397–1409, 2012.

[20] Heiko Hirschmuller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on pattern analysis and machine intelligence*, 30(2):328–341, 2007.

[21] Asmaa Hosni, Christoph Rhemann, Michael Bleyer, Carsten Rother, and Margrit Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):504–511, 2012.

[22] Wei Hua and Youshen Xia. Low-light image enhancement based on joint generative adversarial network and image quality assessment. In *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6. IEEE, 2018.

[23] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE transactions on image processing*, 22(3):1032–1041, 2012.

[24] Bo Jiang, Glenn A Woodell, and Daniel J Jobson. Novel multi-scale retinex with color restoration on graphics processing unit. *Journal of Real-Time Image Processing*, 10(2):239–253, 2015.

[25] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997.

[26] William E Kelley Jr. The evolution of laparoscopy and the revolution in surgery in the decade of the 1990s. *JSLS: Journal of the Society of Laparoendoscopic Surgeons*, 12(4):351, 2008.

[27] Asim Khan, Muhammad Umar Karim Khan, and Chong-Min Kyung. Intensity guided cost metric for fast stereo matching under radiometric variations. *Optics Express*, 26(4):4096–4111, 2018.

[28] Javed Khan, Aamir Saeed Malik, Nidal Kamel, and Sarat Chandra Dass. Highlights removal using reflected energy and histogram analysis. *Journal of Engineering Research*, 7(1), 2019.

[29] Yeong-Taeg Kim. Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE transactions on Consumer Electronics*, 43(1):1–8, 1997.

[30] Ron Kimmel, Michael Elad, Doron Shaked, Renato Keshet, and Irwin Sobel. A variational framework for retinex. *International Journal of computer vision*, 52(1):7–23, 2003.

[31] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[32] Ranyang Li, Junjun Pan, Yaqing Si, Bin Yan, Yong Hu, and Hong Qin. Specular reflections removal for endoscopic image sequences with adaptive-rpca decomposition. *IEEE transactions on medical imaging*, 39(2):328–340, 2019.

[33] Zhetong Liang, Jun Xu, David Zhang, Zisheng Cao, and Lei Zhang. A hybrid l1-l0 layer decomposition model for tone mapping. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4758–4766, 2018.

[34] Xiongbiao Luo, Hui-Qing Zeng, Ying Wan, Xiao-Bin Zhang, Yan-Ping Du, and Terry M Peters. Endoscopic vision augmentation using multiscale bilateral-weighted retinex for robotic surgery. *IEEE transactions on medical imaging*, 38(12):2863–2874, 2019.

[35] Xiongbiao Luo, Hui-Qing Zeng, Ying Wan, Xiao-Bin Zhang, Yan-Ping Du, and Terry M Peters. Endoscopic vision augmentation using multiscale bilateral-weighted retinex for robotic surgery. *IEEE transactions on medical imaging*, 38(12):2863–2874, 2019.

[36] Lena Maier-Hein, Anja Groch, Adrien Bartoli, Sebastian Bodenstedt, G Boissonnat, P-L Chang, NT Clancy, Daniel S Elson, Sven Haase, Eric Heim, et al. Comparative validation of single-shot optical techniques for laparoscopic 3-d surface reconstruction. *IEEE transactions on medical imaging*, 33(10):1913–1930, 2014.

[37] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a completely blind image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[38] YARON Munz, K Moorthy, A Dosis, JD Hernandez, S Bann, F Bello, S Martin, A Darzi, and T Rockall. The benfits of stereoscopic vision in robotic-assisted performance on bench models. *Surgical Endoscopy And Other Interventional Techniques*, 18(4):611–616, 2004.

[39] Bernd Münzer, Klaus Schoeffmann, and Laszlo Böszörmenyi. Content-based processing and analysis of endoscopic images and videos: A survey. *Multimedia Tools and Applications*, 77(1):1323–1362, 2018.

[40] Bernd Münzer, Klaus Schoeffmann, and Laszlo Böszörmenyi. Content-based processing and analysis of endoscopic images and videos: A survey. *Multimedia Tools and Applications*, 77(1):1323–1362, 2018.

[41] Karen Panetta, Sos Agaian, Yicong Zhou, and Eric J Wharton. Parameterized logarithmic framework for image enhancement. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 41(2):460–473, 2010.

[42] Karen A Panetta, Eric J Wharton, and Sos S Agaian. Human visual system-based image enhancement and logarithmic contrast measure. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 38(1):174–188, 2008.

[43] Seonhee Park, Soohwan Yu, Byeongho Moon, Seungyong Ko, and Joonki Paik. Low-light image enhancement using variational optimization-based retinex model. *IEEE Transactions on Consumer Electronics*, 63(2):178–184, 2017.

[44] Long Quan, Heung-Yeung Shum, et al. Highlight removal by illumination-constrained inpainting. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 164–169. IEEE, 2003.

[45] Zhitao Rao, Tingfa Xu, Jiqiang Luo, Jie Guo, Guokai Shi, and Hongqing Wang. Non-uniform illumination endoscopic imaging enhancement via anti-degraded model and l 1

l 2-based variational retinex. *EURASIP Journal on Wireless Communications and Networking*, 2017(1):1–11, 2017.

[46] Weihong Ren, Jiandong Tian, and Yandong Tang. Specular reflection separation with color-lines constraint. *IEEE Transactions on Image Processing*, 26(5):2327–2337, 2017.

[47] Wenqi Ren, Sifei Liu, Lin Ma, Qianqian Xu, Xiangyu Xu, Xiaochun Cao, Junping Du, and Ming-Hsuan Yang. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing*, 28(9):4364–4375, 2019.

[48] Charles-Auguste Saint-Pierre, Jonathan Boisvert, Guy Grimard, and Farida Cheriet. Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images. *Machine Vision and Applications*, 22(1):171–180, 2011.

[49] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47(1-3):7–42, 2002.

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

[51] Syed MZ Abbas Shah, Stephen Marshall, and Paul Murray. Removal of specular reflections from image sequences using feature correspondences. *Machine Vision and Applications*, 28(3-4):409–420, 2017.

[52] Hui-Liang Shen and Qing-Yuan Cai. Simple and efficient method for specularity removal in an image. *Applied optics*, 48(14):2711–2719, 2009.

[53] Hui-Liang Shen and Zhi-Huan Zheng. Real-time highlight removal using intensity ratio. *Applied optics*, 52(19):4483–4493, 2013.

[54] Jian Sun, Nan-Ning Zheng, and Heung-Yeung Shum. Stereo matching using belief propagation. *IEEE Transactions on pattern analysis and machine intelligence*, 25(7):787–800, 2003.

[55] Jinli Suo, Dongsheng An, Xiangyang Ji, Haoqian Wang, and Qionghai Dai. Fast and high quality highlight removal from a single image. *IEEE Transactions on Image Processing*, 25(11):5441–5454, 2016.

[56] Robby T Tan and Katsushi Ikeuchi. Separating reflection components of textured surfaces using a single image. In *Digitally Archiving Cultural Objects*, pages 353–384. Springer, 2008.

[57] Robby T Tan, Ko Nishino, and Katsushi Ikeuchi. Separating reflection components based on chromaticity and noise analysis. *IEEE transactions on pattern analysis and machine intelligence*, 26(10):1373–1379, 2004.

[58] Michael W Tao, Jong-Chyi Su, Ting-Chun Wang, Jitendra Malik, and Ravi Ramamoorthi. Depth estimation and specular removal for glossy surfaces using point and line consistency with light-field cameras. *IEEE transactions on pattern analysis and machine intelligence*, 38(6):1155–1169, 2015.

[59] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*, pages 839–846. IEEE, 1998.

[60] Fan Wang, Samia Ainouz, Caroline Petitjean, and Abdelaziz Bensrhair. Polarization-based specularity removal method with global energy minimization. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1983–1987. IEEE, 2016.

[61] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013.

[62] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[63] Wenhuan Wu, Hong Zhu, Shunyuan Yu, and Jing Shi. Stereo matching with fusing adaptive support weights. *IEEE Access*, 7:61960–61974, 2019.

[64] Qingqing Yang, Pan Ji, Dongxiao Li, Shaojun Yao, and Ming Zhang. Fast stereo matching using adaptive guided filtering. *Image and Vision Computing*, 32(3):202–211, 2014.

[65] Qingxiong Yang. Hardware-efficient bilateral filtering for stereo matching. *IEEE transactions on pattern analysis and machine intelligence*, 36(5):1026–1032, 2013.

[66] Qingxiong Yang. Stereo matching using tree filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(4):834–846, 2014.

[67] Qingxiong Yang, Jinhui Tang, and Narendra Ahuja. Efficient and robust specular highlight removal. *IEEE transactions on pattern analysis and machine intelligence*, 37(6):1304–1311, 2014.

[68] Qingxiong Yang, Shengnan Wang, Narendra Ahuja, and Ruigang Yang. A uniform framework for estimating illumination chromaticity, correspondence, and specular reflection. *IEEE Transactions on Image Processing*, 20(1):53–63, 2010.

[69] Menglong Ye, Edward Johns, Ankur Handa, Lin Zhang, Philip Pratt, and Guang-Zhong Yang. Self-supervised siamese learning on stereo image pairs for depth estimation in robotic surgery. *arXiv preprint arXiv:1705.08260*, 2017.

[70] Kuk-Jin Yoon and In So Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.

[71] Kang Zhang, Yuqiang Fang, Dongbo Min, Lifeng Sun, Shiqiang Yang, Shuicheng Yan, and Qi Tian. Cross-scale cost aggregation for stereo matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1590–1597, 2014.

[72] Dominique Zosso, Giang Tran, and Stanley J Osher. Non-local retinex—a unifying framework and beyond. *SIAM Journal on Imaging Sciences*, 8(2):787–826, 2015.

Chapter 2

Parametric-based Vision Enhancement

This chapter includes material adapted from:

Xia, W., Chen, E.C., Pautler, S.E., Peters, T.M. "Endoscopic image enhancement with noise suppression." Healthcare technology letters 5.5 (2018): 154-157.

2.1 Introduction

Stereoscopic endoscopes have been used increasingly in minimally invasive surgery to visualize the organ surface and manipulate various surgical tools. However, insufficient and irregular light sources become major challenges for endoscopic surgery. In addition, due to quantization errors, the low-light regions include metrication artifacts. These problems unavoidably deteriorate the clear visualization of both the organ being operated on and its surroundings. They may prevent the surgeon from clearly observing important anatomical structures. Hence there is a need to develop image-processing approaches to eliminate such degradations.

2.1.1 Related work

Many image enhancement methods were presented to deal with different issues for nature images. Most of them focus on contrast enhancement and dynamic range compression but less

on illumination enhancement for endoscopic images. These methods do not address the main concern mentioned above, and thus are not effective on endoscopic images.

Classical approaches including the gamma correction-based method [4] and Retinex theory-based method [8] provide the fundamental of the image enhancement algorithms. In gamma correction, the pixels with low intensity values are mapped to higher intensity values following a non-linear projection operator. This method can effectively improve the visibility of low-light regions but suffer from contrast degradation and visual inconsistency. In Retinex theory, the perceived image is often modelled with the illumination and reflectance components, where the illumination is assumed to be piece-wise linear. Earlier algorithms such as single-scale Retinex [7] and multi-scale Retinex algorithms [3, 12] use a center-surround function to mimic the illumination gain of human visual system. These methods can enhance the lightness and image contrast by post processing techniques such as color restoration and histogram equalization. However, they suffer from over-enhancement and heavy color distortions where the image naturalness is degraded. To avoid over-enhancement and reduce color degradation, Wang et al. [15] proposed a image enhancement method by enhancing contrast and preserving the naturalness of the illumination. Recently, Guo et al. [5] proposed an effective low-light image enhancement algorithm, where an improved illumination map is estimated by imposing a structure prior on the maximum values in red, green, and blue channels.

Although these existing image enhancement methods are able to improve the illumination of dark images, there existing some of major challenge to be overcome. Because image enhancement is a scaling operation by nature, most of current image enhancement methods usually suffer from noise amplification in low-light region. Recently, Su and Jung [13] tried to address this problem by introducing a two-step perceptual enhancement algorithm to suppress camera noise in low-light images, but their algorithm effectiveness on endoscopic images still needs to be improved.

2.1.2   Contributions

This chapter proposes an effective method for endoscopic image enhancement by identifying different illumination regions and designing the enhancement design criteria for desired image quality. The proposed method can enhance the visibility of the low light region of endoscopic images, preserve the naturalness of the image, and reduce noise/artifact amplification. The proposed method thus improves the image quality of endoscopic images, and can be used as a pre-processing step of other image processing algorithms.

2.2   Methods

Consider the Retinex model defined as:

$$I(\mathbf{x}) = f_L(\mathbf{x}) \times f_R(\mathbf{x}) \tag{2.1}$$

where $f_R(\mathbf{x})$ represents the natural reflectance of the true scene at pixel $\mathbf{x}$, $f_L(\mathbf{x})$ the illumination map, and $\times$ denotes the element-wise productor operator. Given the observed endoscopic image $I(\mathbf{x})$, there are two ways. One way is to estimate an approximate illumination $f_L^*(\mathbf{x})$ such that the reflectance $f_R(\mathbf{x})$ can be well recovered by using $I(\mathbf{x})/f_L^*(\mathbf{x})$, where / denotes the element-wise division operator. Alternatively, (3.1) can be rewritten as:

$$f_R^*(\mathbf{x}) = I(\mathbf{x}) \times E(\mathbf{x}), \tag{2.2}$$

where $E(\mathbf{x}) = 1/f_L^*(\mathbf{x})$ represents the enhancement factor. Instead of estimating $f_L(\mathbf{x})$, another way is to estimate the enhancement factor $E(\mathbf{x})$ for an enhanced image $f_R^*(\mathbf{x})$ so that it can achieve a visual quality that is expected for $f_R(\mathbf{x})$ and this chapter focuses on the second approach.

For the endoscopic image enhancement, effective illumination region identification and three threshold criteria for desired image quality are given. (**1**) For image region with good

illumination, it aims to preserve its visual appearance and local contrast. (**2**) For image region with low illumination but intact details, it aims to improve its luminance, as well as enhancing local contrast. (**3**) For regions with extremely low illumination and lossy details, it aims to improve its luminance, but suppress amplification of local changes that are mainly contributed by noise and quantization errors. The workflow of the proposed method is displayed in Fig. 2.1.

2.2.1 Illumination region identification

The proposed scheme first partition the input endoscopic image into three regions: well-lit, low-light, and lossy. Let $I(\mathbf{x}) = (I_r(\mathbf{x}), I_g(\mathbf{x}), I_b(\mathbf{x}))$ and let $V(\mathbf{x})$ be the V-space element through converting $I(\mathbf{x})$ from RGB space to HSV space by HSV transformation. The illumination region sets are defined as:

$$
\begin{cases}
\Omega_{WL}(\mathbf{x}) = \{\mathbf{x}|V(\mathbf{x}) > \tau_1\} \\
\Omega_{Lo}(\mathbf{x}) = \{\mathbf{x}|V(\mathbf{x}) \leq \tau_2, mI(\mathbf{x}) < \epsilon\} \quad , \\
\Omega_{LL}(\mathbf{x}) = \{\mathbf{x}|else\}
\end{cases}
\tag{2.3}
$$

where $\Omega_{WL}(\mathbf{x})$ denotes the well-lit region, $\Omega_{LL}(\mathbf{x})$ denotes the low-light region, $\Omega_{Lo}(\mathbf{x})$ denotes the lossy region, $V(\mathbf{x})$ represents the V-space obtained, and $mI(\mathbf{x})$ represents the well-known dark channel image, defined as: $mI(\mathbf{x}) = \min_{c \in \{r,g,b\}} I_c(\mathbf{x})$. Three threshold parameters τ_1, τ_2, and ϵ satisfy $1 > \tau_1 > \tau_2 > \epsilon > 0$.

2.2.2 Image layer enhancement

After the three illumination regions are identified, the captured image $I(\mathbf{x})$ is decomposed into a base layer and a detail layer, where the base layer $B(\mathbf{x})$ would represent the smooth varying luminance of the image, and the detailed layer $D(\mathbf{x})$ would capture local details and contrasts. The two layers are processed and enhanced differently in order to achieve the expected image quality. Similar techniques have been successfully employed in tone-mapping algorithms such

as [3] for user interactive detail enhancement for artistic purposes. To extract the base and detail layers, the edge preserving smooth filter is used as below:

$$
\begin{cases}
B(\mathbf{x}) = F\{I(\mathbf{x})\} \\
D(\mathbf{x}) = I(\mathbf{x}) - F\{I(\mathbf{x})\}
\end{cases}, \tag{2.4}
$$

where $F\{x\}$ is an edge preserving filter. This chapter employs the tree filter [1] to extract detail and base layers since the tree filter has linear time computational speed and strong edge-preserving smoothing power.

For the base layer, a pseudo enhancement factor $\hat{E}_B(\mathbf{x})$ can be constructed directly to ensure $\hat{E}_B(\mathbf{x}) \times B(\mathbf{x})$, the overall luminance of enhanced image $f_R^*(\mathbf{x})$, will be close to τ_1. The base layer represents the smoothly varying luminance, which is generally free of noise and artifacts. Thus, strong scaling factors are used to enhance its low-light region without considering noise amplification. In this case, the enhancement factor analog to the classic gamma correction formulation is constructed, where pixels with an intensity value less than τ_1 are non-linearly mapped to a higher intensity value close to τ_1. $\hat{E}_B(\mathbf{x})$ is given as the following:

$$
\hat{E}_B(\mathbf{x}) = \begin{cases}
1, \mathbf{x} \in \Omega_{WL}(\mathbf{x}) \\
\frac{\tau_1}{\tau_1^\gamma} V_B(\mathbf{x})^{(\gamma-1)}, \mathbf{x} \in \Omega_{LL}(\mathbf{x}) \bigcup \Omega_{Lo}(\mathbf{x})
\end{cases}, \tag{2.5}
$$

where $V_B(\mathbf{x})$ is the V-space of the base layer $B(\mathbf{x})$ and γ is a positive parameter less than 1, which is also known as the gamma compression coefficient. The lower the gamma, the higher the compression, and as a result, the image appears brighter. In general, this process also compresses the local contrast and reduces local detail. However, since the base layer is smooth and is expected to have a uniform distribution, the gamma compression does not cause any problem.

Similarly, a pseudo enhancement factor $\hat{E}_D(\mathbf{x})$ for the detail layer is designed. Due to the extremely low-signal-to-noise ratio and high degree of quantization error present in $\Omega_{Lo}(\mathbf{x})$,

the local changes observed in this region are mostly contributed by noise and artifacts. As a result, unlike in gamma correction with a high gain for low intensity pixels, a linear gain is used for low intensity pixels in order to suppress noise amplification while maintaining the natural appearance of enhanced image. The construction of $\hat{E}_D(\mathbf{x})$ is given as the following:

$$\hat{E}_D(\mathbf{x}) = \begin{cases} 1, \mathbf{x} \in \Omega_{WL}(\mathbf{x}) \\ \hat{E}_B(\mathbf{x}), \mathbf{x} \in \Omega_{LL}(\mathbf{x}) \\ min(\beta, \hat{E}_B(\mathbf{x})), \mathbf{x} \in \Omega_{Lo}(\mathbf{x}) \end{cases}, \tag{2.6}$$

where β is a constant parameter, representing the linear enhancement factor.

To enforce the piece-wise linear assumption of the illumination $f_L(\mathbf{x})$, we smooth the pseudo enhancement factor $\hat{E}_B(\mathbf{x})$ and $\hat{E}_D(\mathbf{x})$ by using a Gaussian function to yield $E_B(\mathbf{x})$ and $E_D(\mathbf{x})$, respectively:

$$\begin{cases} E_B(\mathbf{x}) = \hat{E}_B(\mathbf{x}) * h_1 \\ E_D(\mathbf{x}) = \hat{E}_D(\mathbf{x}) * h_2 \end{cases}, \tag{2.7}$$

where h_1 and h_2 are Gaussian kernels where h_2 has a smaller size than h_1, and $*$ denotes the discrete linear convolution operation. Finally, each enhancement factor is applied to both the base and detail layers, respectively. The enhanced image is obtained by:

$$f_R^*(\mathbf{x}) = E_B(\mathbf{x}) \times B(\mathbf{x}) + E_D(\mathbf{x}) \times D(\mathbf{x}). \tag{2.8}$$

The overall method is described in Algorithm 1

2.2.3 Algorithm Complexity

To analyze the algorithm complexity, let input image size be m×n. From Algorithm 1, we see that steps 2,4,5,6,7 can be computed analytically by equations (2.3),(2.5),(2.6),(2.7), and (2.8),
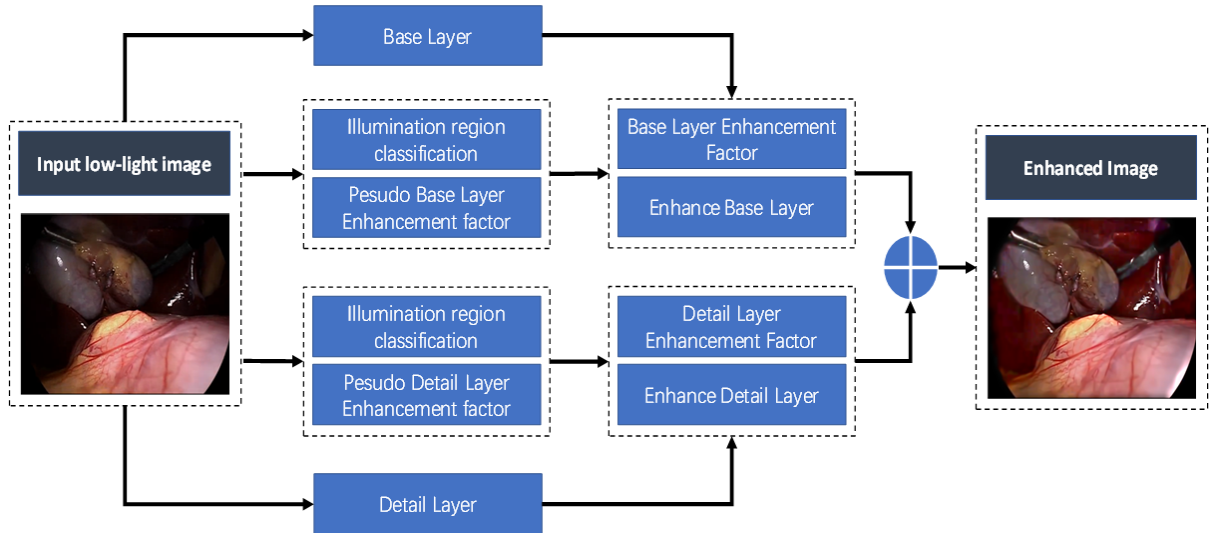
Figure 2.1: Proposed algorithm workflow for endoscopic image enhancement

---

Algorithm 1 for image enhancement

Input:

Observed image $I(\mathbf{x})$, parameters $\tau_1, \tau_2, \epsilon, \gamma$, and β

Step 1: Compute HSV transform on $I(\mathbf{x})$;
Step 2: Compute $\Omega_{WL}(\mathbf{x}), \Omega_{Lo}(\mathbf{x})$, and $\Omega_{LL}(\mathbf{x})$ by (2.3);
Step 3: Compute $B(\mathbf{x})$ and $D(\mathbf{x})$ by (2.4);
Step 4: Compute $\hat{E}_B(\mathbf{x})$ by (2.5);
Step 5: Compute $\hat{E}_D(\mathbf{x})$ by (2.6);
Step 6: Compute $E_B(\mathbf{x})$ and $E_D(\mathbf{x})$ by (2.7);
Step 7: Compute $f_R^*(\mathbf{x})$ by (2.8);

Output: $f_R^*(\mathbf{x})$.

---

respectively. First, Step 1 has the algorithm complexity being $O(mn)$ due to HSV transform. Next, Step 2 can be computed by using a conditional threshold function, which requires $3(mn)$ arithmetic operations. Similarly, Steps 4 and 5 also requires $3(mn)$ operations. Step 6 and 7 require $2(mn)$ operations and $3(mn)$ operations, respectively. Finally, Step 3 has the algorithm complexity being $O(mn)$ since the tree filter has a linear computational time [6]. Thus, the total complexity of Algorithm 1 is $O(mn)$.

2.3 Results

This section will evaluate the effectiveness of the proposed method. Testing video data acquired from robotic-assisted laparoscopic radical prostatectomy is collected by using the da Vinci Si surgical system in the St. Joseph Hospital, London,Canada. Our algorithm is implemented in MATLAB R2018b on a PC with a 2.60GHz Intel Pentium Dual Core Processor.

The proposed algorithm is compared with a group of state of the-art conventional algorithms in term of subjective and objective assessments. To perform the proposed algorithm, the experimental parameters are set as $\tau_1 = 0.4$, $\tau_2 = 0.05$, $\epsilon = 0.01$, $\gamma = 0.2$, and $\beta = 3$, respectively. Among these parameters, $\tau_1 = 0.4$ is used to separate well-lit region and low-light region, and $\tau_2 = 0.05$ and $\epsilon = 0.01$ as two reasonably small parameters to identify lossy region according to (3.3). $\gamma = 0.2$ and $\beta = 3$ are enhancement parameters, where γ is chosen for the non-linear enhancement of the low-light region and β is chosen for the linear enhancement of lossy region. The values of γ and β are determined empirically, as they seem to provide consistent and effective enhancements without visible detail compression and noise amplification.

2.3.1 Qualitative Validation

To illustrate the algorithm performance qualitatively, a visual comparison is performed. Fig.2.2-Fig.2.4 display the enhancement results of three endoscopic images obtained by using eight algorithms: our method, Guo et al.[5], Su et al.[13], Wang et al.[15], Park et al.[11], Wang et al.[14], Liang et al.[9], and Cai et al.[2]. It is observed that our method effectively enhances the visibility of the low-light region and reveal hidden surgical tools and organ surface information, while preventing noise amplification and providing a more natural surgical scene for all images. In comparison, Fig.2.2(c)(d)(f)(i) failed to enhance the dark regions and Fig.2.2(e)(g)(h) have amplified block artifacts and noise. Similarly, Fig.2.3 (c)(d)(f)(i) did not enhance the dark areas successfully and Fig.2.3 have amplified noise in the enhanced image regions. Finally, Fig.2.4(c)(d)(f)(i) have unenhanced dark regions under the surgical tool and Fig.2.4(e)(g)(h)

have low contrast and color cast issues.

2.3.2   Quantitative Validation

To validate the effectiveness of our method, we use two blind evaluation metrics: the naturalness image quality evaluator (NIQE)[10] and the illumination uniformity (IU) which can describe the local consistency of enhanced images. By using the maximum of three color channels of observed image, the local consistency information of enhanced images is computed by

$$\hat{f}_L^*(\mathbf{x}) = mean_{\mathbf{y}\in\Omega(\mathbf{x})} max_{c\in\{r,g,b\}} f_c^*(\mathbf{x}) \tag{2.9}$$

where $f_c^*(\mathbf{x})$ is the color channel of the enhanced image. An IU metric is defined as the standard deviation of $\hat{f}_L^*(\mathbf{x})$. According to the definition of NIQE or IU, lower NIQE or IU value can imply a higher image quality.

Table 2.1 lists quantitative comparison of NIQE results and Table 2.2 lists quantitative comparison of IU results, respectively. Table 2.2 gives the average illumination uniformity index χ of the original input images and the enhanced images obtained by four different algorithms. From this comparison, it is seen that the proposed method and NPEA yield the lowest χ, and thus provide better illumination uniformity for the enhanced images. Table 2.1 gives the mean NIQE and standard deviations for all enhanced images given by four different enhancement algorithms. The proposed method gets a superior NIQE of 2.46 ± 0.12, indicating that our results have consistently higher image quality, compared with the other methods. The higher naturalness index values, yielded by the other methods, indicate increases in noise and blur during enhancement process.

Finally, this section also studies the objective assessment of endoscopic video sequences with various illumination acquired during robotic assisted laparoscopic radical prostatectomy using the da Vinci Si surgical system. The proposed algorithm is compared with related enhancement algorithms: Guo[5], Su[13], Wang[15], Park[11], Wang[14], Liang[9], and Cai[2].

Table 2.1: Quantitative comparison of NIQE

Method	Ours	Guo[5]	Su[13]	Wang[15]	Park[11]	Wang[14]	Liang[9]	Cai[2]
Fig.2.2	2.40	3.19	3.19	2.83	3.35	3.46	2.67	2.65
Fig.2.3	2.44	3.08	3.25	2.80	3.03	3.34	3.09	2.55
Fig.2.4	4.11	4.69	4.10	4.61	4.92	4.97	4.47	4.74

Table 2.2: Quantitative comparison of IUI

Method	Ours	Guo[5]	Su[13]	Wang[15]	Park[11]	Wang[14]	Liang[9]	Cai[2]
Fig.2.2	0.28	0.37	0.41	0.30	0.32	0.36	0.37	0.32
Fig.2.3	0.30	0.39	0.40	0.32	0.34	0.36	0.36	0.34
Fig.2.4	0.20	0.27	0.42	0.20	0.25	0.17	0.15	0.24

For comparative convenience, we denote them as M1(ours), M2[5], M3[13], M4[15], M5[11], M6[14],M7[9], and M8[2], respectively. All the methods were tested on 250 frames from the endoscopic video sequence. Fig. 2.5 plots the IUI results obtained by using the compared methods M1, M2, M3, M4, M5, M6, M7, and M8, respectively. Furthermore, Fig. 2.6 plots the NIQE results obtained by using the compared methods M1, M2, M3, M4, M5, M6, M7, and M8, respectively. From both Fig. 2.5 and Fig. 2.6, it is seen that our algorithm is superior to other algorithms in terms of IUI and NIQE assessment.

2.4 Discussion and conclusion

This chapter proposes a parametric method (non-optimization method) for endoscopic image enhancement. The proposed method first identifies three different illumination regions and then processes the illumination and detail layers separately to meet the enhancement design criteria for desired image quality. The proposed method can enhancing low-light images without amplifying the noise during image enhancement process. Therefore, it has a noise suppression performance, compared current image enhancement methods. According to the experiments with 230 endoscopic images, the proposed algorithm yielded an average of NIQE being 2.46 and an average of IU being 0.22, quantitatively demonstrating superior performance than other
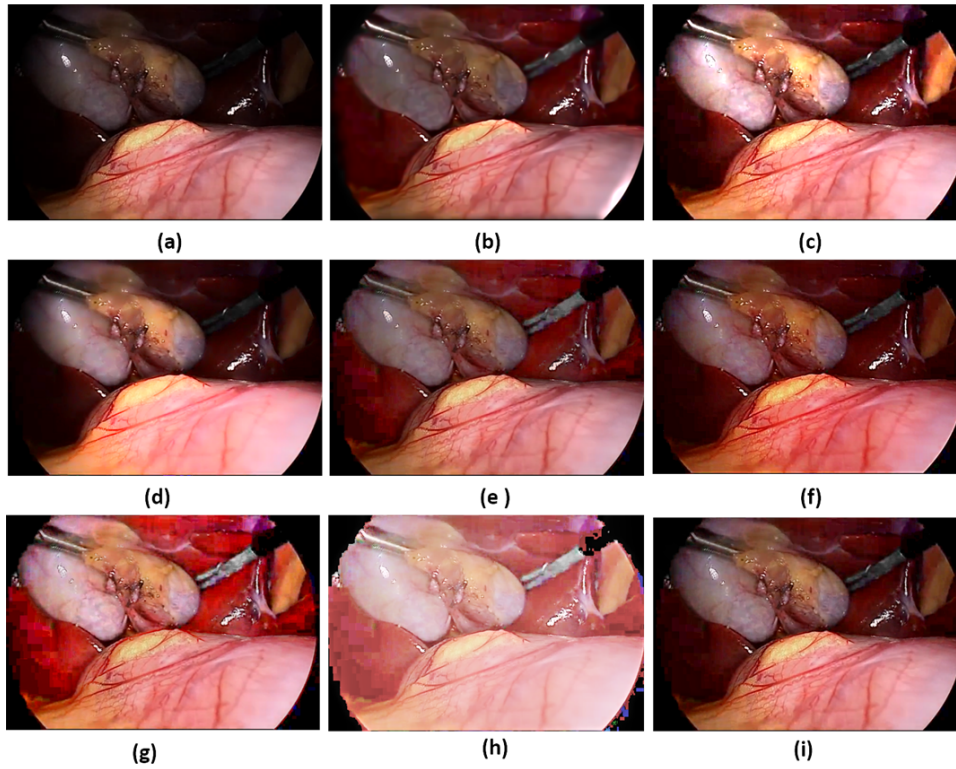
Figure 2.2: Enhanced results for surgical laparoscopic image (a) Unenhanced image, (b) Ours, (c) Ref.[5], (d) Ref.[13], (e) Ref.[15], (f) Ref.[11],(g)Ref.[14], (h)Ref.[9], (i) Ref.[2]
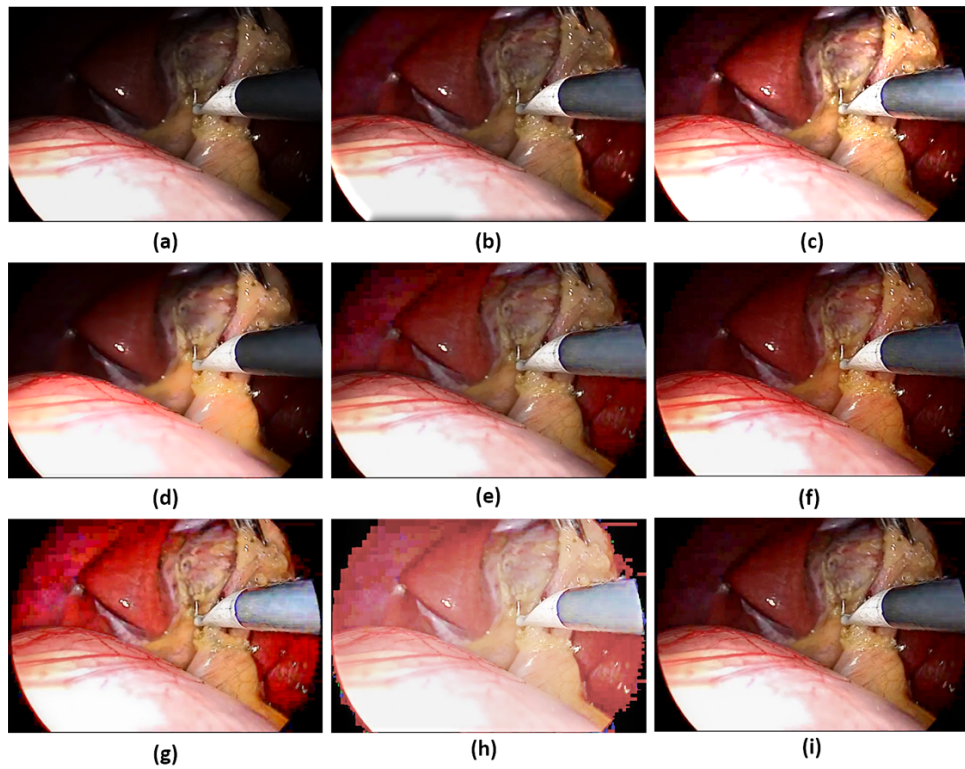
Figure 2.3: Enhanced results for surgical laparoscopic image (a) Unenhanced image, (b) Ours, (c) Ref.[5], (d) Ref.[13], (e) Ref.[15], (f) Ref.[11],(g)Ref.[14], (h)Ref.[9], (i) Ref.[2]
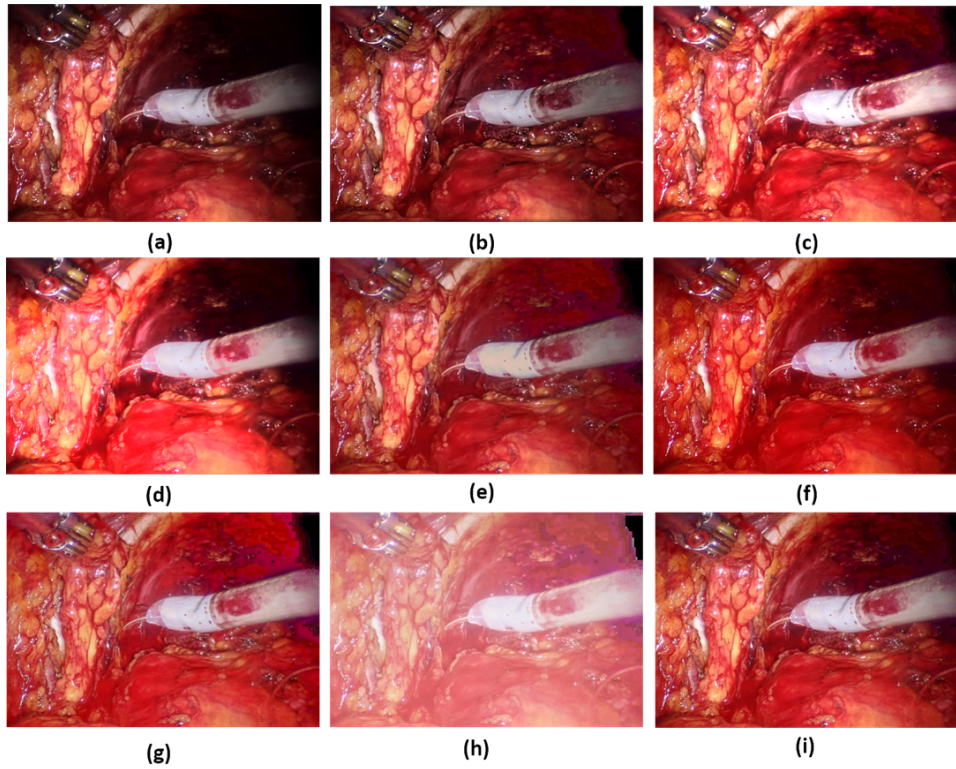
Figure 2.4: Enhanced results for surgical laparoscopic image (a) Unenhanced image, (b) Ours, (c) Ref.[5], (d) Ref.[13], (e) Ref.[15], (f) Ref.[11],(g)Ref.[14], (h)Ref.[9], (i) Ref.[2]
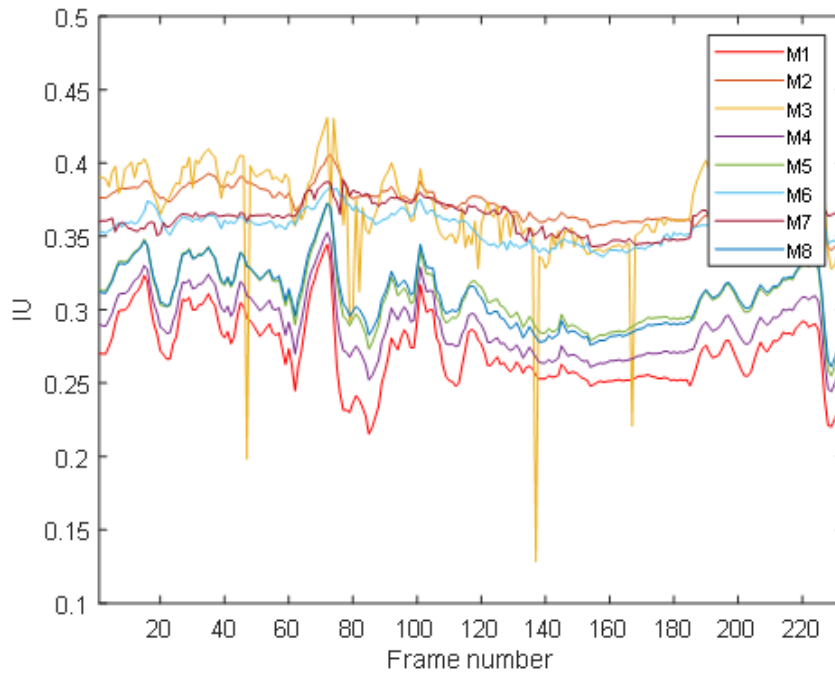
Figure 2.5: IU results of the processed endoscopic video images of using methods: M1(ours), M2[5], M3[13], M4[15], M5[11], M6[14],M7[9], M8[2].

state-of-the-art algorithms. By visual inspection, the proposed method is able to maintain the contrasts and colors in the well-lit image regions, while significantly improving the visibility of the low-light regions. All the results confirm that the proposed method has superior performance without amplifying underlying noise/artifacts in the low-light regions, while other existent algorithms amplified visible noise/artifacts.

The proposed algorithm has a potential limitation on optimizing design parameters. It is need for providing optimized design parameters for surgical scenarios. In addition, the proposed algorithm may cause stiching artifacts near the image boarder, due to the combining of enhancement factors from different illumination regions. On average, when performing a 480×854 endoscopic image with resolution, the computational time of our algorithm needs 1.02s. The computational time will be greatly reduced through parallel implementation in C++ using a GPU and enabling it to run in real time. In addition, the experimental parameters should be reduced and optimized as soon as possible. These issues will be the future research.
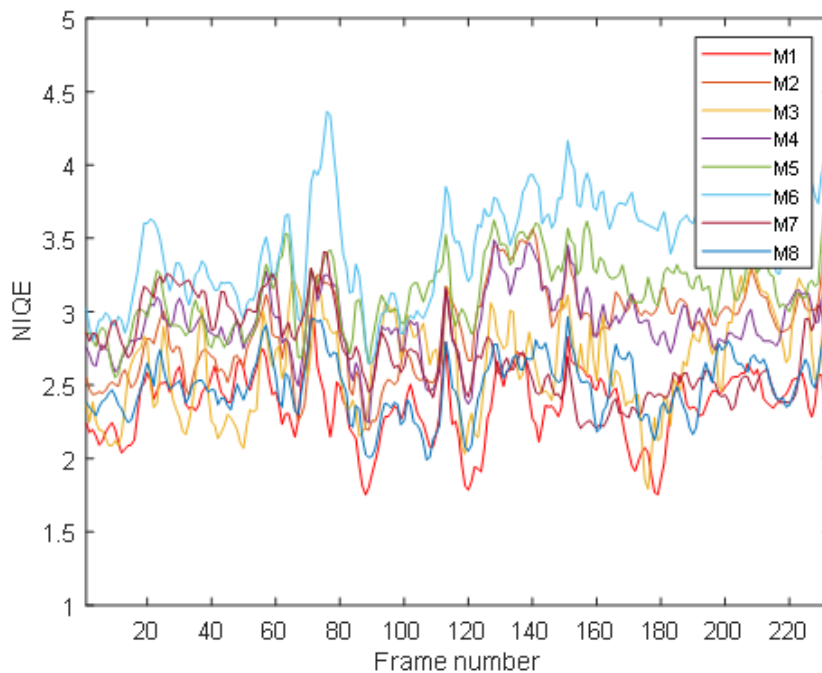
Figure 2.6:   NIQE results of the processed endoscopic video images of using methods: M1(ours), M2[5], M3[13], M4[15], M5[11], M6[14],M7[9], M8[2].

Bibliography

[1] Linchao Bao, Yibing Song, Qingxiong Yang, Hao Yuan, and Gang Wang. Tree filtering: Efficient structure-preserving smoothing with a minimum spanning tree. *IEEE Transactions on Image Processing*, 23(2):555–569, 2013.

[2] Bolun Cai, Xianming Xu, Kailing Guo, Kui Jia, Bin Hu, and Dacheng Tao. A joint intrinsic-extrinsic prior model for retinex. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4000–4009, 2017.

[3] Gabriel Eilertsen, Rafał K Mantiuk, and Jonas Unger. Real-time noise-aware tone mapping. *ACM Transactions on Graphics (TOG)*, 34(6):1–15, 2015.

[4] Hany Farid. Blind inverse gamma correction. *IEEE transactions on image processing*, 10(10):1428–1433, 2001.

[5] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016.

[6] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE transactions on image processing*, 22(3):1032–1041, 2012.

[7] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997.

[8] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[9] Zhetong Liang, Jun Xu, David Zhang, Zisheng Cao, and Lei Zhang. A hybrid l1-l0 layer decomposition model for tone mapping. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4758–4766, 2018.

[10] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a completely blind image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012.

[11] Seonhee Park, Soohwan Yu, Byeongho Moon, Seungyong Ko, and Joonki Paik. Low-light image enhancement using variational optimization-based retinex model. *IEEE Transactions on Consumer Electronics*, 63(2):178–184, 2017.

[12] Zia-ur Rahman, Daniel J Jobson, and Glenn A Woodell. Multi-scale retinex for color image enhancement. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 3, pages 1003–1006. IEEE, 1996.

[13] Haonan Su and Cheolkon Jung. Perceptual enhancement of low light images based on two-step noise suppression. *IEEE Access*, 6:7005–7018, 2018.

[14] Shuhang Wang and Gang Luo. Naturalness preserved image enhancement using a priori

multi-layer lightness statistics. *IEEE transactions on image processing*, 27(2):938–948, 2017.

[15] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013.

Chapter 3

Optimization-based Vision Enhancement

This chapter includes material adapted from:

Xia, W., Chen, E.C., Pautler, S.E., Peters, T.M. "Simplified Retinex Image Enhancement Using Prior Refined Information Fusion." IEEE Transactions on Cybernetics (Submitted)

3.1   Introduction

Robotic laparoscopy uses stereoscopic laparoscope to intuitively visualize the organ surface and directly guide or manipulate various surgical instruments inside the body. Because of the hostile imaging environment inside the human body, captured laparoscopic images often suffer from low visibility and non-uniform illumination caused by a concentrated light source located at the distal end of the laparoscope. This in turn may cause incorrect targeting, surgical risk, and extended operating time during laparoscopic surgery. Although Retinex optimization methods were developed to deal with non-uniform illumination nature images in decades, surgical video image enhancement still needs improving due to issues including colour cast, naturalness, blurred texture, and extensive computational time.

3.1.1   Related work

Similar to the discussion of previous chapter, a low-light image can be approximately modeled as the product of its reflectance and illumination components:

$$I(\mathbf{x}) = f_L(\mathbf{x}) \times f_R(\mathbf{x}) \tag{3.1}$$

where $I(\mathbf{x})$ is the low-light image (input image), $f_R(\mathbf{x})$ represents the reflectance component, and $f_L(\mathbf{x})$ the illuminance component. According to [11], these components should subject to the constraints $f_L(\mathbf{x}) \geq I(\mathbf{x})$ and $0 < f_R(\mathbf{x}) \leq 1$. The goal is to simultaneously estimate the illuminance and reflectance from the observed image.

The Retinex optimization-based enhancement method has a potential advantage for enhancing low-light regions of non-uniform illumination images while preserving the image contrast. To simultaneously estimate the reflectance and illumination components, Fu et al. [8] proposed solving the following Retinex optimization problem:

$$\min_{f_L, f_R} \|I(\mathbf{x}) - f_L(\mathbf{x}) \times f_R(\mathbf{x})\|_F^2 + \lambda_1 \|\nabla f_L(\mathbf{x})\|_F^2 + \lambda_2 \|\nabla f_R(\mathbf{x})\|_1 \tag{3.2}$$
$$+ \lambda_3 \|f_L(\mathbf{x}) - f_L^0(\mathbf{x})\|_F^2$$

where $\| \cdot \|_F$ and $\| \cdot \|_1$ denote the Frobenius norm and l_1 norm. ∇f_L is the first order differential operator of f_L and ∇f_R is the first order differential operator of f_R. The first term is the data fidelity term. The second and third terms are the total variational (TV) terms on the illuminance and reflectance terms, respectively. The fourth term is the refined version of the illuminance component. $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0$ are regularization parameters given in advance.

Following the work of (3.2), Rao et al. [16] proposed solving the following generalized

Retinex optimization problem:

$$\min_{f_L, f_R} \|I(\mathbf{x}) - f_L(\mathbf{x}) \times f_R(\mathbf{x})\|_F^2 + \lambda_1 \|\hat{I}(\mathbf{x}) - f_L(\mathbf{x}) \times f_R(\mathbf{x})\|_1$$
$$+ \lambda_2 \|\nabla f_L(\mathbf{x})\|_1 + \lambda_3 \|\nabla f_R(\mathbf{x})\|_1 + \lambda_4 \|f_L(\mathbf{x}) - f_L^0(\mathbf{x})\|_F^2 \tag{3.3}$$

where the additional second term is used to reduce the impact of non-Gaussian noise.

To strengthen structural information related to reflectance, Li et al. [12] considered an improved gradient map of the observed image to control the gradient of the reflectance by solving the following Retinex optimization problem:

$$\min_{f_L, f_R} \|I(\mathbf{x}) - f_L(\mathbf{x}) \times f_R(\mathbf{x})\|_F^2 + \lambda_1 \|\nabla f_L(\mathbf{x})\|_1 + \lambda_2 \|\nabla f_R(\mathbf{x}) - G\|_F^2 \tag{3.4}$$

where $\lambda_1 > 0, \lambda_2 > 0$ are the regularization parameters, and G is a guidance matrix.

To reduce computation time, Park et al. [15] presented a Retinex-based optimization method employing the l_2 regularization norm. Guo et al. [9] proposed solving the following weighted l_1-*norm* optimization problem:

$$\min_{f_L} \quad \|f_L(\mathbf{x}) - f_L^0(\mathbf{x})\|_F^2 + \alpha \|W \nabla f_L(\mathbf{x})\|_1 \tag{3.5}$$

where $\|W \nabla f_L(\mathbf{x})\|_1$ is used to keep the smoothness of illumination and W is a weighting matrix function based on the structural similarity between the input and output images.

For the structural similarity image, Aujol et al. [2] first introduced a total variation method for image structure:

$$S^*(x) = \arg \min_S \frac{1}{2\rho} \|I(\mathbf{x}) - S(\mathbf{x})\|_F^2 + \|\nabla S(\mathbf{x})\|_1 \tag{3.6}$$

where the first term is to extract similar structures from the input image $I(\mathbf{x})$, the second is the TV regularizer, and $\rho > 0$ is a regularization parameter. Farbman et al. [7] presented a weighted total variation method to describe the image structure:

$$S^*(x) = \arg \min_S \frac{1}{2\rho} \|I(\mathbf{x}) - S(\mathbf{x})\|_F^2 + \|W \nabla S(\mathbf{x})\|_1 \tag{3.7}$$

To further enhance the contrast between texture and structure, especially for visually salient regions, Xu et al. [22] presented an adaptively weighted total variation method for the structure image:

$$S^*(x) = \arg \min_S \frac{1}{2\rho} \|I(\mathbf{x}) - S(\mathbf{x})\|_F^2 + \|W(I)\nabla S(\mathbf{x})\|_1 \tag{3.8}$$

where each element of $W(I)$ is an adaptive weight defined by the Gaussian kernel:

$$exp(-\frac{(x_p - x_q)^2 + (y_p - y_q)^2}{2\sigma^2})$$

where $\mathbf{x}_p = (x_p, y_p)$ denotes the pixel location and σ is used to control the spatial scale of the window.

Most of the conventional Retinex optimization methods require a postprocessing step to reduce the over-enhancement of the reflectance. To deal with this problem, prior enhanced illumination information is incorporated into Retinex optimization models. Fu et al. [8] used Gaussian filtering of the observed image as an illumination prior and Park et al. [15] employed the observed image with gamma correction. Rao et al. [16] used the average of the observed image as illumination prior and Du and Xia[6] used a structure similarity image. Guo et al. [9] used the maximum of three colour channels of the observed image as the illumination prior.

Retinex-based optimization methods minimize a non-convex objective function and thus the alternating direction method of multipliers(ADMM) is used. Since their objective functions include TV or l_1-norm regularization terms, the convergence condition of ADMM is in general not satisfied. Moreover, current Retinex-based optimization methods do not consider prior enhanced reflectance information and thus fail to limit the range of the reflectance. Therefore, the naturalness of non-uniform illumination images cannot be effectively preserved [9].

3.1.2   Contributions

This chapter proposes a simplified Retinex optimization method with prior information fusion. The proposed method can significantly enhance dark region while preserving naturalness, texture details, and image structures, since our approach can integrate both the prior enhanced illumination and reflectance into the optimization process. Second, the proposed algorithm has low computational complexity with a fast computation speed, due to the replacement of the total variation regularization term with two l_2-norm regularization terms. Moreover, the convergence condition of the proposed algorithm is satisfied. Finally, the experimental results of a forced-choice study further confirm that the proposed method is preferred by clinicians, demonstrating that this work is of practical relevance when dealing with illumination variations in robotic-assisted laparoscopic surgery, digitally augmenting the laparoscopic video stream to provide the surgeon with a wider and more immersive field of view.

3.2   Methods

For the convergence condition of ADMM and effective estimation of the reflectance element, the TV regularization term is replaced with two prior information regularization terms. A simplified Retinex optimization method is introduced to solve the following optimization problem:

$$\min_{f_L, f_R} \|I(\mathbf{x}) - f_L(\mathbf{x}) \times f_R(\mathbf{x})\|_F^2 + \lambda_1 \|f_R(\mathbf{x}) - \hat{f}_R(\mathbf{x})\|_F^2$$
$$+ \lambda_2 \|f_L(\mathbf{x}) - \hat{f}_L(\mathbf{x})\|_F^2 \tag{3.9}$$

where the first term relates to data fidelity, $\|f_R(\mathbf{x}) - \hat{f}_R(\mathbf{x})\|_F^2$ is the prior enhanced reflectance term, represented by the difference between the reflectance estimation and the known prior of the reflectance $\hat{f}_R(\mathbf{x})$. $\|f_L(\mathbf{x}) - \hat{f}_L(\mathbf{x})\|_F^2$ is the prior enhanced illuminance term, represented by the difference between the illuminance estimation and the known prior of the refined illuminance $\hat{f}_L(\mathbf{x})$.

To estimate illuminance and reflectance components, (3.9) is first split into two sub-problems:

$$\min_{f_R} \ \|I(\mathbf{x}) - f_L(\mathbf{x}) \times f_R(\mathbf{x})\|_F^2 + \lambda_1 \|f_R(\mathbf{x}) - \hat{f}_R(\mathbf{x})\|_F^2 \tag{3.10}$$

and

$$\min_{f_L} \ \|I(\mathbf{x}) - f_L(\mathbf{x}) \times f_R(\mathbf{x})\|_F^2 + \lambda_2 \|f_L(\mathbf{x}) - \hat{f}_L(\mathbf{x})\|_F^2 \tag{3.11}$$

The above two subproblems can be further rewritten as:

$$\min_{f_R} \ \|I(\mathbf{x})/(f_L(\mathbf{x}) + \delta) - f_R(\mathbf{x})\|_F^2 + \lambda_1 \|f_R(\mathbf{x}) - \hat{f}_R(\mathbf{x})\|_F^2 \tag{3.12}$$

and

$$\min_{f_L} \ \|I(\mathbf{x})/(f_R(\mathbf{x}) + \delta) - f_L(\mathbf{x})\|_F^2 + \lambda_2 \|f_L(\mathbf{x}) - \hat{f}_L(\mathbf{x})\|_F^2 \tag{3.13}$$

where $\delta > 0$ is a small constant number.

Because the objective function of (3.12) is convex, the optimality condition of (3.12) is given by:

$$(1 + \lambda_1) f_R(\mathbf{x}) - I(\mathbf{x})/(f_L(\mathbf{x}) + \delta) - \lambda_1 \hat{f}_R(\mathbf{x}) = 0$$

Similarly, the optimality condition of (3.13) is given by :

$$(1 + \lambda_2) f_L(\mathbf{x}) - I(\mathbf{x})/(f_R(\mathbf{x}) + \delta) - \lambda_2 \hat{f}_L(\mathbf{x}) = 0$$

Therefore, the optimal solution of (3.12) is given by

$$f_R^*(\mathbf{x}) = \frac{1}{(1 + \lambda_1)} \{I(\mathbf{x})/(f_L(\mathbf{x}) + \delta) + \lambda_1 \hat{f}_R(\mathbf{x})\}$$

and the optimal solution of (3.13) is given by

$$f_L^*(\mathbf{x}) = \frac{1}{(1 + \lambda_2)} \{I(\mathbf{x})/(f_R(\mathbf{x}) + \delta) + \lambda_2 \hat{f}_L(\mathbf{x})\}$$

As a follow-up to the iterative ADMM method, we introduce the following alternating iteration scheme:

$$f_R^{(k+1)}(\mathbf{x}) = \frac{1}{(1 + \lambda_1)}\{I(\mathbf{x})/(f_L^{(k)}(\mathbf{x}) + \delta) + \lambda_1 \hat{f}_R(\mathbf{x})\} \qquad (3.14)$$

and

$$f_L^{(k+1)}(\mathbf{x}) = \frac{1}{(1 + \lambda_2)}\{I(\mathbf{x})/(f_R^{(k)}(\mathbf{x}) + \delta) + \lambda_2 \hat{f}_L(\mathbf{x})\} \qquad (3.15)$$

To illustrate the convergence of the proposed alternating iteration method, the objective function of (3.9) is defined as $\Phi(f_R(\mathbf{x}), f_L(\mathbf{x}))$. Because $\Phi(f_R(\mathbf{x}), f_L(\mathbf{x}))$ is a convex function for any given $f_R(\mathbf{x})$ or $f_L(\mathbf{x})$, the sequence $\{(f_R^{(k)}(\mathbf{x}), f_L^{(k)}(\mathbf{x}))\}$ produced by both (3.14) and (3.15) satisfies:

$$f_R^{(k+1)}(\mathbf{x}) \approx \arg \min_{f_R} \Phi(f_R(\mathbf{x}), f_L^k(\mathbf{x}))$$

and

$$f_L^{(k+1)}(\mathbf{x}) \approx \arg \min_{f_L} \Phi(f_R^k(\mathbf{x}), f_L(\mathbf{x}))$$

provided δ is sufficiently small. From the result given in [3], it is seen that the sequence $\{(f_R^{(k)}(\mathbf{x}), f_L^{(k)}(\mathbf{x}))\}$ can approach an optimal solution of (3.9).

3.2.1   Prior refined information

To preserve the naturalness, textural details, and structure of the image, prior refined illumination and reflectance information can play an important role in optimally estimating the reflectance and illumination components. Therefore, the prior refined information should characterize the textural and edge structure of the input image. According to the structure similarity approach [22], we use the prior refined illumination map:

$$\hat{f}_L(\mathbf{x}) = max\{S^*(\mathbf{x}), I(\mathbf{x})\} \qquad (3.16)$$

where $S^*(\mathbf{x})$ is a structure image of the input image $I(\mathbf{x})$, described in (3.8).

Based on the structure component information, the prior refined reflectance map is taken

as:

$$\hat{f}_R(\mathbf{x}) = \frac{I(\mathbf{x})}{\hat{f}_L^*(\mathbf{x})} \tag{3.17}$$

where $\hat{f}_L^*(\mathbf{x}) = max\{S^*(\mathbf{x}) + \delta, I(\mathbf{x})\}$ and the parameter $\delta > 0$ is defined in (3.13).

Based on both (3.16) and (3.17), it is easy to see that the refined illumination and reflectance maps can contain the structure component of the input image. Moreover, $\hat{f}_L(\mathbf{x}) \geq I(\mathbf{x})$ and $\hat{f}_R(\mathbf{x}) \leq 1$. As a result, the naturalness of non-uniformly illuminated images may be effectively preserved [19].

3.2.2 Algorithm implementation and complexity

The proposed image enhancement method consists of two main steps: an initial prior refinement step and an iterative updating step, as described in Algorithm 2, where $\gamma = 2.2$ and the initial illuminance element is taken as $\hat{f}_L(\mathbf{x})$, and

$$E_{error} = \frac{\|f_R^{(k+1)} - f_R^{(k)}\|_2}{\|f_R^{(k)}\|_2} + \frac{\|f_L^{(k+1)} - f_L^{(k)}\|_2}{\|f_L^{(k)}\|_2},$$

where $\| \cdot \|_2$ denotes l_2 norm. The initial refined step includes the refined illuminance and reflectance maps. Fig. 3.1 further displays the flowchart of Algorithm 2.

To analyze algorithm's complexity, let the size of the input image be $m \times n$ and the maximum iteration step is K. Then the refined illuminance map can be obtained by solving the linear system with a sparse Laplacian matrix. According to the analysis presented in [7], the complexity of the algorithm is of $O(\sigma mn)$, and the iterative updating step has a complexity of $6Kmn$. So, the total complexity of the proposed method is $6Kmn + O(\sigma mn)$. Since the window size, σ, is within $[1, 8]$ in general, the total complexity of the proposed method can be approximated as $O(6(K + 2)mn)$. Furthermore, because the proposed algorithm is based on l_2-norm minimization without the TV term, the maximum iteration step usually takes less than $6 - 8$ steps before reaching the optimal result. Therefore, the total complexity of the proposed algorithm is $O(mn)$.

Algorithm 2 for image enhancement

Input:

Observed image $\mathbf{I}$, parameters $\lambda_1, \lambda_2, \delta$, and maximum iteration step K

Initial refined step:

Compute prior refined illuminance map $\hat{f}_L$ by (3.16)

Compute prior refined reflectance map $\hat{f}_R$ by (3.17)

Iterative updating step:

Do for $k = 1, ...$

Step 1: update $f_R^{(k)}$ by (3.14);

Step 2: update $f_L^{(k)}$ by (3.15);

Step 3: if $E_{error} < \epsilon$ or $k \geq K$, stop iteration. Otherwise, $k := k + 1$ and go to step 1

Output: $f_{enhanced} = f_R^{(K)} \circ (f_L^{(K)})^{\frac{1}{\gamma}}$.

3.2.3 Colour balance

To prevent the colour cast phenomenon, two enhanced images in RGB and HSV spaces are fused. Let $I(\mathbf{x}) = (I_r(\mathbf{x}), I_g(\mathbf{x}), I_b(\mathbf{x}))$ and let $V(\mathbf{x})$ be the V-space element, obtained by converting $I(\mathbf{x})$ from RGB space to HSV space by HSV transformation. The final enhanced image is given by:

$$f^*_{enhanced}(\mathbf{x}) = \theta f^1_{enhanced}(\mathbf{x}) + (1 - \theta)f^2_{enhanced}(\mathbf{x}) \tag{3.18}$$

where $f^1_{enhanced}(\mathbf{x})$ is the enhanced image by performing Algorithm 2 on the RGB-space input image, $f^2_{enhanced}(\mathbf{x})$ is the enhanced image by performing Algorithm 2 on the V-space input image, and $\theta \in [0, 1]$ is a fusion coefficient.

3.3 Results

Testing video data acquired from robotic-assisted laparoscopic radical prostatectomy was collected by using the da Vinci Si surgical system in the St. Joseph Hospital, London,Canada. The proposed algorithm is implemented in MATLAB R2018b on a PC with a 2.60GHz Intel Pentium Dual Core Processor.

The proposed algorithm is compared with current state-of-the-art enhancement algorithms in terms of subjective and objective assessments and the implementations of the existing algo-
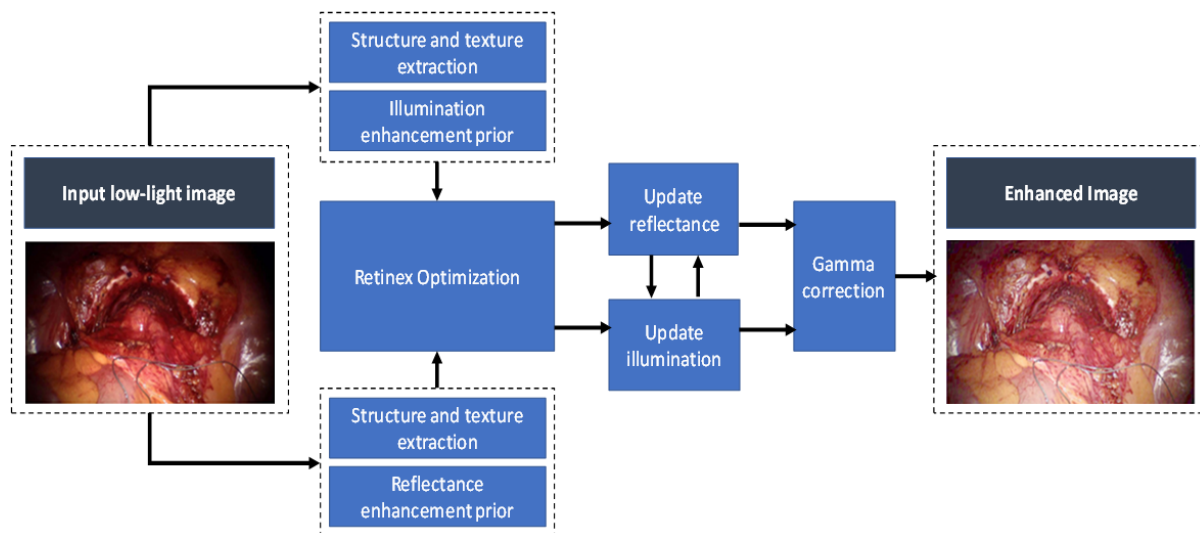
Figure 3.1: Scheme of proposed enhancement method

rithms are available in the authors' websites. These methods include Retinex-based optimization approaches presented by Li et al. [12], Guo et al. [9], Rao et al. [16], Park et al. [15], Cai et al. [4], Fu et al. [8], Zosso et al. [23], and other available approaches: the multiscale Retinex method by Luo et al. [14], the multi-layer decomposition method by Wang and Luo [18], the naturalness preserved enhancement method by Wang et al. [19], the tone mapping method by Liang et al. [13], the two-step perceptual method by Su and Jung [17], and the gamma correction-based method by Huang et al. [10]. The experimental parameters in the proposed algorithm are set in the following discussion. Although these parameters are not necessarily optimal, they are representative of what constitute as sufficient and effective parameters for most surgical images.

3.3.1 Subjective assessment

First, three input laparoscopic images appeared in [14] are studied, and are reproduced in Fig.3.2(a)-Fig.3.4(a). The experimental parameters of the proposed algorithm are set as $\lambda_1 = \lambda_2 = 1$, $\delta = 0.0005$, and $K = 8$. For the input image 3.2(a), Fig.3.2(b)-Fig.3.2(l) displays

a visual comparison of enhanced laparoscopic images by comparing our algorithm with ten existing enhancement algorithms: [12], [9], [18], [15], [8], [23], [19], [14], [13], and [10]. From Fig.3.2 it is seen that our method can provide better visual quality than other methods in terms of image naturalness, organ colour, and textural details. In contrast, Fig.3.2(b)(c)(e) exhibit poor enhancement in lighting and detail. Fig.3.2(d)(f)(g)(h)(i)(j) have varying degrees of image colour distortion and Fig.2.3(k) has blurred texture and structures. For the input image 3.3(a), Fig.3.3(b)-(l) display a visual comparison of enhanced laparoscopic images obtained by our algorithm and other enhancement algorithms. Similarly, 3.3(b)(e)(c) do not have sufficient enhancement in lighting and details. 3.3(d)(f)(g)(h)(i)(j) have different degrees of colour cast and 3.3(k) has halo artifact and blurred textures. For the input image 3.4(a), Fig.3.4(b)-(l) also display the enhanced images from different methods. For the same reason as above, application of the proposed algorithm results in better enhanced images with good detail preservation and realistic appearance under the white light.

Next, three input laparoscopic images from the 2019 MICCAI Challenge [1] are studied, shown in Fig. 3.5(a), Fig.3.6(a), and Fig.3.7(a), respectively. The experimental parameters of the proposed algorithm are set as $\lambda_1 = 1, \lambda_2 = 8, \delta = 0.1$, and $K = 8$. For the input image 3.5(a), Fig.3.5(b)-Fig.3.5(l) display a visual comparison of the laparoscopic images enhanced by our algorithm and other enhancement algorithms: [8], [9], [15], [12], [16], [13], [18], [4], [17], and [10]. From Fig. 3.5 is seen that Figs. 3.5(e)-(k) all have colour cast due to changing the overall colour of the image. Fig.3.5(b)(c) have unenhanced liver regions and the their contrasts are too high causing unrealistic appearances. Fig.3.5(d) and ours outperform the others while our enhanced image has slightly higher contrast. For the input image 3.6(a), a visual comparison of the enhanced laparoscopic images are shown in Fig.3.6(b)-(l). From Fig.3.6 it is seen that Fig.3.6(b)(c) have sub-optimal illumination enhancement and Fig.3.6(g)(h)(i)(k) exhibit serious colour cast. While Figs.3.6(d)(f)(j) and ours have similar overall appearances, ours has better enhancement, revealing more details in the dark region. Similar to the figures above, Fig.3.7(b)-(l) display a visual comparison of enhanced laparoscopic images by different

methods. Once again, it is easy to see that Fig.3.7(b)(c) have unenhanced image regions and Fig.3.7(e)(g)(h)(i)(k) have serious colour cast, altering the organ colours. 3.7(d)(f)(j) and ours outperform the others, while ours has better illumination enhancement and details.

Figure 3.2: Enhanced results for surgical laparoscopic image (a) Input, (b) Ref. [10], (c) Ref. [19], (d) Ref. [9], (e) Ref. [8], (f) Ref. [15], (g) Ref. [12], (h) Ref. [23], (i) Ref. [13],(j)Ref. [18], (k)Ref. [14], (l) Ours

3.3.2 Objective assessment

In addition to the subjective visual perception evaluation, it is also necessary to use index measures to test the performance of the proposed enhancement method. Since no ground truth for surgical laparoscopic images is available, the evaluation of the enhanced images are generally through blind image qualitative assessments methods [5]-[20]. In [14], four medical image

metrics are suggested for evaluating enhanced surgical images. These include 1) the sharpness metric, characterizing how much structural information of contours and boundaries on images, 2) the naturalness metric describing how natural laparoscopic vision appears, 3) the contrast metric η depicting the difference in luminance of regions of interest, and 4) the hybrid metric representing an overall balance of the previous metrics. This section evaluates the quality of enhanced images by using these four blind metrics without any reference images. The higher average metric values indicates the higher image quality of the enhanced laparoscopic image.

Figure 3.3: Enhanced results for surgical laparoscopic image (a) Input, (b) Ref. [10], (c) Ref. [19], (d) Ref. [9], (e) Ref. [8], (f) Ref. [15], (g) Ref. [12], (h) Ref. [23], (i) Ref. [13],(j)Ref. [18], (k)Ref. [14], (l) Ours

For the convenience of discussion , Tables I and II summarize the average results of the four metric values. Table I lists the results of three input laparoscopic images corresponding to Fig. 3.2, Fig. 3.3, and Fig. 3.4. Table II lists results of three input laparoscopic images

corresponding to Fig. 3.5, Fig. 3.6, and Fig. 3.7. From Table I, it is found that for both Fig. 3.2 and Fig. 3.3, the proposed method has a higher average of all four medical image metrics than other algorithms. For Fig. 3.5, the proposed method also has a higher average of four medical image metrics than other algorithms, except for Guo's method [25] having a slightly higher average than ours. Furthermore, the proposed method has a higher total average of three images than the other methods, including Guo's method, implying that the proposed method has a better overall objective performance than the other methods in terms of blind image metrics. Similarly, from Table II it is seen that for Figs. 3.5, 3.6, and 3.7, the proposed method has a higher average of four medical image metrics than other algorithms, which demonstrates the advantage of our algorithm in terms of objective performance. In summary, the proposed algorithm has the potential to achieve a good quality in the objective assessment.

Furthermore, this section also studies the objective assessment of laparoscopic video sequences with various illumination acquired during robotic-assisted laparoscopic radical prostatectomy using the da Vinci Si surgical system. The proposed algorithm is compared with five Retinex optimization algorithms: Fu et al.(2015,[8]), Guo et al.(2017,[9]), Park et al.(2017,[15]), Li et al.(2018,[12]), Rao et al.(2017,[16]). For comparative convenience, let them be denoted by M_1, M_2, M_3, M_4, M_5, and M_6 (ours), respectively. All these methods were tested on 1000 frames from the laparoscopic video sequence. Fig. 3.8 plots the average results of four medical image metrics of enhanced laparoscopic video images by using the compared methods M_1čiter19, M_2 [9], M_3 [15], M_4 [12], M_5 [16], and M_6(ours), demonstrating that our assessment score is higher than the others. Moreover, Fig. 3.9 also shows that our algorithm has relatively fast computational speed.

3.3.3 Forced-choice preference testing

For further quantitative validation, a forced-choice user preference experiment is studied by using 25 low light video images collected from robotic-assisted laparoscopic radical prostatectomy.

Figure 3.4: Enhanced results for surgical laparoscopic image (a) Input, (b) Ref. [10], (c) Ref. [19], (d) Ref. [9], (e) Ref. [8], (f) Ref. [15], (g) Ref. [12], (h) Ref. [23], (i) Ref. [13],(j)Ref. [18], (k)Ref. [14], (l) Ours

Figure 3.5: Enhanced results for surgical laparoscopic image (a) Input, (b) Ref. [10], (c) Ref. [17], (d) Ref. [8], (e) Ref. [15], (f) Ref. [18]), (g) Ref. [12], (h) Ref. [9], (i) Ref. [13],(j) Ref. [4], (k)Ref. [16], (l) Ours
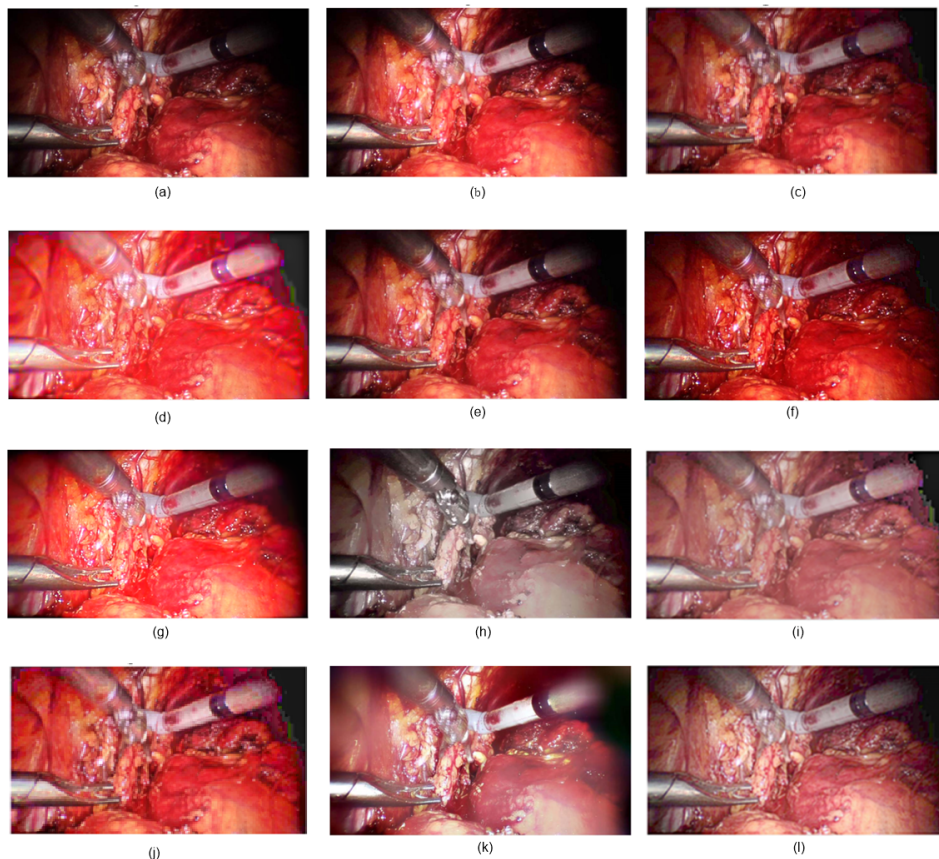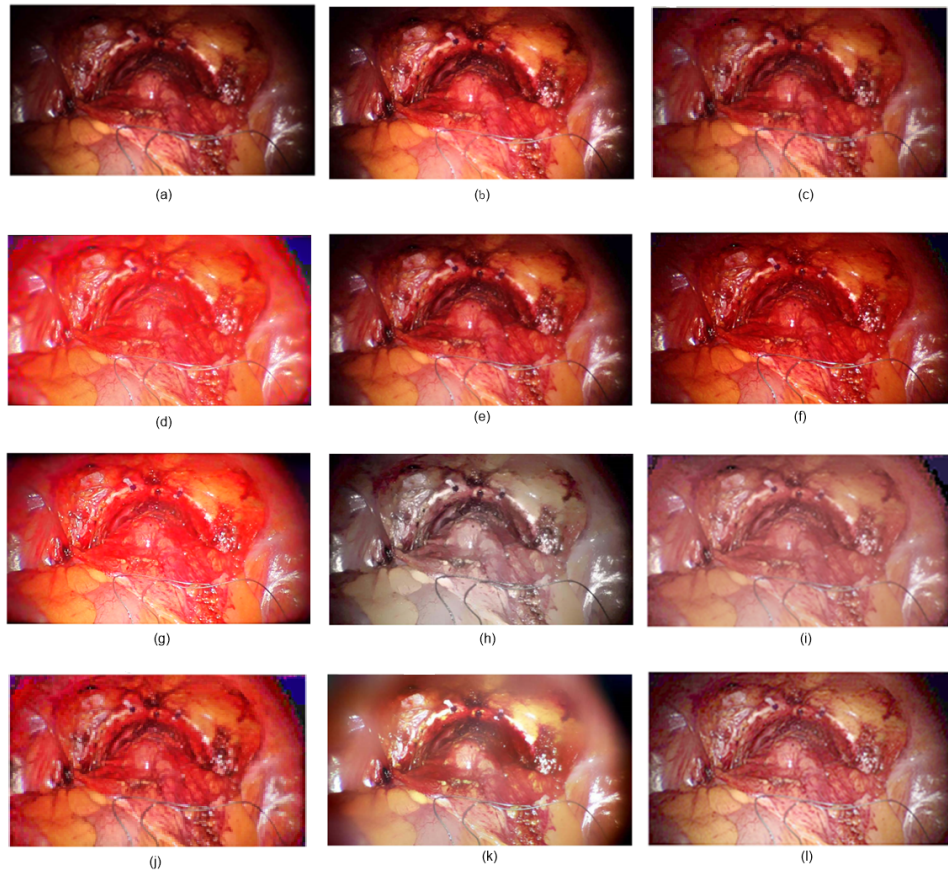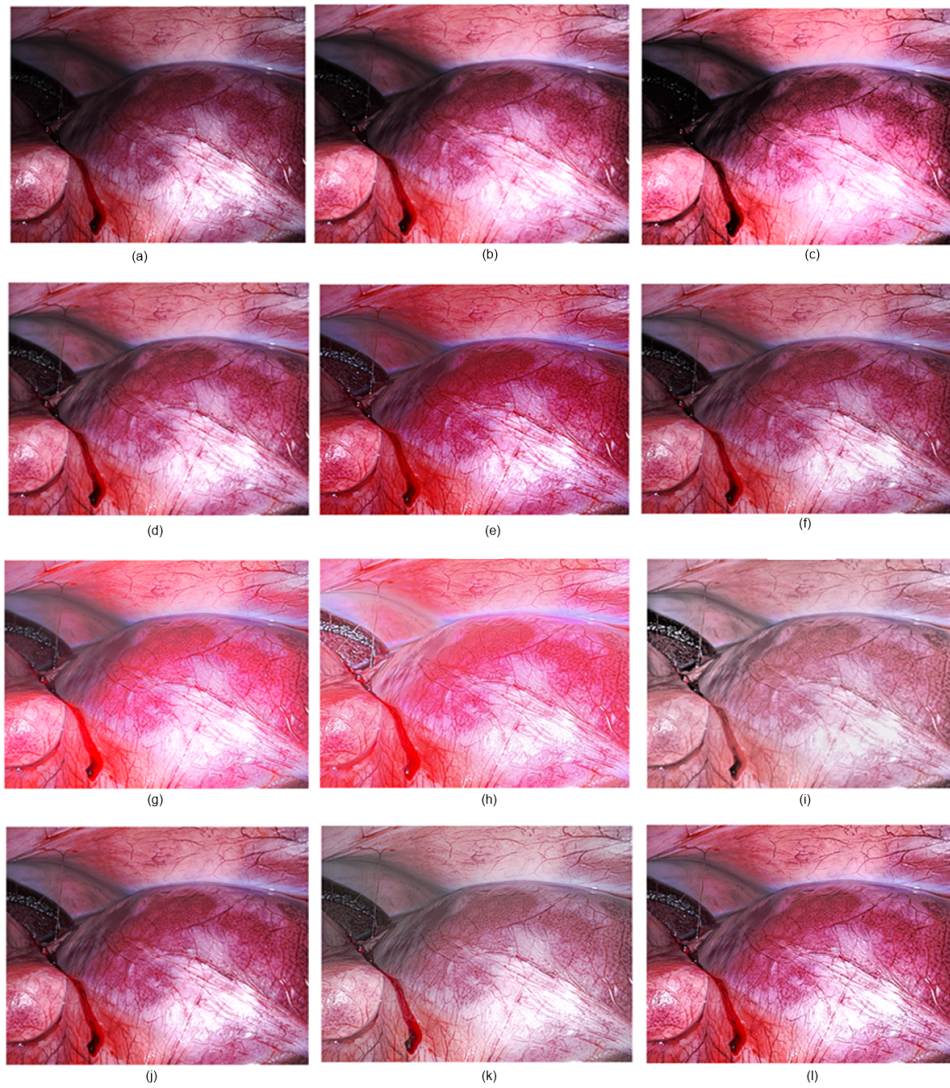
Figure 3.6: Enhanced results for surgical laparoscopic image (a) Input, (b) Ref. [10], (c) Ref. [17], (d) Ref. [8], (e) Ref. [15], (f) Ref. [18]), (g) Ref. [12], (h) Ref. [9], (i) Ref. [13],(j) Ref. [4], (k)Ref. [16], (l) Ours

Table 3.1: Quantitative comparison of enhanced images by different methods

Method	Ref.[8]	Ref.[9]	Ref.[15]	Ref.[12]	Ref.[18]	Ref.[23]	Ref.[14]	Ref.[19]	Ref.[13]	Ref.[10]	Ours
Fig. 3.2	0.42	0.64	0.37	0.57	0.25	0.59	0.61	0.40	0.26	0.45	0.65
Fig. 3.3	0.43	0.66	0.37	0.62	0.25	0.59	0.61	0.40	0.26	0.45	0.65
Fig. 3.4	0.42	0.63	0.38	0.59	0.25	0.55	0.62	0.41	0.26	0.47	0.64
Average	0.42	0.64	0.37	0.59	0.25	0.58	0.61	0.40	0.26	0.46	0.65

Under the protocol approved by the Research Ethics Board of Western University, fourteen clinicians with varying degrees of laparoscopy experiences are recruited to perform the forced-choice assessment of the images. Among the participants, eight were surgeons/fellows with over five years of experience, and six were medical students with less than five years of experience.

The input video images were first processed by the proposed method and five conventional enhancement methods: Li et al. [12], Rao et al. [16], Fu et al. [8], Liang et al. [13], and Huang et al. [10]. The clinicians then independently reviewed these input images, along with versions processed by all of the processing approaches considered here, including our own. As a result, twenty-five sets of testing video images were created, where each set included six enhanced images and the input image. For each trial, the user was asked to select their most favored image among the given image set, where the unenhanced input image was also provided as the baseline reference. The above procedures were repeated for all 25 image sets, and the percentage of each algorithm being the most preferred choice was used as the user's preference score.

According to the feedback from these clinical participants, for the twenty fog-free laparoscopic images, our approach produced the highest score, received an average vote percentage of 33%, while the second-highest method by Fu et al. received an average vote percentage of 19% as shown in Fig. 3.10. A statistical analysis using Wilcoxon's rank test[21] demonstrated significant difference between the proposed method and Fu's method ($p = 0.04$). When the five foggy images are included, the proposed method still produced the highest score, receiving an average vote percentage of 29%, while the second-highest method by Fu et al. received an average vote percentage of 21% as shown in Fig. 3.11. Although our average score is higher,

Table 3.2: Quantitative comparison of enhanced images by different methods

Method	Ref.[8]	Ref.[9]	Ref.[15]	Ref.[12]	Ref.[16]	Ref.[13]	Ref.[18]	Ref.[4]	Ref.[17]	Ref.[10]	Ours
Fig. 3.5	0.44	0.26	0.58	0.31	0.36	0.26	0.26	0.26	0.26	0.48	0.51
Fig. 3.6	0.66	0.47	0.60	0.58	0.56	0.25	0.25	0.24	0.24	0.66	0.67
Fig. 3.7	0.65	0.42	0.63	0.49	0.49	0.25	0.25	0.25	0.25	0.65	0.64
Average	0.55	0.38	0.60	0.46	0.47	0.25	0.25	0.25	0.25	0.60	0.61

the Wilcoxon's rank test gave no significant difference ($p = 0.25$).

3.4 Discussion and conclusion

This chapter proposes a simplified Retinex optimization method with prior enhanced information fusion. This approach minimizes a Retinex optimization model, which consists of one error term and two prior enhanced regularization terms. Having used prior enhanced illumination and reflectance information, which contains image naturalness, structure, and texture, the proposed method can significantly enhance dark regions while preserving naturalness and texture structures. Because we replace TV regularization term with two prior enhanced regularization terms, our algorithm has low computational complexity and fast computation performance. Moreover, the convergence condition of the proposed algorithm is guaranteed. By fusing RGB-space and HSV-space enhancement information, the proposed method can effectively prevent colour cast phenomenon. Finally, experimental results demonstrate that the proposed algorithm is more effective than state-of-the-art algorithms in terms of vision augmentation, performance index, and computation time. Forced-choice preference experimental results further confirm that the proposed method is preferred by clinicians. Therefore, this work is of practical surgical importance in dealing with illumination variations in robotic-assisted laparoscopic prostatectomy, digitally augmenting the laparoscopic video stream, to provide the surgeon with a wider useful field of view.

Since the proposed algorithm minimizes an objective function with prior enhanced illumination and reflectance terms, its effectiveness depends on the accuracy of the prior knowledge terms. In addition, from forced-choice preference results it is observed that the proposed al-
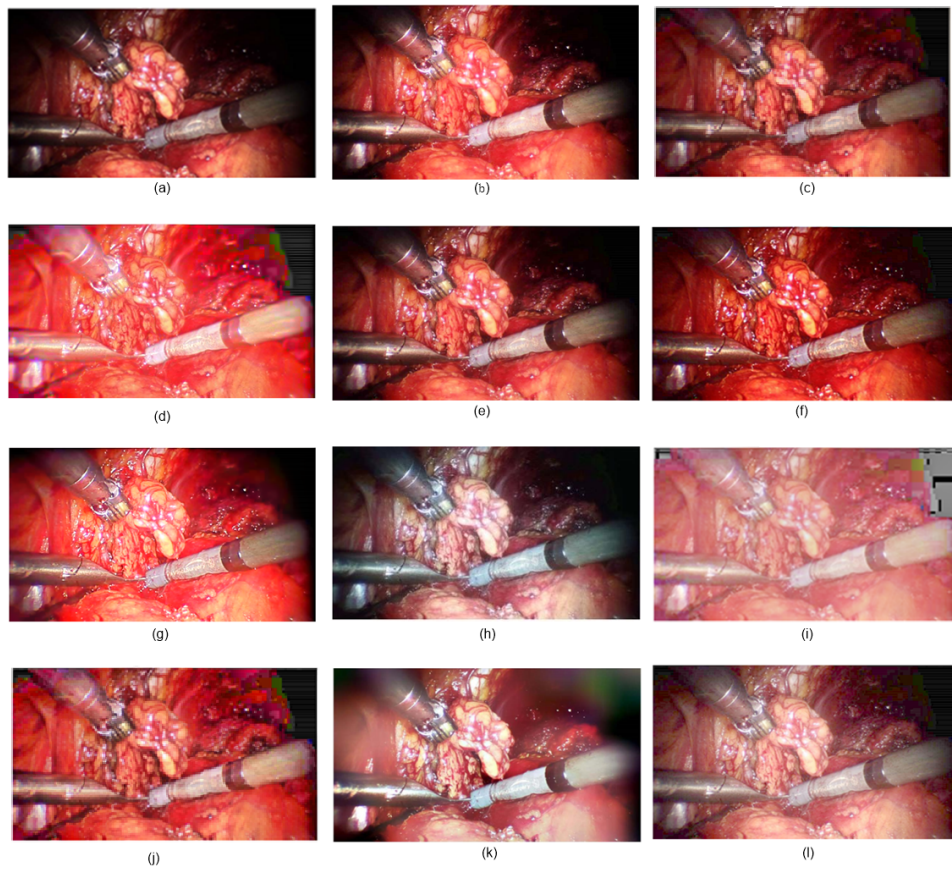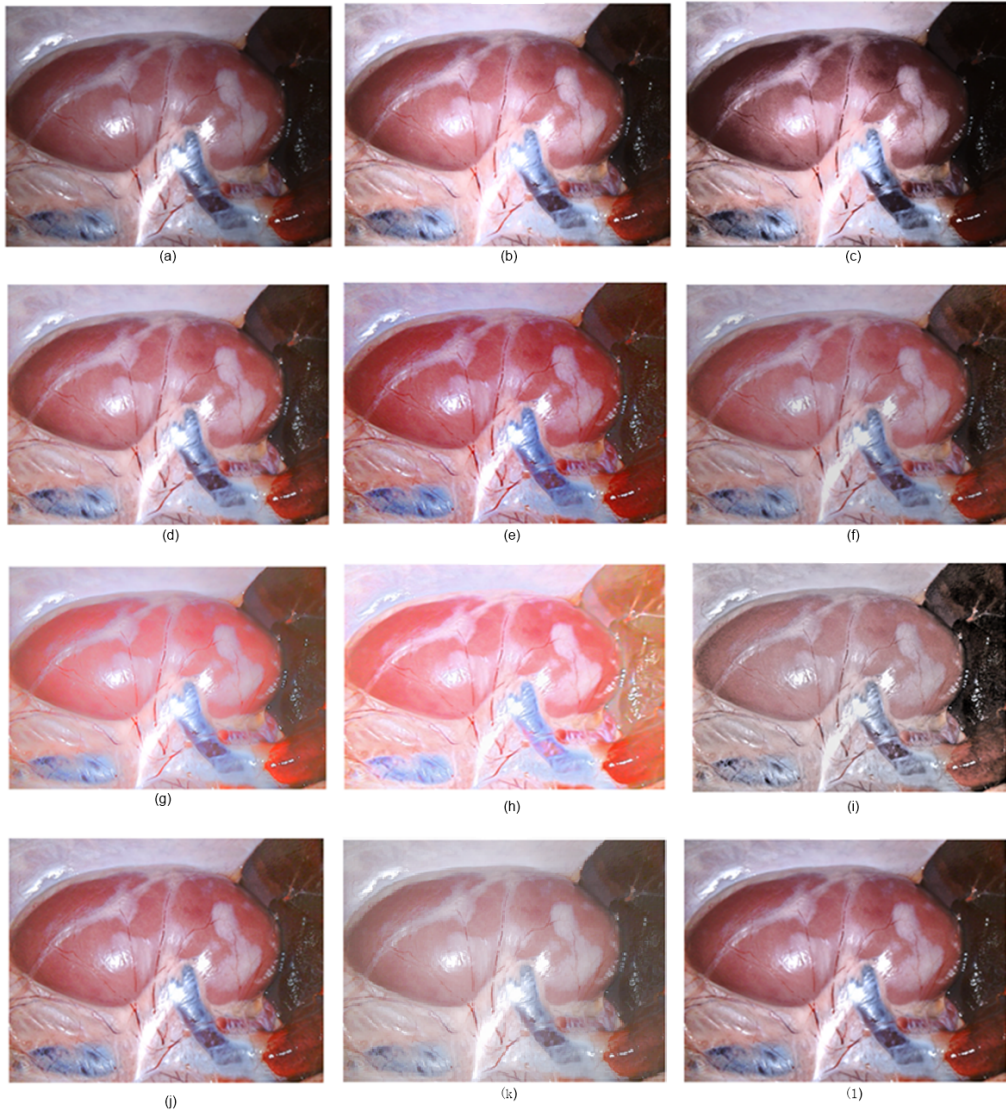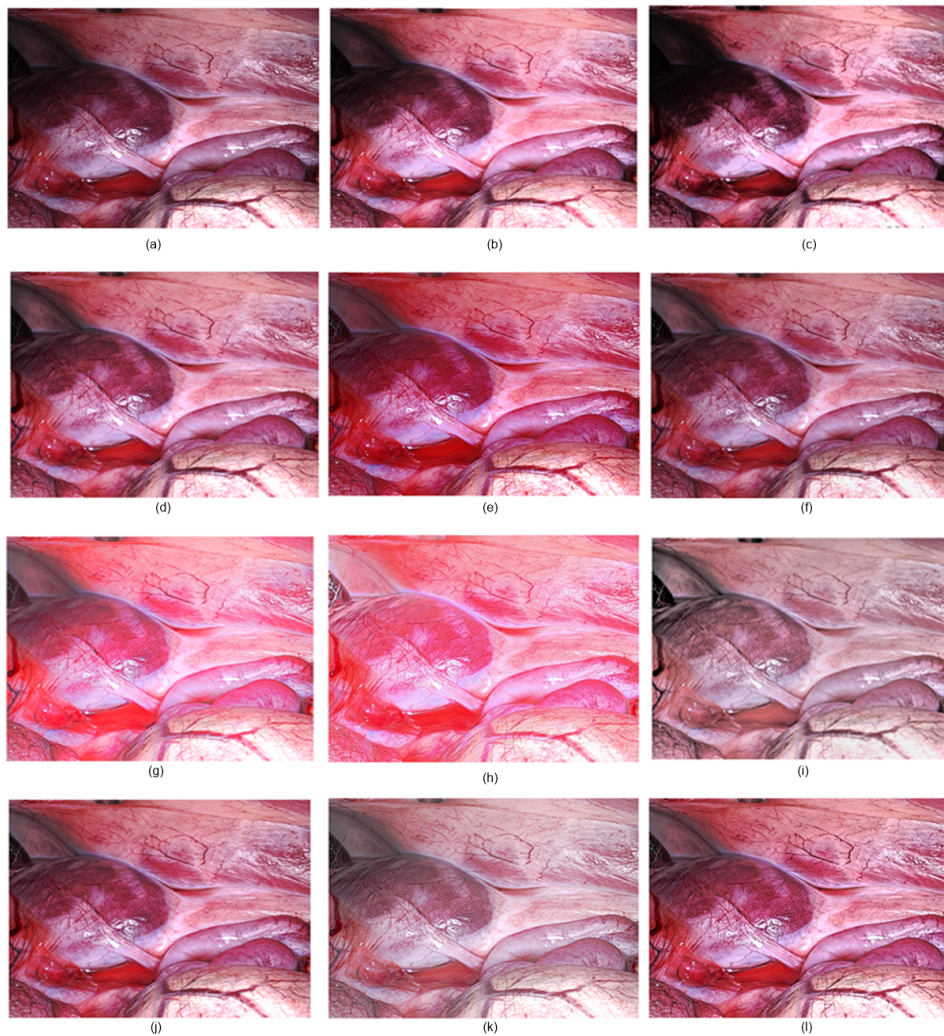
Figure 3.7:   Enhanced results for surgical laparoscopic image (a) Input, (b) Ref. [10], (c) Ref. [17], (d) Ref. [8], (e) Ref. [15], (f) Ref. [18]), (g) Ref. [12], (h) Ref. [9], (i) Ref. [13],(j) Ref. [4], (k)Ref. [16], (l) Ours
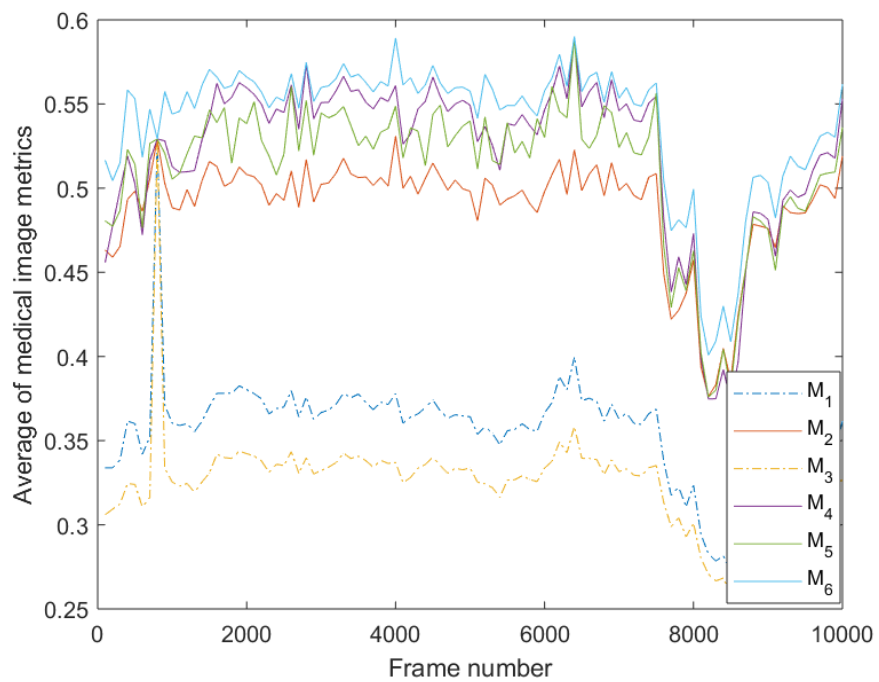
Figure 3.8: Quantitative objective assessment of the processed laparoscopic video images of using methods: $M_1[8]$, $M_2[9]$, $M_3[15]$, $M_4[12]$, $M_5[16]$, and M_6(ours).
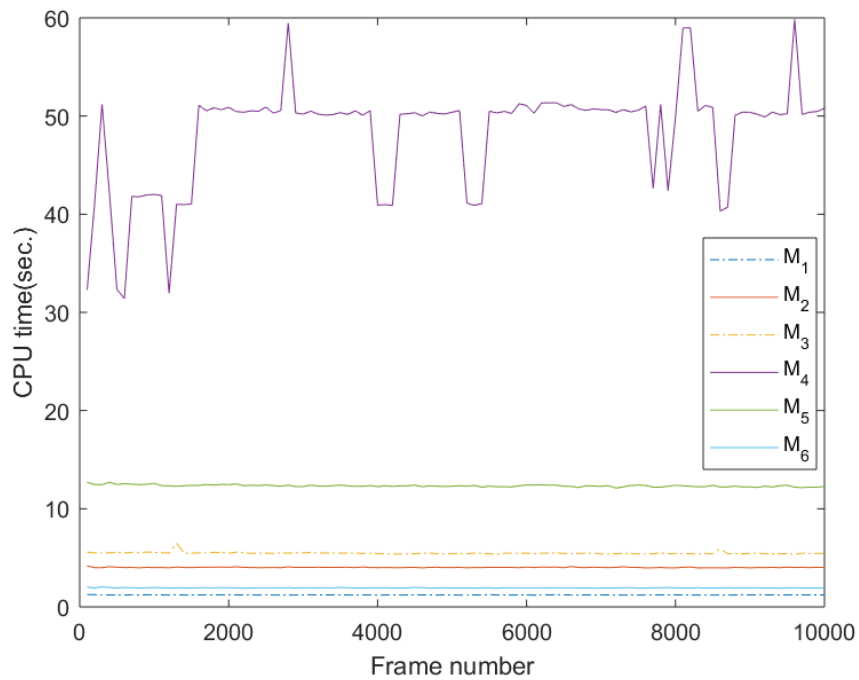
Figure 3.9: Comparison of computing time of using methods: M_1[8], M_2[9], M_3[15], M_4[12], M_5[16], and M_6(ours).
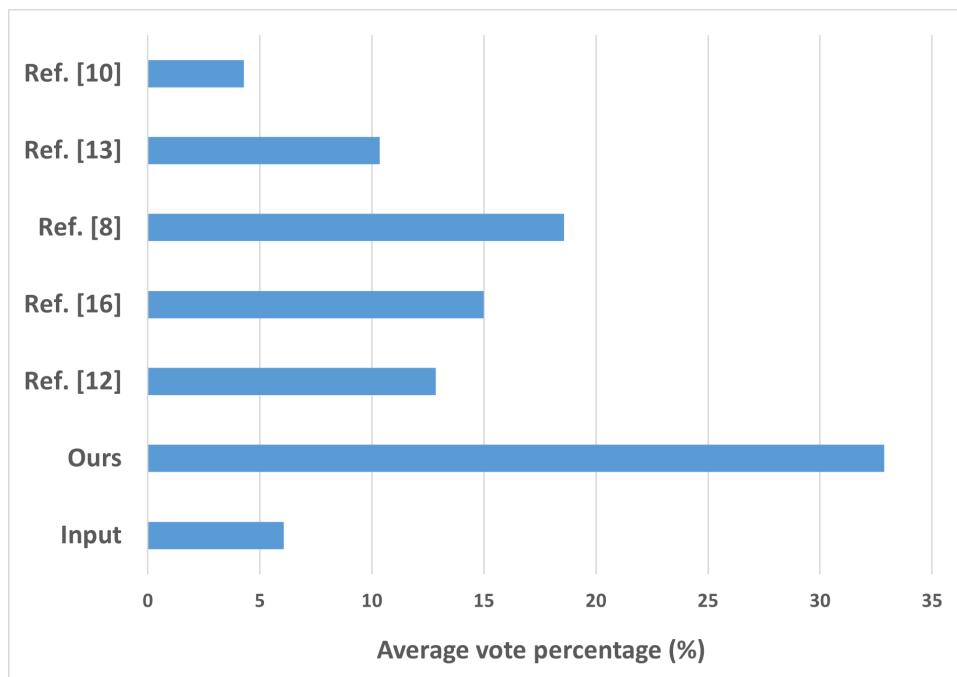
Figure 3.10: Results of forced-choice user study of 20 enhanced fog-free laparoscopic images using the average vote percentages.
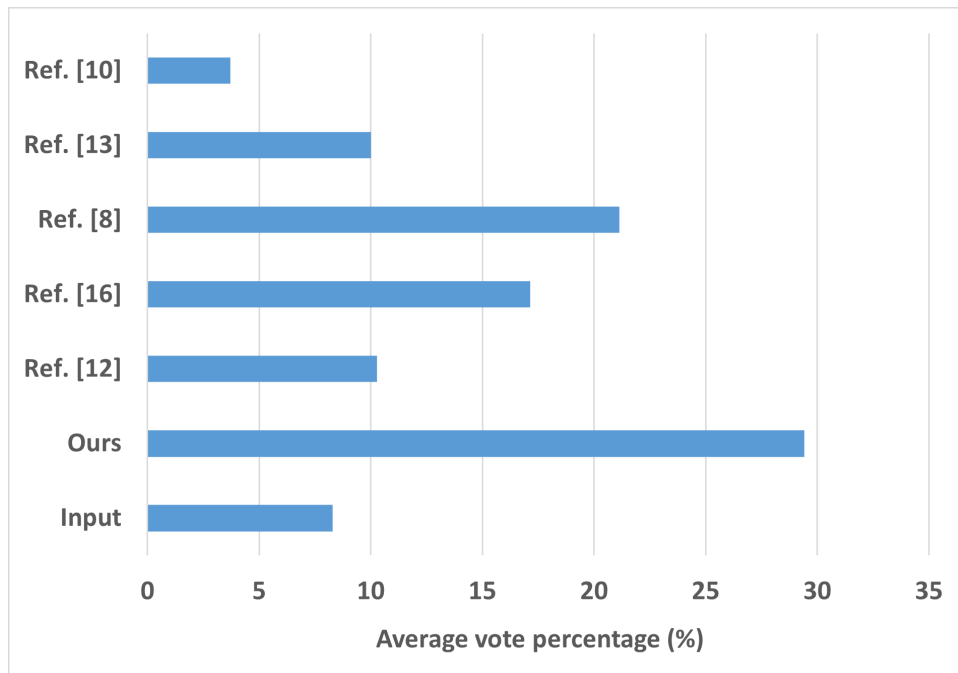
Figure 3.11: Results of forced-choice user study of 25 enhanced laparoscopic images using the average vote percentages.

gorithm is less effective on foggy images, due to the l_2-norm optimization model. At present, the proposed method is running in MATLAB. To further speed up computation for real-time applications, the procedure can be highly parallelized and amenable to GPU implementation. In addition, the experimental parameters need to be further optimized. These issues will be the subject of future research.

Bibliography

[1] Max Allan, Jonathan Mcleod, Congcong Wang, Jean Claude Rosenthal, Zhenglei Hu, Niklas Gard, Peter Eisert, Ke Xue Fu, Trevor Zeffiro, Wenyao Xia, Zhanshi Zhu, Huoling Luo, Fucang Jia, Xiran Zhang, Xiaohong Li, Lalith Sharan, Tom Kurmann, Sebastian Schmid, Raphael Sznitman, Dimitris Psychogyios, Mahdi Azizian, Danail Stoyanov, Lena Maier-Hein, and Stefanie Speidel. Stereo correspondence and reconstruction of endoscopic data challenge, 2021.

[2] Jean-François Aujol, Guy Gilboa, Tony Chan, and Stanley Osher. Structure-texture image decompositionmodeling, algorithms, and parameter selection. *International journal of computer vision*, 67(1):111–136, 2006.

[3] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[4] Bolun Cai, Xianming Xu, Kailing Guo, Kui Jia, Bin Hu, and Dacheng Tao. A joint intrinsic-extrinsic prior model for retinex. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4000–4009, 2017.

[5] Turgay Celik and Tardi Tjahjadi. Automatic image equalization and contrast enhancement using gaussian mixture modeling. *IEEE Transactions on Image Processing*, 21(1):145–156, 2011.

[6] Xiaoyu Du and Youshen Xia. Natural images enhancement using structure extraction and retinex. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 408–420. Springer, 2020.

[7] Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Transactions on Graphics (TOG)*, 27(3):1–10, 2008.

[8] Xueyang Fu, Yinghao Liao, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. *IEEE Transactions on Image Processing*, 24(12):4965–4977, 2015.

[9] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016.

[10] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE transactions on image processing*, 22(3):1032–1041, 2012.

[11] Ron Kimmel, Michael Elad, Doron Shaked, Renato Keshet, and Irwin Sobel. A variational framework for retinex. *International Journal of computer vision*, 52(1):7–23, 2003.

[12] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018.

[13] Zhetong Liang, Jun Xu, David Zhang, Zisheng Cao, and Lei Zhang. A hybrid l1-l0 layer decomposition model for tone mapping. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4758–4766, 2018.

[14] Xiongbiao Luo, Hui-Qing Zeng, Ying Wan, Xiao-Bin Zhang, Yan-Ping Du, and Terry M Peters. Endoscopic vision augmentation using multiscale bilateral-weighted retinex for robotic surgery. *IEEE transactions on medical imaging*, 38(12):2863–2874, 2019.

[15] Seonhee Park, Soohwan Yu, Byeongho Moon, Seungyong Ko, and Joonki Paik. Low-light image enhancement using variational optimization-based retinex model. *IEEE Transactions on Consumer Electronics*, 63(2):178–184, 2017.

[16] Zhitao Rao, Tingfa Xu, Jiqiang Luo, Jie Guo, Guokai Shi, and Hongqing Wang. Non-uniform illumination endoscopic imaging enhancement via anti-degraded model and l 1 l 2-based variational retinex. *EURASIP Journal on Wireless Communications and Networking*, 2017(1):1–11, 2017.

[17] Haonan Su and Cheolkon Jung. Perceptual enhancement of low light images based on two-step noise suppression. *IEEE Access*, 6:7005–7018, 2018.

[18] Shuhang Wang and Gang Luo. Naturalness preserved image enhancement using a priori multi-layer lightness statistics. *IEEE transactions on image processing*, 27(2):938–948, 2017.

[19] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, 22(9):3538–3548, 2013.

[20] Shuigen Wang, Chenwei Deng, Weisi Lin, Guang-Bin Huang, and Baojun Zhao. Nmf-based image quality assessment using extreme learning machine. *IEEE transactions on cybernetics*, 47(1):232–243, 2016.

[21] Frank Wilcoxon. Individual comparisons by ranking methods. In *Breakthroughs in statistics*, pages 196–202. Springer, 1992.

[22] Li Xu, Qiong Yan, Yang Xia, and Jiaya Jia. Structure extraction from texture via relative total variation. *ACM transactions on graphics (TOG)*, 31(6):1–10, 2012.

[23] Dominique Zosso, Giang Tran, and Stanley J Osher. Non-local retinex—a unifying framework and beyond. *SIAM Journal on Imaging Sciences*, 8(2):787–826, 2015.

Chapter 4

Image Highlight Removal Method

This chapter includes material adapted from:

Xia, W., Chen, E.C., Pautler, S.E., Peters, T.M. "A global optimization method for specular highlight removal from a single image." IEEE Access 7 (2019): 125976-125990.

4.1 Introduction

Due to the proximity of the camera, light source, and organ surfaces, the captured images often suffer from specular highlights. The presence of specular highlight is a critical issue for both natural and medical images such as those produced by endoscopes, which can lead to erroneous visual tracking, stereo reconstruction, and image segmentation. Specular highlight removal from a single image is necessary for image analysis and applications. Due to the differences between natural and medical image scenes, existing methods have only been effective on natural images or medical images with textureless regions. In addition, regions with specular highlights may contain vital information relating to the organ such as colour and textures. Therefore, it is desirable to remove specular reflections while preserving the original colour and texture details of the organ surface.

4.1.1 Related work

The dichromatic reflection model [18] is widely used for studying reflection components of a natural colour image, whose intensity $I(\mathbf{x})$ at pixel $\mathbf{x} = (x, y)$ can be described as the combination of diffuse reflection $I_D(\mathbf{x})$ and specular reflection $I_S(\mathbf{x})$:

$$I(\mathbf{x}) = I_D(\mathbf{x}) + I_S(\mathbf{x}) = m_d(\mathbf{x})\Lambda(\mathbf{x}) + m_s(\mathbf{x})\Gamma(\mathbf{x}), \tag{4.1}$$

where $\Lambda(\mathbf{x})$ and $\Gamma(\mathbf{x})$ are the chromaticity of the diffuse reflection and illumination (specular) reflection and $m_d(\mathbf{x})$ and $m_s(\mathbf{x})$ are the diffuse and specular reflection coefficients (parameters), respectively. In RGB colour space, $I(\mathbf{x}) = [I_r(\mathbf{x}), I_g(\mathbf{x}), I_b(\mathbf{x})]^T$ is the colour intensity vector at pixel $\mathbf{x}$, $\Lambda(\mathbf{x}) = [\Lambda_r(\mathbf{x}), \Lambda_g(\mathbf{x}), \Lambda_b(\mathbf{x})]^T$ is the diffuse chromaticity vector, and $\Gamma(\mathbf{x}) = [\Gamma_r(\mathbf{x}), \Gamma_g(\mathbf{x}), \Gamma_b(\mathbf{x})]^T$ is the illumination chromaticity vector. The goal of specular highlight removal is to obtain a specular-free image estimate, based on received highlight image $I(\mathbf{x})$. Because both diffuse chromaticity and reflection coefficients are unknown, specular highlight removal is an ill-posed and blind separation problem.

Highlight removal methods can be divided into main categories: dichromatic reflection model-based and inpainting-based. Many dichromatic reflection model-based methods have been reported to be effective on natural images [18]-[26]. Considerable efforts were devoted to non-optimization modeling techniques, including pseudo specular-free image, colour clustering, bilateral filter, and intensity ratio, respectively. In recently, the inpainting-based method was used for specular highlight removal of medical images. Tan et al. [14] first introduced an inpainting method for highlight removal. Thereafter, joint detection and inpainting-based methods were studied. Oh et al. [13] proposed one HSV space-based method for detecting specular highlights by using two threshold sets:

$$S(\mathbf{x}) < T_s, \ V(\mathbf{x}) > T_v. \tag{4.2}$$

where the highlight image was segmented into two areas, absolute and relative bright areas. Naturally, this detection method is effective but the detected relative bright areas may include white surfaces. Stehle [21] used a inpainting technique with texture filling, while Meslouhi et al. [12] employed a reflection enhancement technique to improve highlight detection accuracy, and an inpainting-based algorithm for specular reflection removal in colposcopic images. Saint-Pierre et al. [16] presented an inpainting technique to correct specular reflections in thoracoscopic images. Arnold et al. [1] proposed a RGB space-based detection method. These inpainting-based methods are only effective on endoscopic images with very small highlight regions. However, in other cases, the texture details in the inpainted region are often lost due to neighborhood interpolation limitations.

Optimization modeling-based methods for specular highlight removal from natural images have been developed in recent years. Most of them are based on the variation of the dichromatic reflection model:

$$\hat{I}(\mathbf{x}) = \alpha(\mathbf{x})\Lambda(\mathbf{x}) + (1 - \alpha(\mathbf{x}))\Gamma(\mathbf{x})$$

where

$$\hat{I}(\mathbf{x}) = \frac{I(\mathbf{x})}{\Sigma_{c\in\{r,g,b\}}I_c(\mathbf{x})}, \alpha(\mathbf{x}) = \frac{m_d(\mathbf{x})}{m_d(\mathbf{x}) + m_s(\mathbf{x})}.$$

In particular, Zhao et al. [27] proposed solving the following optimization problem:

$$\min_{(\alpha,\Lambda)} \sum_x f_1(\alpha(\mathbf{x}), \Lambda(\mathbf{x})) + \beta_1 f_2(\Lambda(\mathbf{x})) + \beta_2 f_3(\alpha(\mathbf{x}), \Lambda(\mathbf{x})) \tag{4.3}$$

where $f_1 = \|\hat{I}(\mathbf{x}) - \alpha(\mathbf{x})\Lambda(\mathbf{x}) - (1 - \alpha(\mathbf{x}))\Gamma(\mathbf{x})\|_2^2, f_2 = \sum_{\mathbf{z}} w(\mathbf{x}, \mathbf{z})\|\Lambda(\mathbf{x}) - \Lambda(\mathbf{z})\|_2^2$, and f_3 is the structure similarity function. Kim et al. [8] solved the following problem:

$$\min_{(\alpha,\Lambda,\lambda)} \sum_x f_1(\alpha(\mathbf{x}), \Lambda(\mathbf{x})) + \beta_1\|\nabla(1 - \alpha(\mathbf{x}))\|_2$$
$$+ \|\nabla\Lambda_\lambda(\mathbf{x}))\|_1 + \beta_2\|\lambda(\mathbf{x})\|_0 \tag{4.4}$$

where $\|\cdot\|_0$ denote l_0 norm. Based on the reflection model, an optimization method for the

diffuse and specular components was presented, where the objective function contains a data term as well as two regularization, isotropic and anisotropic smoothness terms. Wei et al. [24] presented an optimization method for specular highlight reduction by assuming that the surface geometry is known. These methods may estimate simultaneously diffuse and and specular reflection. Yet, because of the non-convexity of the associated objective function in these methods, an alternating direction algorithm is in general employed, and thus proper initial points are required due to local minima problem. For robustness to outliers, Guo et al.[6] proposed a highlight removal method by solving the following nuclear-norm and l_1-norm optimization problem:

$$\min_{(W_d, \Gamma, M_s)} \|W_d\|_* + \lambda \|M_s\|_1 + \tau \|W_d\|_1$$

$$Y = \Phi_d W_d + \Gamma M_s, \quad W_d \geq O, \ M_s \geq O$$

(4.5)

where $\|\cdot\|_*$ denotes the nuclear norm, Y is an observed colour matrix, Φ_d is a colour dictionary matrix, W_d is a matrix of weighting coefficients of pixels, Γ is a column vector of illumination chromaticity, and M_s is a row vector of specular highlights. Under the condition of a given colour dictionary matrix, an alternating iteration scheme is presented by using an augmented Lagrange function to solve (4.5). Recently, Li. et al.[10] proposed a principal component analysis-based optimization method for highlight removal. This method has high algorithm complexity and its estimation error can not be minimized due to the matrix decomposition required.

4.1.2 Contributions

This chapter proposes a global optimization method for specular highlight removal from a single image based on a dichromatic reflection model. First, a modified illumination chromaticity with parameter perturbation is introduced to solve the problem of illumination chromaticity on fully saturated highlight images. Second, a new diffuse chromaticity estimate is proposed by a joint RGB-space and HSV-space detection set and an adaptive inpainting technique. Third, theoretical analysis proves that the estimated diffuse chromaticity can approximate the true diffuse

chromaticity. Finally, a double gradient regularization-based convex optimization method for reflection coefficients is presented, and an iteration algorithm is guaranteed to find the optimal reflection coefficients.

4.2 Methods

To overcome the limitation of current highlight removal methods for natural images or medical images, a global optimization method for specular highlight removal from a single image is presented. First, the illumination chromaticity is modified, the hue and saturation in HSV colour space is corrected, and the diffuse chromaticity is accurately estimated by using the inverse transformation of HSV in RGB space. Second, based on both the diffuse chromaticity estimate and modified illumination chromaticity, a double gradient regularization-based convex optimization problem is minimized to estimate diffuse reflection coefficients. Fig. 4.1 displays the workflow of our highlight removal method, which shows the relationship among various components that are discussed in subsequent subsections.

As an illustrative example, Fig. 4.2 displays highlight removal results of both natural and medical images by three methods, where Figs.4.2 (a) and (e) are highlight images. Figs.4.2 (b) and (f) are results by using our method, (c) and (g) are results by using an inpainting-based method [16], and (d) and (h) are results by using a model based-method [15], where the specular free images are shown in the upper row and the separated specular components are shown in the lower row. From Figs.4.2(c) and (g)(upper), it is seen that the inpainting-based method cannot effectively remove highlights on the natural image and it loses the texture details in the inpainted region of the medical image. From Fig.4.2(d)(upper) it is easy to observe that the model-based method is more effective on the nature image but Fig.4.2(d)(lower) displays its incorrect specular components. From Fig.4.2(h)(upper), it is seen that the model-based method is less effective on the medical images due to colour distortion in the highlight region. In contrast, Figs. 4.2(b) and (f) show our highlight removal results. It is seen that the proposed method can
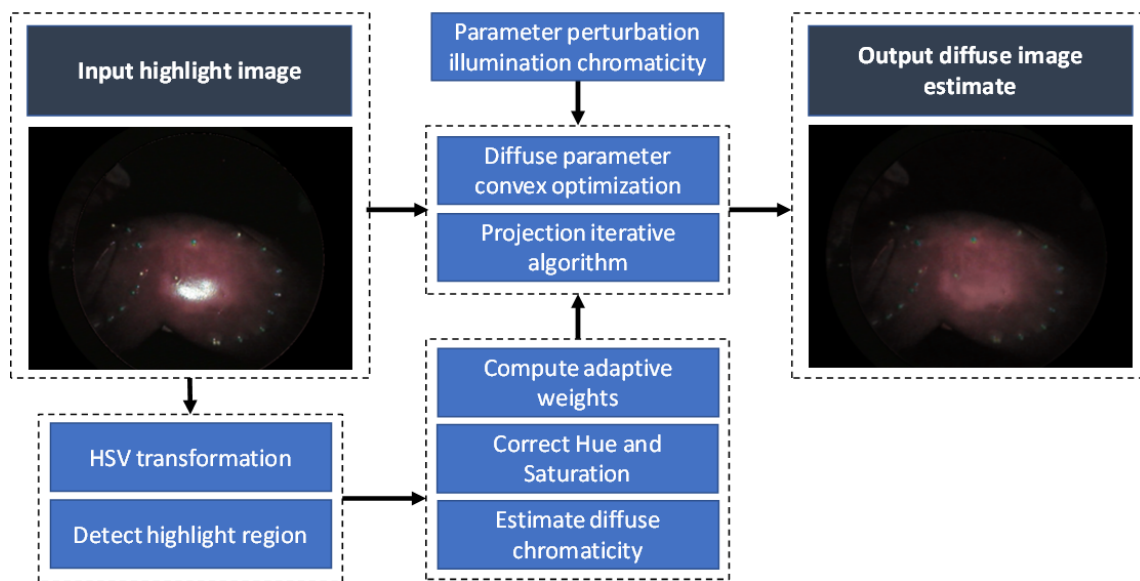
Figure 4.1: Scheme of proposed highlight removal method

effectively remove the highlight from both natural and medical images, while preserving the textures and colour of the original scene.

4.2.1 Modified illumination chromaticity

The current dichromatic reflection model-based methods are less effective on endoscopic images, since the illumination chromaticity is often assumed to be uniform for a given RGB image such that $\Gamma_r(\mathbf{x}) = \Gamma_g(\mathbf{x}) = \Gamma_b(\mathbf{x})$. Most of these approaches normalize the dichromatic reflection model by dividing the summation of its observed colour image channels so that $\Gamma(\mathbf{x}) = [1/3, 1/3, 1/3]^T$ [10]. It is also common in the literature to have unnormalized $I(\mathbf{x})$ in the range $[0, 1]$, where $\Gamma(\mathbf{x}) = [1, 1, 1]^T$, and $0 \leq m_d(\mathbf{x}), m_s(\mathbf{x}) \leq 1$[18]. However, because of the proximity of the light source to the organ surface, the specular highlights in endoscopic images are much stronger than those in natural images, such that some of the highlight pixels are saturated at maximum intensity. In other words, $\max\{I_r(\mathbf{x}), I_g(\mathbf{x}), I_b(\mathbf{x})\} = 1$.

Without the loss of generality, let us consider unnormalized endoscopic images. Let X_0 denote the fully saturated highlight region. Then $I(\mathbf{x}) = [1, \eta, \theta]^T$ for any $\mathbf{x} \in X_0$ since $I_r(\mathbf{x}) \geq$
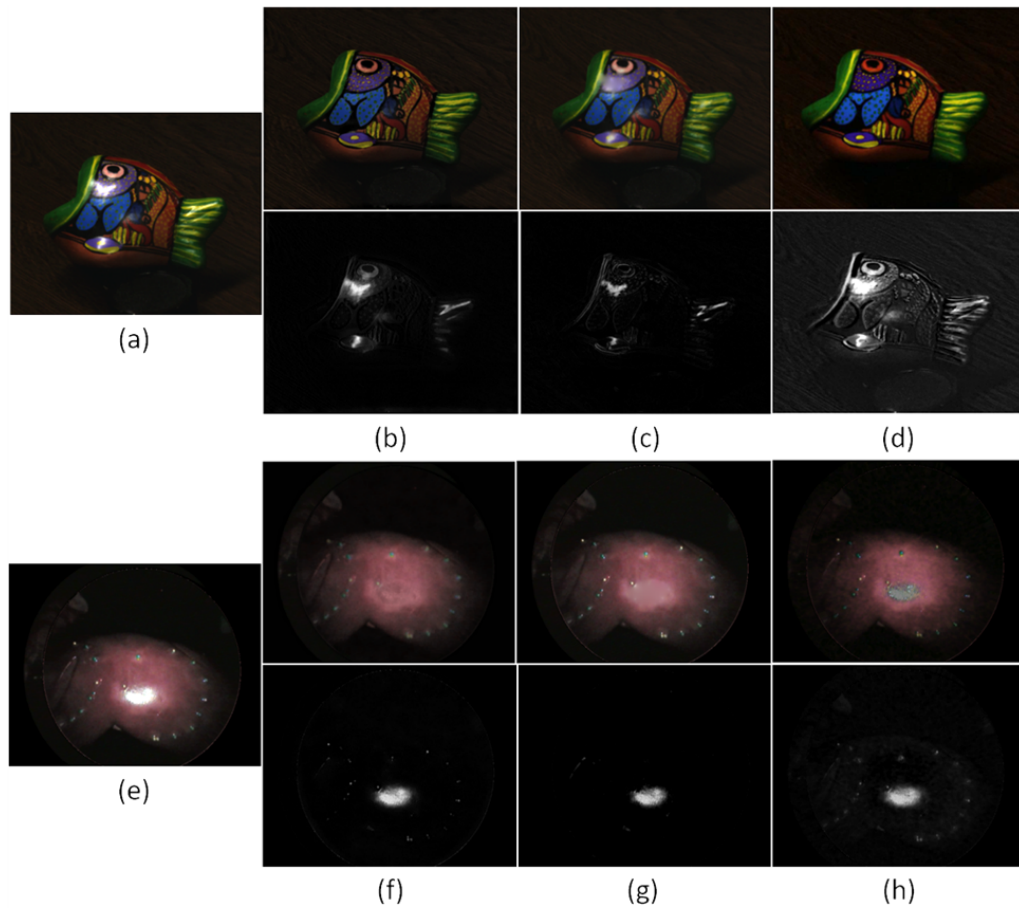
Figure 4.2: Highlight removal results of both natural and medical images. (a) and (e) input images, (b) and (f) our method, (c) and (g) inpainting-based method [16], and (d) and (h) model based-method [15], where the specular free images are displayed in upper row and the separated specular components are shown in lower row.

$max\{I_g(\mathbf{x}), I_b(\mathbf{x})\}$, where $0 \leq \eta \leq 1$ and $0 \leq \theta \leq 1$. As a special case, $I(\mathbf{x}) = [1, 1, 1]^T$ for all $\mathbf{x} \in X_0$. From model (4.1) it follows that

$$(1 - m_s(\mathbf{x}))\Gamma(\mathbf{x}) = m_d(\mathbf{x})\Lambda(\mathbf{x}). \tag{4.6}$$

Since in vivo tissue is seldom pure white due to the presence of hemoglobin, the case that $\Lambda_r = \Lambda_g = \Lambda_b$ is mostly invalid for the endoscopic images. As a result, from $\Gamma(\mathbf{x}) = [1, 1, 1]^T$ or $\Gamma(\mathbf{x}) = [1/3, 1/3, 1/3]^T$ we get $m_d(\mathbf{x}) = 0$ for any $\mathbf{x} \in X_0$ and thus $I_D(\mathbf{x}) = m_d(\mathbf{x})\Lambda(\mathbf{x}) = 0$, which would force the restored specular-free image surface to contain a black hole on X_0.

To avoid the problem of conventional illumination chromaticity on the fully saturated specular reflection region in the endoscopic images, we introduce the following modified illumination chromaticity with parameter perturbation:

$$\Gamma_\epsilon(\mathbf{x}) = [1 - \epsilon, 1, 1]^T \tag{4.7}$$

where $1 > \epsilon > 0$. When the highlight region is very small, ϵ is chosen to be 0. Otherwise, ϵ is in general chosen as the mean intensity of the neighborhood around the highlight region.

4.2.2   Specular highlight detection and colour correction

Design of highlight detection set

To detect specular highlights, we propose using two detection sets in RGB and HSV spaces, respectively. Based on HSV transformation $\varphi(\cdot)$ [35], the observed image $I(\mathbf{x})$ from RGB space to hue is converted into: $H(\mathbf{x})$, saturation: $S(\mathbf{x})$, and value $V(\mathbf{x})$. This section introduces one threshold detection set in HSV space:

$$X_{SV}(\mathbf{x}) = \{\mathbf{x}|S(\mathbf{x}) < \alpha, V(\mathbf{x}) > 1 - \alpha\} \tag{4.8}$$

Using Another threshold detection set in RGB space is given by:

$$X_G(\mathbf{x}) = \{\mathbf{x}|G(\mathbf{x}) \geq \tau\} \tag{4.9}$$

where $\alpha > 0$, $\tau > 0$ are mainly two threshold values, $\hat{I}(\mathbf{x}) = \min_{c \in \{r,g,b\}} I_c(\mathbf{x})$ denotes the dark-channel image, and the gradient magnitude of the dark-channel image is given by

$$G(\mathbf{x}) = \frac{\sqrt{(\hat{I}(x+1,y) - \hat{I}(x-1,y))^2 + (\hat{I}(x,y+1) - \hat{I}(x,y-1))^2}}{2}.$$

Based on (4.8) and (4.9), a joint HSV and RGB-space detection set is defined as:

$$X_{SVG}(\mathbf{x}) = X_{SV}(\mathbf{x}) \bigcup \mathbf{X_G}(\mathbf{x}). \tag{4.10}$$

Correction of hue and saturation

According to (4.10), for any $\mathbf{x} \in X_{SVG}(\mathbf{x})$, the hue and saturation of the observed image in HSV space can be corrected by using the adaptive inpainting update formulas below:

$$\mathbf{H}^*(\mathbf{x}) = \frac{1}{\sum\limits_{u \in \Omega(\mathbf{x})} w_{SVG}(\mathbf{u})} \sum\limits_{u \in \Omega(\mathbf{x})} \mathbf{H}(\mathbf{u}) w_{SVG}(\mathbf{u}), \tag{4.11}$$

$$\mathbf{S}^*(\mathbf{x}) = \frac{1}{\sum\limits_{u \in \Omega(\mathbf{x})} w_{HS}(\mathbf{x}, \mathbf{u})} \sum\limits_{u \in \Omega(\mathbf{x})} \mathbf{S}(\mathbf{u}) w_{HS}(\mathbf{x}, \mathbf{u}) \tag{4.12}$$

where $\Omega(\mathbf{x})$ denotes a window at pixel $\mathbf{x}$ and two adaptive weights are defined as:

$$w_{SVG}(\mathbf{x}) = \begin{cases} \alpha, & \mathbf{x} \in X_{SVG}(\mathbf{x}) \\ 1, & else \end{cases} \tag{4.13}$$

$$w_{HS}(\mathbf{x}, \mathbf{u}) = e^{\left(\frac{-(\mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{u}))^2}{\sigma^2}\right)} e^{-(1 - \mathbf{S}(\mathbf{u}))^2} \tag{4.14}$$

where σ is the spread parameter commonly used in guided bilateral filters, which is often taken as 0.01.

4.2.3   Diffuse chromaticity estimate

Let I_H^* and I_S^* be the corrected hue and saturation of an observed image $I(\mathbf{x})$ in HSV colour space, respectively. The proposed diffuse chromaticity estimate is defined as

$$\Lambda^*(\mathbf{x}) = \varphi^{-1}(I_H^*, I_S^*, 1) \tag{4.15}$$

where $\varphi^{-1}(\cdot)$ is the inverse transformation from HSV colour space to RGB colour space.

4.2.4   Convex optimization for diffuse reflection

According to (4.15), the nonlinear dichromatic reflection model (4.1) can be rewritten as the following linear dichromatic reflection model:

$$I(\mathbf{x}) = m_d(\mathbf{x})\Lambda^*(\mathbf{x}) + m_s(\mathbf{x})\Gamma_\epsilon(\mathbf{x}) + \delta(\mathbf{x}) \tag{4.16}$$

where $\delta(\mathbf{x})$ is the model error.

To estimate diffuse reflection parameters, the model error defined in (4.16) should be minimized:

$$E_1(m_d(\mathbf{x}), m_s(\mathbf{x})) = \frac{1}{2}\|m_d(\mathbf{x})\Lambda^*(\mathbf{x}) + m_s(\mathbf{x})\Gamma^* - I(\mathbf{x})\|_2^2 \tag{4.17}$$

where $\|\cdot\|_2$ denotes the l_2 norm and $\Gamma^* = \Gamma_\epsilon$. On the other side, the regularization terms must be considered to overcome the ill-posed problem. Because the specular-free image is smooth, its gradient changes slowly and thus one l_2-norm regularization term is needed. The specular image mainly consists of large bright region and its gradient tends to be sparse, so another l_1-norm regularization term is also needed. Since the diffuse reflection and illumination

chromaticity are given, the following double gradient regularization term should be minimized:

$$E_2(m_d(\mathbf{x}), m_s(\mathbf{x})) = \beta_1 \|\nabla m_d(\mathbf{x})\|_2^2 + \beta_2 \|\nabla m_s(\mathbf{x})\|_1 \tag{4.18}$$

where $\|\cdot\|_1$ denotes l_1 norm, β_1 and β_2 are the regularization parameters, and $\nabla m_d(\mathbf{x})$, $\nabla m_s(\mathbf{x})$ are the gradient of $m_d(\mathbf{x})$, $m_s(\mathbf{x})$, respectively. Because the two gradient regularization terms can describe the smoothness and non-smoothness of diffuse and specular components, it is useful for texture detail preservation [2][17].

Now, by incorporating both (4.17) and (4.18), a global optimization method is proposed for diffuse reflection coefficient estimate:

$$\min_{m_d, m_s} \sum_x E_1(m_d(\mathbf{x}), m_s(\mathbf{x})) + E_2(m_d(\mathbf{x}), m_s(\mathbf{x}))$$

$$s.t. \qquad 0 \le m_d(\mathbf{x}) \le 1, 0 \le m_s(\mathbf{x}) \le 1 \tag{4.19}$$

4.3 Performance analysis and algorithm implementation

This section analyze the performance of the proposed method and algorithm implementation.

4.3.1 Performance analysis

First, as for the modified illumination chromaticity, the following Proposition 1 holds:

Proposition 1. Let the diffuse chromaticity $\Lambda(\mathbf{x})$ be converted into hue: $\Lambda_H(\mathbf{x})$, saturation: $\Lambda_S(\mathbf{x})$, and value $\Lambda_V(\mathbf{x})$ in HSV space. If $\Gamma_\epsilon(\mathbf{x})$ is taken in model (4.1) and $\Lambda_V(\mathbf{x}) = 1$, then $m_d(\mathbf{x}) > 0$ for any $\mathbf{x} \in X_0$.

Proof. Substituting $\Gamma_\epsilon(\mathbf{x})$ into (4.1), we have

$$I(\mathbf{x}) = m_d(\mathbf{x})\Lambda(\mathbf{x}) + m_s(\mathbf{x})\Gamma_\epsilon(\mathbf{x}). \tag{4.20}$$

Then $I_g(\mathbf{x}) = m_d(\mathbf{x})\Lambda_g(\mathbf{x}) + m_s(\mathbf{x})$, $I_b(\mathbf{x}) = m_d(\mathbf{x})\Lambda_b(\mathbf{x}) + m_s(\mathbf{x})$, and $I_r(\mathbf{x}) = m_d(\mathbf{x})\Lambda_r(\mathbf{x}) +$

$m_s(\mathbf{x})(1 - \epsilon)$. If $I(\mathbf{x}) = [1, 1, 1]^T$ for any $\mathbf{x} \in X_0$, then

$$
\begin{cases}
1 - m_s(\mathbf{x})(1 - \epsilon) = m_d(\mathbf{x})\Lambda_r(\mathbf{x}) \\
1 - m_s(\mathbf{x}) = m_d(\mathbf{x})\Lambda_g(\mathbf{x}) \\
1 - m_s(\mathbf{x}) = m_d(\mathbf{x})\Lambda_b(\mathbf{x})
\end{cases}
$$

Thus $m_d(\mathbf{x})max\{\Lambda_r(\mathbf{x}), \Lambda_g(\mathbf{x}), \Lambda_b(\mathbf{x})\} = max\{1 - m_s(\mathbf{x}), 1 - m_s(\mathbf{x})(1 - \epsilon)\} = 1 - m_s(\mathbf{x})(1 - \epsilon)$.

Because $\Lambda_V(\mathbf{x}) = max\{\Lambda_r(\mathbf{x}), \Lambda_g(\mathbf{x}), \Lambda_b(\mathbf{x})\}$, and hence $m_d(\mathbf{x})\Lambda_V(\mathbf{x}) = 1 - m_s(\mathbf{x})(1 - \epsilon)$. From

$\Lambda_V(\mathbf{x}) = 1$ it follows that $m_d(\mathbf{x}) = 1 - m_s(\mathbf{x})(1 - \epsilon) > 0$ since $m_s(\mathbf{x})(1 - \epsilon) < 1$. If $I(\mathbf{x}) = [1, \eta, \theta]^T$

for any $\mathbf{x} \in X_0$ where $0 \le \eta < 1$ and $0 \le \theta < 1$, then for any $\mathbf{x} \in X_0$

$$
m_d(\mathbf{x}) = 1 - m_s(\mathbf{x})(1 - \epsilon) > 0, \forall \epsilon \in (0, 1)
$$

since $1 - m_s(\mathbf{x})(1 - \epsilon) \ge max\{\eta - m_s(\mathbf{x}), \theta - m_s(\mathbf{x})\}$. $\square$

Proposition 1 addresses that $\Gamma_\epsilon(\mathbf{x})$ can effectively handle the black hole problem in the

saturated highlight regions. In addition, from the analysis of Proposition 1 it is seen that in

the case that $I(\mathbf{x}) = [1, 1, 1]^T$ on the saturated highlight regions, Proposition 1 still holds for

$\Gamma_\epsilon(\mathbf{x}) = [1, 1 - \epsilon, 1]^T$ or $[1, 1, 1 - \epsilon]^T$. However, in other cases that $I(\mathbf{x}) = [1, \eta, \theta]^T$ where

$0 \le \eta < 1$ and $0 \le \theta < 1$, it is difficult for us to take a proper ϵ such that $m_d(\mathbf{x}) > 0$.

Next, to analyze the approximation of the proposed diffuse chromaticity estimate, two use-

ful properties of HSV transformation are given as below:

Proposition 2. Let $\varphi(\cdot)$ be an HSV transformation which converts an image in RGB colour

space to an image in HSV colour space. Then for any free colour value k

$$
\varphi(kR, kG, kB) = (H, S, kV) \tag{4.21}
$$

Proof. According to the standardized colourimetric transformation [20], the three Hexcone

formulas can convert RGB colour space into HSV colour space: $V = max\{R, G, B\}$, $S =$

$(V - min\{R, G, B\}/V$, and

$$H = \begin{cases} 0, & if\ V = min\{R, G, B\} \\ (60^o \times \frac{G-B}{V-min\{R,G,B\}} + 360^o)mod360^o, & ifV = R \\ (60^o \times \frac{G-B}{V-min\{R,G,B\}} + 360^o), & if\ V = G \\ (60^o \times \frac{G-B}{V-min\{R,G,B\}} + 240^o), & if\ V = B \end{cases} \tag{4.22}$$

It is seen that $kV = max\{kR, kG, kB\}$ and $(kV-min\{kR, kG, kB\})/(kV) = k(V-min\{R, G, B\})/(kV) = S$. Again from (4.22) we have

$$\frac{(kG - kB)}{kV - min\{kR, kG, kB\}} = \frac{(G - B)}{(V - min\{R, G, B\})}.$$

Thus (4.22) still holds for any (kR, kG, kB) and kV. It follows (4.21). $\quad \square$

Proposition 3. Let $\varphi^{-1}(\cdot)$ be an inverse transformation from HSV colour space to RGB colour space. Then

$$V\varphi^{-1}(H, S, 1) = \varphi^{-1}(H, S, V) \tag{4.23}$$

Proof. From Proposition 2 it is seen that for any free colour pixel $\mu > 0$

$$\varphi(\mu R, \mu G, \mu B) = (H, S, \mu V) \tag{4.24}$$

Then $\mu(R, G, B) = \varphi^{-1}(H, S, \mu V)$ and thus

$$(R, G, B) = \{\varphi^{-1}(H, S, \mu V)\}/\mu \tag{4.25}$$

Let $\mu = 1/V$. Then

$$(R, G, B) = V\varphi^{-1}(H, S, 1) \tag{4.26}$$

On the other side,

$$(R, G, B) = \varphi^{-1}(H, S, V) \qquad (4.27)$$

From both (4.26) and (4.27) it follows (4.23).   □

Based on Propositions 2 and 3, the following Theorem 1 addresses that the estimated diffuse chromaticity can approximate true diffuse chromaticity:

Theorem 1. Let I_H^* and I_S^* be corrected hue and saturation respectively of the observed image $I(\mathbf{x})$ in HSV space, respectively. If $I_H^* \approx I_{DH}$ and $I_S^* \approx I_{DS}$ where I_{DH} and I_{DS} are hue and saturation elements of diffuse reflection in HSV space, respectively, then

$$I_{DV}\Lambda^*(\mathbf{x}) \approx m_d(\mathbf{x})\Lambda(\mathbf{x}) \qquad (4.28)$$

where $\Lambda(\mathbf{x})$ is the diffuse chromaticity and $0 \leq I_{DV} \leq 1$ is the value element of diffuse reflection in HSV space.

Proof. Let the diffuse reflection vector be

$$I_D(\mathbf{x}) = (I_{Dr}, I_{Dg}, I_{Db}) = m_d(\mathbf{x})\Lambda(\mathbf{x}). \qquad (4.29)$$

Using HSV transformation we have

$$\varphi(I_{Dr}, I_{Dg}, I_{Db}) = (I_{DH}, I_{DS}, I_{DV}) \qquad (4.30)$$

From Proposition 3 it is known that $\varphi^{-1}(I_{DH}, I_{DS}, 1) = \{\varphi^{-1}(I_{DH}, I_{DS}, I_{DV})\}/I_{DV}$. Then

$$\varphi^{-1}(I_H^*, I_S^*, 1) \approx \{\varphi^{-1}(I_{DH}, I_{DS}, I_{DV})\}/I_{DV}.$$

Note that $\varphi^{-1}(I_{DH}, I_{DS}, I_{DV}) = \varphi^{-1}(\varphi(I_{Dr}, I_{Dg}, I_{Db})) = (I_{Dr}, I_{Dg}, I_{Db})$. Then

$$\varphi^{-1}(I_H^*, I_S^*, 1) \approx (I_{Dr}, I_{Dg}, I_{Db})/I_{DV}.$$

Using $\Lambda^*(\mathbf{x}) = \varphi^{-1}(I_H^*, I_S^*, 1)$ and (4.30), it is obtained that

$$I_{DV}\Lambda^*(\mathbf{x}) \approx (I_{Dr}, I_{Dg}, I_{Db}) = m_d(\mathbf{x})\Lambda(\mathbf{x}).$$

On the other side, because $I_D(\mathbf{x})$ and $I_S(\mathbf{x})$ are nonnegative, $0 \leq I(\mathbf{x}) \leq 1$ implies that $0 \leq I_D(\mathbf{x}) \leq 1$. That is, $0 \leq I_{Dr}, I_{Dg}, I_{Db} \leq 1$. Note that $I_{DV} = max\{I_{Dr}, I_{Dg}, I_{Db}\}$. Then $0 \leq I_{DV} \leq 1$.

$\square$

4.3.2 Algorithm implementation

To implement the proposed optimization method, (4.19) is reformulated as (4.31) of the following Theorem:

Theorem 2. Solving (4.19) is equivalent to solving

$$\begin{cases} m_d(\mathbf{x}) = P_{\Omega_1}[m_d(\mathbf{x}) - \frac{\partial E}{\partial m_d}] \\ m_s(\mathbf{x}) = P_{\Omega_2}[m_s(\mathbf{x}) - \frac{\partial E}{\partial m_s}] \\ \mathbf{p}(\mathbf{x}) = P_{\Omega_3}[\mathbf{p}(\mathbf{x}) - \frac{\partial E}{\partial \mathbf{p}}], \quad \frac{\partial E}{\partial \mathbf{q}} = 0 \end{cases} \quad (4.31)$$

where $\mathbf{p} \in R^2$ and $\mathbf{q} \in R^2$ are Lagrange multiplier vectors, and $P_{\Omega_1}(\cdot), P_{\Omega_2}(\cdot), P_{\Omega_3}(\cdot)$ are the projection operators on sets $\Omega_1 = \{m_d | 0 \leq m_d \leq 1\}$, $\Omega_2 = \{m_s | 0 \leq m_s \leq 1\}$, and $\Omega_3 = \{\mathbf{p} = (p_1, p_2) \,|\, |p_i| \leq 1 \ (i = 1, 2)\}$, respectively, and

$$P_{\Omega_i}(z) = arg \min_{u \in \Omega_i} \|u - z\|_2 \ (i = 1, 2, 3)$$

Proof. Note that $\mathbf{y}^T\mathbf{y}/2 = \max_{\mathbf{q}}\{\mathbf{q}^T\mathbf{y} - \mathbf{q}^T\mathbf{q}/2\}$ and $\|\mathbf{y}\|_1 = \max_{\|\mathbf{p}\|_\infty \leq 1}\{\mathbf{p}^T\mathbf{y}\}$ for any vector

$\mathbf{y} \in R^2$ where $\|\mathbf{p}\|_\infty = max\{|p_1|, |p_2|\}$. Then

$$E_2(m_d(\mathbf{x}), m_s(\mathbf{x})) = \tag{4.32}$$

$$\max_{\mathbf{p}, \mathbf{q}} \beta_1(\mathbf{q}^T \nabla m_d(\mathbf{x}) - \|\mathbf{q}\|_2^2/2) + \beta_2 \mathbf{p}^T \nabla m_s(\mathbf{x})$$

where $\mathbf{q}, \mathbf{p} \in R^2$ and $\|\mathbf{p}\|_\infty \leq 1$. Let

$$E(m_d(\mathbf{x}), m_s(\mathbf{x}), \mathbf{p}, \mathbf{q}) = E_1(m_d(\mathbf{x}), m_s(\mathbf{x})) \tag{4.33}$$

$$+\beta_1 \mathbf{q}^T \nabla m_d(\mathbf{x}) - \beta_1 \|\mathbf{q}\|_2^2/2 + \beta_2 \mathbf{p}^T \nabla m_s(\mathbf{x}).$$

Then solving (4.19) is equivalent to solving:

$$\min_{m_d, m_s} \max_{\mathbf{p}, \mathbf{q}} \ E(m_d(\mathbf{x}), m_s(\mathbf{x}), \mathbf{p}(\mathbf{x}), \mathbf{q}(\mathbf{x}))$$

$$s.t. \ \ (m_d(\mathbf{x}), m_s(\mathbf{x})) \in \Omega_1 \times \Omega_2 \tag{4.34}$$

$$(\mathbf{p}(\mathbf{x}), \mathbf{q}(\mathbf{x})) \in \Omega_3 \times R^2.$$

According to the saddle point theorem [4], $(m_d^*, m_s^*, \mathbf{p}^*, \mathbf{q}^*)$ is a global minimum point of (4.34) if and only if for any $(m_d, m_s, \mathbf{p}, \mathbf{q}) \in \Omega_1 \times \Omega_2 \times \Omega_3 \times R^2$

$$E(m_d^*, m_s^*, \mathbf{p}, \mathbf{q}) \leq E(m_d^*, m_s^*, \mathbf{p}^*, \mathbf{q}^*) \leq E(m_d, m_s, \mathbf{p}^*, \mathbf{q}^*). \tag{4.35}$$

From the right inequality of (4.35), (m_d^*, m_s^*) is a minimum point of $E(m_d, m_s, \mathbf{p}^*, \mathbf{q}^*)$. From [9] it follows that (m_d^*, m_s^*) satisfies:

$$(m_d - m_d^*)\frac{\partial E}{\partial m_d} \geq 0, \ \forall m_d \in \Omega_1. \tag{4.36}$$

and

$$(m_s - m_s^*)\frac{\partial E}{\partial m_s} \geq 0, \ \forall m_s \in \Omega_2. \tag{4.37}$$

Similarly, from the left inequality of (4.36), we see that $(\mathbf{p}^*, \mathbf{q}^*)$ is a minimum point of $E(m_d^*, m_s^*, \mathbf{p}, \mathbf{q})$.

It follows that

$$(\mathbf{p} - \mathbf{p}*)^T \frac{\partial E}{\partial \mathbf{p}} \geq 0, \quad \forall \mathbf{p} \in \Omega_3 \tag{4.38}$$

and

$$(\mathbf{q} - \mathbf{q}^*)^T \frac{\partial E}{\partial \mathbf{q}} \geq 0, \quad \forall \mathbf{q} \in R^2. \tag{4.39}$$

According to the projection Theorem [9], (4.36) equals $m_d(\mathbf{x}) = P_{\Omega_1}[m_d(\mathbf{x}) - \frac{\partial E}{\partial m_d}]$, (4.37) equals $m_s(\mathbf{x}) = P_{\Omega_2}[m_s(\mathbf{x}) - \frac{\partial E}{\partial m_s}]$, (4.38) equals $\mathbf{p}(\mathbf{x}) = P_{\Omega_3}[\mathbf{p}(\mathbf{x}) - \frac{\partial E}{\partial \mathbf{p}}]$, and (4.39) equals $\frac{\partial E}{\partial \mathbf{q}} = 0$.

Algorithm 3 for specular highlight removal

1: Input: specular highlight image $I(\mathbf{x})$
2: Compute $\{H(\mathbf{x}), S(\mathbf{x}), V(\mathbf{x})\}$ by HSV transformation
3: Compute highlight detection sets by (4.8) and (4.9)
4: Compute Hue estimate $H^*(\mathbf{x})$ by (4.11)
5: Compute Saturation estimate $S^*(\mathbf{x})$ by (4.12)
6: Compute diffuse chromaticity estimate by (4.15)
7: Compute $m_d^*(\mathbf{x})$ by updating algorithm (4.41)
8: Compute $I_D^*(\mathbf{x}) = m_d^*(\mathbf{x})\Lambda^*(\mathbf{x})$
9: Output: Specular-free image estimate $I_D^*(\mathbf{x})$

On one side, the gradient in (4.31) can be computed by

$$\frac{\partial E}{\partial \mathbf{p}} = \beta_2 \nabla m_s(\mathbf{x}), \quad \frac{\partial E}{\partial \mathbf{q}} = \beta_1 (\nabla m_d(\mathbf{x}) - \mathbf{q}(\mathbf{x})).$$

On the other side, the Gauss-Ostrogradsky theorem implies that

$$\mathbf{q}^T \nabla m_d(\mathbf{x}) = -m_d(\mathbf{x})div(\mathbf{q}), \mathbf{p}^T \nabla m_s(\mathbf{x}) = -m_s(\mathbf{x})div(\mathbf{p}) \tag{4.40}$$

where $div(\mathbf{p}(\mathbf{x})) = \partial p_1(\mathbf{x})/\partial x + \partial p_2(\mathbf{x})/\partial y$ is the divergence of vector field $\mathbf{p}(\mathbf{x})$. Then

$$\frac{\partial E}{\partial m_d} = (m_d(\mathbf{x})\Lambda^*(\mathbf{x}) + m_s(\mathbf{x})\Gamma^* - I(\mathbf{x}))^T \Lambda^*(\mathbf{x}) - \beta_1 div(\mathbf{q}(\mathbf{x}))$$

and

$$\frac{\partial E}{\partial m_s} = (m_d(\mathbf{x})\Lambda^*(\mathbf{x}) + m_s(\mathbf{x})\Gamma^* - I(\mathbf{x}))^T\Gamma^* - \beta_2 div(\mathbf{p}(\mathbf{x})).$$

Therefore, the following projection gradient algorithm for solving (4.19) is introduced by com-

bining the gradient representations above with (4.31):

$$
\begin{cases}
m_d^{k+1}(\mathbf{x}) = P_{\Omega_1}[m_d^k(\mathbf{x}) - h_k\frac{\partial E}{\partial m_d}(m_d^k(\mathbf{x}), m_s^k(\mathbf{x}), \mathbf{q}^k(\mathbf{x}))] \\[2mm]
m_s^{k+1}(\mathbf{x}) = P_{\Omega_2}[m_s^k(\mathbf{x}) - h_k\frac{\partial E}{\partial m_s}(m_d^k(\mathbf{x}), m_s^k(\mathbf{x}), \mathbf{p}^k(\mathbf{x}))] \\[2mm]
\mathbf{p}^{k+1}(\mathbf{x}) = P_{\Omega_3}[\mathbf{p}^k(\mathbf{x}) - h_k\beta_2\nabla m_s^k(\mathbf{x})] \\[2mm]
\mathbf{q}^{k+1}(\mathbf{x}) = \mathbf{q}^k(\mathbf{x}) - h_k\beta_1(\nabla m_d^k(\mathbf{x}) - \mathbf{q}^k(\mathbf{x}))
\end{cases}
\tag{4.41}
$$

where $h_k > 0$ is the step length. According to the analysis given in [4], the projection iterative

algorithm can converge globally to the optimal solution of (4.19) provided that the step length is

appropriately small. Finally, a specular highlight removal algorithm is described in Algorithm

3.

4.4 Results

In this section, the experimental results are presented to show the efficiency and effectiveness

of the proposed method on testing natural and medical images. For natural images, the ana-

lytic and benchmark images from previous studies [15] are used. For the medical images, the

benchmark images from the well-known TMI dataset [11] and real endoscopic images acquired

from the local hospital are used. For specular highlight removal testing, the proposed method

is compared with five dichromatic reflection model-based methods: Shen's [19], Yang's [25],

Suo's [22], Ren's [15], and Souza's [5] methods, and two inpainting-based methods: Saint-

Pierre's [16] and Bernal's methods [3]. The proposed algorithm for endoscopic images takes

design parameters $\alpha = 0.05$, $\beta_1 \in [1, 5]$, $\beta_2 = 0.1$, and $\tau \in [0.01, 0.1]$, and for natural images

takes design parameters $\alpha = 0.1$, $\tau = 0.01$, $\beta_1 = 3$, and $\beta_2 = 0.1$. Our algorithm is implemented

in MATLAB R2018b on a PC with a 2.60GHz Intel Pentium Dual Core Processor.

4.4.1 Natural image assessment

Analytic image testing

To analyze the proposed method quantitatively, the proposed method is compared with the state-of-the art methods on four natural synthetic images (Masks, Cups, Fruit, and Animals) that have known ground truth. Figs. 4.3(b)-(g) demonstrate highlight removal results by using seven methods. From the highlight removal results, it is seen that our results are visually similar to those of Shen [19], Yang [25], Suo [22], Ren [15], and Souza [45], and they are all better than Bernal's approach [3]. Furthermore, to quantitatively evaluate the highlight removal performance, the peak signal-to-noise ratio (PSNR) and structural similarity image measurement (SSIM) are considered as objective measures [23]. Table II lists the computed results by the seven methods, which shows that the proposed method has competitive performance for highlight removal and structural information preservation, compared with the other six approaches.

To explore the performance of the proposed method in the presence of noise, the proposed method and the other six methods are performed on the four natural images to which different types of noise: salt & pepper, speckle, and Gaussian noise, were added. PSNR was used as an evaluation metric. For the evaluation convenience, the average PSNR of four images before and after removal of specular reflection is taken. Let salt & pepper noise have noise density of 0.01, speckle noise have variance of 0.01, and Gaussian white noise have a standard deviation of $\sigma = 3$. Fig. 4.4(a) displays their comparative results. Let salt & pepper noise have noise density of 0.02, speckle noise have variance of 0.02, and Gaussian white noise have a standard deviation of $\sigma = 6$. Fig. 4.4(b) displays their comparative results. In Figs. 4.4(a) and 4.4(b), the proposed method has an improvement in the PSNR of the processed images for speckle and Gaussian noise. On the other hand, the proposed method's improvement is less than that achieved by Bernal's method due to the use of the median filter during inpainting process.

Table 4.1: Quantitative comparison of separation results

Images	Masks		Cups		Fruit		Animals	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
[3]	23.75	0.79	33.15	0.94	33.73	0.96	32.36	0.98
[5]	34.90	0.95	39.50	0.97	39.40	0.97	37.50	0.97
[19]	34.12	0.91	39.20	0.96	38.90	0.97	37.25	0.98
[25]	32.20	0.92	37.63	0.95	37.54	0.97	36.46	0.97
[15]	34.54	0.93	38.01	0.96	37.75	0.97	36.89	0.98
[22]	34.20	0.93	37.27	0.95	40.40	0.97	35.26	0.96
Our method	36.57	0.96	39.87	0.97	39.98	0.96	37.50	0.98

Benchmark image testing

To validate the proposed method, four benchmark natural images (Fish, Wood, Lotus, and Watermelon) are studied since these images have unknown ground truth [19]. Fig. 4.5 displays the highlight removal results obtained by the seven methods. In terms of visual inspection, our method produces promising results, which indicates its effectiveness on the natural scenes. Similar to the other five dichromatic reflection model-based methods [19], [25], [15], [22], [5], our method is comparable in highlight removal, colour recovery, and detail preservation. Compared with the inpainting-based method [3], the proposed method maintains the lightness and colour fidelity of image background, which is shown in the fish (first row) and the lotus (third row) images in Fig. 4.5.

4.4.2   Medical image assessment

Benchmark image testing

In this section, two groups of benchmark images with saturated highlights and complex textures from the TMI dataset [11] are studied. Fig. 4.6 demonstrates highlight removal results by five methods on five benchmark images. Highlight removal comparisons are shown in Fig. 4.6(b)-(f). As depicted in Fig. 4.6(b), our approach removes specular highlights from all endoscopic images and preserves the original texture details and colours on organ surfaces. From local

Figure 4.3: Highlight removal results for analytic natural images: (a) Input images, (b) Ours, (c) Ref. [19], (d) Ref. [15], (e) Ref. [25], (f) Ref. [3],(g) Ref. [22],(h) Ref. [5]

details shown in Fig. 4.6, it is seen that the other four methods are capable of removing specular highlights, but fail to preserve textures and colours. From Figs. 4.6(c) and (f), it is seen that Saint-Pierre's [16] and Bernal's methods [3] produce over-smooth images on the highlight regions. From Fig. 4.6 (d) and (e), it is seen that the methods of both Yang et al. [25] and Ren et al. [15] cause heavy degradation and colour distortion.

Fig. 4.7 demonstrates highlight removal results of other five benchmark images. Highlight removal comparisons are shown in Fig. 4.7(b)-(e). In Fig. 4.7(b), our method removes specular highlights from all endoscopic images and preserves the original texture details and colour on organ surfaces. Similarly, from Figs. 4.7(c) and (f) we see that both Saint-Pierre's and Bernal's methods result in overly smoothed images on the highlight regions. From Figs. 4.7 (d) and (e), we see that the methods of both Yang et al. [25] and Ren et al. [15] cause heavy degradation and colour distortion.

Finally, as for computation time, the computational time of our method on our machine is 35 secs for a 600x700 benchmark endoscopic image, while the times for the approaches of Ren et al. is 9 secs; Saint-Pierre et al. is 120secs; Bernal et al. is 20 secs, and Yang et al. is 20 secs.
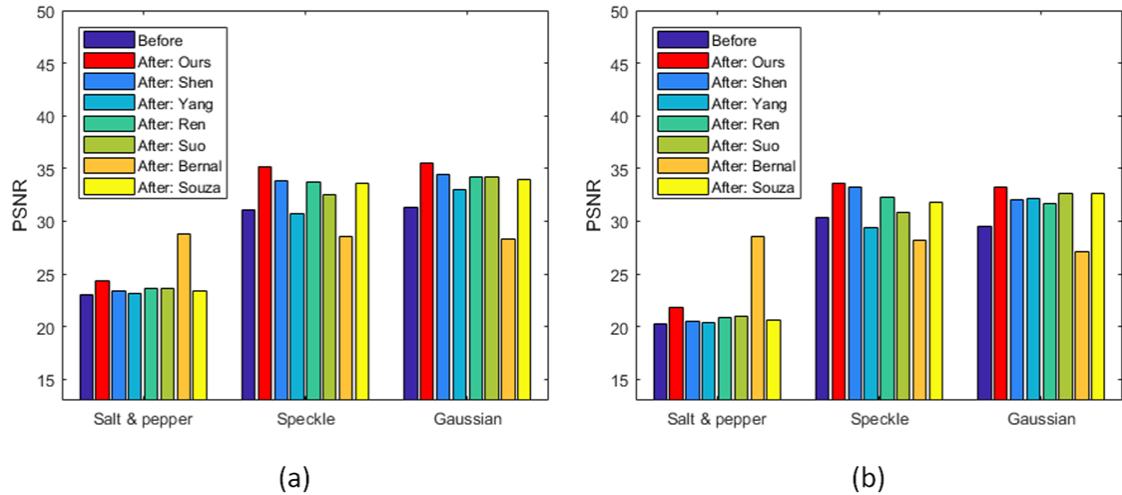
Figure 4.4:   Highlight removal performance comparison in noisy cases,(a) noise density = 0.01 for salt and pepper noise, noise variance = 0.01 for speckle noise, and $\sigma = 3$ for Gaussian noise, (b) noise density = 0.02 for salt and pepper noise, noise variance = 0.02 for speckle noise, and $\sigma = 6$ for Gaussian noise

Real image testing

In this section, specular highlight removal testing on real laparoscopic images with saturated highlights and complex textures acquired from endoscopic surgeries is studied. For a comparison, five real laparoscopic images are shown in Fig. 4.8(a). Figs. 4.8(b)-(f) display specular highlight removal results by five methods. It is seen that the proposed method removes specular highlights from all laparoscopic images and preserves original texture details and colour on organ surfaces shown in Fig. 4.8(b). In contrast, the saturated highlight regions are distorted by the methods of Shen [19], Yang [25], Ren [15], and Bernal [3], shown in Fig. 4.8(c), (d), (e), and (f) respectively.

Forced-choice preference testing

For further quantitative validation, a forced-choice preference experiment is performed on 20 specular highlight images acquired from real 2-D laparoscopic surgeries. Each image was processed by the proposed method and other four highlight removal methods [19] [15] [25] [3].
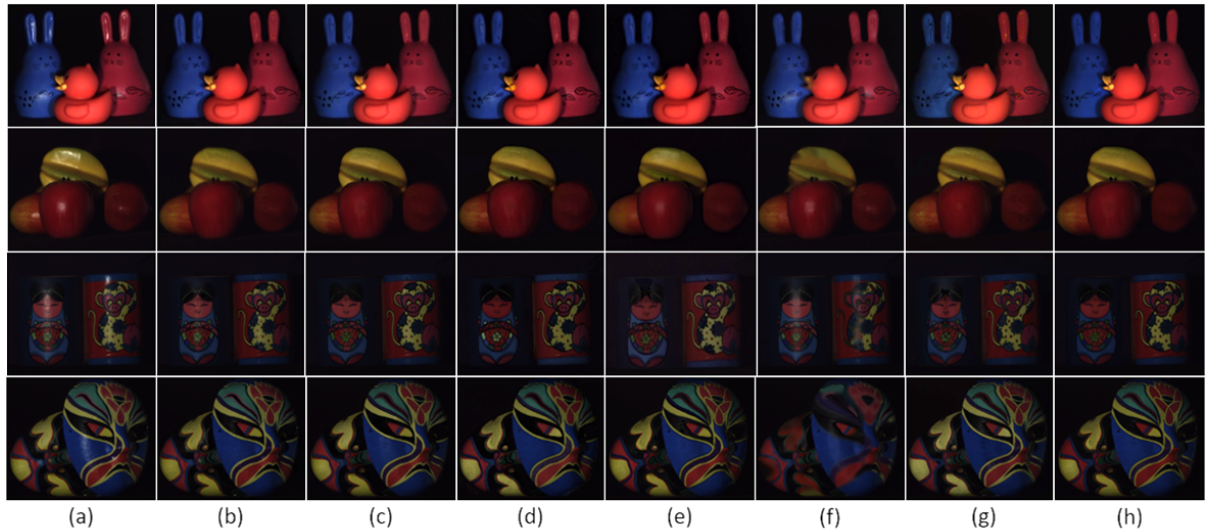
Figure 4.5: Highlight removal results for benchmark natural images: (a) Input images, (b) Ours, (c) Ref. [19], (d) Ref. [15], (e) Ref. [25], (f) Ref. [3],(g) Ref. [22],(h) Ref. [5]

Twenty sets of testing images were created, where each set included 5 processed images. This user study was performed under a protocol approved by the Research Ethics Board of Western University. Twenty clinicians with varying degrees of laparoscopy experience created the test group. The participants included 10 surgeons/fellows with over 5 years of experience and 10 residents in training with less than 5 years of experience. For each trial, two processed images randomly drawn from each image set were shown to the participants, who were blinded to the processing method, and were asked for their preferences. To reduce the number of trials, the previously preferred image remained in the next trial and was paired with an un-compared image from the image set. This process was repeated until all 5 images within the image set were compared and the last remaining preferred image was recorded. The above procedures were repeated for all 20 image sets, and the percentage of each algorithm with the most preferred choice was used as the user's preference score. According to the feedback from the clinical participants, the proposed approach produced the highest score, receiving an average approval rating of 76%, demonstrating significant favor from the participants, while the other four highlight removal methods received less than 24% approval rating.

Figure 4.6: Highlight removal results for benchmark endoscopic images: : (a) Input images, (b) Ours, (c) Ref. [16], (d) Ref. [15], (e) Ref. [25], (f) Ref. [3]

4.4.3   Application to stereo reconstruction

This section studies an application of the proposed method to stereo reconstruction. 20 benchmark endoscopic images from the "Distance dataset" with known ground truth [11] are tested. For a comparison, the cost volume filtering method [7] for stereo reconstruction is performed, based on the unprocessed dataset, the processed dataset by the proposed method, and by the inpainting-based method [3], respectively. The reconstructions based on the proposed method gave the average root mean square distance (RMSD) as 0.69 mm. In contrast, the reconstruction based on [3] yields an RMSD of 1.24 mm, and the unprocessed dataset 1.10 mm. Furthermore, Fig. 4.9 displays the comparison of disparity results, where Fig. 4.9(b) is the disparity map without highlight removal, Fig. 4.9(c) displays the result of following our highlight
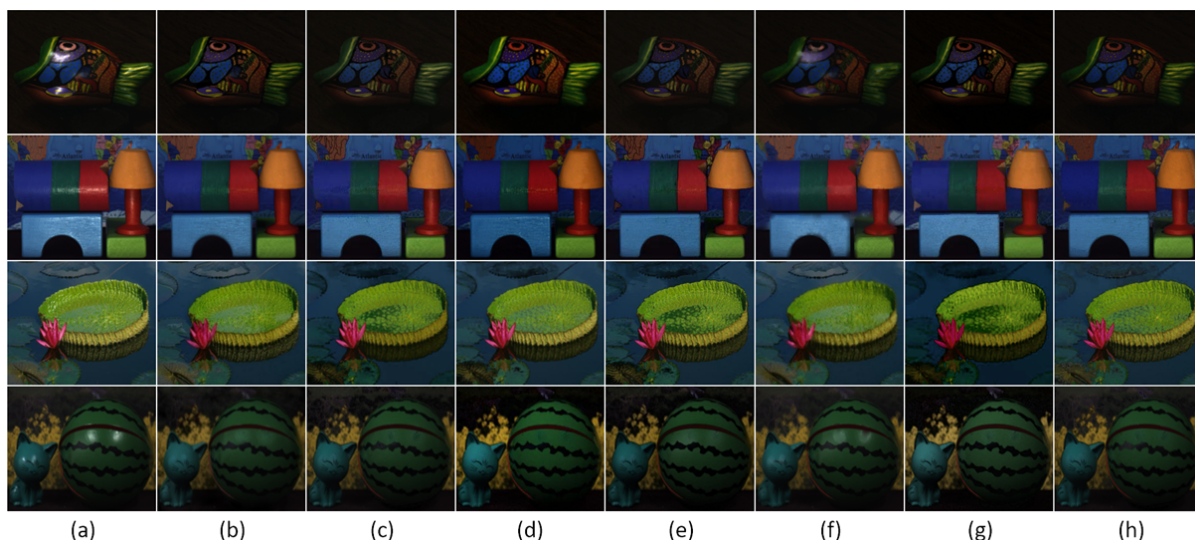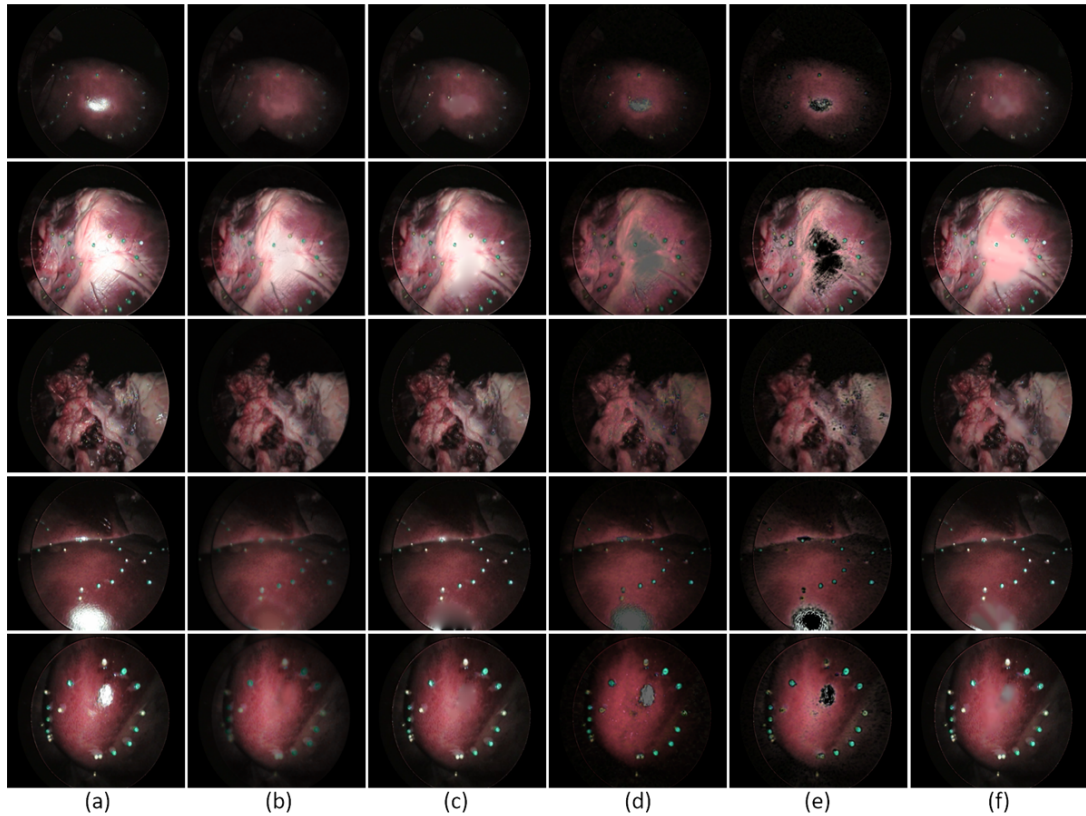
Figure 4.7: Highlight removal results for benchmark endoscopic images: (a) Input images, (b) Ours, (c) Ref. [16], (d) Ref. [15], (e) Ref. [25], (f) Ref. [3]

removal approach. Fig. 4.9(d) is the result of the following inpainting-based method. The red contours of the region of interest indicate that the proposed method gave a smooth disparity map, while others have nonsmooth disparity maps that may lead to inaccurate reconstruction of the 3D surface. It shows that the proposed highlight removal approach has the potential to significantly enhance stereoscopic reconstruction accuracy.

4.5 Discussion and conclusion

This chapter proposes a global optimization method for specular highlight removal from a single image, based on the dichromatic reflection model that is widely used in natural images. For effectiveness on medical images, this method takes advantage of the fact that such im-
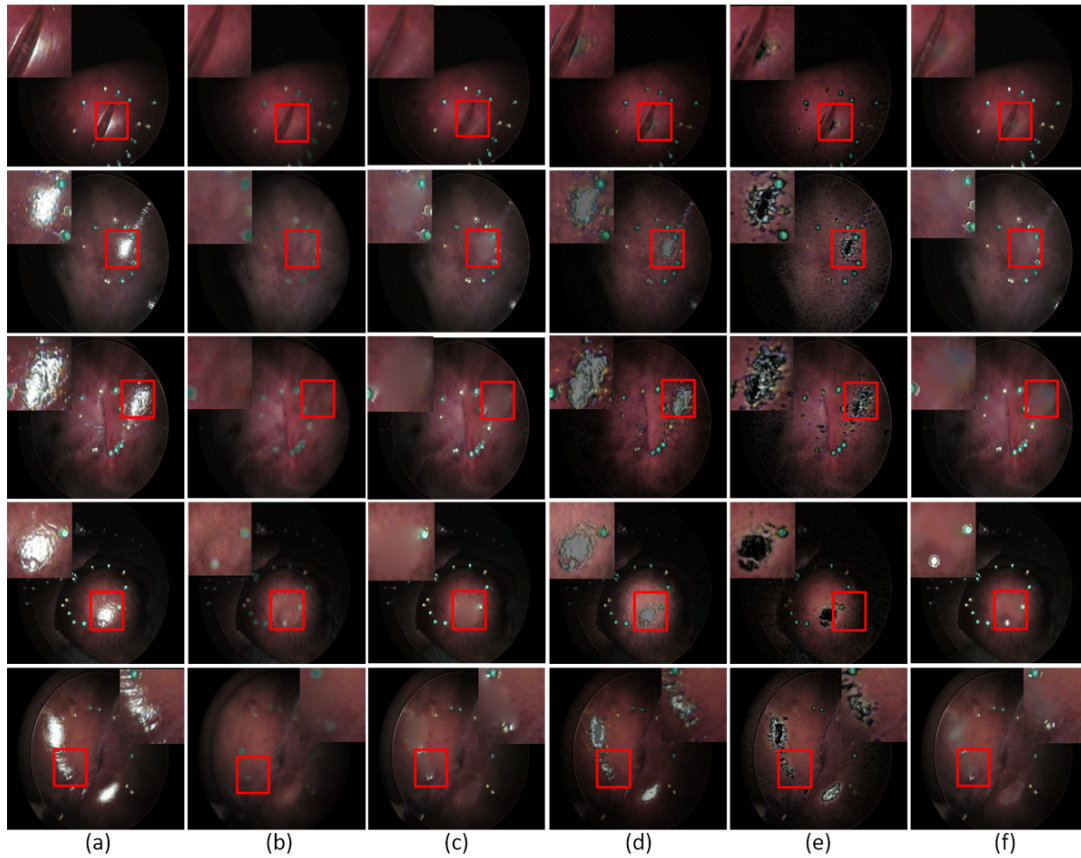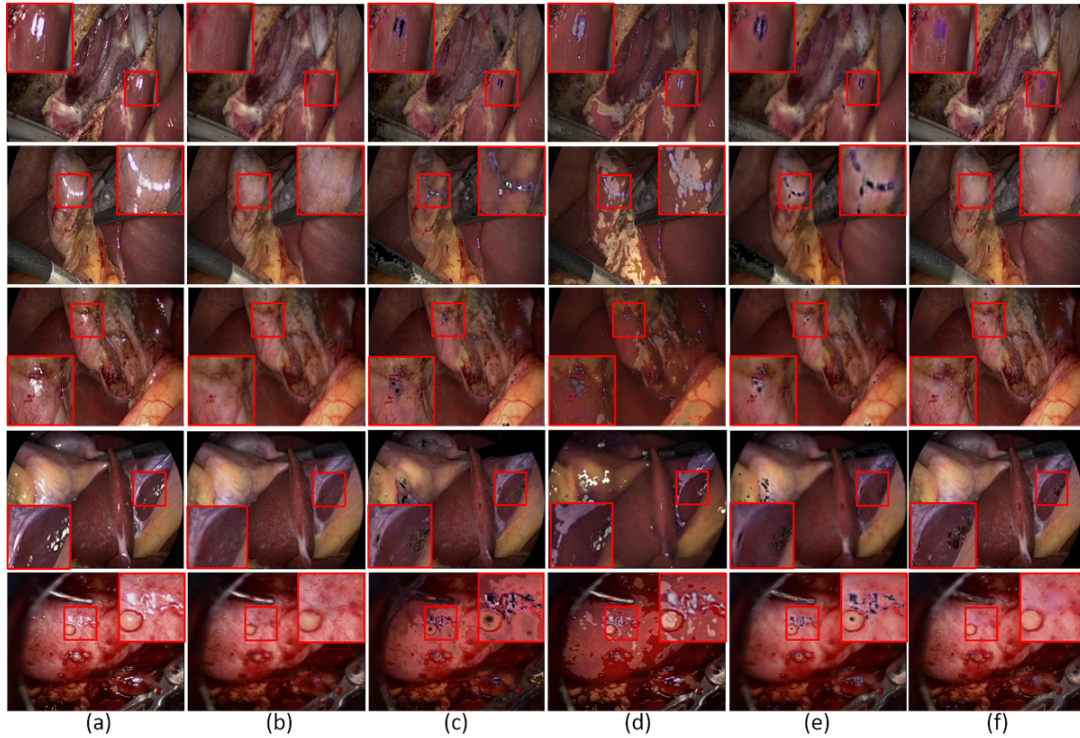
Figure 4.8: Highlight removal results for real laparoscopic images: (a) Input images, (b) Ours, (c) Ref. [19], (d) Ref. [15], (e) Ref. [25], (f) Ref. [3]

ages have very small colour variation due to the lack of blue and green components, which leads to smooth varying hue in HSV space. Also, the saturation layer is smoothly varying, and which doesn't contain any intensity edges in RGB space. Because of these properties, more accurate hue and saturation estimates are obtained by using an adaptive inpainting technique. Theoretical analysis shows that estimated diffuse chromaticity can approximate true diffuse chromaticity so that the diffuse refection estimation error can be minimized. Based on the estimated diffuse colour and modified illumination colour, the specular highlight removal problem is reformulated as a convex optimization problem with double regularization. In contrast, current optimization modeling-based methods for specular highlight removal may suffer from local minimum problem. Since two regularization terms can describe diffuse and specular component features, the proposed highlight removal method can not only remove highlights, but also preserve texture details. The quantitative experimental results demonstrate that the pro-
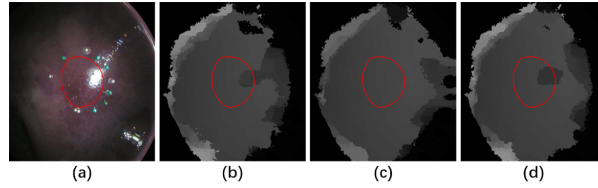
Figure 4.9: Disparity comparison results: (a) Input left image, (b) Disparity map without highlight removal, (c) Disparity map by our highlight removal algorithm (d) Disparity map by highlight removal method [3]

posed method is effective on benchmark and real endoscopic images as well as natural images, in terms of image quality and detail preservation, compared with several competing methods. Moreover, forced-choice preference experimental results confirm that the proposed method is accepted clinically. Furthermore, having effectively removed highlights from endoscopic images, the proposed method significantly enhances stereo-reconstruction surface accuracy and thus can provide surgeons more accurate depth information for automated surgical guidance.

According to the theoretical analysis, the effectiveness of the proposed algorithm depends on the colour correction, which may be less effective for regions with complex colour variations. Also, since the proposed method consists of both the detection technique and the optimization technique, it has relatively high computational cost, comparing to other non-optimization methods for specular highlight removal. Therefore, accelerating the proposed method is needed for real-time processing. Instead of l_1-norm optimization, l_2-norm optimization procedures could improve computational efficiency. In addition, temporal information may be helpful for colour correction and lead to enhancing accuracy of the proposed algorithm. These issues will be the subject of future research.

Bibliography

[1] Mirko Arnold, Anarta Ghosh, Stefan Ameling, and Gerard Lacey. Automatic segmentation and inpainting of specular highlights for endoscopic imaging. *EURASIP Journal on Image and Video Processing*, 2010:1–12, 2010.

[2] Jean-François Aujol, Guy Gilboa, Tony Chan, and Stanley Osher. Structure-texture image decompositionmodeling, algorithms, and parameter selection. *International journal of computer vision*, 67(1):111–136, 2006.

[3] Jorge Bernal, Javier Sánchez, and Fernando Vilarino. Impact of image preprocessing methods on polyp localization in colonoscopy frames. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7350–7354. IEEE, 2013.

[4] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[5] Antonio Carlos dos Santos Souza, Márcio Cerqueira de Farias Macedo, Verônica Paixão do Nascimento, and Bruno Santos Oliveira. Real-time high-quality specular highlight removal using efficient pixel clustering. In *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 56–63. IEEE, 2018.

[6] Jie Guo, Zuojian Zhou, and Limin Wang. Single image highlight removal with a sparse and low-rank reflection model. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 268–283, 2018.

[7] Asmaa Hosni, Christoph Rhemann, Michael Bleyer, Carsten Rother, and Margrit Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):504–511, 2012.

[8] Hyeongwoo Kim, Hailin Jin, Sunil Hadap, and Inso Kweon. Specular reflection separation using dark channel prior. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1460–1467, 2013.

[9] David Kinderlehrer and Guido Stampacchia. *An introduction to variational inequalities and their applications*. SIAM, 2000.

[10] Ranyang Li, Junjun Pan, Yaqing Si, Bin Yan, Yong Hu, and Hong Qin. Specular reflections removal for endoscopic image sequences with adaptive-rpca decomposition. *IEEE transactions on medical imaging*, 39(2):328–340, 2019.

[11] Lena Maier-Hein, Anja Groch, Adrien Bartoli, Sebastian Bodenstedt, G Boissonnat, P-L Chang, NT Clancy, Daniel S Elson, Sven Haase, Eric Heim, et al. Comparative validation of single-shot optical techniques for laparoscopic 3-d surface reconstruction. *IEEE transactions on medical imaging*, 33(10):1913–1930, 2014.

[12] Othmane Meslouhi, Mustapha Kardouchi, Hakim Allali, Taoufiq Gadi, and Yassir Benkaddour. Automatic detection and inpainting of specular reflections for colposcopic images. *Open Computer Science*, 1(3):341–354, 2011.

[13] JungHwan Oh, Sae Hwang, JeongKyu Lee, Wallapak Tavanapong, Johnny Wong, and Piet C de Groen. Informative frame classification for endoscopy video. *Medical Image Analysis*, 11(2):110–127, 2007.

[14] Long Quan, Heung-Yeung Shum, et al. Highlight removal by illumination-constrained inpainting. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 164–169. IEEE, 2003.

[15] Weihong Ren, Jiandong Tian, and Yandong Tang. Specular reflection separation with color-lines constraint. *IEEE Transactions on Image Processing*, 26(5):2327–2337, 2017.

[16] Charles-Auguste Saint-Pierre, Jonathan Boisvert, Guy Grimard, and Farida Cheriet. Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images. *Machine Vision and Applications*, 22(1):171–180, 2011.

[17] Hayden Schaeffer and Stanley Osher. A low patch-rank interpretation of texture. *SIAM Journal on Imaging Sciences*, 6(1):226–262, 2013.

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

[19] Hui-Liang Shen and Zhi-Huan Zheng. Real-time highlight removal using intensity ratio. *Applied optics*, 52(19):4483–4493, 2013.

[20] Alvy Ray Smith. Color gamut transform pairs. *ACM Siggraph Computer Graphics*, 12(3):12–19, 1978.

[21] Thomas Stehle. Removal of specular reflections in endoscopic images. *Acta Polytechnica*, 46(4), 2006.

[22] Jinli Suo, Dongsheng An, Xiangyang Ji, Haoqian Wang, and Qionghai Dai. Fast and high quality highlight removal from a single image. *IEEE Transactions on Image Processing*, 25(11):5441–5454, 2016.

[23] Robby T Tan, Ko Nishino, and Katsushi Ikeuchi. Separating reflection components based on chromaticity and noise analysis. *IEEE transactions on pattern analysis and machine intelligence*, 26(10):1373–1379, 2004.

[24] Xing Wei, Xiaobin Xu, Jiawei Zhang, and Yihong Gong. Specular highlight reduction with known surface geometry. *Computer Vision and Image Understanding*, 168:132–144, 2018.

[25] Qingxiong Yang, Jinhui Tang, and Narendra Ahuja. Efficient and robust specular highlight removal. *IEEE transactions on pattern analysis and machine intelligence*, 37(6):1304–1311, 2014.

[26] Qingxiong Yang, Shengnan Wang, Narendra Ahuja, and Ruigang Yang. A uniform framework for estimating illumination chromaticity, correspondence, and specular reflection. *IEEE Transactions on Image Processing*, 20(1):53–63, 2010.

[27] Yongqiang Zhao, Qunnie Peng, Jize Xue, and Seong G Kong. Specular reflection removal using local structural similarity and chromaticity consistency. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 3397–3401. IEEE, 2015.

Chapter 5

Stereo matching

This chapter includes material adapted from:

5.1   Introduction

In endoscopic/laparoscopic surgery, the surgeon is manipulating the tools while looking at the surgical scene displayed in the LCD monitor. Since the video displayed on the monitor is two-dimensional, the surgeons need to rely on their experience to accurately and safely perform the surgery due to lacking the depth information. Stereo matching is a common technique which can establish correspondences between images captured by the left and right cameras. Based on these correspondences, the distances between the corresponding left and right feature points can be computed, known as disparities. By using the estimated disparities, we can obtain the depth information. Therefore, stereo matching can provide depth information to surgeons that can increase the safety of surgical procedures. Many stereo matching methods have been reported to work well for natural images, but for laparoscopic images, they are still limited by illumination differences, weak texture content, specular highlights, and occlusions.

5.1.1 Related work

The optimization approach has the advantage of utilizing different regularization terms for ensuring a smooth disparity map with sub-pixel accuracy, and is another option for minimizing the disparity energy function. However, it is less popular in stereo matching, since the energy function is non-linear and non-convex. In other computer vision applications such as the optical flow, a similar variational model is employed and solved using the linearization technique. For example, Horn and Schunck [9] first presented an l_2 norm optimization method to find corresponding pixels between adjacent video frames in optical flow, and to improve robustness to non-Gaussian noise and and edge preservation, Zach et al. [23] presented a TV-L1 norm optimization method. Although these methods can be applied to stereo matching, they require 1) good disparity initialization, and 2) input images that have been equalized with respect to illumination for the linearization to be effective, especially for scenes with large disparities.

To provide an acceptable initialization, Revaud et al. [20] proposed to use a structure edge detection approach (SED) [21] for feature point detection and DeepMatching(DM) [6] for feature matching, to create an initialization to compute the optical flow. Bao et al. [2] proposed an improved non-local patch matching technique with a bilateral weighted matching cost for optical flow initialization, and recently, Lie et al. [14] proposed a sparse feature point-based disparity initialization method by using block-matching on SIFT and Canny features. However, these feature detection methods often fail to detect feature points on smooth tissues with weak textures, which leading to an inaccurate disparity map initialization. In addition, all existing feature matching methods fail in the presence of strong specular highlights.

Image processing steps are often applied to the images prior to stereo matching to remove streak-like artifacts while obtaining smooth disparity. These techniques include left-right consistency check [19] [13], super-pixel interpolation [5] [8], and weighted median filtering [10] [17].

5.1.2   Contributions

This chapter proposes a robust edge-preserving stereo matching method for laparoscopic image pairs, which integrates sparse feature matching, image illumination equalization, and disparity optimization. First, the sparse feature matching approach can effectively detect and match the feature points near image textures and edges, especially in the case of surgical tools, organ tissues, and surfaces with strong specular highlights, allowing an initial disparity map to be robustly estimated. Second, the illumination equalization approach can minimize the illumination differences between left and right laparoscopic images, thereby leading to enhancing the accuracy of the disparity map estimation. Third, by using the good disparity initialization and image illumination correction, an accurate disparity map can be estimated by a refined TV-L1 disparity optimization model.

5.2   Methods

For robust and accurate disparity estimation, this section proposes a novel stereo matching method that combines three techniques: sparse feature matching, illumination correction, and refined TV-L1 disparity optimization. Fig. 5.1 shows the flowchart of our proposed method for stereo matching laparoscopic image pairs.

5.2.1   Optimal sparse feature matching

Instead of conventional SIFT and Canny feature point detection [14], this section presents a feature point detection scheme based on image gradient changes on a given epipolar line:

$$x_{left} = \{x_i | \, |\frac{\partial^2 L_{left}(x_i)}{\partial x^2}| > \tau (i = 1...n)\}, \tag{5.1}$$

$$x_{right} = \{x_i | \, |\frac{\partial^2 L_{right}(x_i)}{\partial x^2}| > \tau (i = 1...m)\} \tag{5.2}$$

where x_{left} and x_{right} represent the set of detected feature points, and L_{left} and L_{right} represent pixels corresponding to horizontal lines in the left and right images respectively. Since the edges and textures are often associated with large gradient changes, it is reasonable to find these corresponding feature points by using second order derivative threshold. For an illustrative example, let the input left image be shown in Fig.5.2(a). Fig.5.2(b) plots extracted feature points from a horizontal scanline, based on (5.1) with the parameter $\tau = 0.3$. It is seen that these feature points are sparse and mainly distributed on maximum, minimum, and inflection points of the pixel intensity curve, denoted by asterisks. From Fig.5.2(b) it is observed that these feature points with four cases correspond to the edges and textures on the horizontal scanline of Fig.5.2(a). Moreover, because these points are sparse, the time to compute the initial dense disparity map can be greatly reduced.

Next, unlike the Birtchfield cost [3], we match feature points on the epipolar line by minimizing the sum of horizontal neighborhood differences of each pair of left and right feature points:

$$E(i, j) = \sum_{\Delta=0}^{\kappa} |\delta_{i,j}(\Delta)|^2 \tag{5.3}$$

where i and j are indices of the left and right feature points, $\Delta \in [0, \kappa]$ represents the neighborhood window of feature points, and $\delta_{ij}(\Delta)$ represents the horizontal distance between the neighborhoods of the feature points i and j, as illustrated in Fig.5.3(a) and Fig.5.3(b).

Following the work of Lucas and Kanade [15], the horizontal difference δ_{ij} is the minimizer of the following cost function:

$$C(\delta_{ij}) = \sum_{\Delta} [L_{left}(x_{left}(i) + \Delta + \delta_{ij}(\Delta)) \\ - L_{right}(x_{right}(j) + \Delta)]^2 + \delta_{ij}(\Delta)^2 \tag{5.4}$$

where $x_{left}(i)$ represents the ith detected feature point on the left scanline and $x_{right}(j)$ the jth such point on the right scanline. Using the linear approximation of L_{left}, the minimization of

Eqn. (5.4) can be converted into the following cost function:

$$\hat{C}(\delta_{ij}) = \sum_{\Delta}[L_{left}(x_{left}(i) + \Delta) + \delta_{ij}\frac{\partial L_{left}(x_{left}(i) + \Delta)}{\partial x}$$
$$- L_{right}(x_{right}(j) + \Delta)]^2 + \delta_{ij}(\Delta)^2. \tag{5.5}$$

Let $\partial\hat{C}/\partial\delta_{ij} = 0$. Then the horizontal difference can be approximately estimated by:

$$\delta_{ij}^* = \frac{L_{left}(x_{left}(i) + \Delta) - L_{right}(x_{right}(j) + \Delta)}{\frac{\partial L_{left}(x_{left}(i)+\Delta)}{\partial x})^2 + 1} \times \frac{\partial L_{left}(x_{left}(i) + \Delta)}{\partial x}. \tag{5.6}$$

From Eqn. (5.6), it follows the approximate expression of (5.8) as:

$$\hat{E}(i, j) = \sum_{\Delta=0}^{\kappa} |\delta_{i,j}^*(\Delta)|^2 \tag{5.7}$$

Finally, the two horizontal difference costs with different one-sided windows are fused to avoid the occlusion and specular highlights. For the positive window $\Delta^+ \in [0, \kappa]$, a positive cost is defined as:

$$\hat{E}^+(i, j) = \sum_{\Delta^+=0}^{\kappa} |\delta_{i,j}^*(\Delta^+)|^2, \tag{5.8}$$

while for the negative window $\Delta^- \in [-\kappa, 0]$, a negative cost is defined as:

$$\hat{E}^-(i, j) = \sum_{\Delta^-=-\kappa}^{0} |\delta_{i,j}^*(\Delta^-)|^2 \tag{5.9}$$

Since the occlusion can only occur on either the positive or negative side, combining Eqn. (5.8) and Eqn. (5.9) together can obtain the optimal matching cost of the feature points:

$$E^*(i, j) = min(\hat{E}^+(i, j), \hat{E}^-(i, j)) + \alpha|\hat{E}^+(i, j) - \hat{E}^-(i, j)| \tag{5.10}$$

,where $|\cdot|$ denotes absolute value.

For each sparse feature point on the left scanline denoted by $x_{left}(i)$, the optimal matched feature point on the right scanline is given by:

$$x_{right}(j_i^*) = x_{right}(arg \ \min_j(E^*(i, j)))$$
(5.11)

The disparity for each feature point on the left scanline is estimated by:

$$d^*(i) = x_{right}(j_i^*) - x_{left}(i).$$
(5.12)

As an illustrative example, Fig.5.4(b) plots optimal matching result of sparse feature points on the right scanline of input left and right images.

5.2.2   Illumination equalization and correction

Illumination differences between left and right images occur frequently in various environments and thus deteriorate the quality of stereo image pairs. For robustness to illumination differences, most of current stereo matching methods use similarity measure, colour restoration, and illumination enhancement techniques [16]. However, these approaches are not guaranteed to equalize illumination between left and right images.

To enhance the accuracy of the disparity optimization, an illumination equalization technique is proposed to correct illumination differences between left and right images. Let $(x_{left}(i), y(i))$ and $(x_{right}(i), y(i))$ denote a pair of matched feature points on left and right images, respectively, and let $\{d^*(i)\}$ denote the corresponding sparse disparity. Then a sparse illumination correction factor can be computed by:

$$s(\mathbf{x}_{left}(i), \mathbf{y}(i)) = \frac{I_{right}(\mathbf{x}_{left}(i) + d^*(i), \mathbf{y}(i))}{I_{left}(\mathbf{x}_{left}(i), \mathbf{y}(i))}$$
(5.13)

Based on Eqn. (5.13), a left image with illumination correction at matched feature points can

be built as:

$$\bar{I}_{left}(\mathbf{x}_{left}(i), \mathbf{y}(i)) = s(\mathbf{x}_{left}(i), \mathbf{y}(i))I_{left}(\mathbf{x}_{left}(i), \mathbf{y}(i)) \tag{5.14}$$

Substituting Eqn. (5.13) into Eqn. (5.14), it follows that

$$\bar{I}_{left}(\mathbf{x}_{left}(i), \mathbf{y}(i)) = I_{right}(x_{right}(j_i^*), \mathbf{y}(i)). \tag{5.15}$$

Thus, the proposed illumination equalization approach minimizes the illumination difference between left and right images at the matched feature points.

5.2.3   Initial disparity estimate and refined left image

The initial disparity values at matched feature points can be obtained by employing the technique outlined in Eqn. (5.12), while the unknown disparities of other pixels can be linearly interpolated from the disparities of the matched feature points. Furthermore, the initial disparity map can be refined using the weighted median filter [11]:

$$D_0 = \begin{cases} D_{[\sum_{i=l}^{l} \hat{r}_i + 1]/2} & \text{if } \sum_{i=l}^{l} \hat{r}_i \text{ is odd} \\ [D_{\sum_{i=l}^{l} \hat{r}_i/2+1} + D_{\sum_{i=l}^{l} \hat{r}_i/2}]/2 & \text{if } \sum_{i=l}^{l} \hat{r}_i \text{ is even} \end{cases} \tag{5.16}$$

where $\hat{r}_i$ denotes the sorted bilateral weight vector based on unsorted weight function in a given window, expressed as:

$$\begin{aligned} r(x + \eta_x, y + \eta_y) &= exp(-\frac{|\eta_x|^2 + |\eta_y|^2}{\sigma_s^2}) \\ &\quad exp(-\frac{\|I_{left}(x + \eta_x, y + \eta_y) - I_{left}(x, y)\|_2^2}{\sigma_r^2}) \end{aligned} \tag{5.17}$$

where $\| \cdot \|_2$ denotes l_2 norm, $|\eta_x| \leq \kappa$ and $|\eta_y| \leq \kappa$, κ represents the radius centered at location (x, y), and σ_s and σ_r are the spatial and range influence parameters, respectively. Usually, σ_s is chosen as half of the filter window size and σ_r is chosen as 0.01.

Based on Eqn. (5.14), a refined left image with illumination correction can be given by:

$$\bar{I}_{left}(\mathbf{x}, \mathbf{y}) = \bar{s}(\mathbf{x}, \mathbf{y}) I_{left}(\mathbf{x}, \mathbf{y}) \tag{5.18}$$

where $\bar{s}(\mathbf{x}, \mathbf{y})$ is the dense illumination factor created by linear interpolation and smoothed by the weighted median filter.
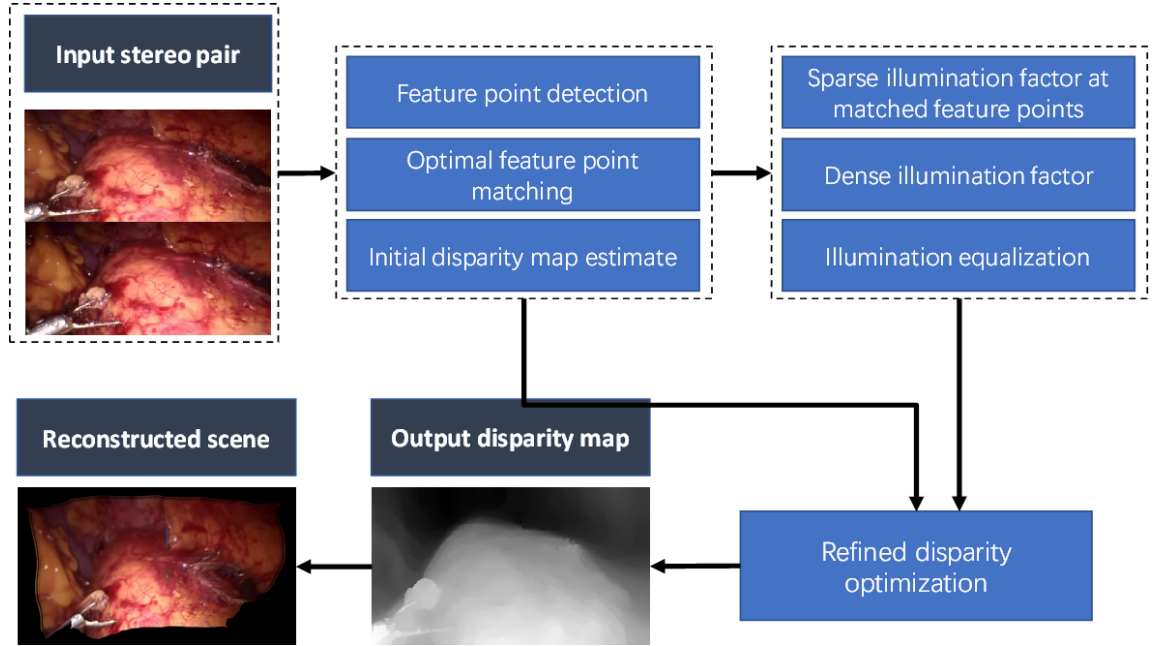
Figure 5.1: The flowchart of the proposed stereo matching method for disparity map estimation and surface reconstruction

5.2.4   Refined TV-L1 disparity optimization

Total variational (TV) optimization methods have been successfully used to calculate the optical flow between two image frames. For the robustness against non-Gaussian noise and and edge preservation, Zach et al. [23] presented a TV-L1 norm optimization method which solves:

$$\min_{D} \int_{\Omega} (|I_{left}(\mathbf{z}) - I_{right}(\mathbf{z} + D(\mathbf{z}))| + |\nabla D|) d\mathbf{z}. \tag{5.19}$$
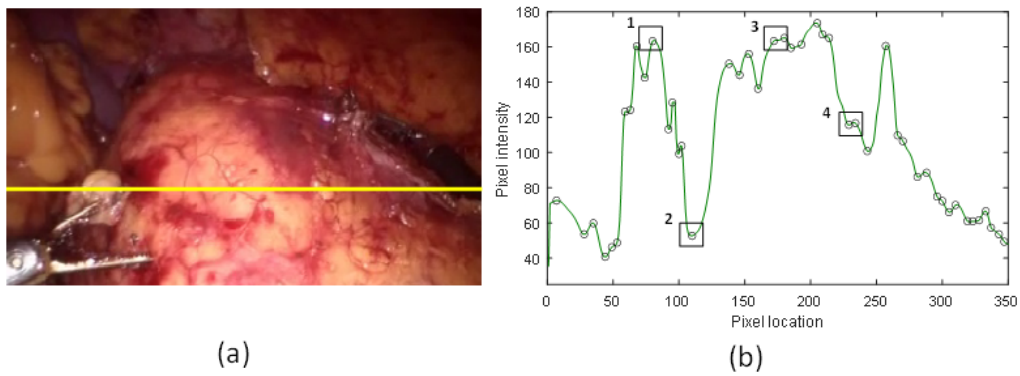
Figure 5.2:   An example of detected feature points: (a) input left image (b) extracted feature points on horizontal scanline.
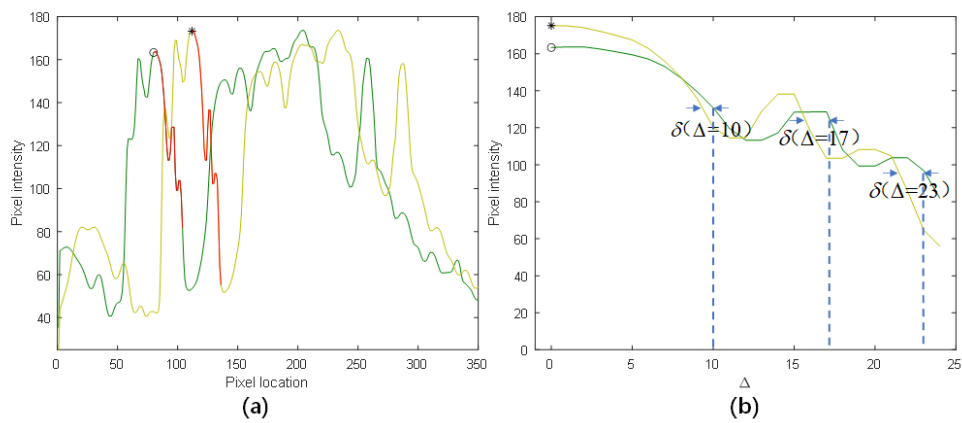
Figure 5.3:   (a) an example of the detected feature points, marked as the circle and asterisk, and their neighborhoods, marked in red, (b) an illustration of the horizontal differences between the neighborhoods of the left and right feature points.
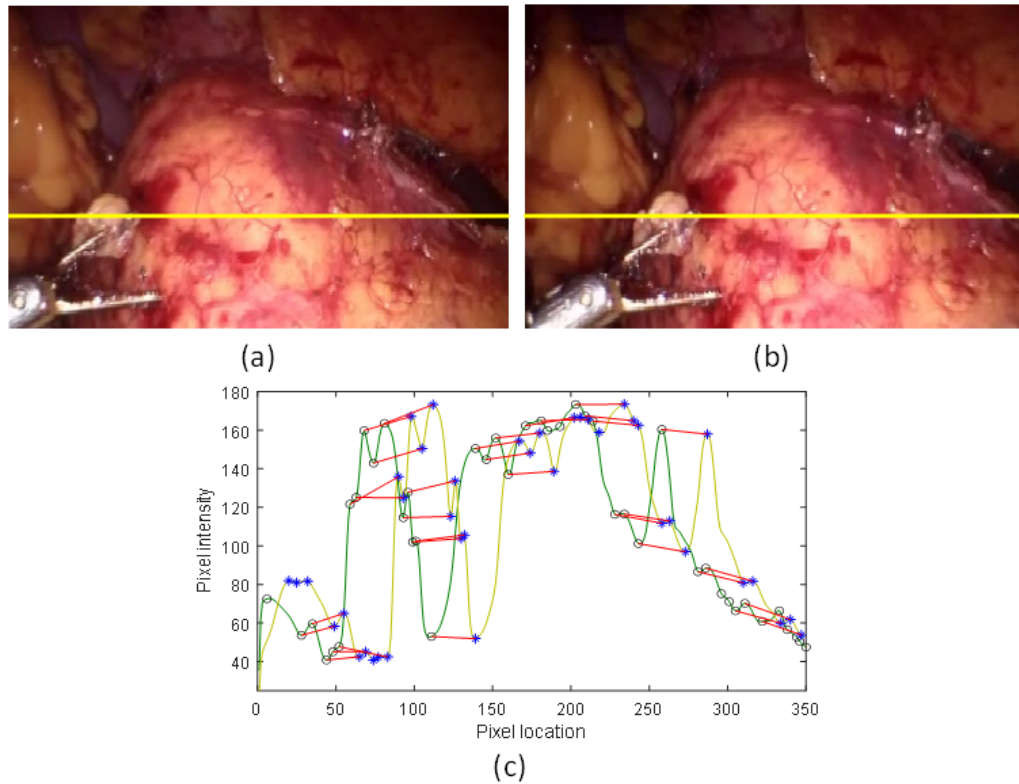
Figure 5.4:   An example of sparse feature point matching: (a) input left image (b) input left image (c) matched feature point on horizontal scanline

where $D(\mathbf{z})$ is the disparity map to be estimated, ∇D is the gradient of D, and $I_{left}(\mathbf{z})$ and $I_{right}(\mathbf{z})$ are the left and right images at $\mathbf{z} = (x, y)$.

Because the problem Eqn. (5.19) is nonlinear and nonconvex, a linear approximation for $I_{right}(\mathbf{z} + D(\mathbf{z}))$ is necessary. As a result, a good initial disparity map and equalized illumination are usually necessary to rapidly compute an optimal accurate disparity. By integrating our illumination correction and initial disparity map into the TV-L1 optimization process, this section proposes a refined disparity optimization method, which solves:

$$\min_D \int_\Omega (|\hat{I}_{left}(\mathbf{z}) - I_{right}(\mathbf{z} + D(\mathbf{z}))| + |\nabla D|)d\mathbf{z} \tag{5.20}$$

where $\hat{I}_{left}(\mathbf{z})$ is the refined left image defined in (5.16). The corresponding discrete TV-L1 optimization problem with illumination correction:

$$\min_D \sum_{\mathbf{z} \in \Omega} (|\hat{I}_{left}(\mathbf{z}) - I_{right}(\mathbf{z} + D(\mathbf{z}))| + |\nabla D|) \tag{5.21}$$

Since Eqn. (5.14) can be effectively solved by using solution approach [23], the disparity map can be well estimated.

As an illustrative example of the refined disparity optimization method, the endoscopic image pairs shown in Fig. 5.5(a) are considered. The disparity optimization problem defined in Eqn. (5.21) is solved by using different initial disparity maps given by three methods: the proposed sparse matching approach, the SIFT operator-based sparse matching approach [14], and the CVF approach [11], respectively. Fig. 5.5(b), Fig. 5.5(c), and Fig. 5.5(d) display different initial disparity maps by these three approaches, respectively. We see that our method produces the best disparity initialization, especially in regions of specular highlight, while Fig.5.5(e), Fig.5.5(f), and Fig.5.5(g) displays three final disparity maps by solving (5.21) with the corresponding initial disparity maps. From Fig.5.5, it is clear that our initialization yields a high quality optimized disparity map, compared to the results that are based on initializations from SIFT and CVF approaches that result in large disparity errors in the highlighted region.

As another illustrative example, the endoscopic image pairs shown in Fig. 5.6(a) and (b) are considered. Fig. 5.6 displays disparity optimization results with illumination correction, where Fig. 5.6(d) is a disparity map obtained by solving Eqn. (5.21) with illumination correction and (c) is a disparity map obtained by solving Eqn. (5.19) without illumination correction. It us clear that illumination correction improves the smoothness and accuracy of the final disparity map.

Finally, combining the proposed sparse feature matching, illumination correction, and refined disparity optimization steps, the proposed stereo matching algorithm for disparity map estimation is summarized in Algorithm 4 and Fig. 5.1 displays its workflow.

Algorithm 4 for disparity map estimation

- Input: image pairs $I_{left}(\mathbf{z})$ and $I_{right}(\mathbf{z})$ and parameters: $\tau, \alpha, \kappa, \sigma_r$

- Compute left and right feature points by (5.1) and (5.2)

- Compute positive and negative matching costs by (5.8) and (5.9)

- Compute fused matching cost by (5.10)

- Compute matched feature point pairs by (5.11)

- Compute matched sparse disparity by (5.12)

- Compute dense illumination correction factor by (5.13)

- Compute refined left image by (5.18)

- Compute dense disparity by solving (5.21)

- Output: disparity map estimate

5.3   Results

This section illustrates the performance of the proposed stereo matching method, by carrying out experiments on both biological phantoms and real laparoscopic surgical images. Our algorithm is implemented in MATLAB R2018b executed on a PC with a 2.60GHz In-
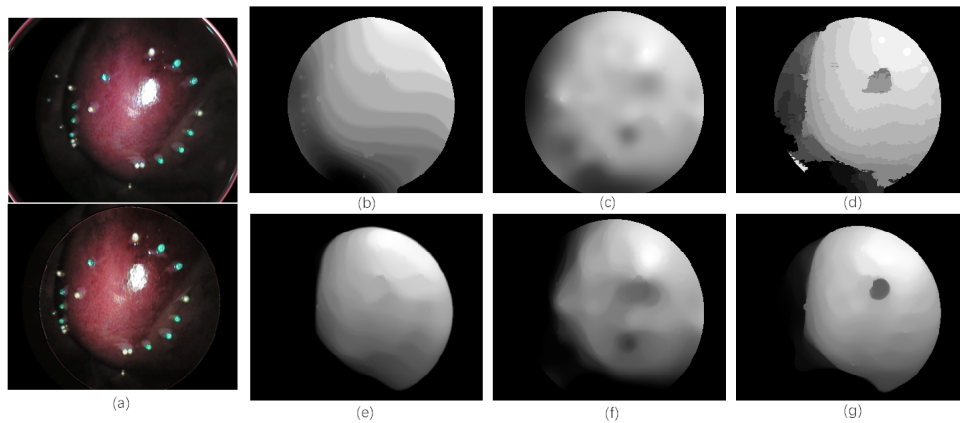
Figure 5.5:   An example of disparity map results resulted from different initial disparity maps by three stereo matching approaches. (a) Input image, (b) Our initial disparity map, (c) Initial disparity map by SIFT [14], (d) Initial disparity map by CVF [11], (e) Our optimized disparity map (f) Optimized disparity map by SIFT, (g) Optimized disparity map by CVF.
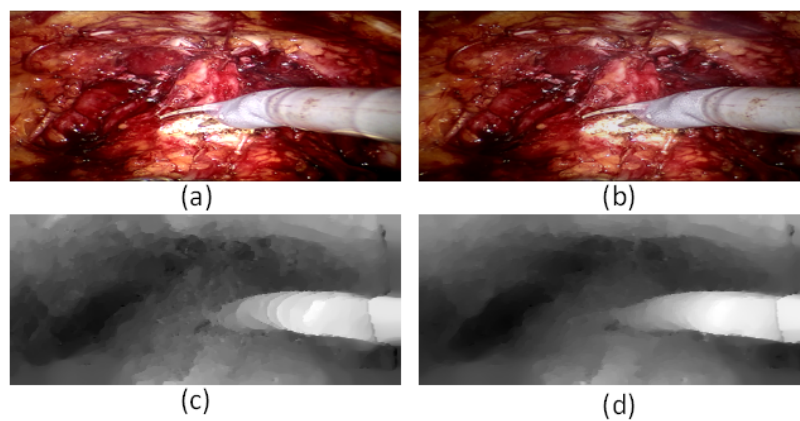
Figure 5.6:   An example of disparity optimization with illumination correction. (a) Input left image, (b) Input right image, (c) Disparity map with illumination correction, (d) Disparity map without illumination correction.

tel Pentium Dual Core Processor. Our experiments are based on the following parameters: $\tau = 0.5, \alpha = 0.1, \kappa = 25$, and $\sigma_r = 0.01$. Although these parameters are not optimized analytically, they are sufficient and effective for most images. The subjective and objective assessments are used for the verification of the proposed method.

5.3.1 Subjective assessment

First, it is tested that three laparoscopic video images acquired from robotic-assisted laparoscopic radical prostatectomy collected using the da Vinci Si robotic surgical system. The proposed stereo matching approach is compared with currently available stereo matching approaches: (c) Bao et al. [2] (d) Chang et al. [4] (e) Yang et al. [22] (f) Hosini et al. [11] (g) Khan et al. [12] (h) Heo et al. [7] (i) Zinner et al. [24].

Fig. 5.7(b)-Fig. 5.7(i) displays the disparity maps estimated by the eight different approaches above, applied to the laparoscopic image shown in Fig. 5.7(a). Fig. 5.7 demonstrates qualitatively that the proposed method produces the best disparity map in terms of overall smoothness and edge sharpness. In comparison, the disparity maps from Fig. 5.7(d) -Fig. 5.7(i) are unable to accurately outline the surgical tool. In addition, as a result of occlusions, there are distinct bad pixels and disparity errors near the edges, and while Bao's result given in Fig. 5.7(c) is comparable with ours, there are nevertheless erroneous disparity inconsistencies on smooth textured surfaces. Fig. 5.8(b)-Fig. 5.8(i) display the reconstructed surgical scene obtained using disparity maps obtained by the competing stereo correspondence methods. From Fig. 5.8 it is seen that the proposed method yields a visually smooth and intuitive reconstruction of the surgical scene. In comparison, reconstructions from methods shown in Fig. 5.8(c)-Fig. 5.8(i) are not smooth and contain various surface errors related directly to the disparity errors introduces by each of these approaches.

Fig. 5.9(b)-Fig. 5.9(i) and Fig. 5.11(b)-Fig. 5.11(i) demonstrate more examples of disparity map comparisons amongst the eight different approaches. From both Fig. 5.9(b) and Fig. 5.11(b), the proposed method produces superior disparity maps in terms of overall smooth-

ness, edge sharpness, and the delineation of thin structures. In addition, the depth differences between surgical tools and tissues can be clearly visualized. In comparison, the disparity maps from the other methods have distinct disparity errors and unclear outlines of the tools and tissues. In Fig. 5.9(c) and Fig. 5.11(c), Bao's results demonstrate good accuracy for soft tissue but also large erroneous disparity values on the surgical tools, possibly due to textureless surfaces. Similarly, Fig. 5.10 and Fig. 5.12 display the surface reconstruction results corresponding to their disparity maps in Fig. 5.9 and Fig. 5.11, respectively. From Fig. 5.10(b) and Fig. 5.12(b), it is observed that the proposed method produces visually smooth and intuitive reconstructions of the surgical scenes. In comparison, reconstructions from the other methods contain various surface errors, such as spikes and holes, especially on regions containing the textureless surgical tool. The borders of the reconstructed surgical scene seem to be distorted, causing the image to appear unnatural.

To demonstrate that the proposed method is also effective on laparoscopic video images with large specular highlights, a laparoscopic video image from MICCAI 2019 challenge[1], shown in Fig.5.13(a) with a large specular highlight on the organ surface, is studied. Fig.5.13(b)-Fig.5.13(i) display the disparity maps estimated by the eight different approaches discussed in the above table when applied to this image. It is seen that the proposed method results in a smooth disparity map with sharp edge, despite the large specular highlight on the organ surface. From Fig.5.13(c),Fig.5.13(d), and Fig.5.13(f)-Fig.5.13(i), it is observed that the estimated disparity maps have different degrees of degradation near the highlight region. While Fig.5.13(e) demonstrates a relatively good disparity estimation in the highlight region, the disparity map itself is not smooth, leading to inaccurate surface reconstruction. Furthermore, Fig.5.14 displays the surface reconstruction results corresponding to their disparity maps in Fig.5.13. It is easy to see that the proposed method can accurately reconstruct the shape and the surroundings of the porcine kidney, especially on the highlighted surface. In comparison, Fig.5.14(c),Fig.5.14(d), and Fig.5.14(f)-Fig.5.14(i) all result in distinct reconstruction errors on the highlighted surface. For example, Fig.5.14(e) shows a kidney surface that is distorted in the region of the specular

Table 5.1: Quantitative comparison of SSI

Method	Ours	Bao [2]	Yang [22]	Hosni [11]	Chang [4]	Khan [12]	Heo [7]	Zinner [24]
Fig.5.7(a)	0.92	0.92	0.84	0.87	0.79	0.84	0.85	0.84
Fig.5.9(a)	0.91	0.90	083	0.83	0.77	0.82	0.83	0.83
Fig.5.11(a)	0.88	0.80	0.75	0.74	0.70	0.74	0.74	0.72
Fig.5.13(a)	0.91	0.90	0.89	0.86	0.88	0.89	0.88	0.88

reflection with several small holes.

Finally, four examples of 3D reconstructed surgical scene results described above can be viewed from link: https://uwoca-my.sharepoint.com/:f:/g/personal/wxia43_uwo_ca/EitUts4Qp8xOkvoI1NIC-ysBCHktbQ0fRHpgIVqSpQXe0Q?e=yLPxcy

5.3.2 Objective assessment

First, the processed real endoscopic stereo images discussed in the previous section is analyzed quantitatively. Since no ground truth is available for the surgical images, the quality of the estimated disparity maps is evaluated by using the structural similarity index (SSI) value on [0, 1] between the left image and the warped right image. The closer the SSI is to 1, the better the disparity map is estimated. Tables I lists a quantitative comparison of blind SSI, which demonstrates that the proposed method is superior to the other approaches.

Second, this quantitative testing is performed on 40 benchmark endoscopic stereo images from the "TMI Distance Dataset" and "TMI Angle Dataset" with ground truth surface reconstructions [18], where 20 stereo images with varying orientations are taken from the Angle dataset and other 20 stereo images with varying distances of the endoscope relative to the organ surfaces are taken from the Distance dataset.

For quantitative comparison, the average absolute surface reconstruction error measure [18] is employed. Table II lists the results of the surface reconstruction errors from eight different methods on the 20 images from the Distance dataset, where the surface reconstruction error is measured in millimeters (mm). Table III lists the results of surface reconstruction errors from the eight different methods on the 20 images from the Angle dataset. From Tables II and III, it

Table 5.2: Comparison of reconstruction error using Distance Dataset (mm)

Method	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8	Image 9	Image 10
Ours	0.51	0.57	0.43	0.27	0.50	0.48	0.50	1.3	0.44	1.09
Bao [2]	5.35	1.80	2.83	0.84	1.14	4.81	0.33	1.7	2.41	0.91
Hosni [11]	2.48	0.62	0.23	1.13	0.32	1.89	0.94	0.63	1.72	1.89
Chang [4]	2.01	0.94	1.04	1.42	1.77	2.67	0.84	2.22	7.20	1.71
Yang [22]	3.89	0.92	1.14	0.43	1.05	3.12	0.80	2.50	6.01	1.60
Khan [12]	0.59	0.86	0.69	0.97	0.85	1.08	0.89	1.29	2.44	1.31
Heo [7]	0.99	0.93	1.36	1.01	0.66	3.25	1.22	1.23	1.72	1.24
Zinner [24]	0.64	0.75	0.72	0.75	1.56	2.78	0.79	1.39	1.45	2.25
Method	Image 11	Image 12	Image 13	Image 14	Image 15	Image 16	Image 17	Image 18	Image 19	Image 20
Ours	0.89	0.91	0.67	1.38	0.83	1.20	0.75	0.36	0.71	0.61
Bao [2]	1.43	1.03	1.27	4.12	2.02	3.35	0.91	0.49	0.53	0.85
Hosni [11]	0.95	0.97	0.64	0.89	1.93	3.33	1.29	0.48	1.13	0.73
Chang [4]	1.21	1.00	1.05	2.07	1.96	2.02	1.35	1.12	1.04	1.15
Yang [22]	0.99	1.07	0.82	1.87	3.04	1.47	0.99	0.84	1.18	1.13
Khan [12]	1.58	1.17	1.25	1.86	1.49	2.54	1.50	1.04	1.44	1.45
Heo [7]	1.66	1.52	2.11	2.32	1.55	2.29	1.44	1.12	1.25	2.09
Zinner [24]	1.67	1.3	1.23	1.88	1.78	2.08	1.52	1.08	1.33	1.69

is observed that the proposed method results in smaller surface reconstruction errors than the competing techniques in most cases.

Furthermore, Fig. 5.15 displays the box-plot results of surface reconstruction error on the 20 images of Distance dataset, including median, first and third quartiles, minimum and maximum values. Similarly, Fig. 5.16 displays the box-plot result of surface reconstruction error on on the 20 images of the Angle dataset.

From both Fig.5.15 and Fig.5.16, it is seen that our method has the lowest median value for both D-Angle and D-Distance data sets among eight methods. It demonstrates that the proposed method is superior to other methods in terms of the reconstruction accuracy. The proposed method also has the lowest first and third quartiles and thus has a robust performance. Moreover, the maximum distance of the proposed method is far lower than the other approaches, which further demonstrates that this new method can achieve a good precision.

5.4   Discussion and conclusion

This chapter proposes a robust and accurate stereo matching method for reconstructing surfaces from stereoscopic laparoscopic image pairs. To achieve this goal, three techniques: fast sparse feature matching, image illumination equalization, and refined TV-L1 disparity opti-

Table 5.3: Comparison of reconstruction error using Angle Dataset (mm)

| Method	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8	Image 9	Image 10
Ours	0.89	0.91	0.67	1.38	0.83	0.60	0.83	0.45	0.45	0.50
Bao [2]	1.43	1.03	1.27	4.12	2.02	2.21	1.32	2.94	2.80	2.23
Hosni [11]	0.95	0.97	0.64	0.89	1.93	2.51	1.39	1.08	1.45	1.51
Chang [4]	1.2	1	1.05	2.07	1.96	1.93	1.76	2.85	3.65	3.71
Yang [22]	0.99	1.07	0.82	1.87	3.04	5.34	1.44	2.06	2.78	2.32
Khan [12]	1.58	1.17	1.25	1.86	1.49	1.74	1.56	2.19	1.75	1.46
Heo [7]	1.66	1.52	2.11	2.32	1.55	1.92	2.18	2.32	1.52	1.38
Zinner [24]	1.67	1.3	1.23	1.88	1.78	1.77	1.47	1.9	1.65	1.87
Method	Image 11	Image 12	Image 13	Image 14	Image 15	Image 16	Image 17	Image 18	Image 19	Image 20
Ours	0.73	0.84	0.34	0.78	0.33	0.31	1.3	0.67	0.81	0.57
Bao [2]	1.32	3.18	2.88	1.11	3.53	5.66	2.28	2.05	2.07	0.92
Hosni [11]	1.02	1.02	1.02	1.02	0.71	2.43	3.71	2.97	0.55	2.36
Chang [4]	1.04	2.27	1.18	1.36	0.93	2.14	1.93	3.87	1.23	1.02
Yang [22]	0.98	1.46	1.08	1.07	1.3	4.13	2.78	2.02	0.67	0.78
Khan [12]	1.30	1.30	1.40	1.69	1.18	0.94	1.70	2.13	1.77	1.24
Heo [7]	1.48	2.19	3.45	2.8	3.39	1.38	2.11	1.92	2.89	0.80
Zinner [24]	1.41	1.66	1.65	1.71	1.06	1.19	1.58	2.06	3.31	1.96
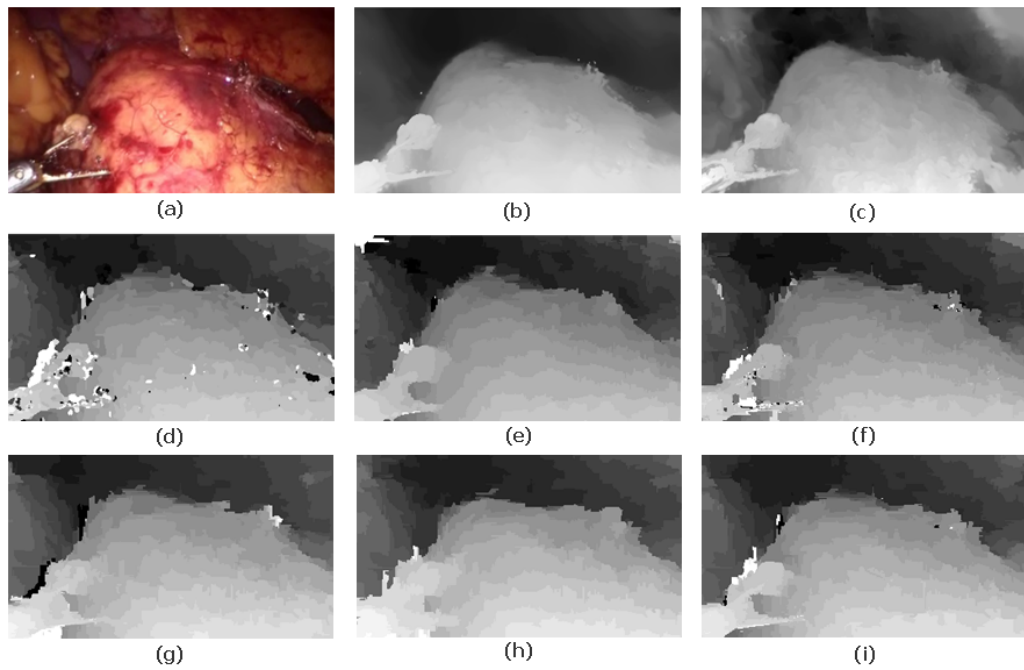
Figure 5.7: Disparity map results obtained by different stereo correspondence methods on surgical laparoscopic image. (a) Input left image, (b) ours, (c) Ref. [2], (d) Ref. [4], (e) Ref. [22], (f) Ref. [11], (g) Ref. [12], (h) Ref. [7],(i) Ref. [24]

Figure 5.8: Reconstructed surgical scene in accordance with various estimated disparity maps in Fig. 2 (a) Input left image, (b) ours, (c) Ref. [2], (d) Ref. [4], (e) Ref. [22], (f) Ref. [11], (g) Ref. [12], (h) Ref. [7],(i) Ref. [24]

Figure 5.9: Disparity map results obtained by different stereo correspondence methods on surgical laparoscopic image. (a) Input left image, (b) ours, (c) Ref. [2], (d) Ref. [4], (e) Ref. [22], (f) Ref. [11], (g) Ref. [12], (h) Ref. [7],(i) Ref. [24]
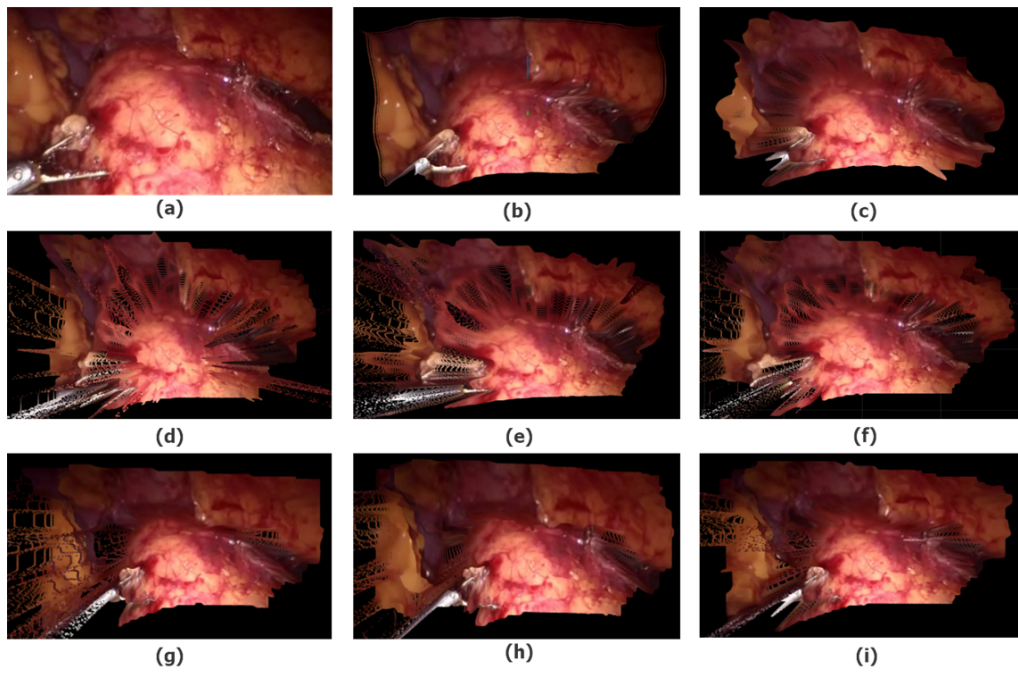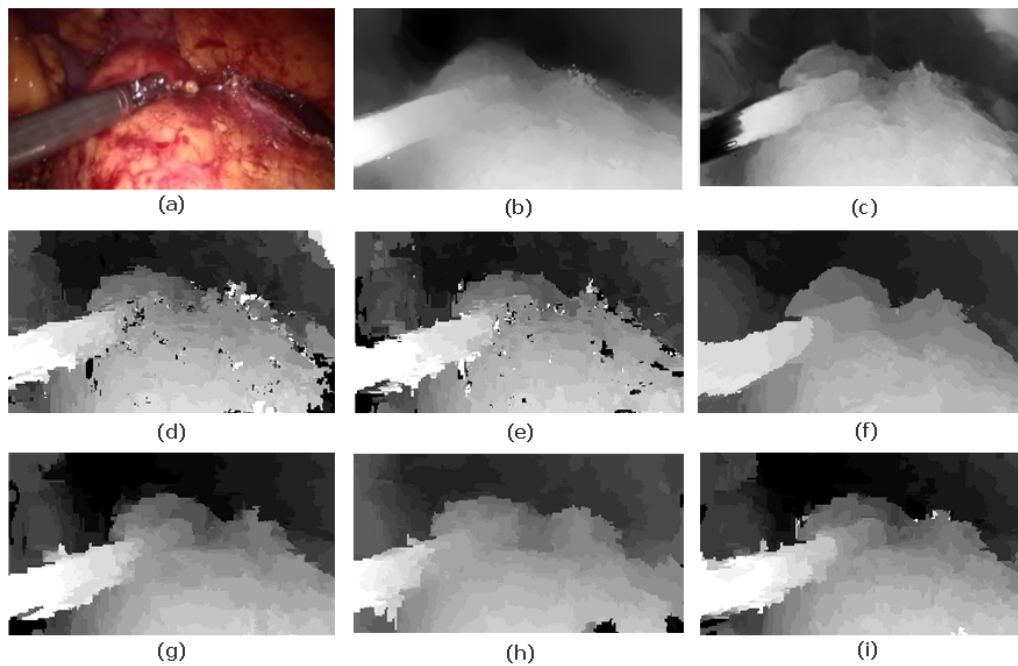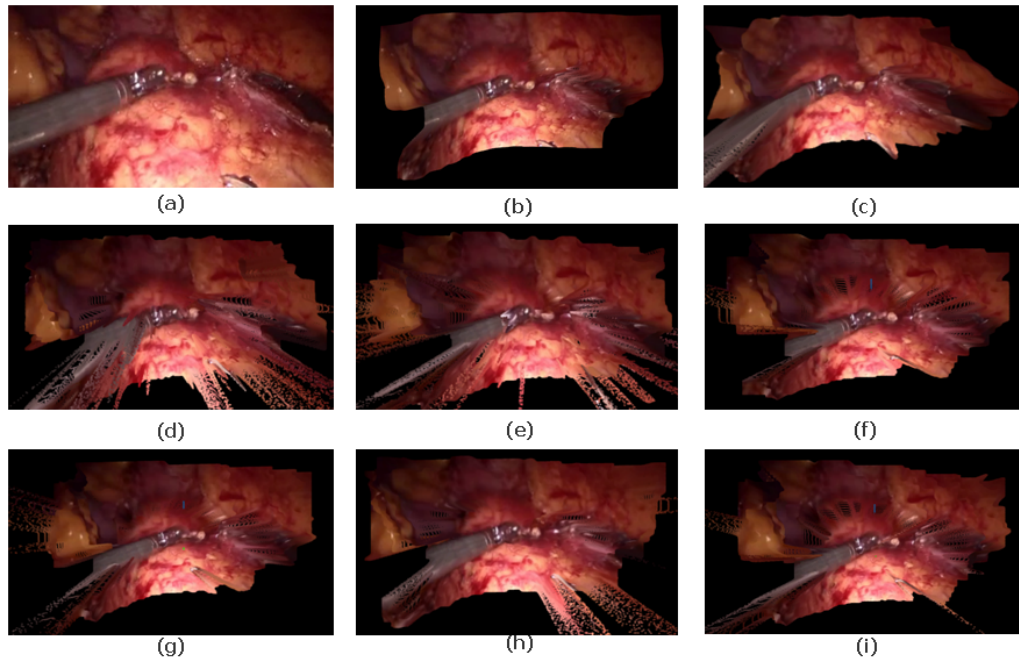
Figure 5.10: Reconstructed surgical scene in accordance with various estimated disparity maps in Fig. 3. (a) Input left image, (b) ours, (c) Ref. [2], (d) Ref. [4], (e) Ref. [22], (f) Ref. [11], (g) Ref. [12], (h) Ref. [7],(i) Ref. [24]

mization, are introduced. Based on the sparse feature matching approach, optimal matching of sparse feature points can be achieved, specially in the scenario where the images include surgical tools, organ tissues, and surfaces with strong specular highlights. As a result, the initial disparity estimate demonstrates superior accuracy than other conventional initial disparity estimates, and the accuracy of the resulting final disparity map is greatly improved. The illumination difference between left and right images can be minimized by the proposed illumination equalization, which in turn improves the disparity estimation accuracy. After both refined initial disparity estimate and illumination correction are integrated into the disparity optimization process, a more accurate disparity map can be estimated. Therefore, this work is of practical surgical importance in providing depth information and surface reconstruction of laparoscopic image pairs.

The proposed algorithm may over smooth disparity map on the valley-shaped surfaces, which could occur on the connected surfaces between the adjacent organs. As a result, the esti-

Figure 5.11:   Disparity map results obtained by different stereo correspondence methods on surgical laparoscopic image. (a) Input left image, (b) ours, (c) Ref. [2], (d) Ref. [4], (e) Ref. [22], (f) Ref. [11], (g) Ref. [12], (h) Ref. [7],(i) Ref. [24]

Figure 5.12: Reconstructed surgical scene in accordance with various estimated disparity maps in Fig. 3. (a) Input left image, (b) ours, (c) Ref. [2], (d) Ref. [4], (e) Ref. [22], (f) Ref. [11], (g) Ref. [12], (h) Ref. [7],(i) Ref. [24]
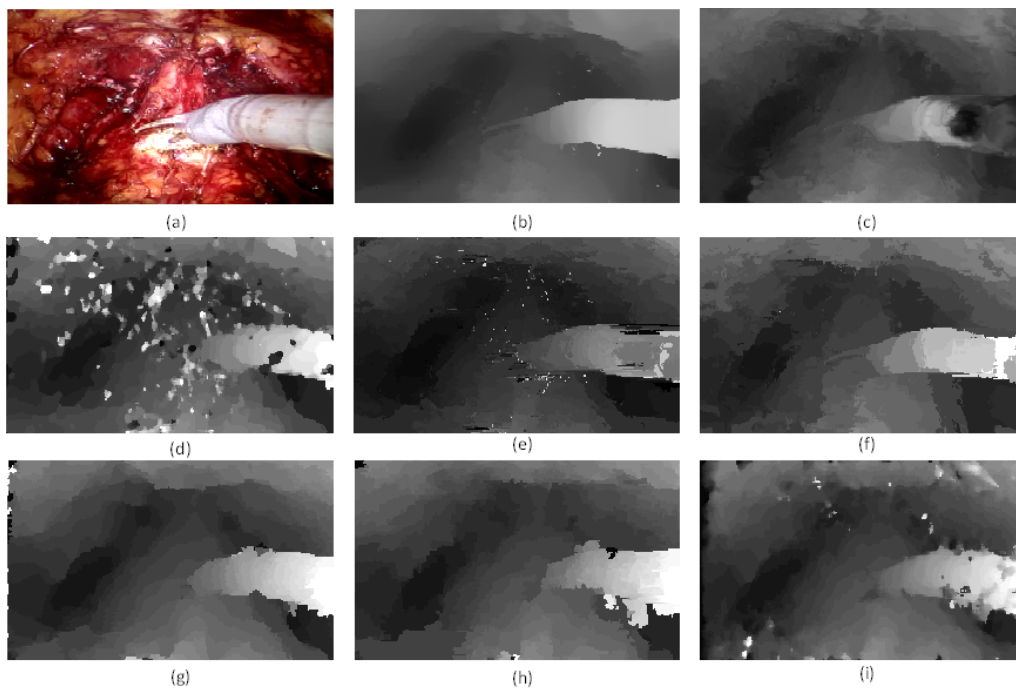
Figure 5.13:   Disparity map results obtained by different stereo correspondence methods on surgical laparoscopic image with with specular highlight. (a) Input left image, (b) ours, (c) Ref. [2], (d) Ref. [4], (e) Ref. [22], (f) Ref. [11], (g) Ref. [12], (h) Ref. [7],(i) Ref. [24]
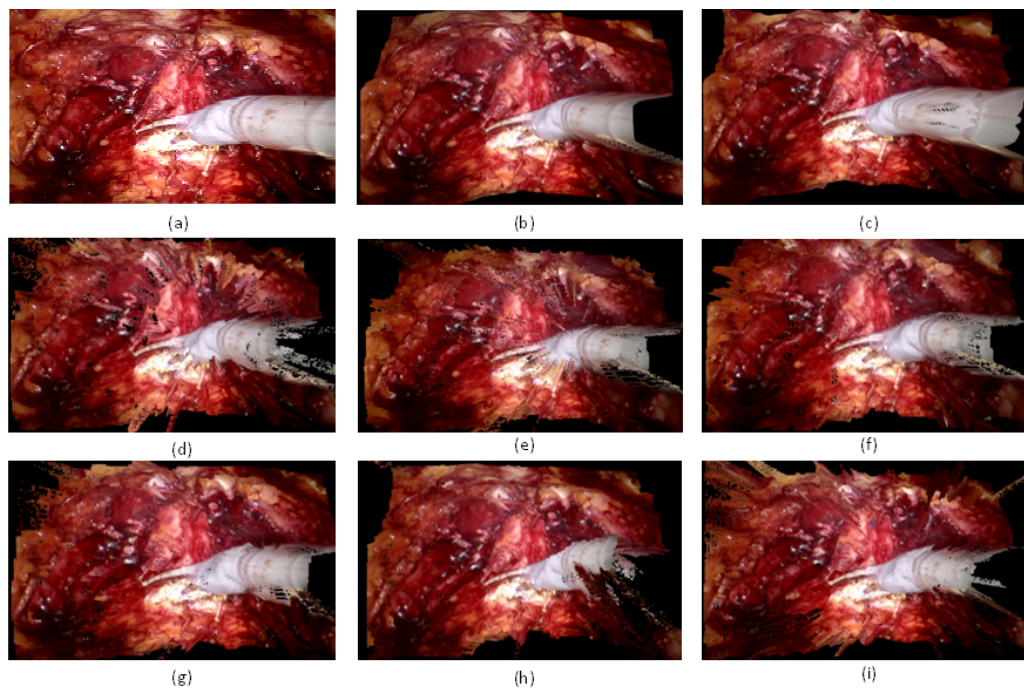
Figure 5.14: Reconstructed surgical scene in accordance with various estimated disparity maps in Fig. 7. (a) Input left image, (b) ours, (c) Ref. [2], (d) Ref. [4], (e) Ref. [22], (f) Ref. [11], (g) Ref. [12], (h) Ref. [7],(i) Ref. [24]
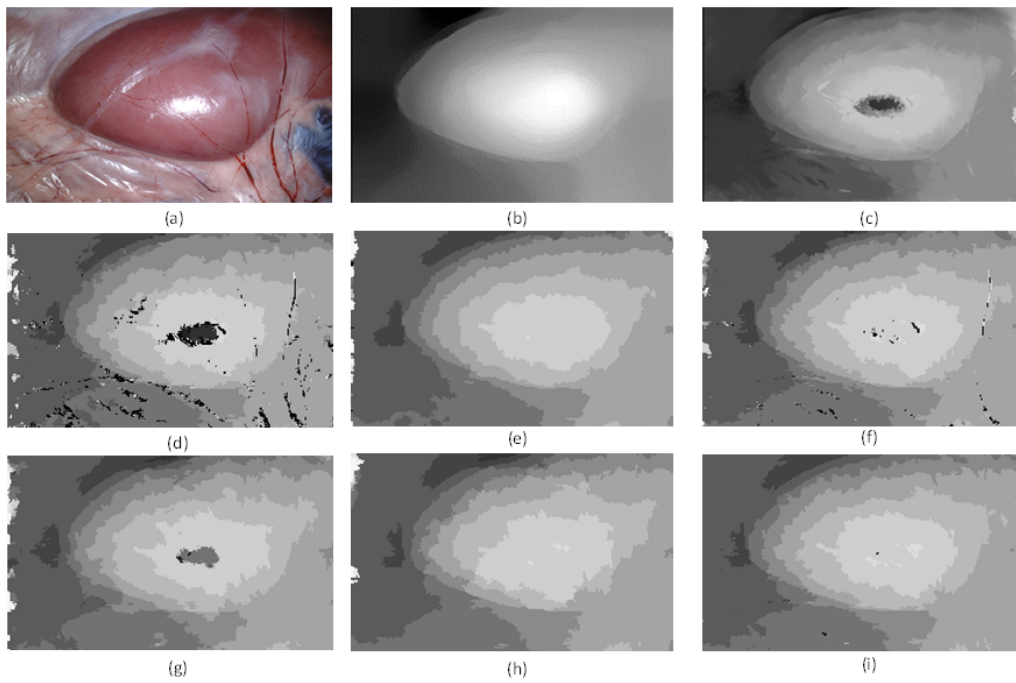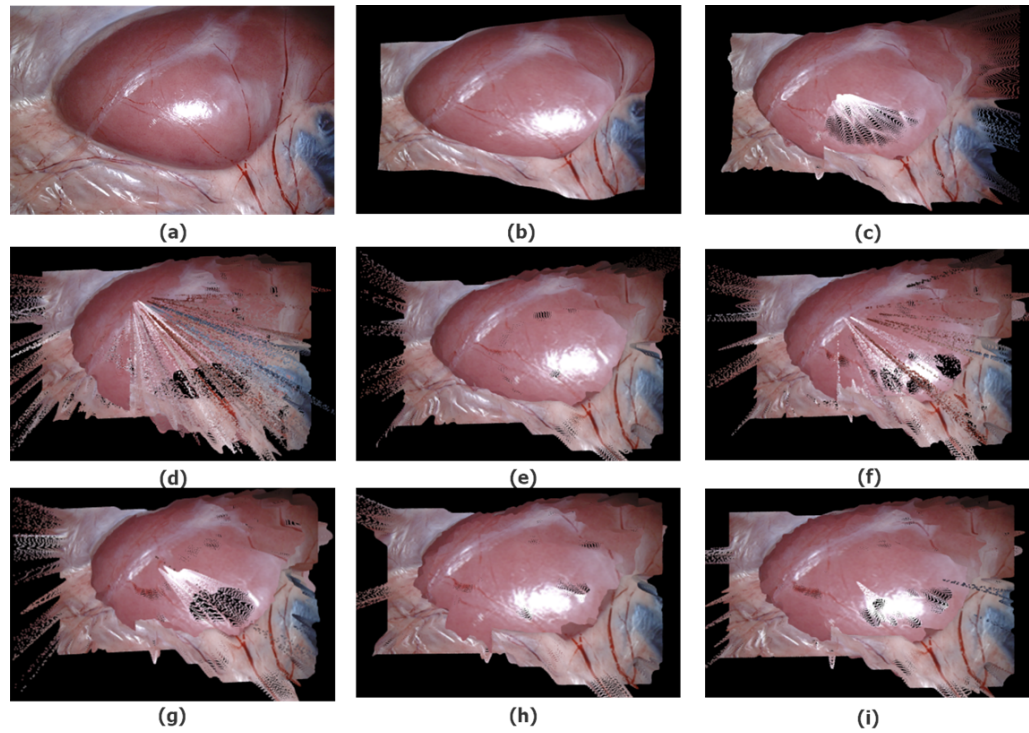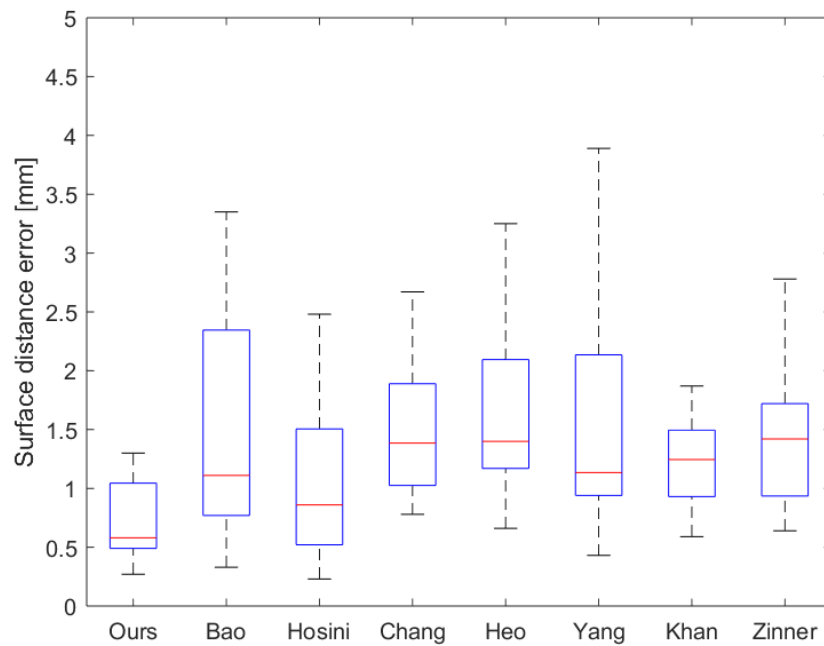
Figure 5.15: Box plot of surface reconstruction error on on the 20 images of D-Distance dataset, according to Table II for eight different methods.

mated disparity map could produce less accurate surface reconstruction, compared to some of the state-of-the-art methods. On the other side, optimized experimental parameters are useful for avoiding oversmoothing disparity map. These issues will be the subject of future research.

Bibliography

[1] Max Allan, Jonathan Mcleod, Congcong Wang, Jean Claude Rosenthal, Zhenglei Hu, Niklas Gard, Peter Eisert, Ke Xue Fu, Trevor Zeffiro, Wenyao Xia, Zhanshi Zhu, Huoling Luo, Fucang Jia, Xiran Zhang, Xiaohong Li, Lalith Sharan, Tom Kurmann, Sebastian Schmid, Raphael Sznitman, Dimitris Psychogyios, Mahdi Azizian, Danail Stoyanov, Lena Maier-Hein, and Stefanie Speidel. Stereo correspondence and reconstruction of endoscopic data challenge, 2021.

[2] Linchao Bao, Qingxiong Yang, and Hailin Jin. Fast edge-preserving patchmatch for large
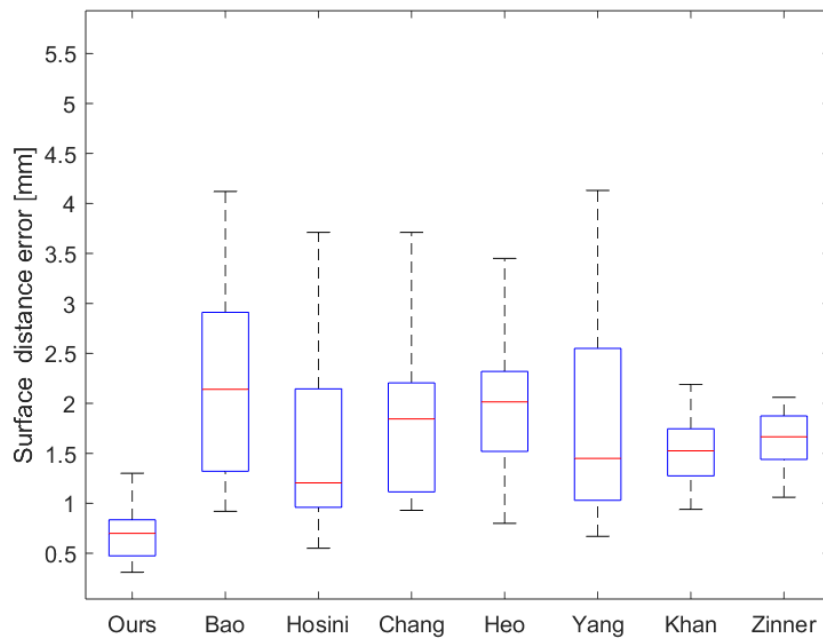
Figure 5.16: Box plot of surface reconstruction error on on the 20 images of D-Angle dataset, according to Table III for eight different methods.

displacement optical flow. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3534–3541, 2014.

[3] Stan Birchfield and Carlo Tomasi. Depth discontinuities by pixel-to-pixel stereo. *International Journal of Computer Vision*, 35(3):269–293, 1999.

[4] Yong-Jun Chang and Yo-Sung Ho. Pixel-based adaptive normalized cross correlation for illumination invariant stereo matching. *Electronic Imaging*, 2017(5):124–129, 2017.

[5] Feiyang Cheng, Hong Zhang, Mingui Sun, and Ding Yuan. Cross-trees, edge and super-pixel priors-based cost aggregation for stereo matching. *Pattern recognition*, 48(7):2269–2278, 2015.

[6] Piotr Dollár and C Lawrence Zitnick. Structured forests for fast edge detection. In *Pro-

ceedings of the IEEE international conference on computer vision, pages 1841–1848, 2013.

[7] Yong Seok Heo, Kyong Mu Lee, and Sang Uk Lee. Robust stereo matching using adaptive normalized cross-correlation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):807–822, 2010.

[8] Heiko Hirschmüller, Peter R Innocent, and Jon Garibaldi. Real-time correlation-based stereo vision with reduced border errors. *International Journal of Computer Vision*, 47(1-3):229–246, 2002.

[9] Berthold KP Horn and Brian G Schunck. Determining optical flow. In *Techniques and Applications of Image Understanding*, volume 281, pages 319–331. International Society for Optics and Photonics, 1981.

[10] Asmaa Hosni, Christoph Rhemann, Michael Bleyer, Carsten Rother, and Margrit Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):504–511, 2012.

[11] Asmaa Hosni, Christoph Rhemann, Michael Bleyer, Carsten Rother, and Margrit Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(2):504–511, 2012.

[12] Asim Khan, Muhammad Umar Karim Khan, and Chong-Min Kyung. Intensity guided cost metric for fast stereo matching under radiometric variations. *Optics Express*, 26(4):4096–4111, 2018.

[13] Cheng Lei, Jason Selzer, and Yee-Hong Yang. Region-tree based stereo using dynamic programming optimization. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2378–2385. IEEE, 2006.

[14] Wen-Nung Lie, Hsi-Hung Huang, Shih-Wei Huang, and Kai-Che Liu. Using sparse-point disparity estimation and spatial propagation to construct dense disparity map for stereo endoscopic images. In *Pacific-Rim Symposium on Image and Video Technology*, pages 100–109. Springer, 2017.

[15] Bruce D Lucas, Takeo Kanade, et al. An iterative image registration technique with an application to stereo vision. 1981.

[16] Xiongbiao Luo, A Jonathan McLeod, Uditha L Jayarathne, and Terry M Peters. Multi-scale retinex aggregation to enable robust dense stereo correspondence. In *2015 International Conference on 3D Vision*, pages 407–415. IEEE, 2015.

[17] Ziyang Ma, Kaiming He, Yichen Wei, Jian Sun, and Enhua Wu. Constant time weighted median filtering for stereo matching and beyond. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 49–56, 2013.

[18] Lena Maier-Hein, Anja Groch, Adrien Bartoli, Sebastian Bodenstedt, G Boissonnat, P-L Chang, NT Clancy, Daniel S Elson, Sven Haase, Eric Heim, et al. Comparative validation of single-shot optical techniques for laparoscopic 3-d surface reconstruction. *IEEE transactions on medical imaging*, 33(10):1913–1930, 2014.

[19] Karsten Mühlmann, Dennis Maier, Jürgen Hesser, and Reinhard Männer. Calculating dense disparity maps from color stereo images, an efficient implementation. *International Journal of Computer Vision*, 47(1-3):79–88, 2002.

[20] Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, and Cordelia Schmid. Epicflow: Edge-preserving interpolation of correspondences for optical flow. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1164–1172, 2015.

[21] Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui, and Cordelia Schmid. Deepmatching: Hierarchical deformable dense matching. *International Journal of Computer Vision*, 120(3):300–323, 2016.

[22] Qingxiong Yang. Stereo matching using tree filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(4):834–846, 2014.

[23] Christopher Zach, Thomas Pock, and Horst Bischof. A duality based approach for real-time tv-l 1 optical flow. In *Joint pattern recognition symposium*, pages 214–223. Springer, 2007.

[24] Christian Zinner, Martin Humenberger, Kristian Ambrosch, and Wilfried Kubinger. An optimized software-based implementation of a census-based stereo matching algorithm. In *International Symposium on Visual Computing*, pages 216–227. Springer, 2008.

Chapter 6

Conclusions

This thesis focuses on image-processing algorithms for laparoscopic vision recovery and stereo matching. The goal in this study is to improve the visual quality of laparoscopic images and to provide depth information to the surgeons, in order to prevent inadvertent injuries and reducing operating time.

Chapter 2 proposes a non-optimization approach to enhance the low-light regions without amplifying the noise during image enhancement process. To achieve this, three illumination regions are first identified: the well-lit region, the low-light region, and the lossy region. Then different enhancement strategies are applied to different regions to meet the enhancement design criteria for desired image quality. For well-lit region, no image enhancement is required, and all details are preserved. For low-light regions, non-linear enhancement is used to improve both the illumination and details. For lossy-region, linear enhancement is used for low intensity pixels in order to suppress noise amplification while maintaining the natural appearance of enhanced image. Overall, this chapter addressed the noise amplification problem during the image enhancement process.

In Chapter 3, a simplified Retinex optimization method is proposed to efficiently enhance the low-light laparoscopic images. The objective function of this method consists of the data-fidelity error term, prior enhanced illumination term, and prior enhanced reflectance term. By

replacing the traditional TV regularization terms with the two prior enhanced l_2-norm terms, the algorithm has low computational complexity and satisfy the convergence condition. Since the prior information contains the structure and texture, the proposed method can enhance the low-light image while preserving naturalness and texture details. In addition, the color cast phenomenon is corrected by RGB-space and HSV-space information fusion. In this chapter, the proposed method is also compared with related enhancement methods, in terms of quantitative evaluation, qualitative evaluation, and computational time. A user study with the clinicians further demonstrates that this method is preferred by clinicians, especially for fog-free laparoscopic images.

Chapter 4 addresses the issue of specular highlight in endoscopic/laparoscopic images. Based on the dichromatic reflection model, a global optimization method is proposed to remove specular highlight in endoscopic images. Utilizing the small color variations in endoscopic images, an effective color correction technique using adaptive inpainting technique is introduced. Based on the estimated diffuse color and illumination color, the specular highlight removal problem is reformulated as a convex optimization problem with double regularization, which has a significant advantage over the existing methods in terms of convergence, accuracy, and robustness. Moreover, experimental results show that the proposed highlight removal method is effective on both natural images and endoscopic images. Furthermore, this chapter also demonstrates that specular highlight removal can improve the accuracy of stereo matching. A user study is carried out and the proposed method is highly preferred by the clinicians.

Stereo matching technique can be used to estimate the disparity map of input stereo image pairs which can provide depth information and reconstruction 3D surgical scene. Chapter 5 proposes a robust and accurate stereo matching method for stereo laparoscopic images, by introducing three techniques: sparse feature matching, image illumination equalization, and refined disparity optimization. Based on our sparse feature matching approach, optimal matching of sparse feature points can be obtained, especially for stereo image pairs with surgical tools, organ tissues, and surfaces with strong specular highlights. Therefore, the estimated

initial disparity map is superior to conventional methods and the accuracy of the resulting final disparity map is improved. The illumination difference between left and right images can be minimized by the proposed illumination equalization, which in turn improves the disparity estimation accuracy. After both refined initial disparity estimate and illumination correction are integrated into the disparity optimization process, a more accurate disparity map can be obtained. Despite the seemingly complicated workflow, the initialization process can be parallelly implemented and when given a good initialization, the optimization can quickly converge to the optimal result within 2-3 iterations, leading to great computational efficiency. This Chapter also compare the proposed method with state-of-the-art stereo matching algorithms. Experiment results show that the proposed algorithm is superior to conventional algorithms in terms of visual comparison and surface reconstruction.

For real-time implementation, the first step is to convert the existing algorithm from MATLAB environment to C++ environment, which will significantly boost the computation speed (10-30 times faster). In addition, GPU programming can be used to further boost the computational speed by parallel processing. These implementations should improve the computation speed of the proposed image enhancement algorithm, highlight removal algorithm, and stereo-matching algorithm to real-time standard.

I hope that by developing the algorithms described in this thesis, visualization and guidance for minimally invasive procedures can be improved. This thesis presented several methods to overcome laparoscopic imaging problems, such as noise, insufficient illumination, low-light regions, specular highlights, and lack of depth information. These advances have the potential to improve the visualization and related applications, which in turn help the surgeons to improve surgical outcomes. The proposed vision enhancement method can reveal organ details, bleeding regions, and surgical tools in the low-light areas, and this algorithm can help surgeons perform surgeries more safely and efficiently. Highlight removal technique can remove the bright specular highlight from laparoscopic images, which can not only improve the visualization for surgeons, but also the reliability of other computer-assisted algorithms, such

as tracking, and registration. Finally, stereo matching can create a 3D surgical scene from 2D images, which provides valuable depth information to the surgeons. In addition, the reconstructed 3D surface allows the surgeon to observe the scene from a different vantage point, and it also allows the registration between laparoscopic images and pre-operative images to improve image guidance. My focus for future work is to create real-time implementations of the proposed algorithms and make them more suitable for clinical applications.

Curriculum Vitae

Name:	Wenyao Xia
Post-Secondary Education and Degrees:	University of Toronto Toronto, Ontario, Canada B.Sc. Applied Science in Engineering Science, 2007-2011
	University of Toronto Toronto, Ontario, Canada M.Eng. Electrical and Computer Engineering, 2011-2013
	Western University London, Ontario, Canada Ph.D. Medical Biophysics, 2016-2021
Honours and Awards:	Western Graduate Research Scholarship Western University 2016-2020
	Ontario Graduate Scholarship Western University 2019-2020
Related Work Experience:	Graduate Research Assistant in Biomedical Engineering Western University 2016 - 2020
	Technical Analyst Princess Margaret Hospital, Toronto, Canada 2013-2015

Journal Publications:

Xia, W., Chen, E.C., Pautler, S.E., Peters, T.M. "A Robust Edge-Preserving Stereo Matching Method for Laparoscopic Images." IEEE Transactions on Medical Imaging, (in submission)

Xia, W., Chen, E.C., Pautler, S.E., Peters, T.M. "Simplified Retinex Image Enhancement Using Prior Refined Information Fusion." IEEE Transactions on Cybernetics, (in submission)

Xia, W., Chen, E.C., Pautler, S.E., Peters, T.M. "A global optimization method for specular highlight removal from a single image." IEEE Access 7 (2019): 125976-125990.

Xia, W., Chen, E.C., Pautler, S.E., Peters, T.M. "Endoscopic image enhancement with noise suppression." Healthcare technology letters 5.5 (2018): 154-157.

Nystrom, N. N., Hamilton, A. M., **Xia, W.**, Liu, S., Scholl, T. J., Ronald, J. A. "Longitudinal visualization of viable cancer cell intratumoral distribution in mouse models using oatp1a1-enhanced magnetic resonance imaging." Investigative radiology 54.5 (2019)

Xia, W., Moore, J., Chen, E.C., Xu, Y., Ginty, O., Bainbridge, D., Peters, T.M. "Signal dropout correction-based ultrasound segmentation for diastolic mitral valve modeling." Journal of Medical Imaging 5.2 (2018): 021214.

Xia, W., Breen, S. "Image Registration Assessment in Radiotherapy Image Guidance Based on Control Chart Monitoring." Journal of Medical Imaging 5.2 (2018): 021221.

Ginty, O., Moore, J., Xu, Y., **Xia, W.**, Fujii, S., Bainbridge, D., Peters, T.M., Kiaii, B., and Chu, M. "Dynamic Patient-Specific Three-Dimensional Simulation of Mitral Repair: Can We Practice Mitral Repair Preoperatively?." Innovations 13.1 (2018): 11–22.

Ginty, O., Moore, J.,**Xia, W.**, Bainbridge, D., Peters, T.M., Kiaii, B., and Chu, M.”Dynamic Patient-Specific Mitral Valve Models for Surgery Simulation.” Canadian Journal of Cardiology 33.10 (2017): S32.

Conference Proceedings:

Xia, W., Fortin, M., Ahn, J., Rivaz, H., Battie, M.C., Peters, T.M. and Xiao, Y. ”Automatic paraspinal muscle segmentation in patients with lumbar pathology using deep convolutional neural network.” In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019. 318–325.

Geurten, J., **Xia, W.**, Jayarathne, U., Peters, T.M. and Chen, E.C. ”Endoscopic laser surface scanner for minimally invasive abdominal surgeries.” In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018. 143–150.

Ginty,O., Moore, J., **Xia, W.**, Bainbridge, D., and Peters, T. ”Patient-Specific Indirectly 3D Printed Mitral Valves for Pre-Operative Surgical Modelling”. In SPIE Medical Imaging, 2017.

Xia, W., Hatzinakos, D. ”An efficient projected subgradient algorithm for blind image deconvolution using an L1-TV cost function.” IEEE International Conference on Image Processing, Orlando, Oct. 2012. 3045–3048

Research Abstract:

Xia, W., Chen, E., Peters, T.M. ”Specular highlight removal for endoscopic images. 16th Annual Imaging Network Ontario Symposium”, Toronto, Mar. 2018

Xia, W., Moore, J., Chen, E., Xu, Y., Ginty, O., Bainbridge, D., Peters, T.M. "Automatic 3D ultrasound image segmentation approach with spatial coherency constraint for patient specific dynamic mitral valve model". 18th London Health Research Day, London, Mar. 2017

Xia, W., Moore, J., Chen, E., Xu, Y., Ginty, O., Bainbridge, D., Peters, T.M. "A combinational 3D ultrasound image segmentation approach with spatial coherency constraint for patient specific dynamic mitral valve model." 15th Annual Imaging Network Ontario Symposium, London, Mar. 2017

Xia, W., Breen, S., Wang, A., Jaffray, D. "Software development for image guidance on the magnetic resonance-guided radiation therapy system (update)." World Congress on Medical Physics and Biomedical Engineering, Toronto, Jun. 2015.

Xia, W., Breen, S., Wang, A., Winter, J. Jaffray, D. "Software Development for Image Guidance on the Magnetic Resonance-Guided Radiation Therapy (MRgRTTM) System." 12th Annual Imaging Network Ontario Symposium, Toronto, Mar. 2014