Microsurgical Laser Scalpel Based on Spectroscopic Feedback: The "Smart Scalpel"

by

Keng Hui Lim

B.S., Mechanical Engineering (1999)
Imperial College of Science, Technology and Medicine

Submitted to the Department of Mechanical Engineering in Partial Fulfillment of the Requirements for the Degree of Master of Science in Mechanical Engineering

at the

Massachusetts Institute of Technology

February 2001

© Massachusetts Institute of Technology
All rights reserved

Signature of Author......................................................

Department of Mechanical Engineering
January 18, 2001

Certified by..............................................................

Ian W. Hunter
Hatsopoulos Professor of Mechanical Engineering
Thesis Supervisor

Accepted by..............................................................

Ain A. Sonin
Chairman, Department Committee on Graduate Students
Microsurgical Laser Scalpel Based on Spectroscopic Feedback: The "Smart Scalpel"

by

Keng Hui Lim

Submitted to the Department of Mechanical Engineering
on January 19, 2001 in Partial Fulfillment of the Requirements for the Degree of Master of Science in Mechanical Engineering

Abstract

To improve the effectiveness of microsurgical techniques, we have developed a semi-autonomous robotic surgical tool (called the "Smart Scalpel") as an alternative approach to the treatment of vascular lesions. The Smart Scalpel employs optical reflectance spectroscopy and computer vision to identify and selectively target blood vessels with a focused treatment laser. Since the laser beam only heats along the blood vessels, collateral damage to adjacent tissue is substantially minimized. The Smart Scalpel also employs rapid real-time feedback analysis for on-line modification of the treatment parameters, quantification of treatment efficacy and compensation for motion tremor. These capabilities allow precise control over the energy dosage to achieve optimal treatment result. This thesis presents the design of a prototype instrument, quantification of system performance, and methods of image analysis. The thesis will also present results of animal testing and preliminary human studies.

Thesis Supervisor: Ian W. Hunter

Title: Hatsopoulos Professor of Mechanical Engineering
Acknowledgements

When I first visited the Bioinstrumentation lab, I thought that all MIT labs looked like this. I was wrong. It turns out to be the best-equipped and dynamic lab I’ve known. It also became my first home.

Foremost, I would like to thank my advisor, Ian Hunter, who has inspired me with his creativity, breadth of knowledge and generosity (yes, Ian is especially generous at buying lab equipments). I am grateful to him for giving me the guidance and independence to explore the project. I also must thank Dr. Colin Brenan for his excellent mentorship and guidance. His advice both at and outside work is invaluable.

I would also like to thank my fun and workaholic lab mates Patrick Anquetil, Bryan Crane, Robert David, Matthew Graham, Tanya Kanigan, John Madden, Peter Madden, Sylvain Martel, Luke Sosnowski and James Tangorra. My interactions with them both inspired me and opened up my mind. I looked forward to the Thursday lab meeting every week not only for the food, but also to learn from every one of them. Thanks also to the lively bunch from the Hogan group, Max Berniker, Steve Buerger and Brandon Rohrer, who we share the mezzanine with. Go Newman!!

Wellman Labs at MGH became my second home later on into the project. I am privileged to work with Dr. Rox Anderson, whose insights into the clinical practicalities of the Smart Scalpel are invaluable. I also received great advice and help from Bill Farinelli and Rieko Tachihara. This project would only be half as successful if not for them.

Life in MIT would be intolerable without ice hockey, beer, 373K and my friends. Besides my buddies in the lab, I am also fortunate to know great friends like Chee Wei, George, Kai, Maggie and Meow Seen. We have many good times together, and I will always treasure their friendship. And this is also to all my friends back in Singapore.

Finally, I would like to thank my parents for their love and support. They have always supported my decisions, even though they really want me to go home. I miss them greatly, and look forward to going home soon.

MIT is a truly amazing place where everyone I met is very highly talented. It is truly a great experience to immense myself among these exceptional individuals.
Contents

CHAPTER 1 ................................................................................................................................ 12

1.1 SMART SCALPEL CONCEPT ............................................................................................... 13
1.2 SMART SCALPEL APPLICATIONS .................................................................................. 14

CHAPTER 2 ................................................................................................................................ 16

2.1 CURRENT CLINICAL PRACTICE ....................................................................................... 16
2.2 AUTOMATED PWS REMOVAL ............................................................................................ 17
   2.2.1 Attempts at Automated PWS Removal ........................................................................ 18
   2.2.2 Smart Scalpel approach ............................................................................................. 18
2.3 TREMOR TRACKING ......................................................................................................... 19
   2.3.1 Nyquist Sampling Criteria ......................................................................................... 20
2.4 PROJECT OBJECTIVES ..................................................................................................... 20
2.5 SMART SCALPEL AS PLATFORM TECHNOLOGY ............................................................. 21

CHAPTER 3 ................................................................................................................................ 22

3.1 SKIN TISSUE ANATOMY .................................................................................................... 22
   3.1.1 Normal Skin Structure .............................................................................................. 22
   3.1.2 PWS Structure .......................................................................................................... 23
3.2 CHARACTERISTICS OF PWS .......................................................................................... 24
3.3 LASER TREATMENT .......................................................................................................... 25
   3.3.1 Laser wavelength $\lambda_L$ .......................................................................................... 26
   3.3.2 Radiant Exposure ....................................................................................................... 27
   3.3.3 Pulse Width $T_p$ ....................................................................................................... 28
3.4 SUMMARY .......................................................................................................................... 29
CHAPTER 4

4.1 OPTICAL PROPERTIES OF HUMAN SKIN

4.1.1 Specular Reflectance from Skin Surface

4.1.2 Diffuse Reflectance (Backscattered Light)

4.1.3 Tissue Scattering and Absorption

4.2 REFRACTIVE INDEX MATCHING

4.3 ORTHOGONAL POLARIZATION SPECTROSCOPY (OPS)

4.4 OPTICAL REFLECTANCE SPECTROSCOPY (ORS)

4.4.1 Spectroscopic Identification of Blood Vessels

4.4.2 Demonstration of Spectroscopic Identification Technique

4.4.3 Assessment of Treatment Efficacy

4.5 SUMMARY

CHAPTER 5

5.1 PROTOTYPE 1 - DUAL LIGHT ILLUMINATION

5.1.1 Optical System

5.1.2 Laser Subsystem

5.1.2.1 Energy Loss from Laser

5.1.2.2 Laser Beam Shutter

5.1.3 Instrumentation

5.2 PROTOTYPE 2 - DESIGN FOR MINIATURIZATION

5.2.1 Optical Designs

5.2.2 Summary and Current Design

CHAPTER 6

6.1 LINE DETECTION TECHNIQUES

6.2 DIFFERENTIAL GEOMETRIC PROPERTIES OF A LINE

6.3 GAUSSIAN CONVOLUTION

6.3.1 Estimating Image Derivatives

6.3.2 Image Smoothing and 2D Convolution

6.3.3 Scale-Space Parameter (Resolution)

6.3.4 Implementing Gaussian Convolution
6.3.4.1 Direct Convolution (FIR).................................................. 55
6.3.4.2 Recursive Filtering (IIR).................................................. 56
6.3.5 Efficient Discrete Derivatives Approximation.......................... 57
6.4 RIDGE EDGE DETECTION .................................................... 58
   6.4.1 Directional Derivatives.................................................. 58
   6.4.2 Extreme Points.......................................................... 59
6.5 ALGORITHM IMPLEMENTATION............................................. 59
   6.5.1 Taylor Polynomial Approximation................................... 59
   6.5.2 Hysteresis Thresholding................................................ 61
   6.5.3 Result of the Algorithm................................................ 62
6.6 DETECTION OF LINE WIDTH.................................................. 62
   6.6.1 Edge Detection (Canny Filter)......................................... 64
      6.6.1.1 Gradient and Direction........................................... 64
      6.6.1.2 Non-maximum Suppression (NMS)........................... 65
      6.6.1.3 Hysteresis Thresholding......................................... 66
   6.6.2 Moving in the Direction of Line Normal........................... 66
   6.6.3 Localization of Edge Pixels........................................... 67
   6.6.4 Results of Width Detection............................................. 67
6.7 IMAGE NOISE ........................................................................ 68
   6.7.1 Adding Gaussian Noise to an Image................................ 69
6.8 TESTING THE ALGORITHMS AGAINST NOISE....................... 69
6.9 SPEED OF IMAGE ANALYSIS.................................................. 71
6.10 SUMMARY............................................................................. 72

CHAPTER 7 .................................................................................. 73

7.1 TIMING ANALYSIS AND BANDWIDTH................................. 73
7.2 REGION DIVISION................................................................. 75
   7.2.1 Equal Regional Division................................................. 75
   7.2.2 Regional Division based on number of targets.................. 76
7.3 IMAGE PROCESSING STRATEGY ............................................. 76
   7.3.1 Step 1: First Pass Imaging............................................. 76
   7.3.2 Step 2: Grow Area from Seed Pixels.............................. 77
List of Figures

Figure 1.1: Control system block diagram of the Smart Scalpel. .............................................................. 13
Figure 2.1: Smart Scalpel as a platform technology .................................................................................. 21
Figure 3.1: Layers of the skin (http://www.pbandelskincare.com) ............................................................. 23
Figure 3.2(a) (b): Relation of vessel measurements to increasing sub-epidermal depth ......................... 24
Figure 3.3: Absorption spectra of hemoglobin. .......................................................................................... 26
Figure 4.1: Schematic representation of the light components at tissue interface. .................................. 30
Figure 4.2(a) Beamsplitter used to redirect light. (b) Oblique lighting. .................................................... 34
Figure 4.3: Absorbance spectra of human blood and melanin. .................................................................. 35
Figure 4.4: Optical system for blood vessels imaging. .............................................................................. 36
Figure 4.5(a) and (b): Reflectance images of mouse ear at 577 nm and 650 nm....................................... 37
Figure 4.6: Measured reflectance spectra of a reddish-purple PWS [14]. .................................................. 38
Figure 5.1: Prototype 1 optical and instrumentation setup. ....................................................................... 40
Figure 5.2: Beamsplitter arrangement. ...................................................................................................... 41
Figure 5.3: Beam “shutter”. ..................................................................................................................... 44
Figure 5.4: Optical design of Prototype 2 ................................................................................................ 45
Figure 5.5: Tunable filter design. ............................................................................................................. 46
Figure 5.6: LEDs illumination. ............................................................................................................... 47
Figure 5.7: Smart Scalpel prototype. ........................................................................................................ 48
Figure 6.1: ID Parabolic Line Profile. ......................................................................................................... 51
Figure 6.2: Gradient vanishes at the center of the line. ............................................................................ 51
Figure 6.3: Large negative value at the center of the line. ....................................................................... 51
Figure 6.4: (a) Original Image. (b) 2nd Derivative image. (c) After hysteresis thresholding. .................. 62
Figure 6.5: Method of line width detection ............................................................................................... 63
Figure 6.6: Support of the non-maximum suppression operator. ................................................. 65
Figure 6.7: Test image.............................................................................................................. 68
Figure 7.1: Six regions in image field .................................................................................... 75
Figure 7.2(a) Blood vessels. (b) Vessel coordinates (seeds) extracted. ............................. 77
Figure 7.3: Grown area from seed pixels ............................................................................... 77
Figure 7.4: Region division .................................................................................................... 78
Figure 7.5(a) Raster scan. (b) Line tracing ........................................................................ 82
Figure 8.1: Image reticle ........................................................................................................ 85
Figure 8.2: MTF of red and green channel at center of image .............................................. 86
Figure 8.3: MTF as a function of lateral displacement .......................................................... 87
Figure 8.4: MTF as a function of axial distance ..................................................................... 88
Figure 8.5: Optical extinction coefficient for milk-water concentration [41] ..................... 89
Figure 8.6: MTF as a function of milk concentration ............................................................. 90
Figure 8.7: Mask .................................................................................................................... 92
Figure 8.8: Experimental setup ............................................................................................ 92
Figure 8.9: Tracking transfer function .................................................................................. 93
Figure 9.1: (a) Image of blood vessels. (b) Image processing results ............................... 97
Figure 9.2(a): Images of ectactic leg veins. (b) Image processing results ....................... 97
Figure 9.3(a): Blood vessels before treatment (illumination at 577 nm) ....................... 99
Figure 9.4(a) Blood vessels before treatment (illumination at 577 nm) ......................... 100
Figure 9.5: Histology results ............................................................................................... 102
Figure 9.6: Close-up view showing coagulated blood vessels ........................................... 102
List of Tables

Table 3.1: *Typical lasers used in dermatology.* ................................................................. 27
Table 4.1: *Optical properties of tissue.* ............................................................................. 32
Table 6.1: *Results of tests against different characteristics of noise.* .............................. 70
Table 6.2: *Processing speeds of various operations.* ......................................................... 71
Table 8.1: *Camera sensor specifications.* .......................................................................... 84
Table 9.1: *Treatment laser parameters.* ........................................................................... 98
Chapter 1

Introduction

The use of medical robots is gaining popularity in hospitals. The field of surgical robotics offers a number of attractive features. It has potential to provide better quality patient care by improving current procedures, allowing additional capabilities that are currently unavailable, and improving the success of surgeries. A small number of surgical robots have already found clinical application in the fields of neurosurgery, cardiology, orthopedic and prostate surgery. In addition to these robots, which have reached the stage of clinical trial, many more prototype systems have been developed in research laboratories.

An interesting use of robots is in the field of microsurgery. Many microsurgical procedures require a high degree of physical dexterity, accuracy, and control, which may degrade with physician fatigue. This problem can be partially alleviated through the use of a microsurgical tool to aid in tissue identification and removal. In order to achieve the high accuracy required, closed loop feedback can be incorporated into the control system. A feedback control system allows any change in the physical system or external perturbation to be compensated in real time, thus minimizing the error. Although this technique is widespread throughout many engineering disciplines, it is virtually nonexistent in surgical instrumentation.
1.1 Smart Scalpel Concept

Our embodiment of this feedback concept is a device called the “Smart Scalpel” (Figure 1.1).

![Diagram of Smart Scalpel system](image)

**Figure 1.1:** Control system block diagram of the Smart Scalpel.

In the Smart Scalpel system, tissue physical properties as well as expert knowledge from the physician are inputs to a computer describing the attributes of healthy tissues versus the diseased tissues to be excised. The system then non-invasively measures the physio-chemical properties (morphological, biochemical, mechanical, electrical, etc.) of a tissue volume and compares this information with the computer model to identify the diseased tissues. This result is then used for continuous feedback control of the targeting and treatment of the diseased tissues with a spatially localized energy source, leaving the surrounding healthy tissues intact. This automated scanning and targeting sequence is repeated until all the diseased tissues have been treated. A handheld micromanipulator (probe) serves as the interface between the patient and the Smart Scalpel, providing a convenient way to extract information about the tissue state and to direct the therapeutic energy to the appropriate targets [1-3].
Techniques to measure one or more physical attributes of the tissue include reflectance spectroscopy, polarization spectroscopy, infrared tomography, optical coherence tomography, magnetic resonance imaging, Raman spectroscopy, fluorescence spectroscopy, fluorescence polarization detection, mechanical impedance measurements, and electrical impedance measurements. Examples of directed energy sources include photon beam (laser), electron beam, localized electric field, directed acoustic energy, and inertia cutting (low frequency mechanical energy).

The many desirable attributes of the Smart Scalpel have the potential not only to improve performance in current microsurgical procedures, but also to facilitate the development of new treatments not yet feasible with existing technology. The accuracy and reliability of present day procedures can be enhanced through finer and more consistent diagnostic and motion controls, and collateral damage can be minimized through selective targeting and quantitative, real-time assessment of the treatment efficacy. This system of real-time feedback has great potential to increase patient comfort, shorten patient recovery times, and decrease the overall medical cost. In addition, the system can potentially be part of a tele-operated surgical system.

1.2 Smart Scalpel Applications

The Smart Scalpel is most powerful in its concept and there are many possible medical applications for it. Some of these include:

- Dermatology ➔ laser-based removal of vascular lesions, veins, hair, and tattoo.
- Surface skin ablation using laser, e.g. burn patients.
- Selective identification and treatment of cancerous tissue. Cancerous tissues are known to have unique spectroscopic properties that enable one to distinguish between tumors and normal tissue. Also, the tissue can be dyed with compounds such that the cancer cells will register a unique spectroscopic signature.
• Photorefractive keratectomy. This involves the use of an excimer laser to sculpt the corneal surface, altering the surface optical architecture to correct refractive errors.

In the field of laser-based surgery, attempts have been made to utilize the Smart Scalpel concept to enhance clinical outcomes. Some of these instruments are described in [4-7]. These instruments are similar to the Smart Scalpel in that the feedback signal is a direct measurement of tissue properties. However, the Smart Scalpel will incorporate the ability to treat, and aims for better detection capability and bandwidth than these instruments.
Chapter 2

Smart Scalpel in Dermatology

Dermatology has been identified as an initial application of the Smart Scalpel. The areas identified are the removal of nevus flammus, which includes port wine stain and telangiectasias, and the removal of superficial leg veins, psoriasis, and hair. The main focus for now will be on the treatment of port wine stains (PWS).

PWS is a congenital, vascular malformation of the dermis, which is estimated to occur in five children per thousand births. On a macroscopic scale, PWS appears as a pink, red, or purple lesion that commonly involves the forehead, face, occiput, and nuchal regions. Since malignant transformation does not occur, the indication for treatment is primarily based on cosmetic considerations and psychosocial stress.

2.1 Current Clinical Practice

Current PWS therapy involves the illumination of approximately 10 mm diameter regions of skin with the output of a 585 nm pulsed dye laser (pulse width 0.4 ms) having a fluence level between $4 \times 10^4 - 8 \times 10^4 \text{ J/m}^2$ [8]. Selective absorption of the laser energy by the blood results in thermal necrosis of blood vessels and irreversible damage. Over time the body absorbs these
vessels, and the lesion fades or completely disappears. In practice, selective thermolysis also damages the tissue surrounding the blood vessels. Since there is no discrimination between the blood vessels and healthy tissues, the collateral damage results in pain for the patient and tissue scarring. The procedure is slow (> 1 hour per session and > 8 sessions per treatment) and treatment response is poor when the blood vessels are deep and relatively small. Finally, aside from PWS appearance, no feedback is generated to assess treatment efficacy or appropriateness of the photon dosage.

2.2 Automated PWS Removal

Besides being tissue non-selective, most of the laser treatment methods used today share a common disadvantage: all require substantial expertise to achieve reliable and reproducible results. There are reported problems such as hypertrophic scarring and inhomogeneity of blanching due to inhomogeneous energy delivery and dosage. In addition, these manual methods are tedious and time-consuming, and usually require many sessions over a period of months before clinically acceptable results are achieved.

During PWS treatment, the laser dosimetry should remain constant. However, conventional method manually moves the laser delivery handpiece over the PWS, and as a result, this movement is operator-dependent and duplication of initial treatment motion is not possible with subsequent applications. Movement of the beam out of focus and variation in the angle of the beam modify the spot size. The speed of the handpiece moving across the lesion is also not controlled. Therefore, each of these inconsistencies may result in very inaccurate coverage due to poor dosimetry.

An automated delivery system can produce more efficient and better controlled treatments than free hand methods. It also allows longer endurance than humans in the same task. Many experts have demonstrated that the precise control of dosimetry is an important criteria for successful laser therapy [9]. Thus, automated treatment devices will surely improve the treatment results by using lasers in an optimal fashion.
2.2.1 Attempts at Automated PWS Removal

Two categories of automated treatment devices have been developed: scanners and automated handpieces [9]. In the first category, the two systems developed are the Multiscan™ and the Scanall™. These instruments automatically scan the laser spot over the PWS using mirrors. Multiscan™ produces a 5 mm spot diameter, and the Scanall™ system uses a 0.5 mm spot diameter. The velocity of scanning (50 to 300 mm/s) is adjusted to provide the appropriate incident fluence. Both scanners work without contacting the skin, and Scanall™ uses a video camera to obtain an image of the lesion, and this image is mapped to the computer to limit the laser treatment to the PWS.

Hexascan™ and CC-scan™ are automated handpieces controlled by a scanner mechanism composed of two stepper motors [9]. The motors are controlled by a microprocessor so they can be programmed to scan the laser across the treatment area in a variety of predefined patterns. These patterns are arranged so that no laser spot is placed adjacent to another spot, thus limiting thermal injury. The Hexascan™ handpiece contains a power meter and a shutter to ensure that there is ample cooling between two adjacent spots. Comparing the results between the conventional "free hand point-by-point" technique and the Hexascan™ reveals a factor of two increase in the percentage of good clinical results. Treatment duration was reduced to 20% of the point-by-point technique, and a drastic reduction in hypertrophic scarring was observed.

In all four automated scanning systems the goal is greater uniformity in laser delivery than manual techniques. Still, laser energy is delivered to the entire region of the PWS. No effort is made to discriminate between blood vessels, which need to be treated, and the surrounding tissue, where unnecessary collateral damage occurs.

2.2.2 Smart Scalpel approach

The Smart Scalpel is a radical departure from these automated approaches. Rather than scanning the laser across the entire tissue, we are using the blood vessel reflectance signal to
identify the blood vessels to be treated and deliver the laser energy only onto those vessels. In addition, this signal is monitored throughout the treatment to assess the treatment efficacy and to adjust the laser parameters (e.g. laser power and pulse width) accordingly.

Due to the flexibility of the Smart Scalpel, any laser can be attached to the system via optical fiber. Therefore, the dermatologist can have the flexibility to use the laser wavelength he deems fit. For more effective treatment, the Smart Scalpel is also designed to recognize the dimensions of the blood vessels in real-time in order to utilize the optimal laser dosage. This capability is very important, and is impossible with the current treatment techniques. In order to compensate for relative motion between the imaged region and the focused laser beam, the imaging and targeting sequence is performed at a high rate.

The major design components of the PWS Smart Scalpel are (1) a non-invasive method to selectively identify and locate blood vessels within the PWS, (2) a means to deliver the laser energy to the target spatial location, (3) a strategy for heating the blood vessels, and (4) a method to estimate the optimal laser dosage during treatment. The Smart Scalpel system aims for high spatial resolution to see the fine blood vessels, and high bandwidth to eliminate problems associated with relative motion between the instrument and the patient.

2.3 Tremor Tracking

Tremor has been identified as the highest frequency source of relative motion between the Smart Scalpel and a resting patient. Tremor is an approximately rhythmic, roughly sinusoidal involuntary movement. It may be either pathologic or physiologic; physiologic tremor is a normal variant that happens in every person ([10-12]).

Physiologic tremor occurs when a person tries to maintain a posture. Any voluntary muscle activation leads to a low-amplitude, fine tremor. The frequency of physiologic tremor varies throughout different parts of the body. For example, normal elbow tremor has a frequency of 3 to 5 Hz while wrist tremor is between 8 to 12 Hz. Tremor typically reduces to a smaller
level (frequency and/or amplitude) when a person is at rest, e.g. when the limb is fully supported against gravity and the muscles are not voluntarily activated.

During laser treatment with the Smart Scalpel, the patient will be in a resting posture with the body supported or even constrained to restrict movement. The treatment probe will be held securely against the patient’s body and this will reduce relative motion since the probe will move together with the body. With these measures implemented together, it is thus reasonable to assume that the displacement frequency will be as small as 3 Hz, and with an amplitude displacement of 370 μm (this tremor displacement is taken from an amplitude experiment conducted by Elble [11]).

2.3.1 Nyquist Sampling Criteria

Nyquist Sampling Theorem states that no information is lost if a waveform of highest frequency \( v \) is sampled faster than the Nyquist frequency, where

\[
 f_{\text{Nyquist}} = 2v .
\]  

(2.1)

Therefore, if the tremor frequency is approximately 3 Hz, then the Smart Scalpel should ideally be sampling at a minimum frequency of 6 Hz (system bandwidth).

2.4 Project Objectives

The aim of this project is to design a Smart Scalpel system to meet the specifications discussed earlier. The objectives of this project are:

1. Design strategies for the laser treatment process.
2. Design and build an optical system for imaging and laser targeting.
3. Seek algorithms to extract the blood vessels and their dimensions in images.
4. Test and improve the system hardware/software to meet the required resolution and bandwidth specifications.

5. Test the system on animal models to assess the treatment efficacy.

6. Perform clinical trials on human patients.

7. Miniaturization.

2.5 Smart Scalpel as Platform Technology

In dermatological applications, the ultimate aim is to use the Smart Scalpel as a platform instrumentation for microsurgery.

As illustrated in the figure above, the Smart Scalpel can be a highly modular instrument, giving the physician a high degree of flexibility in configuring a system best suited for treatment of a particular disease condition. The physician can select a treatment algorithm and “plug-in” a laser module into the optical fiber interface for optimal treatment of a diversity of dermatological conditions.
Chapter 3

Vascular Lesions

Port Wine Stain (PWS) represents one type of congenital malformation involving mature capillaries. These “hemangiomas” are members of a larger group of congenital lesions termed nevus flammeus, which commonly involve the forehead, face, occiput, and nuchal regions.

3.1 Skin Tissue Anatomy

3.1.1 Normal Skin Structure

Optically, the skin can be considered to consist of four distinct layers: the epidermis, the upper dermis layer, the lower dermis layer, and the subcutaneous fat (Figure 3.1).

The epidermis is approximately 100 μm thick and is composed of closely packed epithelial cells forming microscopically distinct layers. The cells on the surface of the skin, known as the stratum corneum, are fully keratinized dead cells. Beneath the epidermis, the upper dermis layer (400 μm thick) contains a system of capillary loops. The capillaries are supplied by the superficial arteriolar plexus formed by vessels 50 μm in diameter (in normal skin), which in turn originate from arteries entering the lower dermis. This superficial plexus is one of the
primary factors in skin coloration. The *lower dermis* (400 µm) is composed of collagen bundles, elastic tissue, sweat glands, follicular structures, and blood vessels. Beneath the lower dermis is the *subcutaneous fat*. The major absorbing or scattering entity in each layer is melanin, blood and collagen.

![Layers of the skin](http://www.pbandelskincare.com)

**Figure 3.1: Layers of the skin (http://www.pbandelskincare.com)**

### 3.1.2 PWS Structure

PWS consists of a large number of ectactic (dilated) blood vessels (capillary-to-venule sized) in the superficial dermal capillary, resulting in a bluish-red skin blemish. The pathological structure of PWS can be classified into four group types: constricted, intermediate, dilated and deeply located.

In the *constricted* type, the blood vessels differ little from the normal pattern, and their number is not more than in normal skin. In the *dilated* type, the vessels are considerably enlarged (up to 400 µm) and often contain red cells. In some cases, this alteration in the vessels is found scattered throughout the upper and lower dermis (*deeply located* type). The *intermediate* type shows histological features common to the constricted and dilated types.
PWS blood vessels can enlarge with age. If left untreated, they may become enlarged and elevate the skin, causing the surface to take on a cobblestone appearance. Occasionally, the vessels can create "exaggerated growth" or hypertrophy, which can lead to impairment of functions such as seeing and breathing.

### 3.2 Characteristics of PWS

The central abnormalities characterizing PWS are an increase in vessel number (vascular profiles) and ectasia. Barsky et al. [13] have observed the following characteristics of PWS in a study involving 100 patients. Vascular area refers to the percentage of dermal area composed of blood vessels.

![Graphs](image)

**Figure 3.2(a) (b): Relation of vessel measurements to increasing sub-epidermal depth.**

The results showed that vessel number sharply decreases with depth and the majority of vessels are located in the immediate sub-epidermal dermis. The mean vessel depth is $460 \pm 170$ μm, with a median depth of $310$ μm. This means that the penetration depth of the imaging light and treatment laser has to reach beyond $310$ μm for the treatment to be effective.

The graphs also revealed that the vascular area at most depths are below 8%. This means that healthy tissue actually make up the majority of the PWS region, unless the blood vessels at
different depths criss-cross so much that the effective area becomes large. Therefore, the current treatment method of illuminating the entire region with laser is very inefficient since only a small percentage of the region needs to be treated. This observation serves to reinforce the need for an automated instrument that only targets blood vessels.

3.3 Laser Treatment

During laser exposure, absorption convert radiant energy into heat within each target in the exposure field. Essentially, any mammalian tissue heated to 70 °C to 100 °C would suffer protein denaturation, leading to "coagulation necrosis". Coagulation necrosis is useful for causing hemostasis due to the denaturation of plasma proteins and the closing of vessels. The vessels then collapse and reabsorb by the body, which results in lesion blanching (color lightening).

Using appropriate dosimetry, laser treatment of PWS can achieve "selective photothermolysis", which induces selective thermal damage of abnormal blood vessels and minimizes the risk of scarring. Although results of clinical studies are encouraging, a study in 1997 showed that only a small proportion of patients (10-20%) obtained 100% fading of their PWS, even after undergoing multiple treatments. Van Gemert et al. [14] suggested that the primary reason for the inadequate clinical results was that all PWS are treated with virtually identical laser parameters, which may not be appropriate on an individual patient basis. Due to the complexity of PWS pathology, it is difficult to use one set of laser treatment parameters to treat all types of PWS lesions. Now that laser systems are available that allow user-specified selection of irradiation parameters, the optimal parameters can be selected based on the anatomy and physiology of the PWS.

The three main parameters necessary to achieve selective photothermolysis are (1) a wavelength that is preferentially absorbed by the desired targeted structure, (2) sufficient laser fluence (energy per unit area in J/m²) to reach a damaging temperature in the targeted structure, and (3) an exposure duration less than or equal to the thermal relaxation time of the target.
3.3.1 Laser wavelength $\lambda_L$

The optimal laser wavelength $\lambda_L$ is defined as the wavelength that penetrates to blood vessel depths to cause substantial hemoglobin absorption and damage, while causing minimal damage to other healthy tissues.

In general, $\lambda_L$ should coincide or be close to the absorption spectral peak of hemoglobin (Figure 3.3). There are four major absorption peaks in the visible to IR region at 415 nm, 540 nm, 577 nm and 940 nm. These wavelengths are centered at an isobestic point of oxy- and deoxy-hemoglobin (whole blood).

![Figure 3.3: Absorption spectra of hemoglobin.](image)

The higher the absorption coefficient, the greater the degrees of energy transfer. However, the depth of tissue penetration is also an important factor. Scattering probability by tissue varies as $1/\lambda^4$ (Rayleigh scattering), and major tissue chromophores tend to have greater absorption at shorter wavelength. Therefore, longer wavelengths have greater penetration depths. For this reason, despite the fact that 415 nm has the greatest blood absorption, there is too much
scattering for the photons to reach most blood vessels. Most of the popular treatment wavelengths currently used in clinics lie around the two middle absorption bands ([8] & [15]).

Current clinical applications also use wavelengths near the fourth absorption band, such as the Nd:YAG laser at 1064 nm. These lasers are used for treating deep vessels, and since the hemoglobin absorption is lower at those wavelengths, the laser energy has to be greater for effective heating.

### 3.3.2 Radiant Exposure

It is important to know the damage threshold fluence sufficient to effect selective photothermolysis. The laser fluence \((J/m^2)\) is difficult to establish by theoretical modeling because of epidermal melanin absorption, multiple scattering events within the skin, and the fact that blood vessels are located at different dermal depths. In general, treatments using fluence of \(4 \times 10^4\) to \(8 \times 10^4\) \(J/m^2\) are used for vascular lesions. Table 3.1 gives a summary of the types of lasers in dermatology and their specifications [8].

<table>
<thead>
<tr>
<th>Laser</th>
<th>Wavelength (nm)</th>
<th>Treat Fluence ((x 10^4 J/m^2))</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argon</td>
<td>488, 514</td>
<td>1-10</td>
<td>CW</td>
</tr>
<tr>
<td>Classic KTP</td>
<td>532</td>
<td>10-40</td>
<td>CW</td>
</tr>
<tr>
<td>Cu or Cu-Br</td>
<td>512-578</td>
<td>1-10</td>
<td>CW</td>
</tr>
<tr>
<td>Krypton</td>
<td>570</td>
<td>1-10</td>
<td>CW</td>
</tr>
<tr>
<td>Pulse dye (yellow)</td>
<td>585</td>
<td>4-8</td>
<td>Pulsed</td>
</tr>
<tr>
<td>Derm-KTP</td>
<td>532</td>
<td>2-20</td>
<td>Pulsed</td>
</tr>
<tr>
<td>Pulsed dye (green)</td>
<td>510</td>
<td>3-5</td>
<td>Pulsed</td>
</tr>
<tr>
<td>Q-sw. Nd:YAG – green - infrared</td>
<td>532, 1064</td>
<td>3-5, 4-10</td>
<td>Pulsed</td>
</tr>
<tr>
<td>Q-sw. Ruby (red)</td>
<td>694</td>
<td>4-10</td>
<td>Pulsed</td>
</tr>
<tr>
<td>Q-sw. Alexandrite (infrared)</td>
<td>755</td>
<td>4-10</td>
<td>Pulsed</td>
</tr>
</tbody>
</table>

Table 3.1: Typical lasers used in dermatology.
Current pulsed lasers do not interface with an external trigger source for pulse control. On the other hand, the pulse width of CW lasers can be easily controlled using optical modulators or deflectors, which can be integrated into the Smart Scalpel system.

3.3.3 Pulse Width $T_p$

The pulse width (duration) governs the spatial confinement of the thermal energy within the targeted vessel. Ideally, the pulse width $T_p$ should be compatible with the diameter $d$ of the vessel and be about equal to the thermal relaxation time $T_d$ for that dimension. The thermal relaxation time, defined as the time required for the instantaneous temperature generated inside the blood vessel after exposure to the laser pulse to decrease by 50%, is given by,

$$\tau_r = \frac{d^2}{16\alpha}, \quad (3.1)$$

where $\alpha$ is the thermal diffusivity. For blood composed of approximately 70% water, the thermal diffusivity $\alpha$ is $1.2 \times 10^{-7}$ m$^2$/s [16]. Therefore, typical values for $\tau_r$ are $\approx 1$ ms for $d = 50$ µm and $\approx 10$ ms for $d = 150$ µm. These values are in agreement with the 1 to 10 ms pulse width requirement measured by Dierickx et al. [17].

If $T_p \gg T_r$, heat diffuses outside the vessel during the laser exposure, reducing the target specificity, and can cause extra thermal damage to surround tissue. A very short pulse, $T_p \ll T_r$, will generate a high-peak intravascular temperature rise, leading to localized explosive vaporization of tissue water, or to photoacoustic shock waves, which will result in vessel rupture. Hemorrhage leading to purpura (bruising) then takes place, and in most cases, repair mechanisms may revascularize the tissue since the vessels are not completely obliterated. In the ideal situation where $T_p \approx T_r$, the thermal energy is just enough to diffuse to the vessel wall and coagulate it. Here, selective photothermolysis takes place where the blood vessels "cook" to result in tissue necrosis and collapse.
It is impossible for a single pulse width to fit the thermal relaxation time of all the blood vessels. This is one of the limitations of the current laser treatment devices. To overcome this, the Smart Scalpel is designed to recognize the diameter of the blood vessels during treatment and adjust the laser pulse width in real time. This capability is necessary to achieve superior clinical results.

3.4 Summary

Port wine stain consists of a large number of ectactic blood vessels near the skin surface. During laser treatment, the blood absorbs the light energy which causes the blood vessels to be damaged and reabsorbed by the body. The 3 important parameters that determine the optimal laser dosage are wavelength, fluence and pulse width.
Chapter 4

Tissue Optics

In order to design the Smart Scalpel optical system, it is important to understand the interactions between light and skin tissues. This chapter will discuss some of the relevant issues.

4.1 Optical Properties of Human Skin

There are four components of light at a tissue interface (see Figure 4.1):

- **Incident Light** (Polarized)
- **Specular Reflectance** (Polarized)
- **Diffuse Reflectance or Backscattered Light** (Depolarized)
- **Diffuse Transmittance**

Figure 4.1: Schematic representation of the light components at tissue interface.
(1) specular reflectance arising from the surface,
(2) diffuse reflectance due to the light backscattered from within the tissue,
(3) diffuse transmittance in the tissue, where absorption and scattering take place,
(4) internal reflection of the scattered light that strikes the tissue’s interior surface.

4.1.1 Specular Reflectance from Skin Surface

A small fraction (4 to 7%) of an incident radiation is reflected from the skin surface due to the change in refractive index between air \((n_d = 1.0)\) and the stratum corneum \((n_d \approx 1.55)\). Skin surface is not continuously flat, but contains multiple folds with air spaces between them. Therefore, these surfaces presents additional optical interface for specular reflectance. Taken together, these specular reflections account for the “glare” from skin and contain the visual cues related to surface texture. Since we are not interested in these surface details, it is necessary to filter them away during imaging.

4.1.2 Diffuse Reflectance (Backscattered Light)

The light that is not initially reflected enters the tissue, where it is absorbed and/or scattered. Scattering causes a change in the direction of light propagation, and is the only physical process by which this light can be returned through the skin surface to participate in skin reflectance. The backscattered light component carries all visual cues related to internal skin structures such as blood vessels or infiltrates, but carries few or no visual cues related to the surface texture.

4.1.3 Tissue Scattering and Absorption

Skin tissues are highly turbid medium and photons undergo many scattering and absorption events as they travel though the tissues. Scattering results from physical
inhomogeneities in the medium, which lead to inhomogeneities in the refractive index. The scattering and absorption processes taken together essentially determine the penetration of radiation into the skin, as well as the remittance of scattered radiation from the skin.

Three main optical parameters that govern the behavior of light-tissue interaction are the scattering coefficient $\mu_s$, the absorption coefficient $\mu_a$, and the anisotropy coefficient $g$. The anisotropy coefficient is a measure of the asymmetry of a single scattering pattern, and varies between isotropic scattering ($g = 0$) and complete forward scattering ($g = 1$). Skin tissues are typically highly forward scattering ($g$ between 0.7 to 0.9), and this allows greater penetration of light.

From these three parameters, the reduced scattering coefficient $\mu'_s = \mu_s (1 - g)$, which describes the forward scattering light loss, and the total optical extinction coefficient $\mu_t = (\mu'_s + \mu_a)$ can be calculated. The reciprocal of the total optical extinction coefficient is sometimes referred to as the transport mean free path (MFP), or the distance that a photon travels before being scattered:

$$MFP = \frac{1}{\mu_a + \mu_t (1 - g)}.$$  \hspace{1cm} (4.1)

These optical properties at 633 nm (derived from diffusion theory, data obtained from Cheong et al. [18]) are summarized in Table 4.1. These values are also close to that measured by van Germert et al. [19], Verkruysse et al. [20] and Schmitt et al. [21].

<table>
<thead>
<tr>
<th>Tissue</th>
<th>$\lambda$ (nm)</th>
<th>$\mu_a$ (mm$^{-1}$)</th>
<th>$\mu_s$ (mm$^{-1}$)</th>
<th>$g$</th>
<th>$\mu'_s$ (mm$^{-1}$)</th>
<th>$\mu_t$ (mm$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human dermis</td>
<td>633</td>
<td>0.27</td>
<td>18.7</td>
<td>0.81</td>
<td>3.553</td>
<td>3.823</td>
</tr>
</tbody>
</table>

Table 4.1: Optical properties of tissue.

The epidermis has greater scattering and absorption coefficients of $\mu_s \approx 47$ mm$^{-1}$ and $\mu_a \approx 1.9$ mm$^{-1}$ respectively. In a multi-layer model of the skin tissue, it is normally assumed that $\mu'_s$ is
the same for both the epidermis and dermis. This assumption is justified because, firstly $\mu_s$ does not vary drastically between most tissue types at the same wavelength, and secondly the epidermis is much thinner than the dermis. Therefore, it is often further assumed that the optical extinction coefficient $\mu_t$ of the tissue has a single value (homogeneity).

4.2 Refractive Index Matching

An optical window pushed against the skin provides a uniform surface for imaging. Specular reflectance that arises from the glass-skin interface can be reduced to values within the range for normal skin by applying index matching compounds capable of spreading to and filling the spaces in the skin. These index matching compounds must have $n_d$ close to that of the stratum corneum, exhibit flat transmission spectrums, and have low absorptions.

Another important observation is that the decrease in skin reflectance is unrelated to optical absorption by the applied compounds since they have low absorption coefficients. Therefore, a greater fraction of the incident radiation must now penetrate the plaque.

4.3 Orthogonal Polarization Spectroscopy (OPS)

Multiple scattering scrambles the polarization of a polarized incident light. In contrast, specular reflectance preserves the polarization plane of the incident light (Figure 4.1). Therefore, viewing skin through a linear analyzer separates the two components of tissue reflectance. When the planes of polarization are parallel, images with enhanced surface details are obtained and these are useful for examination of texture, elevation and scale. When the polarization planes are orthogonal, only depolarized photons scattered within the tissue can pass the analyzer, and an enhanced view of vasculature, inflammatory and pigmented lesions is obtained. OPS imaging produces high contrast images otherwise inaccessible by standard intravital microscopy techniques [22].
Figure 4.2 illustrates two methods to implement OPS imaging. Total elimination of the specular component is achieved when the polarizer and analyzer filters are arranged at right angles to each other. This is because the polarization plane of the incident light is parallel to the plane of the analyzer.

Figure 4.2(a) Beamsplitter used to redirect light.  (b) Oblique lighting.

It should be noted that the oblique lighting in Figure 4.2(b) may produce unwanted shadows from the raised or depressed surfaces on the object.

4.4 Optical Reflectance Spectroscopy (ORS)

Optical reflectance spectroscopy (ORS) uses diffused white-light irradiation to detect the backscattered (reflected) light from the PWS skin at visible and near-infrared wavelengths. The reflected spectrum at specific wavelengths contains information of the lesion, and is dependent on the skin optical and anatomical properties.

4.4.1 Spectroscopic Identification of Blood Vessels

ORS combined with OPS are used in the Smart Scalpel to non-invasively identify blood vessels in the tissue. The PWS (and skin in general) contains two dominant chromophores:
mela{n}in resident in the epidermis and hema{globin} found in the blood. Figure 4.3 shows that whole blood exhibits two strong absorption bands at 540 nm and 577 nm due to a high concentration of hemoglobin chromophore [23]. The other two absorption peaks are at 415 nm and 940 nm. In comparison, the relative melanin absorption spectrum generally decreases with increasing wavelength, while its overall magnitude scales with melanin concentration [24].

From Figure 4.3, it can be seen that normalization of the absorption spectrum at 577 nm (where blood absorption dominates) with respect to the melanin spectrum at 650 nm (where melanin absorption dominates) generates a high contrast spectrum. Therefore, subtraction (or ratio) of PWS images taken at these two wavelengths generates a high contrast image from which blood vessels are uniquely identified and their locations determined.

In addition to absorption, the scattering properties of skin must be considered in the selection of an appropriate imaging scheme. Longer wavelength lights are highly forward scattering, thus allowing both substantial tissue penetration and high remittance of light scattered out of the tissue after penetration ([25] & [26]). Therefore, although the greatest hemoglobin absorption peak is at 415 nm, this wavelength is too short to penetrate to the depth of PWS blood vessels. At 415 nm, the penetration depth, defined as the distance at which the energy incident on
the skin is extinguished by 37% (1/e), is 100 μm [25]. On the other hand, the penetration depth of 600 nm light is 550 μm, which is close to the mean vessel depth of PWS.

### 4.4.2 Demonstration of Spectroscopic Identification Technique

The optical system used for the proof of concept is illustrated in Figure 4.4. The object imaged was the back of a mouse ear. Polarized light was focused on the skin surface and the reflected light was collected by an objective (Olympus, 0.4 NA, ∞-corrected, 10x) that collimated the light and passed it through an analyzer. The light was then passed through a liquid crystal tunable filter (CRI Varispec tunable filter, VS-05/30-HC-20-S). The wavelengths used are 577 nm for the blood vessel image and 650 nm for the normalization image (FWHM 40 nm). The light that exited the tunable filter was focused onto a CCD camera (Pulnix TM-9701) by an achromat to minimize aberration.

![Optical system for blood vessels imaging](image)

Figure 4.4: *Optical system for blood vessels imaging.*

Figure 4.5 shows the blood vessel contrast enhancement made possible with this two-wavelength imaging scheme. Figure 4.5(a) is the image of the mouse ear taken at the blood absorption wavelength of 577 nm. Besides the presence of the blood vessels, the circular hair follicles are also clearly visible. Figure 4.5(b) is an image of the same region acquired with non-absorbing 650 nm light, and Figure 4.5(c) is the image after background subtraction and dynamic
range expansion. There is a high degree of background suppression in image (c), resulting in an enhanced blood vessel image contrast.

![Image](a) ![Image](b) ![Image](c)

Figure 4.5(a) and (b): Reflectance images of mouse ear at 577 nm and 650 nm.

(c) Background subtraction and dynamic range expansion.

The background subtraction in image (c) can be performed by either one of the following pixel operations before dynamic range expansion:

\[
I_c(x, y) = \frac{I_A(x, y)}{I_B(x, y)}, \text{ or } \\
I_c(x, y) = I_A(x, y) - I_B(x, y). 
\]

(4.2)  
(4.3)

In general, both methods yield almost the same high contrast image, but sometimes one of them produces a better contrast than the other. The disadvantage of the ratio method is that it is slower and trickier to implement.
4.4.3 Assessment of Treatment Efficacy

ORS-OPS can be used to determine quantitatively the degree of PWS blanching following each laser treatment. Most importantly, ORS-OPS can also be used to discriminate the treated PWS vessels from the untreated ones. Van Gemert et al. [14] measured the reflectance spectra of normal skin and untreated and treated PWS, and found that there are characteristic differences in their spectra enabling easy distinction (see Figure 4.6). Therefore, the treatment efficacy can potentially be assessed in real time during the treatment process to determine whether the laser therapy is successful. Those blood vessels that are identified as successfully treated will not be treated again, while others that are less successful will be re-treated.

![Figure 4.6: Measured reflectance spectra of a reddish-purple PWS [14].](image)

4.5 Summary

The Smart Scalpel employs optical reflectance spectroscopy for identification of blood vessels. To reduce unwanted reflections during imaging, techniques including orthogonal polarization spectroscopy, refractive index matching and oblique illumination are used.
Chapter 5

Smart Scalpel Design

In this chapter, the optical designs and instrumentations of the Smart Scalpel is discussed.

5.1 Prototype 1 - Dual Light Illumination

The optical and instrumentation setup is shown in Figure 5.1.

5.1.1 Optical System

To achieve high frequency switching, two Perkin-Elmer strobe lights (MVS 7020, PerkinElmer, Wellesley, MA) are used as illuminators. These are optical fiber coupled strobes with a bandwidth of 20 Hz, and both the flashing and intensity can be controlled with external triggers. Large optical fibers (12.7 mm diameter, Dolan-Jenner, Lawrence, MA) are used to deliver the light from the strobe, which are then focused using convex lens.

The green filter is a 577 nm bandpass filter, and the red filter is a 650 nm bandpass filter (FWHM ≈ 60 nm). Neutral density filters (ND) are used to match the intensities from both
channels. To implement OPS, polarizers are placed at the optical fiber exit and CCD camera. The CCD camera used is Pulnix TM-9701 (Pulnix, Sunnyvale, CA) with a resolution of 755 by 484 pixels and a frame rate of 15 frames per second. The camera lens is a 55 mm Computar lens (f2.8, Computar, Commack, NY), which is set at maximum magnification to give a field of view of 17 mm by 13 mm.

![Diagram](image_url)

Figure 5.1: Prototype 1 optical and instrumentation setup.
To ensure that the object is always at the same distance to the camera, a sapphire window is positioned at the focus distance, and the object to be imaged will be held against it. Due to its excellent material properties and hardness, the sapphire is very thin (1 mm thick), and has a high energy threshold.

Oblique lighting is used to illuminate the object directly. This lighting scheme does not totally eliminate specular reflection due to less efficient polarization, and it also suffers the disadvantage of producing shadows. This becomes a problem especially during two-wavelength image subtraction because the shadows created by both light sources from opposite sides are different. However, this scheme is used because it is not possible to implement the more superior scheme of using beamsplitters (see Figure 5.2).

![Beamsplitter arrangement.](image)

The reason is because the sapphire window reflects a ghost image of the light source (in this case, the face of the optical fiber) onto the camera. The only way to eliminate this reflection is to implement oblique illumination so that the unwanted reflection is out of the camera’s field of view. In retrospect, a benefit of using oblique lighting is that most of the incident light intensity is not lost, whereas a beamsplitter loses 75% of the intensity.

### 5.1.2 Laser Subsystem

The laser subsystem is designed so that the laser can be coupled to the Smart Scalpel via an optical fiber. This allows the Smart Scalpel to operate as an independent unit not influenced
by the location of the laser source. A lens collimation unit is set up to collimate the light that spreads from the end of the optical fiber. The collimated beam then passes through a two-axis galvanometer system (*General Scanning, Bedford, MA*), which directs the beam towards the blood vessels in the field of view. This arrangement converts the mirror rotation at the galvanometers to a linear displacement of the laser spot on the surface of the object. Final focusing is done using a planar-convex lens. The diameter of the focused beam has to be close to that of the blood vessels in order to limit collateral damage.

**5.1.2.1 Energy Loss from Laser**

The laser used is a 514 nm Coherent Innova-100 CW argon laser (*Coherent Inc., Santa Clara, CA*). Before the optical fiber is used, the system is setup with the laser reflected directly into the Smart Scalpel using mirrors. A polarizing beamsplitter (92% reflectance for S-polarization) is then used to reflect the laser onto the object; the P-polarized laser is converted to S- using a half-wave plate in order to achieve maximum reflection. However, when the optical fiber is later incorporated in the system, the laser that emerged from the fiber is depolarized. It is found that a significant amount of energy is lost to polarize the laser again before entering the beamsplitter. An ordinary 50T-50R beamsplitter cube will also not solve the energy loss problem.

A solution is to use a dichroic mirror as the beamsplitter. The dichroic mirror chosen is a high pass transmission filter with a cutoff wavelength at 540 nm. This ensured that all the laser energy ($\lambda < 540$ nm) is reflected, while the 2 illumination wavelengths ($> 540$ nm) are all transmitted.

Although energy is conserved, the use of dichroic mirrors clearly suffers from the disadvantage that laser wavelengths around the 2 illumination wavelengths (from 547 nm to 680 nm) cannot be used. This problem is made worse by the fact that most of the popular treatment lasers used today has wavelengths within that region! For this reason, another candidate for the laser source is the Nd:YAG laser. This laser operates at the infrared-red region of 1064 nm, and is excellent for treating deep vessels.
The reasons why every effort is made to conserve energy are as follows:

1. High energy fluence is required for short exposure times. Having a short pulse width will also increase the system bandwidth.
2. There is energy loss at the optical fiber because glass fiber absorbs light.
3. There is energy loss while collimating the laser that exited from the optical fiber because of the large numerical aperture of the fiber.
4. A cheaper laser module can be used to provide the lower power.

*If a very high power laser is available, then all the energy losses can be tolerated. In this case, a 50T-50R beamsplitter is the best option.*

**5.1.2.2 Laser Beam Shutter**

In order to produce the desired pulse width during treatment, an optical shutter is used to block the laser beam. During the galvanometer transit time between targets, the shutter is also closed to prevent heating of the healthy tissue between blood vessels. Four different shutter instruments can be used.

(1) Electro-optic modulator. The electro-optic modulator has a high bandwidth and works by using the birefringence properties of a liquid crystal to alter the polarization of a beam.

(2) Acousto-optic modulator. The acoustic modulator uses compressions and rarefactions associated with an acoustic wave to form a grating that diffracts the incident light.

(3) Mechanical shutter. The energy threshold for the shutter blades must be able to withstand the high intensity laser energy.

(4) Beam deflectors. The beam deflector can be used as a shutter by deflecting the laser beam towards or away from the optical fiber. The deflector can be a mechanical galvanometer, or an electro-optic deflector, which has a much higher bandwidth.
The shutter chosen for the present Prototype is a mechanical galvanometer. The main reasons are that there is no energy loss, and 100% power modulation can be achieved. The optical setup is illustrated in Figure 5.3.

![Optical Setup Diagram](image)

**Figure 5.3: Beam “shutter”.**

### 5.1.3 Instrumentation

The images from the camera are acquired by a framegrabber (Matrox Meteor II Digital, *Matrox, Dorval, Canada*). The information acquired is processed by a computer, which determines the blood vessel targets and feeds the coordinates to the scanning galvanometers via a digital-analog converter (50 kS/s @ 18 bits/sample) on a data acquisition board (Allios board, *MIT Bioinstrumentation Lab, MA*). The computer also coordinates the strobing (and intensity) sequences, the optical shutter, and the feedback control loops.

### 5.2 Prototype 2 - Design for Miniaturization

The miniaturized system will consist of a handheld probe acting as the Smart Scalpel-to-patient interface. This probe has to be small and light for ease of use and accessibility to all parts of the body. Therefore, the number of components inside the probe has to be kept to a minimum.
In order to mimic the optical layout of the handheld version, Prototype 2 is designed with miniaturization in mind.

5.2.1 Optical Designs

The main design concept is illustrated in Figure 5.4.

![Figure 5.4: Optical design of Prototype 2.](image)

Two strobe units focus the filtered light via a dichroic mirror onto the optical fiber. The dichroic mirror has the cutoff wavelength of 600 nm (in between the 2 wavelengths) and is used to prevent light loss. The light then passes through an optical fiber to the probe. This optical setup allows a single fiber to act as if it is two separate light sources. Inside the probe, the light is reflected by a dielectric mirror (to ensure minimal light loss) onto the sapphire window. Since light from both channels illuminate from the same position, the 2 images obtained will have the
same intensity distribution and shadow details, and this allows image subtraction to be performed accurately.

The second optical fiber brings the laser beam into the probe, where the beam passes through a collimation unit, a 2-axis scanner, a focusing lens and a dichroic mirror. In order to keep the overall size small, a high resolution miniature camera is used. *All other electronics and instruments that control the system are the same as in Prototype 1.*

Another design based on the same concept is shown in Figure 5.5.

![Figure 5.5: Tunable filter design.](image)

In this design, a liquid crystal tunable filter is used to provide high frequency filtering between the 2 wavelengths. In order to match the light intensities at both wavelengths, a variable neutral density filter is placed before the tunable filter. Both these instruments must have high bandwidths and can be controlled digitally.

The final design considered is illumination using light emitting diodes (LEDs) (see Figure 5.6). Today’s InGaAlP LEDs are capable of producing a high amount of luminosity. As solid state devices, they are not subject to the maintenance requirements associated with conventional lamp and shutter technologies, and are resistant to vibration and mechanical shock. Mostly importantly, the pulse width and brightness are easily controlled electronically. Lastly, LEDs are very small and cheap, and can be easily integrated into the probe.
5.2.2 **Summary and Current Design**

Currently, the Smart Scalpel prototype is based on the first design. The reasons are that the strobe lamps can be digitally controlled very easily, and are able to produce a tremendous amount of light. The following is a picture of the Smart Scalpel prototype.

![LEDs illumination diagram](image-url)
Figure 5.7: Smart Scalpel prototype.
Chapter 6

Detecting Blood Vessels in Images

In order to detect the blood vessels in images, a line extraction algorithm is used. An important feature of the algorithm is that it must be fast and extract the mid-line of the blood vessels so that the laser spot can heat the vessels accurately. This section first discusses some of the algorithms considered, followed by detailed analysis of the selected algorithms.

6.1 Line Detection Techniques

Lines are characterized as ridges and valleys in an image. Without first stepping into the domains of ridge analysis, an initial approach for line detection is simply binary thresholding to identify the vessel regions, followed by skeletonization to extract the essential lines ([27] & [28]). Due to the presence of noise and brightness variations, the raw image can be Gaussian filtered followed by contrast balance by subtracting with a median filtered image. The actual thresholding can be performed statistically, regionally, or by relaxation.

It was found that the above approach worked well only for clean images containing high contrast and well-defined blood vessels. However, this was not the case for images taken from the human skin. As a result, the thresholding produced inconsistent results, and the detection
errors were too large. In addition, some of the regions identified did not exactly correspond to the blood vessels, which meant that the skeleton operation did not yield the actual vessel mid-line.

The published schemes on line detection can be classified into three main categories. The first category detects lines by only considering the gray values of the image. Line points are extracted by using purely local criteria, e.g., local gray value differences. Since this will generate a lot of false hypotheses for line points, elaborate and computationally expensive perceptual grouping schemes have to be used to select salient lines in the images.

The second category is to regard lines as objects having parallel edges. In a first step, the local direction of a line is determined for each pixel, and then two edge detection filters are applied perpendicular to the line. Each edge detection filter is tuned to detect either the left or right edge of the line. The responses of each filter are combined in a non-linear way to yield the final response of the operator. The advantage of this approach is that since the edge detection filters are Gaussian derivatives, the procedure can be iterated over scale-space to detect lines of arbitrary widths. However, because special directional edge detection filters have to be constructed that are not separable, the approach is computationally expensive.

In the third category, the image is considered as a height map \( z(x,y) \) and lines are extracted based on differential geometric properties. The basic idea behind these algorithms is to locate the positions of ridges and ravines in the image function. One such technique is to specifically identify the ridge edgels, which are locus of points where the image height is locally maximal in the transverse direction and the gradients are vanishing; this direction is equivalent to the minimum second directional derivative. Due to the high accuracy and computation efficiency, this method was selected for the Smart Scalpel application. The next sections will now analyze the algorithms and methods to improve computation efficiency.

### 6.2 Differential Geometric Properties of a Line

In order to detect lines in one dimension, say with a parabolic profile \( z(x) \), it is sufficient to determine the points where the first derivative \( z'(x) \) vanishes (see the figures below).
However, it is usually convenient to select only salient lines. A useful criterion for salient lines is the magnitude of the second derivative $z''(x)$ at the point where $z'(x) = 0$. In this case, a line point will have $z''(x) << 0$.

![Parabolic Line Profile](image1)

**Figure 6.1:** 1D Parabolic Line Profile.

![First Derivative Profile](image2)

**Figure 6.2:** Gradient vanishes at the center of the line.

![Second Derivative Profile](image3)

**Figure 6.3:** Large negative value at the center of the line.

These properties can be easily extended to two dimensions. In this case, the center of the line is the position where the first directional derivative in the direction perpendicular to the line should vanish, and the second directional derivative should be of large negative value. These terms will be discussed in detail in Section 6.4. It is also observed that the first derivative takes on its maximum absolute value at the edge of the line. Therefore, edges can be detected by locating the points where the first derivative is locally maximal in the direction of the gradient.
6.3 Gaussian Convolution

6.3.1 Estimating Image Derivatives

Since real images contain a significant amount of noise, it is not sufficient to estimate the first and second derivatives based only on the immediate neighbors. A better method to perform the estimates is by convolving (denoted by $\otimes$) the image with the derivatives of a Gaussian kernel:

\[
\frac{\partial F(x,y)}{\partial x} \approx F(x,y) \otimes \frac{\partial G(x)}{\partial x}, \quad (6.1)
\]

\[
\frac{\partial^2 F(x,y)}{\partial x^2} \approx F(x,y) \otimes \frac{\partial^2 G(x)}{\partial x^2}, \quad (6.2)
\]

\[
\frac{\partial^2 F(x,y)}{\partial x \partial y} \approx F(x,y) \otimes \frac{\partial G(x)}{\partial x} \otimes \frac{\partial G(y)}{\partial y}, \quad (6.3)
\]

where $G$ is a one-dimensional Gaussian function and $F(x,y)$ is the image. The Gaussian function and its derivatives are given by:

\[
G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}, \quad (6.4)
\]

\[
\frac{\partial G(x)}{\partial x} = \frac{1}{\sqrt{2\pi}\sigma} \left( -\frac{x}{\sigma^2} \right) e^{-\frac{x^2}{2\sigma^2}}, \quad (6.5)
\]

\[
\frac{\partial^2 G(x)}{\partial x^2} = \frac{1}{\sqrt{2\pi}\sigma} \left( -\frac{x}{\sigma^2} \right)^2 e^{-\frac{x^2}{2\sigma^2}} - \frac{1}{\sqrt{2\pi}\sigma^3} e^{-\frac{x^2}{2\sigma^2}}, \quad (6.6)
\]

where $\sigma$ is the standard deviation, also known as the resolution or scale-space parameter.
6.3.2 Image Smoothing and 2D Convolution

An image contains noise that must be removed prior to derivative estimations and line detection. Since noise is mostly high frequency signals, they can be attenuated using a two-dimensional Gaussian filter. This operation is also known as Gaussian smoothing. In two dimensions, the Gaussian function has a rotationally symmetric profile given by:

\[ G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}. \]  

(6.7)

A two dimensional convolution is fairly simple to implement but is computationally expensive, especially for large width. However, a two dimensional Gaussian function in Fourier domain has the following special property:

\[ G_\sigma(u,v) = e^{-\frac{u^2+v^2}{2\sigma^2}} = e^{-\frac{u^2}{2\sigma^2}} \ast e^{-\frac{v^2}{2\sigma^2}} \]

(6.8)

\[ \Rightarrow e^{\frac{\delta^2}{2\sigma^2}} \delta(y) \otimes e^{\frac{\delta^2}{2\sigma^2}} \delta(x), \]

where \( \delta \) is the impulse response. This means that a convolution with a two dimensional Gaussian can be separated into two orthogonal convolutions with one-dimensional Gaussians. This effectively reduces the number of operations per pixel from \( N^2 \) to \( 2N \). Mathematically, this is equivalent to:

\[ [\text{Image}] \otimes G_\sigma(x,y) = ([\text{Image}] \otimes G_\sigma(x)) \otimes G_\sigma(y), \]

(6.9)

where the one-dimensional Gaussian function \( G_\sigma \) is given by Equation (6.4). Due to this separable property, Gaussian convolutions are computationally efficient compared to convolutions based on other “low-pass" functions.
6.3.3 Scale-Space Parameter (Resolution)

The scale-space parameter $\sigma$ (or resolution) is an important parameter for the following reasons:

a. The choice of the resolution for smoothing creates a tradeoff between retaining the fine details of the line, which may correspond to high frequencies, and suppressing unwanted features due to noise. In general, larger values of $\sigma$ lead to better smoothing of the image but worse localization of the line.

b. The choice of $\sigma$ affects the width of the lines detected. A high resolution $\sigma$ extracts the major contours of an object, while the smaller resolutions extract finer image features. Unfortunately, when the resolution is scaled to fit the major contours, the finer contour details are often lost. Therefore, the value of $\sigma$ has to be chosen carefully to extract blood vessels of a specified width.

c. A Gaussian kernel with large $\sigma$ has a large width and more coefficients. Hence, more mathematical operations have to be performed during each convolution operation and this will slow down the image processing. (A method using recursive filters will be described later that will solve this problem.)

d. The choice of $\sigma$ affects the maximum negative response of the second directional derivative in scale-space [29]. The higher the resolution $\sigma$, the smaller the second derivative will be. Hence, the threshold to select salient lines will have to be set to an accordingly smaller value. These concepts will become clear when they are discussed in the later sections.

Some of the problems described above can be solved by detection at multiple scales of resolution [30] (or detection at repeated convolutions using small resolutions). However, this method is computationally expensive and is not practical in a time-critical application.
6.3.4 Implementing Gaussian Convolution

Two methods of implementing the Gaussian convolution are used.

6.3.4.1 Direct Convolution (FIR)

The direct convolution produces an output sequence $y(i)$ that is the weighted sum of the current and past inputs $x(i)$, where $h_k$ are the coefficients of the convolution kernel of width $N$:

$$y(i) = \sum_{k=0}^{N-1} h_k x(i-k) .$$  \hfill (6.10)

This operator is also known as a Finite Impulse Response (FIR) filter because the impulse response has finite width, i.e. it is non-zero only over a finite interval [31]. FIR suffers from the fact that it is computationally expensive since the width of the kernel $N$ is generally large, and the number of operations per point is proportional to the number of coefficients.

The coefficients of the Gaussian kernel can be obtained by sampling the discrete Gaussian function, and then truncating at suitably small values. However, the filter created exhibits the poor property of falling off slowly away from the center. A better method is to use the coefficients of the binomial expansion. These coefficients can be calculated by:

$$h_k = \frac{1}{2^N} \frac{N!}{k!(N-k)!} , \quad k = 0..N ,$$  \hfill (6.11)

where $N+1$ is the width of the kernel. The standard deviation $\sigma$ of the equivalent Gaussian kernel is approximately $\sqrt{N/4}$. Due to the fact that the binomial filter approximates that of the Gaussian especially at large $N$, the binomial filter can also be separately convolved.
Recursive Filtering (IIR)

An Infinite Impulse Response (IIR) filter produces an output \( y(i) \) that is the weighted sum of the current and past inputs \( x(i) \) and past outputs \( y(i-k) \) [32]. The IIR has an impulse response of infinite width and is implemented recursively. This filtering technique deals with the determination of the coefficients \( a_k \)'s and \( b_k \)'s of the following recursive system of order \( n \):

\[
y(i) = \sum_{k=0}^{m-1} b_k x(i-k) - \sum_{k=1}^{n} a_k y(i-k),
\]

(6.12)

so that the transfer function best approximates, in a least-square criterion, that of the non-recursive FIR system.

The key to the approach of the IIR is the derivation of an exponentially based filter family that closely approximates the Gaussian filters of the same resolution \( \sigma \), and then to implement the filters in a recursive way. The method is discussed in detail in Appendix A.

Dealing with the causal recursive system given by Equation (6.12) instead of the non-recursive system (Equation (6.10)) reduces the number of operations per output element from \( N \) to \( m+n \) (where \( m+n < N \)). In addition, recursive filtering is done with a fixed number of operations per point independent of the resolution \( \sigma \) of the Gaussian function. Therefore, it is computationally efficient and requires much fewer computational steps than direct or frequency domain convolution using the Fast Fourier Transform.

The disadvantage of the IIR filter is that the results are less accurate than a FIR. However, this error can be reduced by using a higher order filter that contains more coefficients but is slower.
6.3.5 Efficient Discrete Derivatives Approximation

Lindeberg [33] has shown that the scale-space properties in the continuous case transfer directly to the discrete domain, and operators that commute before discretization also commute after discretization. One important computational implication of this is that image derivatives can be computed directly from smoothed data, and that will give the same result as convolution with the corresponding derivative kernel. This is based on a linear, shift-invariant system, and the fact that convolution is commutative.

For example, if 5 different derivatives are to be approximated, there is no need to separately convolve the image with 5 large support derivative kernels. The same result can be obtained by smoothing the image with a large support Gaussian kernel, followed by computing the derivatives using 5 small support difference operators. On hindsight, this means that there is no need for re-doing the smoothing part of the transformation for 4 extra times. The difference operators for computing the derivatives (after smoothing) are given by:

\[
\frac{\partial F(x, y)}{\partial x} \approx \frac{1}{2\epsilon} \left[ F(x+1, y) - F(x-1, y) \right], \quad (6.13)
\]

\[
\frac{\partial F(x, y)}{\partial y} \approx \frac{1}{2\epsilon} \left[ F(x, y+1) - F(x, y-1) \right], \quad (6.14)
\]

\[
\frac{\partial^2 F(x, y)}{\partial x^2} \approx \frac{1}{\epsilon^2} \left[ F(x-1, y) - 2F(x, y) + F(x+1, y) \right], \quad (6.15)
\]

\[
\frac{\partial^2 F(x, y)}{\partial y^2} \approx \frac{1}{\epsilon^2} \left[ F(x, y-1) - 2F(x, y) + F(x, y+1) \right], \quad (6.16)
\]

\[
\frac{\partial^2 F(x, y)}{\partial x \partial y} \approx \frac{1}{4\epsilon^2} \left[ F(x+1, y+1) + F(x-1, y-1) - F(x+1, y-1) - F(x-1, y+1) \right]. \quad (6.17)
\]

From all the methods proposed in this section, it can be deduced that the most efficient way of computing the derivatives is to smooth the image using IIR filters in a separable manner, following by derivative approximation using the small support difference equations.
6.4 Ridge Edgel Detection

Ridge edgel detection involves finding local extrema in special directions. Therefore, we need to introduce the concept of directional derivatives, which measures how functions vary when restricted to a subset of its domain. As mentioned in Section 6.2, the zeros of first directional derivatives in the direction perpendicular to the ridge lines are candidate edgel points. This perpendicular direction corresponds to the direction of largest principle curvature, which is also the direction of minimum second directional derivative. In the case of a ridge, this derivative has a large negative value compared to its orthogonal principal curvature, which has magnitude close to zero.

6.4.1 Directional Derivatives

Let \( \hat{n} \) be a vector pointing in the direction of largest principle curvature. Given an image function \( F \), the first directional derivative in the direction \( \hat{n} \) is defined by:

\[
(\hat{n} \cdot \nabla) F(x, y) = \hat{n}_x \frac{\partial}{\partial x} F(x, y) + \hat{n}_y \frac{\partial}{\partial y} F(x, y)
\]  

(6.18)

Similarly, the second directional derivative in the direction \( \hat{n} \) is defined by:

\[
(\hat{n} \cdot \nabla)^2 F(x, y) = \begin{bmatrix} \hat{n}_x & \hat{n}_y \end{bmatrix} \begin{bmatrix} \frac{\partial^2}{\partial x^2} F(x, y) & \frac{\partial}{\partial x \partial y} F(x, y) \\ \frac{\partial}{\partial x \partial y} F(x, y) & \frac{\partial^2}{\partial y^2} F(x, y) \end{bmatrix} \begin{bmatrix} \hat{n}_x \\ \hat{n}_y \end{bmatrix}
\]

(6.19)

where the \( 2 \times 2 \) matrix is known as the Hessian matrix.
The vector \( \hat{n} \) and the magnitude of the second directional derivative can be computed from the eigenvectors and eigenvalues (K) of the Hessian matrix. In the case of a ridge, the eigenvector that corresponds to the minimum eigenvalue (K\(_{\text{min}}\)) is in the direction of the largest principle curvature.

### 6.4.2 Extreme Points

With the directional derivatives thus defined, the image \( F(x,y) \) has a local maximum ([34] & [35]) if the following conditions are true:

- \(|K_{\text{min}}| >> |K_{\text{max}}| \) and Hess\( (F) \) is negative definite.
- \((\hat{n} \cdot \nabla)F(x,y) = 0\)

### 6.5 Algorithm Implementation

This section discusses an algorithm for implementing the ridge detection technique discussed in the previous section.

#### 6.5.1 Taylor Polynomial Approximation

To determine the location of a line, one could use a zero crossing detector to find points where the first directional derivative vanishes. However, this would yield the position of the line only with pixel accuracy. In order to overcome this, this algorithm proposes to detect the lines by locally approximating the image function \( F(x,y) \) by its second order Taylor polynomial ([36] & [29]). The coefficients of the polynomial are determined by convolving the image with the derivatives of a Gaussian kernel. This polynomial is given by:
\[
F(x,y) = R + \left( R_x x + R_y y \right) + \frac{1}{2} \left( R_{xx} x^2 + 2 R_{xy} xy + R_{yy} y^2 \right), \tag{6.20}
\]

where \( R_x, R_y, R_{xx}, R_{xy} \) and \( R_{yy} \) are the locally estimated derivatives at \((x,y)\).

Curvilinear structures in 2D can be modeled as curves \( s(t) \) that exhibit a characteristic 1D line profile in the direction \( n(t) \) perpendicular to the line. Based on ridge analysis, the direction in which the second directional derivative \( s''(t) \) takes on its maximum absolute value will be used as the direction \( n(t) \). This direction can be determined by calculating the eigenvalues and eigenvectors of the Hessian matrix (which is negative definite at a line point):

\[
H(x,y) = \begin{pmatrix}
R_{xx} & R_{xy} \\
R_{xy} & R_{yy}
\end{pmatrix}.
\tag{6.21}
\]

The calculation can be done in a numerically stable and efficient way by using one Jacobi rotation to annihilate the \( r_{xy} \) term. The eigenvector that points in the direction \( n(t) \) is then given by the eigenvector that corresponds to the maximum absolute eigenvalue. Now, let this eigenvector be given by \( (n_x, n_y) \) with \( \| (n_x, n_y) \| = 1 \). A quadratic polynomial is then used to determine whether the first directional derivative along \( (n_x, n_y) \) vanishes within the current pixel. This zero-crossing point is given by:

\[
\begin{pmatrix}
p_x \\
p_y
\end{pmatrix} = \begin{pmatrix}
n_x \\
n_y
\end{pmatrix},
\tag{6.22}
\]

where

\[
t = -\frac{R_x \hat{n}_x + R_y \hat{n}_y}{R_{xx} \hat{n}_x^2 + 2 R_{xy} \hat{n}_x \hat{n}_y + R_{yy} \hat{n}_y^2}.
\tag{6.23}
\]

The point is declared a line point if this position falls within the pixel’s boundaries, i.e. \((p_x, p_y) \in [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]\). The magnitude of the second directional derivative can be used
to select salient lines. The mathematical theories and proofs for the above equations are given in Appendix B.

6.5.2 Hysteresis Thresholding

Pixels that are marked as line points have different magnitudes (or strengths) of the second derivative along \((n_x, n_y)\). Line points with high saliency have higher strengths. Rather than simply selecting a global threshold value to extract the salient line, which is highly inaccurate, a thresholding technique using *hysteresis* is used.

Hysteresis thresholding uses a high threshold \(T_H\) and a low threshold \(T_L\). Any pixel in the image that has a value greater than \(T_H\) is presumed to be a line pixel, and is marked as such immediately. Then, any pixels that are *connected* to this line pixel and that have a value greater than \(T_L\) are also marked as line pixels. For example, if a line pixel is assigned as white (pixel value 255) and non-line pixel as black (pixel value 0), then this algorithm is given mathematically as:

\[
\text{Result}(x,y) = \begin{cases} 
\text{Pixel}(x,y) > T_H & \Rightarrow \text{Result}(x,y) = 255 \\
\text{Pixel}(x,y) > T_L \text{ and } \text{Pixel}(\text{neighbor}) > T_H & \Rightarrow \text{Result}(x,y) = 255 \\
\text{Pixel}(x,y) > T_L \text{ and } \text{Pixel}(\text{neighbor}) < T_H & \Rightarrow \text{Result}(x,y) = 0 \\
\text{Pixel}(x,y) < T_L & \Rightarrow \text{Result}(x,y) = 0 
\end{cases}
\]

(6.24)

The marking of neighbors is performed *recursively* for computational efficiency.

The choice of the high threshold value \(T_H\) depends on the resolution of the Gaussian kernel (see Section 6.3.3). It also depends on other factors such as the brightness values of the lines and background. Therefore, it is desirable to have a method to estimate \(T_H\). It is found that a good approximation is the 99% brightness value of the histogram of the second derivative magnitude image. Several other estimation schemes, such as entropy and statistical, have also been tested on numerous images, but the 99%-histogram method comes closest to producing a
good estimator. It does not always give the optimal value because it depends on the density of lines detected, but it will give a number close to the best value.

### 6.5.3 Result of the Algorithm

A representative result is shown in Figure 6.4. The resolution (σ) of the Gaussian kernel used is 3.0; hysteresis thresholding is carried out at $T_H = 35$ and $T_L = 21$ after an 8-bit dynamic range expansion.

![Figure 6.4](image)

Figure 6.4: (a) Original Image. (b) 2nd Derivative image. (c) After hysteresis thresholding.

There are numerous line detection methods that use similar concept. Some methods even extend the derivative analysis up to 3rd order [37]. In general, the robustness and computational efficiency of these methods vary, depending on what the application. In the detection of blood vessels from skin images, it is found that the polynomial approximation technique gives excellent results at relatively high speeds.

### 6.6 Detection of Line Width

The pulse width of the laser depends on the thermal relaxation time of the blood vessels, which in turn depends on the vessel diameter. Therefore, it is desirable to compute the width of the blood vessel in the image so that the laser parameter can be adjusted during treatment.
One algorithm to detect line width made use of the fact that the scale-space parameter $\sigma$ of the Gaussian influences the scale-normalized response of a line as a function of line width. This algorithm can be performed by iteration through scale-space while selecting the scale that yields the maximum response of the desired line width. However, large number of convolutions makes this computation extremely expensive, especially if one is only interested in lines in a certain range of widths. Furthermore, this approach will only yield a coarse estimate of the line width, since the scale-space is quantized in rather coarse intervals.

A line (blood vessel) is bounded by an edge on each side. Therefore, one method to compute the width of each line point is to search for the edge pixels to the left and right in the direction $n(t)$ perpendicular to the line. The line width is then equal to the total distance to the edge pixels in the left and right direction (Figure 6.5).

![Figure 6.5: Method of line width detection](image)

The Canny edge detector ([38] & [39]) has been identified as the most suitable algorithm for edge finding. The first reason is that this detector has been widely recognized as a standard due to its accuracy. The second reason is that this edge detector uses the Gaussian first derivatives, which are in fact computed as a “by-product” during line detection. Hence, the computationally expensive convolutions do not have to be performed. In addition, the normal
vector \( n(t) \), which is needed to search for the left and right edges, also need not be computed since it was already computed during line detection.

This proposed algorithm not only guarantees at least single-pixel accuracy in line width detection, but also performs the computation efficiently.

### 6.6.1 Edge Detection (Canny Filter)

The Canny edge detector consists of three stages: gradient and direction computation, non-maximum suppression and hysteresis thresholding.

#### 6.6.1.1 Gradient and Direction

As mentioned in Section 6.2, the first derivative takes on its maximum absolute value at the edge of the line. Therefore, edges can be detected by locating the points where this gradient magnitude is locally maximal in the direction of the gradient. The gradient magnitude at a pixel \((x,y)\) is defined as:

\[
G(x, y) = \sqrt{\left( \frac{\partial F(x, y)}{\partial x} \right)^2 + \left( \frac{\partial F(x, y)}{\partial y} \right)^2}. \tag{6.25}
\]

There is no need to compute these partial derivatives since they were already computed during the line detection process. To speed up computation, this gradient magnitude can also be approximated as the sum of the absolute values of the partial derivatives.

The direction of the gradient at each pixel \((x,y)\) can be estimated as:

\[
\theta(x, y) = \tan^{-1}\left( \frac{\partial F(x,y)}{\partial y} / \frac{\partial F(x,y)}{\partial x} \right). \tag{6.26}
\]
In non-maximum suppression (NMS), the edges are thresholded based on the direction of the gradient at each pixel. The basic principle is that the magnitude of the gradient at an edge pixel should be greater than that of its neighbors that lie on opposite sides and along the same gradient direction.

The algorithm uses a nine-pixel neighborhood centered on pixel $P_{x,y}$ (see Figure 6.6). The normal to the edge direction is shown as an arrow, and has components $(u_x, u_y)$. This is the direction that non-maximum suppress on the gradient magnitude needs to be implemented. In general, this direction does not point directly to any pixels $P_{i,j}$, where the gradient magnitude is known. Therefore, it is necessary to estimate the gradient magnitude at a point between the two pixels which lie closest to the line through $P_{x,y}$ in the direction $u$. Two of these points are marked as “+” in the figure below.

If it is assumed that the gradient changes linearly as a function of position, then the gradient magnitude can be interpolated using the vector $u$. In Figure 6.6, the gradient must be interpolated from the two points given by $P_{x,y+1}$ and $P_{x+1,y+1}$. The value of this interpolated gradient is then given by:
\[ G_1 = \frac{u_x}{u_y} G(x + 1, y + 1) + \frac{u_y - u_x}{u_y} G(x, y + 1) \]  
\[ (6.27) \]

where \( G(i,j) \) is calculated from Equation (6.25). Similarly, the interpolated gradient at a point on the opposite side is:

\[ G_2 = \frac{u_x}{u_y} G(x - 1, y - 1) + \frac{u_y - u_x}{u_y} G(x, y - 1) \]  
\[ (6.28) \]

The point \( P_{x,y} \) is marked as an edge if \( G(x,y) > G_1 \) and \( G(x,y) > G_2 \). In general, there are eight major cases to check for, and some shortcuts can be made for efficiency’s sake.

### 6.6.1.3 Hysteresis Thresholding

As in line detection, hysteresis thresholding is implemented to extract the salient edges more accurately. This is done by immediately accepting edge points that have a gradient magnitude larger than \( T_H \) and rejecting points that have a gradient magnitude smaller than \( T_L \). All other edge points are accepted if they are connected to accepted points by a connected path.

### 6.6.2 Moving in the Direction of Line Normal

Once the edges are located, then moving in the positive \( n(t) \) direction until an edge pixel is found will yield “half” the width \( W_L \), and performing the same operation in the reverse direction will yield the other “half” \( W_R \). The width of the line is then given by the sum of the 2 “half” widths \((W_L + W_R)\).

If one half of the width \( W_L \) is not detected due to missing edge pixels, then it can be assumed to have the same width as its opposite half \( W_R \). Also, in the occasional event that both
“half” widths are not detected, then the width of the line can be assumed to be the same as its nearest neighbor.

6.6.3 Localization of Edge Pixels

Localization of edge refers to how close the detected edge pixels are to the actual edge position. In general, the localization of edge pixels depends on the scale ($\sigma$) of the Gaussian convolution kernel, and there is a tradeoff between edge localization and edge detection accuracy. A small value of $\sigma$ leads to good localization but may cause erroneous detection of false edges because the small amount of smoothing fails to filter away most of the noise. On the other hand, a large value of $\sigma$ leads to poor localization because the Gaussian function is too planar at the center to define the edge position accurately. Poor edge localization is of greater concern than poor edge detection because the false edges detected are usually outside the line (blood vessel) boundaries, which does not affect the accuracy of the width detected. In general, a value of $\sigma$ between 1 and 3 produces a good tradeoff.

Recall in Section 6.6 that the Gaussian first derivatives from the line detection can be used in the Canny edge detector. If the $\sigma$ used to detect the lines is greater than 3, then separate convolutions (with $\sigma < 3$) have to be performed again to detect the edges accurately. This will cause the entire algorithm to be much slower, but it is unavoidable.

6.6.4 Results of Width Detection

The algorithm is tested on a series of images, and proved to be successful. One such test image is an “$\infty$” symbol with increasing widths from the center:
The lines are accurately detected at the center of the symbol, and the brightness of the lines represents the width detected. Notice that the detected lines get brighter as they approach the top (and bottom), indicating the detection of larger widths. This result is also verified by analyzing the pixels individually.

### 6.7 Image Noise

All image acquisition processes are subject to noise of some type. There are two types of noise that are of specific interest in image analysis. *Signal-independent* noise is a random set of gray levels that is statistically independent of the image data. They are added to the pixels in the image to give the resulting noisy image. This kind of noise occurs when an image is transmitted electronically from one place to another. If $A$ is a perfect image and $N$ is the noise that occurs during transmission, then the final image $B$ is:

$$B = A + N.$$  \hspace{1cm} (6.29)

$A$ and $N$ are statistically unrelated to each other. The noise image $N$ could have any statistical properties, but a common assumption is that it follows the normal distribution (central-limit theorem) with a mean of zero and some measured or presumed standard deviation. This type of noise is referred to as a *Gaussian* noise.
The second major type of noise is called *signal-dependent* noise. In this case, the level of the noise value at each point in the image is a function of the gray level there. The grain seen in some photographs is an example of this sort of noise, and it is generally harder to deal with than signal-independent noise. Fortunately, it is of less importance, and becomes manageable if the image is sampled properly.

### 6.7.1 Adding Gaussian Noise to an Image

Noise can be artificially added to an image. Deliberately corrupting an image with noise allows us to test the resistance of an image processing operator to noise and assess the performance of various noise filters. Since noise is dominantly high frequency signal, it can be suppressed using low pass filters such as Gaussian or median filters.

To produce a noisy image \( B \) from an original image \( A \), Gaussian noise with a mean \( \mu_n \) and a variance \( \sigma_n^2 \) can be added by:

\[
B = A + (\sigma_n * \text{Rand}) + \mu_n ,
\]

where \( \text{Rand} \) is a normally distributed random number. This task can be performed in Matlab under the function “imnoise( )” [40].

### 6.8 Testing the Algorithms against Noise

Noise was added to an original image to test the effectiveness of the line detection algorithm. The results are shown in the following figures.
<table>
<thead>
<tr>
<th>Characteristic of Noise Added</th>
<th>Image</th>
<th>Result of algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>No noise added</td>
<td><img src="image1" alt="Image" /></td>
<td><img src="result1" alt="Result" /></td>
</tr>
<tr>
<td>$\mu_n = 0.0$</td>
<td><img src="image2" alt="Image" /></td>
<td><img src="result2" alt="Result" /></td>
</tr>
<tr>
<td>$\sigma_n = 0.0$</td>
<td><img src="image3" alt="Image" /></td>
<td><img src="result3" alt="Result" /></td>
</tr>
<tr>
<td>$\mu_n = 0.0$</td>
<td><img src="image4" alt="Image" /></td>
<td><img src="result4" alt="Result" /></td>
</tr>
<tr>
<td>$\sigma_n = 0.0008$</td>
<td><img src="image5" alt="Image" /></td>
<td><img src="result5" alt="Result" /></td>
</tr>
<tr>
<td>$\mu_n = 0.0$</td>
<td><img src="image6" alt="Image" /></td>
<td><img src="result6" alt="Result" /></td>
</tr>
<tr>
<td>$\sigma_n = 0.0002$</td>
<td><img src="image7" alt="Image" /></td>
<td><img src="result7" alt="Result" /></td>
</tr>
<tr>
<td>Histogram rescaled to reduce contrast</td>
<td><img src="image8" alt="Image" /></td>
<td><img src="result8" alt="Result" /></td>
</tr>
</tbody>
</table>

Table 6.1: Results of tests against different characteristics of noise.
It can be seen that the algorithm performed well in the presence of noise. By increasing the resolution (\(\sigma\)) of the Gaussian operator, the noise can be further suppressed before the line detection begins. However, this is done at the expense of losing image details.

It is also observed that changing the image contrast has almost no effect on the accuracy of the algorithm. Most of the lines that are nearly visible to the eyes are still being detected. Therefore, the algorithm is not greatly affected by the lack of contrast.

### 6.9 Speed of Image Analysis

The speed of the algorithm based on the derivative approximation method discussed in Section 6.3.5 is tested on the image in Figure 6.4. The image size is 400 by 400 pixels and the test is performed on an Intel Celeron® 466 MHz computer. The processing times are summarized in Table 6.2 below.

<table>
<thead>
<tr>
<th>No.</th>
<th>Process</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Line Detection: (\sigma = 3.0)</td>
<td>2.83</td>
</tr>
<tr>
<td>2.</td>
<td>Line and Width Detection: (\sigma = 3.0)</td>
<td>4.87</td>
</tr>
<tr>
<td>3.</td>
<td>Line and Width Detection: (\sigma = 4.0)</td>
<td>5.58</td>
</tr>
</tbody>
</table>

Table 6.2: Processing speeds of various operations.

With the presently available computing power, the processing times are not fast enough to achieve the desired bandwidth. The speed can be greatly improved with the use of faster computers; at the time of writing, Intel Pentium® 4 processor at 1.5 GHz is already available. The next chapter presents a method to achieve high bandwidths based on the slow processing power presently available.
6.10 Summary

In this chapter, a method to extract curvilinear structures is presented. The approach is based on the differential geometric properties of an image function. For each pixel, the second order Taylor polynomial is computed by convolving the image with Gaussian filters. Line points (representing capillary centers) are required to have a vanishing gradient and a high curvature in the direction perpendicular to the line. An edge detection method is also used to determine the width of the blood vessels. This is done by computing the distances between the edges that bound the blood vessel. In order to achieve high computation speed, the derivatives are estimated using IIR Gaussian filters and small support difference operators.
Chapter 7

Imaging Strategy

The expensive computation associated with convolution operations limits the bandwidth of the system. This problem can be solved in 3 ways: software, hardware and strategy. As shown in the previous chapter, the software solution is the use of IIR filters for convolution. The hardware solution is simply the use of faster processor and instruments. This section will discuss the third solution, which is the design of a suitable imaging strategy to meet the bandwidth requirement.

7.1 Timing Analysis and Bandwidth

To operate at a bandwidth of 6 Hz, the time required for each feedback loop is 167 ms. If the dwell time of the laser on each target is 10 ms, the laser can only treat a maximum of 16 targets in an image field which contains potentially up to 365 targets (using a camera resolution of $755 \times 484$ pixels and assuming the blood vessels cover 0.1% of the image field). This is equivalent to treating only 4.4% of the blood vessels in 1 operation loop, and without taking into account the time taken for grabbing the images and processing them. In another words, it is impossible for the laser scalpel to treat the entire region in 1 operation loop. Therefore, the
region will have to be divided into many smaller regions, where the treatment for each small region will be carried out within 167 ms.

The bandwidth of the system is governed by the following equation:

\[
Bandwidth = \frac{I}{T_{\text{grab}} + T_{\text{galvo}} + T_{\text{system}} + T_{\text{laser}} + T_{\text{proc}}},
\]

where:

1. \(T_{\text{grab}}\) is the time taken by the program to grab an image. It depends on both the speed of the camera and the frame grabber, and is currently 66.7 ms per grab. 2 grab operations from the green and red channels are required per operation loop, and this adds up to 133 ms;
2. \(T_{\text{galvo}}\) is the step response of the scanning galvanometers, which have settling times on the order of 10 ms;
3. \(T_{\text{system}}\) is the delay due to hardware systems such as the data acquisition board;
4. \(T_{\text{laser}}\) is the time required for the laser to thermally treat \(N\) blood vessel targets. This value depends on the treatment laser pulse width, which is approximately 10 ms per target;
5. \(T_{\text{proc}}\) is the time taken for image processing. It depends on the processing speeds of both the computer and the efficiency of the image analysis algorithm, and constitutes the main bottleneck in the feedback loop due to the high amount of computations required.

The times taken up by \(T_{\text{grab}}, T_{\text{galvo}}\) and \(T_{\text{system}}\) only allow a window of less than 34 ms (in 167 ms) for image processing and laser treatment (\(T_{\text{laser}}\) and \(T_{\text{proc}}\)). Therefore, dividing the image field into smaller regions to process is a good way to speed up the process loop.
7.2 Region Division

7.2.1 Equal Regional Division

A simple division of the image field into equal squares or rectangles is not enough to deal with the speed problem. The situation is described in the example below.

Figure 7.1: Six regions in image field.

Figure 7.1 shows an image field divided into 6 equal regions. Region 4 contains more targets than the other regions while region 6 contains no target at all. This scenario leads to the following observations:

1. Convolution and other image processing operations will be wasted on region 6, which contains no target at all.
2. Image processing need not be performed on the entire regions 1 and 5, which contains very few targets.
3. The processing time for regions with more targets (e.g. region 4) will be longer than those with fewer targets. Thus, the processing times of each regions will be unequal.
4. Regions with more targets are laser treated for longer times.

Therefore, it is clearly shown that equal division will lead to both time wasting and non-uniform operation times for each region. In this case, the blood vessels cannot be consistently tracked. On the other hand, the image field can be divided into much smaller regions so that the
inconsistencies can be scaled down. However, these regions may have to be so small that there will be too many of them to make the whole operation efficient and practical.

### 7.2.2 Regional Division based on number of targets

A better approach is to divide the image field into regions that contain an *equal* number of targets. Parts of the image that contain no target will be ignored. This will solve all the problems encountered in the previous approach. However, in order to do this, the locations of the targets have to be estimated beforehand. This means that the tissue has to be scanned once to determine the approximate target locations before the *actual* treatment with feedback can take place. This method will take a longer treatment time but it guarantees that the feedback bandwidth will be fast.

### 7.3 Image Processing Strategy

The following sub-sections will discuss the steps to implement the regional division scheme.

#### 7.3.1 Step 1: First Pass Imaging

This involves grabbing a red and green image to perform background subtraction and image analysis to identify the blood vessels. The locations of the blood vessels are marked as \( S(x,y) \), which are referred to as the *seed* pixels. This first pass imaging will take approximately a few seconds.
Blood vessels

7.3.2 Step 2: Grow Area from Seed Pixels

Since the patient may move after the first pass imaging, the subsequent positions of the blood vessels may not be in the same position $S(x,y)$. However, these new positions will be very near the original positions if the patient moves little, or if the Smart Scalpel is held securely to the skin. Therefore, an area can be “grown” around the seed pixels, which will represent future vessel positions. These possible targets will be called $T(x,y)$.

In order to determine how much area is to be “grown”, we need to estimate how much the patient may move. As mentioned in Section 2.3, the tremor displacement is approximately 370 μm. Now, assuming that 370 μm represents 10 pixels in the image (actual value has to be calibrated using reticles), then an area 20 pixels wide has to be grown around the seed pixels.

Figure 7.2(a) Blood vessels. (b) Vessel coordinates (seeds) extracted.

Figure 7.3: Grown area from seed pixels.
From this point onwards, all subsequent image processing operations will only be performed on \( T(x,y) \). This will ensure that no operation is wasted on parts of the region that contain no target. For example, if the blue region in Figure 7.3 covers 20% of the image, then the subsequent image processes on the whole image will be 5 times faster.

### 7.3.3 Step 3: Regional Division

The image field will now be divided into rectangular regions that contain the same number of seed pixels \( S(x,y) \). The image is first divided into several equal rows, and then the program pointer will count the number of seed pixels within each row (from left to right) and define a boundary when a specified number \( N \) is reached. In this way, rectangular regions of equal height but unequal width will be defined, where each region contains the same \( N \) number of targets.

![Region division](image)

**Figure 7.4: Region division.**

In order to deal with the case where the blood vessels in one region may move to its neighboring region during treatment, the regions can be made to overlap each other by a number of pixels.
7.3.4 Step 4: Feedback Operation

With the regions now defined, the final step is the feedback operation, where the operation loops are sequentially performed on each region.

The time taken by one operation loop depends on the size of the region it process, which is determined by the number of “seed” targets $N$ in each region. In general, a large value of $N$ will lead to big regions and low bandwidth, while a small value of $N$ will create a large number of regions which will slow down the overall process.

7.4 Laser Pulse Width Based on Blood Vessel Diameter

The treatment is more effective if the laser pulse width is based on the thermal relaxation time of the blood vessels. Larger diameter vessels have longer thermal relaxation times and consequently require longer laser pulse widths. Based on the image algorithm discussed, the diameter of the blood vessels can be approximated by determining the distance between the edges on opposite sides of the vessel. If this algorithm is implemented during the feedback operation, it will reduce the bandwidth due to the extra computations required.

If the system bandwidth is not to be affected, one solution is to perform the width detection during the first pass imaging stage (Section 7.3.1), and “grow” different types of regions that correspond to a specified range of widths. The algorithm is proposed as follows:

1. Assign the width of lines into several groups, where each group has its specified range, e.g. from 10 $\mu$m to 30 $\mu$m.
2. Perform the first pass image processes to determine the lines and their widths.
3. Group the line pixels according to their width, and “grow” different types of regions around them depending on their group. This step is exactly the same as Section 7.3.2 except that the regions grown are tagged to represent different widths.
4. During the feedback operation, there is no need to compute the width anymore; the laser pulse width is adjusted depending on the type of region that the blood vessel is in. In this way, there is no change to the treatment bandwidth.

This algorithm assumes that blood vessels of a particular width stay within their assigned regions during treatment. This assumption is reasonable since the width of blood vessels changes only gradually, except at distributary junctions. This algorithm will become problematic if the blood vessels are located very close together such that the different regions overlap.

### 7.5 Alternative Strategies

#### 7.5.1 Dynamic Region Strategy

The strategy proposed above is used to meet the bandwidth requirement in the present instrument setup. However, if faster computer and instruments are used so that speed is no longer restricted, then the imaging strategy should be changed. Laser treatment should now be performed immediately without implementing a first pass scan because the image processing is able to handle a much larger area. In this case, a dynamic region approach can be employed:

1. At the start of treatment, a reasonably large region is defined at one corner of the image. Image analysis and laser treatment will only be performed inside this region.
2. At the first treatment loop, once \( N \) number of targets and their diameters have been detected in the region, the image analysis will terminate and the laser treatment will begin. The value of \( N \) is chosen to meet the bandwidth requirement.
3. After the \( N \) targets have been treated, the region will now be dynamically shifted to cover the nearest area that has not been processed. Then, the image analysis and laser treatment on \( N \) newly detected targets will begin on the new position.
4. This region will shift systematically until the entire image has been covered.

As before, the purpose of defining a region is for computation efficiency, and the purpose of treating a fixed \( N \) number of targets is to achieve uniform operation times per treatment loop.
7.5.2 Two-Point Tagging

Another strategy to achieve a high treatment bandwidth is two-point tagging of the imaging tissue. If the object is tagged with two markers within the field of view, then they serve as fiducial points and any translation and rotation of the object can be calculated from the relative movement of the line formed. Therefore, if the shape of the blood vessels (and their diameters) with respect to the markers is determined before the feedback-treatment begins, then the time-consuming line detection algorithm does not have to be performed. The only image analysis required is to determine the positions of the markers by finding their centers of mass (Com). This algorithm can be easily performed by thresholding to isolate the markers, followed by a Com finder or morphological skeletonization. Matrix algebra can then be performed to determine the translation and rotation.

This method assumes that the shape of the object is preserved, which is reasonable if the patient is held sufficient still so that the tissue is not stretched or compressed during treatment. Since the field of view is only 17 mm by 13 mm, the markers must be small and placed at very short distances apart. Therefore, this poses the problem of placing these markers over an extensive area of the skin. However, if the lesion area is small enough, then this technique becomes feasible.

7.6 Laser Scanning Strategy

There are two strategies for laser treatment, raster scan and line tracing. Raster scan involves a row-by-row scanning of the regions, and the beam shutter is opened when the scan is directly above the blood vessels. This technique is illustrated in Figure 7.5a where the red line represents the computer program searching for targets from top-down, and the white circles represent the targets. The scan is implemented after all the blood vessel coordinates within the region are determined. Since it may not be necessary to apply laser onto every pixel targets, the treatment process can be speeded up if the laser scan is skipped every few rows and columns.
Unlike raster scan, which is intermittent, line tracing involves tracing the laser beam along blood vessels (Figure 7.5b). The tracing along a particular blood vessel will be continuous as long as the imaging algorithm detects it as a continuous line. Although the line tracing algorithm appears to be more complicated, it does not lose computational efficiency to the raster scan because it can be easily incorporated into the hysteresis thresholding algorithm (Section 6.5.2). This means that the image processing program will locate the line pixels and execute the laser scan simultaneously. In terms of the laser treatment, this scheme is better than the raster scan because the laser will heat along a single blood vessel continuously, thus ensuring that the blood vessel is exposed to the thermal energy for a longer period of time.

In terms of speed, both strategies may outdo each other depending on the distribution of blood vessels. The distribution that allows the galvanometers to rotate through less overall angle will complete the scan faster.

### 7.7 Summary

A regional division strategy is used to overcome the slow image processing and to achieve the high bandwidth requirement of the Smart Scalpel. This technique involves performing an initial scan to locate the positions of the blood vessels, and then dividing the imaging field into smaller regions that are processed individually.
Chapter 8

Smart Scalpel Characterization

This chapter describes the methods of quantifying the performance and optical characteristics of the Smart Scalpel and gives the results.

8.1 Resolution and Modulation Transfer Function

Resolution is a measure of the imaging system’s ability to discern fine details. It can be expressed as the smallest distance between 2 objects that are resolvable. Therefore, resolution can often be represented in terms of the number of line pairs per millimeter (lp/mm). Due to this representation, resolution is often referred in terms of frequency. The inverse of this frequency yields the spacing between the two resolvable objects (in μm). This specification is used for both camera and lens.

The modulation transfer function (MTF) is used to quantify the resolution of the optical system. It shows how well frequency information is transferred from the object to the image. The MTF is computed by taking the ratio of the modulation amplitude at the output to the modulation at the input, and is calculated over a range of spatial frequencies. The modulation amplitude is given by:
\[ Modulation = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}} \tag{8.1} \]

where \( I_{\text{max}} \) is the maximum intensity and \( I_{\text{min}} \) is the minimum intensity in the image.

### 8.2 Camera Resolution

The resolution of the optical system depends on the resolution of the camera and the quality of the lens. Most of the time, the limiting factor is the camera. The camera used in the Smart Scalpel is a Pulnix TM 9701 CCD camera (monochrome) with the following specifications:

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Pixels</td>
<td>768 (H) ( \times ) 484 (V)</td>
</tr>
<tr>
<td>Sensor Size (H) ( \times ) (V)</td>
<td>8.91 mm ( \times ) 6.58 mm</td>
</tr>
<tr>
<td>Cell Size (H) ( \times ) (V)</td>
<td>11.6 ( \mu )m ( \times ) 13.6 ( \mu )m</td>
</tr>
</tbody>
</table>

Table 8.1: Camera sensor specifications.

From these specifications, the resolution of the camera can be calculated, based on a field of view (FOV) of 17 mm by 13 mm, by applying the following formula:

\[ Camera \ Resolution \ (\mu m) = 2 \times Cell \ Size \tag{8.2} \]

\[ System \ Resolution \ (\mu m) = (Camera \ Resolution) \times \frac{FOV}{SensorSize} \tag{8.3} \]

Based on the data given, the horizontal resolution is calculated to be 44.27 \( \mu \)m while the vertical resolution is 53.72 \( \mu \)m. These values represent the finest features that can be resolved.
8.3 MTF of Imaging System

To quantify the resolution, a Heidenhain reticle (Figure 8.1) containing spatial frequencies from 0.25 lp/mm to 228 lp/mm was imaged at both the green and red wavelengths within a field of view of 17 mm by 13 mm. The Smart Scalpel lenses are focused such that the distance of best focus is approximately 500 μm below the contact surface. This is to enable the blood vessels underneath the skin to be focused sharply.

![Image reticle](image.png)

Figure 8.1: Image reticle.

8.3.1 MTF Results at Image Center

The MTF results at the distance of best focus are shown in Figure 8.2. The results indicate that the optical system is a low-pass spatial filter. The 50% attenuation bandwidth of the green channel is 12.5 lp/mm for the horizontal bars and 9.7 lp/mm for the vertical bars, and this corresponds to a horizontal resolution of 40 μm and a vertical resolution of 51.5 μm. The 50% attenuation bandwidth of the red channel is 11.9 lp/mm horizontally and 11.6 lp/mm vertically, which corresponds to a horizontal resolution of 42 μm and a vertical resolution of 43.1 μm. (Note that these values are close to those calculated in Section 8.2).
Modulation Transfer Function (577 nm)

Figure 8.2: MTF of red and green channel at center of image.

It is observed that similar results are obtained for both channels. This is important because the image from the red channel is used as the background subtraction for the image from the green channel.

8.3.2 Flatness of Image Field

Degradation of the image away from its center occurs due to aberration of the lens elements. Therefore, it is necessary to quantify the image quality at the edges to see whether it is
acceptable. This can be determined through measurement of the contrast of one spatial frequency (i.e. one group of bars in the reticle) as a function of lateral offsets from the center. This measurement is done at the distance of best focus. The results for frequency 11.3 lp/mm at each spectral channel are plotted in Figure 8.3. Zero lateral displacement is where the bars concerned are in the center of the imaging field.

The modulations in both channels fall less than 50% from the image center. Therefore, it can be concluded that the field flatness extends across the entire field of view, and that the image quality is acceptable throughout the field of view.
8.3.3 Depth of Field

Since the depth of the blood vessel can vary from near the surface to as deep as 630 μm, it is essential that the imaging system have sufficient depth of focus to resolve those deep vessels. The MTF measurements of the 11.3 lp/mm bars as a function of axial offsets from the contact surface are plotted in Figure 8.4.

![Axial Modulation Transfer Function (577 nm)](image1)

![Axial Modulation Transfer Function (650 nm)](image2)

Figure 8.4: MTF as a function of axial distance.
Based on the 50% attenuation criterion, the depth of focus for the green channel is 870 \( \mu \text{m} \) for the vertical lines and 1260 \( \mu \text{m} \) for the horizontal lines. For the red channel, the depth of focus is 740 \( \mu \text{m} \) for the vertical lines and 870 \( \mu \text{m} \) for the horizontal lines.

### 8.4 Penetration Depth of Imaging System

#### 8.4.1 Phantom Material for Tissue

In order to image the blood vessels deep in the epidermis-dermis region, photons has to undergo scattering and absorption events to reach the vessels and then reflect back to the camera. Therefore, it is necessary to measure the depth of penetration of the imaging system. Homogenized milk has been identified as a suitable phantom media to simulate the optical properties of human tissue. Its optical properties in visible wavelength imaging have been extensively studied. Light scattering in milk is due to casein micelles (proteins) of mean particle size 0.15 \( \mu \text{m} \), which gives milk its white appearance, and also fat globules 1 to 2 \( \mu \text{m} \) in diameter. Variations in the concentration of scatters are achieved through dilution with singly distilled water. The optical extinction coefficient \( \mu_t \) as a function of milk concentration in distilled water (by volume) [41] is given as follows.

![Optical Extinction Coefficient vs Milk Concentration](image)

Figure 8.5: Optical extinction coefficient for milk-water concentration [41].
From Beer's law, the following relationship between milk and skin tissue can be derived:

\[
e^{-\mu_{\text{skin}} z_{\text{skin}}} = \frac{P_{\text{skin}}(z_{\text{skin}})}{P_{0\text{skin}}} = \frac{P_{\text{milk}}(z_{\text{milk}})}{P_{0\text{milk}}} = e^{-\mu_{\text{milk}} z_{\text{milk}}}
\]

\[
\Rightarrow \mu_{\text{skin}} z_{\text{skin}} = \mu_{\text{milk}} z_{\text{milk}}
\]

(8.4)

where \(\mu\) is the optical extinction coefficient and \(z\) is the depth of penetration. Since \(\mu_{\text{skin}}\) is a known property, the depth of penetration in skin \(z_{\text{skin}}\) can be calculated if \(\mu_{\text{milk}}\) and \(z_{\text{milk}}\) are found experimentally.

8.4.2 Experiment

The penetration depth is found by measuring the MTF of the 16 lp/mm bars through different milk concentrations. These measurements are done at both the green and red wavelengths, and at the center of the imaging field. The height of the milk-water solution (\(z_{\text{milk}}\)) is 5.5 mm. The results of the experiment are given in Figure 8.6.

![Modulation Transfer Function](image)

Figure 8.6: MTF as a function of milk concentration.
The 1/e attenuation criterion corresponds to a milk concentration of 1.3\%, which is equivalent to an optical extinction coefficient $\mu_{\text{milk}}$ of 0.7 mm$^{-1}$. Using $\mu_{\text{skin}} = 3.823$ mm$^{-1}$ (from Table 4.1), and $z_{\text{milk}} = 5.5$ mm, the depth of penetration $z_{\text{skin}}$ is found to be 1007 $\mu$m. Since the photons have to travel through the tissue and then back to the camera, the effective depth of penetration ($\frac{1}{2} * z_{\text{skin}}$) is 504 $\mu$m. This penetration depth is sufficient to detect most blood vessels which typically lie in the 310 $\mu$m region.

### 8.5 Bandwidth Testing

Tremor is identified as the main source of relative displacement between the patient and the Smart Scalpel. The Smart Scalpel must have sufficient bandwidth to track the blood vessels during treatment. There are two bandwidths that characterize the system: system bandwidth and tracking bandwidth.

The *system bandwidth* is defined as the number of operation loops executed by the computer program per second. An operation loop includes imaging, processing, and laser targeting. As discussed in Section 7.3, the time taken by one operation loop depends on the size of the region it process, which is determined by the number of “seed” targets $S(x,y)$. The greater the number of targets, the lower the system bandwidth.

The system bandwidth is a measure of how fast the system operates, but does not indicate how well the system can actually track a moving object which is more important. The *tracking bandwidth* is defined as the highest frequency that the object can move without the system losing track of it.

The system bandwidth can be easily read off from the computer. The tracking bandwidth measurement is less direct and the following section describes the experiment to determine this value.
8.5.1 Experiment to Determine Tracking Bandwidth

The imaging object is a 750 μm wide red strip attached to a glass plate (Figure 8.7). The red strip acts both as a target and a mask, whereby the laser beam (~750 μm in diameter) will not pass through the plate when the laser is accurately pointing at the mask. On the other hand, if the targeting is not accurate, the beam will pass straight through.

![Figure 8.7: Mask.](image)

The following instruments in Figure 8.8 are set up. The glass plate is attached to a linear stepper motor stage (*Compumotor, Rohnert Park, CA*). A white screen is placed under the glass plate so that a bright spot is seen when the laser misses its target, while a faint patch (due to diffraction) is seen when the laser spot is directly on the mask. In order to record the tracking performance, a Hamamatsu photo-multiplier tube (*HC120 series, Hamamatsu, Bridgewater, NJ*) is pointed at the white screen to measure the light intensity. A large voltage is recorded when the laser misses its target while a small voltage is recorded when the laser is spot on.

![Figure 8.8: Experimental setup.](image)
During the experiment, the linear stage oscillates the glass plate at increasing frequencies while the light intensities on the white screen are recorded. The glass plate is moved at an approximately sinusoidal motion that obeys the equation:

\[ y = A \sin(2\pi f t), \]  

(8.5)

where \( y \) = displacement, \( A \) = amplitude (370 \( \mu \)m), \( f \) = frequency and \( t \) = time.

The Smart Scalpel computer is a 500 MHz Pentium. The image processing algorithm is set to detect 50 targets in each region (recall the regional division scheme at Section 7.3).

8.5.2 Results

The PMT’s voltage is recorded every time the Smart Scalpel detects a target and “switched on” the laser beam. At the end of the experiment for an oscillation frequency, the average voltage of the measurements is computed. A low average suggests successful tracking, while a high value indicates erroneous tracking. From all these data, a tracking transfer function can be plotted. The results of the experiment are given in the graph below.

Figure 8.9: Tracking transfer function.
The system fails to track most targets at frequencies above 1.7 Hz. The 50% attenuation criteria is chosen to define the tracking bandwidth. From the graph above, the tracking bandwidth is found to be 1.0 Hz. The system bandwidth recorded by the computer is 2.2 Hz.

8.6 Summary

The Smart Scalpel has an effective optical resolution of 51.5 μm, a depth of field of 740 μm, and a reasonably flat imaging field; the depth of light penetration into the skin is 504 μm. The system bandwidth is 2.2 Hz and the tracking bandwidth is 1.0 Hz. It is desired to reach a tracking bandwidth of 6 Hz and this bandwidth can be achieved if better instruments are used.
Chapter 9

Clinical Imaging and Laser Treatment

Human and animal tissue-surface blood vessels are imaged using the Smart Scalpel to test its optical system and the image analysis algorithm. A laser treatment is also performed on an animal model. This chapter reports the results of the experiments.

9.1 Index Matching Gel

To reduce specular reflection during imaging, 3 types of index matching gel are tested. They are (1) Aquasonic Transmission Gel (Parker Laboratories), (2) Vidal Sassoon styling gel\(^1\), and (3) Johnson & Johnson Baby Oil Gel.

It is found that the Johnson & Johnson Baby Oil Gel is the most suitable. This gel is colorless and is the least viscous. During tissue imaging, it is found that the low viscosity allows the gel to spread easily and therefore, any air pockets formed are easy to remove. Spectral analysis indicates that the absorption is almost negligible and the spectral profile is flat.

\(^1\) This is highly recommended by researchers at Wellman Labs (MGH) for confocal tissue imaging.
9.2 Imaging Results from Animal Models

To assess the imaging performance of the Smart Scalpel and the image processing algorithms, images of a rabbit’s ear are taken with the system. The rabbit is a New Zealand rabbit, and hair is removed from the ear before the images are taken.

9.2.1 Anatomy of the Rabbit’s Ear

There is little pigmentation in the ear, and the blood vessels are clearly visible with the naked eye. They appear red against the pale-orange tissue background. The blood vessel network consists of a central artery running along the middle of the ear, and smaller vessels spreading from it. The small blood vessels are typically found near the end and at the edge of the ear. The central artery is approximately $1.6 \text{ mm}$ while the smallest blood vessel is less than $120 \mu\text{m}$.

9.2.2 Imaging Results from Rabbit Ear

An image taken from the end of the ear is shown in Figure 9.1(a). The blood vessels are clearly visible from the image, and the contrast is excellent. The largest vessel on the right is approximately $470 \mu\text{m}$ across, while the medium vessel in the lower part of the image is approximately $250 \mu\text{m}$ across. The very fine vessels are approximately $100 \mu\text{m}$ across. The image processing result is shown in Figure (b). As seen, the algorithm is successful in detecting most of the discernible blood vessels.
9.3 Imaging Results from Human Patients

Imaging results from human patients are also successful. The image shown in Figure 9.2(a) is taken from ectactic blood vessels near the ankle. The vessel is clearly visible against the background. The size of the central large vessel is approximately 450 μm. The image processing result is shown in figure (b), whereby the algorithm is successful in detecting most of the discernible blood vessels.
9.4 Laser Treatment of Animal Model

To test the Smart Scalpel’s close-loop treatment method, laser treatments are performed on a rabbit’s ear. The rabbit is anesthetized and hair is removed from its ear before the treatment. The laser parameters used are as follows:

<table>
<thead>
<tr>
<th>Laser</th>
<th>Coherent Innova 100 CW Argon Laser</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength</td>
<td>514 nm</td>
</tr>
<tr>
<td>Beam diameter</td>
<td>750 μm</td>
</tr>
<tr>
<td>Power</td>
<td>1 watt</td>
</tr>
<tr>
<td>Pulse width</td>
<td>80 ms to 100 ms (CW laser pulsed mechanically)</td>
</tr>
<tr>
<td>Fluence</td>
<td>$18.1 \times 10^4$ J/m$^2$ to $22.6 \times 10^4$ J/m$^2$</td>
</tr>
</tbody>
</table>

Table 9.1: Treatment laser parameters.

Two experiments are performed, the first one using a laser pulse width of 100 ms (fluence $22.6 \times 10^4$ J/m$^2$) and the second one using a pulse width of 80 ms (fluence $18.1 \times 10^4$ J/m$^2$). The skin surface is not cooled prior to the laser treatment; the program to detect the blood vessel diameter is also not used.

9.4.1 Experiment 1 (100 ms pulse width, $22.6 \times 10^4$ J/m$^2$)

Figure 9.3(a) is a green illuminated image (at 577 nm) of the tissue region before the treatment. Figure (b) shows the blood vessel targets identified by the Smart Scalpel on a background-subtracted image. Figure (c) shows the green illuminated image immediately after the treatment.
As seen in figure (c), few of the blood vessels visually disappear immediately after treatment while the rest appear fainter. Dermatologists at MGH have observed that most blood vessels typically disappear immediately after treatment using the conventional laser devices. Therefore, similar observation should take place with the Smart Scalpel treatment if it is successful. The concept of selective photothermolysis discussed in Section 3.3 stresses on selecting the appropriate laser wavelength, fluence and pulse width. Therefore, one possible reason why the first experiment is unsuccessful is because the laser parameters are not optimal.
9.4.2 Experiment 2 (80 ms pulse width, $18.1 \times 10^4 \text{ J/m}^2$)

A second experiment is performed using a shorter pulse width of 80 ms. Figure 9.4(a) to (c) illustrate the results of this experiment.

Figure 9.4(a) Blood vessels before treatment (illumination at 577 nm).

(b) Targets identified by the Smart Scalpel.

(c) Blood vessels immediately after treatment.

The optimal laser parameter seems to be chosen for this experiment. In figure (c), most of the blood vessels completely disappear after the treatment. The same observation is seen 5
minutes after the treatment (images much longer after the treatment are not taken). There is also no sign of side effects such as purpura.

There are several possible theories that may contribute to this disappearance:

- Thermal energy causes the blood vessels to constrict and force the blood out. The blood flow then becomes occluded.
- Coagulated blood inside the vessels has a different spectral property and absorbs light less strongly. Therefore, they appear much fainter.
- Skin tissue damage above the blood vessels causes a change in spectral property such that light cannot penetrate to the blood vessels underneath.
- Damaged blood vessel walls have a different spectral property.

In order to verify the destruction of the blood vessels, punch biopsies are taken 1 hour after the treatment.

9.4.2.1 Histology Results

Histology samples from punch biopsies are prepared with the help of dermatologists at MGH. The histological results shown in Figure 9.5 show agglutinated red blood cell and platelet thrombi confined to the blood vessels. The magnified image in Figure 9.6 presents a close-up view of the blood coagulation and endothelial cell (around vessel wall) damage. These observations prove that the blood vessels have indeed by damaged by the laser energy.

The darker shades in the tissue region (red arrows) represent thermal damage. Notice that the tissue scarring is only confined to straight paths along which blood vessels reside. The rest of the tissues that are not in the path of the blood vessels are not damaged at all. If the same site is to be treated with a conventional laser device, this tissue damage will be seen throughout the entire region. Therefore, this serves to reiterate the fact that the Smart Scalpel is a safer treatment device.
Figure 9.5: Histology results

Figure 9.6: Close-up view showing coagulated blood vessels
9.5 Summary

The Smart Scalpel is able to identify blood vessels in both animal and human. The Smart Scalpel has also been tested successfully on a rabbit ear, whereby blood vessel damage is achieved without damage to the surrounding healthy tissue. The tissue imaging and laser treatment are still in their preliminary stages, and more extensive tests are required.
Chapter 10

Conclusions and Future Work

10.1 Conclusions

The Smart Scalpel represents a new class of medical device which closes the loop between diagnosis and treatment. The system utilizes real-time information about tissue properties to selectively deliver laser energy to the specified blood vessel targets. It allows precise and automatic control over the treatment parameters to achieve optimal treatment results, and also enables physicians to assess the treatment efficacy reliably.

The Smart Scalpel system has been built and characterized. We have demonstrated the operations of its optical system, computer vision system, and laser delivery system to a series of test objects and live tissues successfully. The system bandwidth and tracking bandwidth are currently 2.2 Hz and 1.0 Hz respectively. These bandwidths can be increased if better instruments are used.

The underlying objective of the Smart Scalpel is to achieve greater treatment safety and result. Closed-loop treatment whereby the device selects the tissues to be treated without sacrificing treatment effectiveness is definitely more superior to “blind” treatment. The potential
applications of the Smart Scalpel technology are huge. Already, many applications are obvious, while others are still unknown because we have not conceived of them yet.

10.2 Future Work

10.2.1 Image and Software Libraries

The next process is to collect images of different types of vascular lesions, such as PWS, telangiectasia, spider veins and leg veins. With a library of images, it is then possible to determine which other type of vascular lesions the Smart Scalpel is suitable to treat. In addition, this library can be used to design separate software algorithms, where each algorithm is tailored to treat different types of vascular lesions. These software “packages” will become useful when the Smart Scalpel is used as the platform technology envisioned in Section 2.5.

10.2.2 Laser Treatment

It is also necessary to carry out further laser treatment tests on animal models. Biopsy data needs to be collected to analyze the efficacy of the treatment and to determine the optimal laser parameters. After sufficient data is collected, the next step is to conduct clinical trials on human patients.

10.2.3 System Improvement and Miniaturization

The existing instruments presently limit the system performance. These instruments include the computer, galvanometers, camera and framegrabber. At the time of writing, more superior replacements already exist that can be incorporated into the future prototype. In order to miniaturize the system, the components that need to be replaced are the lenses and mirrors, optical fiber and laser. Finally, the future version of the controlling software should be coded in C program instead of Basic.
10.2.4 Optimal Laser Treatment Strategies

Optimization of PWS laser treatment on an individual patient basis requires development of diagnostic methods to determine epidermal melanin concentration, distribution of vessel number, diameter, depth, and blood flow velocity. There are several non-invasive diagnostic sensing methods that measure important optical-thermal data to determine these PWS anatomy and physiology. These include infrared tomography (IRT), optical low-coherence reflectometry (OLCR) [14] and optical coherent tomography (OCT). These techniques are worth investigating in future because they can be incorporated in the Smart Scalpel to provide complementary information.

10.2.5 Explore other Smart Scalpel Applications

Like vascular lesion treatment, it is worth exploring the use of closed-loop treatment on other medical applications. Some of the potential applications considered are:

- Treatment of varicose veins and spider veins.
- Hair removal.
- Tattoo and pigmentation removal.
- Treating the sebaceous glands for acne.
- Skin resurfacing.
Appendix A

Recursive Gaussian Filter (IIR)

The IIR is a recursive filtering structure that drastically reduces the computational effort required for smoothing and performing the first and second Gaussian derivatives. These operations are done with a fixed number of operations per output point independent of the resolution of the Gaussian function considered. The key to the approach is (1) the use of an exponentially based filter family and (2) the use of a recursive sequence.

This section will not explain the IIR concepts and the formula derivations, but rather give the results and methods of implementation. In addition, only the 4th order Gaussian smoothing filter will be discussed. For more detailed analysis, refer to [32] and [31].

A.1 Recursive Filter

The problem of recursive filter design deals with the determination of the coefficients $a_k$'s and $b_k$'s of a rational transfer function of the form:

$$H_a(z^{-1}) = \frac{\sum_{k=0}^{m-1} b_k z^{-(k-1)}}{1 + \sum_{k=1}^{n} a_k z^{-k}} , \quad \{ \text{A.1} \}$$

which characterizes the following recursive system of order $n$.
\[ y(i) = \sum_{k=0}^{m-1} b_k x(i-k) - \sum_{k=1}^{n} a_k y(i-k), \]  

so that the rational transfer function \( H_a(z^{-1}) \) is exactly, or best approximates, the transfer function of the non-recursive system given by a FIR filter.

### A.2 4th Order Gaussian Filter

In the case where the order \( n \) is equal to 4, an operator \( h_a(n) \) can be written in the general case as:

\[
h_a(n) = \left[ a_0 \cos\left(\frac{w_0}{\sigma} n\right) + a_1 \sin\left(\frac{w_0}{\sigma} n\right)\right] e^{\frac{b_k n}{\sigma}} + \left[ c_0 \cos\left(\frac{w_1}{\sigma} n\right) + c_1 \sin\left(\frac{w_1}{\sigma} n\right)\right] e^{\frac{b_k n}{\sigma}}
\]

In the design of the IIR filter, the first step is to find out the set of coefficients \( a, b, c \) and \( w \) in order that the operator \( h_a(n) \) approximates in the mean square error sense the Gaussian operator.

For the Gaussian smoothing function, the following coefficients are obtained with a normalized mean square error of \( \varepsilon^2 = 8.594099e^{-8} \):

\[
g_a(x) = \left[ 1.68 \cos\left(0.6318 \frac{x}{\sigma}\right) + 3.735 \sin\left(0.6318 \frac{x}{\sigma}\right)\right] e^{-1.783 \frac{x}{\sigma}} \\
- \left[ 0.6803 \cos\left(1.997 \frac{x}{\sigma}\right) + 0.2598 \sin\left(1.997 \frac{x}{\sigma}\right)\right] e^{-1.723 \frac{x}{\sigma}}.
\]

### A.3 Implementation

To apply the design to a non-causal impulse response, we can transform \( h(n) \) into a sum of causal sequences by splitting it into two halves \( h_+(n) \) and \( h_-(n) \). Then, \( h(n) = h_+(n) + h_-(n) \), and with:
\[ h_+(n) = \begin{cases} h(n) & n \geq 0 \\ 0 & n < 0 \end{cases} \quad \{ A.5 \} \]

\[ h_-(n) = \begin{cases} 0 & n \geq 0 \\ h(n) & n < 0 \end{cases} \quad \{ A.6 \} \]

\( h_+(n) \) and \( h_-(n) \) are causal with opposite direction and we can compute a recursive system \( H_+(z^{-1}) \) and \( H_-(z) \) having impulse responses close to \( h_+(n) \) and \( h_-(n) \) respectively.

\[
H_+(z^{-1}) = \frac{n_{00}^+ + n_{11}^+ z^{-1} + n_{22}^+ z^{-2} + n_{33}^+ z^{-3}}{1 + d_{11}^+ z^{-1} + d_{22}^+ z^{-2} + d_{33}^+ z^{-3} + d_{44}^+ z^{-4}},
\{ A.7 \}
\]

\[
H_-(z) = \frac{n_{11}^- z^2 + n_{22}^- z^3 + n_{33}^- z^4 + n_{44}^- z^5}{1 + d_{11}^- z^{-1} + d_{22}^- z^{-2} + d_{33}^- z^{-3} + d_{44}^- z^{-4}}.
\{ A.8 \}
\]

The two \( z \)-transforms in \{ A.7 \} and \{ A.8 \} correspond to two transfer functions of stable 4th order filters recursing from the left to the right for the causal sequence \{ A.5 \}, and from the right to the left for the anticausal sequence \{ A.6 \}. They are implemented by applying the recursive sequence

\[
y_+^k = n_{00}^+ x_k + n_{11}^+ x_{k-1} + n_{22}^+ x_{k-2} + n_{33}^+ x_{k-3} - d_{11}^+ y_+^{k-1} - d_{22}^+ y_+^{k-2} - d_{33}^+ y_+^{k-3} - d_{44}^+ y_+^{k-4},
\]

\[
y_-^k = n_{11}^- x_{k+1} + n_{22}^- x_{k+2} + n_{33}^- x_{k+3} + n_{44}^- x_{k+4} - d_{11}^- y_-^{k+1} - d_{22}^- y_-^{k+2} - d_{33}^- y_-^{k+3} - d_{44}^- y_-^{k+4},
\]

\[
y_k = y_+^k + y_-^k \quad (k=1,\ldots,N).
\{ A.9 \}
\]

where \( x_k \) is the input to the filter and \( y_k \) is the output.

The coefficients \( n_{ij} \) and \( d_{ij} \) can be computed by applying the following equations derived from the \( z \) transform of \{ A.3 \}:
\[ n_{33}^+ = e^{\frac{h_1 - 2h_0}{\sigma}} \left[ c_i \sin\left(\frac{w_l}{\sigma}\right) - c_0 \cos\left(\frac{w_l}{\sigma}\right) \right] + e^{\frac{h_0 - h_1}{\sigma}} \left[ a_i \sin\left(\frac{w_0}{\sigma}\right) - a_0 \cos\left(\frac{w_0}{\sigma}\right) \right], \]

\[ n_{22}^+ = 2e^{\frac{h_0}{\sigma}} \left[ (a_0 + c_0) \cos\left(\frac{w_l}{\sigma}\right) \cos\left(\frac{w_0}{\sigma}\right) - a_i \cos\left(\frac{w_l}{\sigma}\right) \sin\left(\frac{w_0}{\sigma}\right) - c_i \cos\left(\frac{w_0}{\sigma}\right) \sin\left(\frac{w_l}{\sigma}\right) \right] + c_0 e^{-\frac{2h_0}{\sigma}} + a_0 e^{-\frac{h_0}{\sigma}}, \]

\[ n_{11}^+ = e^{\frac{-h_1}{\sigma}} \left[ c_i \sin\left(\frac{w_l}{\sigma}\right) - (c_0 + 2a_0) \cos\left(\frac{w_l}{\sigma}\right) \right] + e^{-\frac{h_0}{\sigma}} \left[ a_i \sin\left(\frac{w_0}{\sigma}\right) - (2c_0 + a_0) \cos\left(\frac{w_0}{\sigma}\right) \right], \]

\[ n_{00}^+ = a_0 + c_0. \]

\[ d_{ii}^+ = e^{-\frac{2h_0 - h_i}{\sigma}}, \]

\[ d_{33}^+ = -2 \cos\left(\frac{w_0}{\sigma}\right) e^{\frac{h_0 - 2h_i}{\sigma}} - 2 \cos\left(\frac{w_l}{\sigma}\right) e^{\frac{h_0 - 2h_i}{\sigma}}, \]

\[ d_{22}^+ = 4 \cos\left(\frac{w_l}{\sigma}\right) \cos\left(\frac{w_0}{\sigma}\right) e^{\frac{h_0 h_i}{\sigma}} + e^{-\frac{2h_i}{\sigma}} + e^{-\frac{2h_0}{\sigma}}, \]

\[ d_{11}^+ = -2e^{-\frac{h_1}{\sigma}} \cos\left(\frac{w_l}{\sigma}\right) - 2e^{-\frac{h_0}{\sigma}} \cos\left(\frac{w_0}{\sigma}\right). \]

\[ d_{ii}^- = d_{ii}^+ \quad i = 1, \ldots, 4, \]

\[ n_{ii}^- = n_{ii}^+ - d_{ii}^+ n_{00}^+ \quad i = 1, \ldots, 3, \]

\[ n_{ii}^- = -d_{ii}^+ n_{00}^+ \quad i = 4. \]

### A.3.1 Normalization

Before implementing the recursive sequence, the IIR filter has to be normalized. To do this, each coefficient in the operator \( h_a(n) \) has to be scaled by a factor which satisfies a given constraint. In the case of the smoothing operator used to approximate the Gaussian filter, the following constraint is required:
\[ N_g \left( \sum_{n=-d_0}^{\infty} h_a(n) \right) = 1, \quad \text{where } N_g \text{ is the scale factor.} \quad \{ \text{A.10} \} \]

This scale factor can be numerically computed and divided to each coefficient before applying the recursive sequence.

**A.4 Computation Speed**

In the case of a 4\textsuperscript{th} order IIR filter, the number of operations per pixel is 31. This is equivalent to implementing a FIR filter with 16 coefficients. In general, the width of the FIR filter used is greater than 25 (equivalent to 49 operations per pixel), and therefore, this justifies the use of the IIR filter. In addition, the speed of IIR convolution is unaffected by the choice of \( \sigma \), while that of the FIR filter increases with \( \sigma \) because the filter becomes wider.

The number of operations reduces to 23 per pixel in the case of 3\textsuperscript{rd} order IIR filter, and 15 per pixel in the case of 2\textsuperscript{nd} order filter. These are achieved, of course, at the expense of accuracy.
Appendix B

Mathematic Theories for Line Detection

B.1 Approximation of an image function by a polynomial

Given an image function $F(x,y)$, its quadratic polynomial approximation is given by:

$$F(x,y) = R + (R_x x + R_y y) + \frac{1}{2}(R_{xx} x^2 + 2R_{xy} xy + R_{yy} y^2) , \quad \{\text{B.1}\}$$

where $R_x$, $R_y$, $R_{xx}$, $R_{xy}$ and $R_{yy}$ are the locally estimated derivatives at $(x,y)$ that are obtained by convolving the image with the appropriate Gaussian kernels.

B.2 Solution of zero crossing

To extract the lines in an image, the solution of zero crossing has to be determined. This is equivalent to finding the solution to the polynomial $\{\text{B.1}\}$ where the first derivative along the unit vector $\mathbf{n}$ goes to zero. $\mathbf{n}$ is chosen such that it is pointing in the direction perpendicular to the line. This solution is also the mathematical proof of Equation (6.22).
From Equation \{ B.1 \},

\[ F(x, y) = R + \left( R_x + R_y \right) y + \frac{1}{2} \left( R_{xx} x^2 + 2 R_{xy} xy + R_{yy} y^2 \right) , \]

\[ \frac{\partial}{\partial x} F(x, y) = R_x + R_{xx} x + R_{xy} y , \] \{ B.2 \}

\[ \frac{\partial}{\partial y} F(x, y) = R_y + R_{yy} y + R_{xy} x . \] \{ B.3 \}

From Equation (6.18),

First Derivative along \( \hat{n} = 0 \)

\[ \hat{n}_x \frac{\partial}{\partial x} F(x, y) + \hat{n}_y \frac{\partial}{\partial y} F(x, y) = 0 \]

Substituting from Equations \{ B.2 \} and \{ B.3 \},

\[ \left( R_x \hat{n}_x + R_y \hat{n}_y \right) \left( R_{xx} \hat{n}_x + R_{xy} \hat{n}_y \right) x + \left( R_{yy} \hat{n}_x + R_{xy} \hat{n}_y \right) y = 0 . \] \{ B.4 \}

Since the solution \((x, y)\) has to lie along \( \hat{n} \), therefore:

\[ (x, y) = \left( t \hat{n}_x, t \hat{n}_y \right) , \] where \( t \) is a constant. \{ B.5 \}

Substituting Equation \{ B.5 \} into \{ B.4 \},

\[ \left( R_x \hat{n}_x + R_y \hat{n}_y \right) \left( R_{xx} \hat{n}_x + R_{xy} \hat{n}_y \right) \hat{n}_x + \left( R_{yy} \hat{n}_x + R_{xy} \hat{n}_y \right) \hat{n}_y = 0 \]

\[ \Rightarrow \quad t = - \frac{R_x \hat{n}_x + R_y \hat{n}_y}{R_{xx} \hat{n}_x^2 + 2 R_{xy} \hat{n}_x \hat{n}_y + R_{yy} \hat{n}_y^2} \]

Therefore, from \{ B.5 \}, the position where the first derivative = 0 is:
\[
x = \left( -\frac{R_x \hat{n}_x + R_y \hat{n}_y}{R_{xx} \hat{n}_x^2 + 2R_{xy} \hat{n}_x \hat{n}_y + R_{yy} \hat{n}_y^2} \right) \hat{n}_x, \quad \text{(Proof for Equation (6.22))}
\]
\[
y = \left( -\frac{R_x \hat{n}_x + R_y \hat{n}_y}{R_{xx} \hat{n}_x^2 + 2R_{xy} \hat{n}_x \hat{n}_y + R_{yy} \hat{n}_y^2} \right) \hat{n}_y.
\]

### B.3 Jacobi Rotation

The eigenvectors and eigenvalues of the Hessian matrix can be found in a numerically stable and efficient way by using one Jacobi rotation [42].

Given a 2 \times 2, real symmetric matrix \( A = \begin{bmatrix} a & b \\ b & d \end{bmatrix} \), the steps to find the eigenvectors and eigenvalues are:

\[
\tilde{t} = \frac{2b}{a - d}, \quad t = \frac{\tilde{t}}{1 + \sqrt{1 + \tilde{t}^2}}, \\
c = \frac{1}{\sqrt{1 + t^2}}, \quad s = ct.
\]

Then the eigenvalues and eigenvectors are given by:

\[
\lambda_1 = (a + tb), \quad \vec{e}_1 = \begin{bmatrix} c \\ s \end{bmatrix}, \quad \{ \text{B.6} \}
\]
\[
\lambda_2 = (d - tb), \quad \vec{e}_2 = \begin{bmatrix} -s \\ c \end{bmatrix}, \quad \{ \text{B.7} \}
\]

where \( \|\vec{e}_1\| \) and \( \|\vec{e}_2\| \) are equal to 1.
Bibliography


Reports*, 1993.

[33] T. Lindeberg, Discrete derivative approximations with scale-space properties: A basis for
low-level feature extraction. *Journal of Mathematical Imaging and Vision*, vol. 3, pp. 349-
376, 1993.

[34] D. Eberly, R. Gardner, B. Morse, S. Pizer, and C. Scharlach, Ridges for image analysis.

*Proceedings of the IEEE Workshop on Applications of Computer Vision*, pp. 205-212,

European Conference on Computer Vision, Lecture Notes in Computer Science*, vol. 1064,
pp. 630-641, 1996.

[37] O. Monga, N. Armande, and P. Montesinos, Thin nets and crest lines: application to
satellite data and medical images. *Computer Vision and Image Understanding*, vol. 67,

[38] J. Canny, A computational approach to edge detection. *IEEE Transaction on Pattern


