

國立臺灣大學電機資訊學院電子工程學研究所

碩士論文

Graduate Institute of Electronics Engineering

College of Electrical Engineering & Computer Science

National Taiwan University

Master Thesis

應用於雙波段紅外線之乳癌檢測盲源分離演算法
Application of Blind Source Separation Algorithms on
Dual-band IR Spectrogram for Breast Cancer Detection

李鎮宇

Chen-Yu Li

指導教授：陳少傑 博士

Advisor: Sao-Jie Chen, Ph.D.

中華民國 105 年 7 月

July, 2016





誌 謝

本論文能夠完成並且順利取得碩士學位，首先要感謝指導教授陳少傑博士，在碩士生涯中，教授提供一套完善的研究學習流程，教授的循循善誘，使得我在研究學習的過程中建立穩固的研究基礎，從中培養獨立思考以及自我要求的研究精神。感謝陳中明教授，提供研究基礎的實驗環境，並給予寶貴的研究經驗，讓研究能夠更為順利。感謝口試委員們不吝撥冗指導，給予寶貴的研究經驗，使得本論文的內容架構更加充實與完善。

研究所的在學期間內受到許多貴人的協助與關照。首先要特別感謝亦師亦友的鑫平學長，從踏進研究生的學習開始，您總是常伴左右並且從中協助我在研究上各種疑難雜症，以及感謝實驗室 102 級的學長們：詩堯學長、晟豪學長、岱鑫學長、鼎鈞學長和士毅學長，你們提供我許許多多的研究經驗、修課資訊和學習資料，感謝學長們的分享與交流，使得我的學習過程在相當完備的基礎下，再創佳績，感謝同為 103 級同學們：哲論、尚儕、力元和明宏，感謝你們一路走來的同舟共濟，一起分擔實驗室以及生活上的大小事務，並且培養在研究團隊中的合作精神，感謝實驗室的學弟們：子源、展翔、岳徵和林俊，感謝你們繼續讓實驗室充滿了朝氣與活力，薪火相傳且繼續傳承實驗室的研究精神。

最後感謝我的祖父母李國和先生和林烏甜女士，以及我的父母李木文先生和梁彩虹女士，祖父母是傳統的務農人家，父母是平凡的上班族群，感謝他們不辭辛勞地撫養與啟蒙，提供我一個衣食無虞的成長環境，秉持他們勤勞務實的耕耘精神，才有我今天個人小小的成就，感恩他們大大的付出，實是屬於他們的成就！

謹以本文獻給所有關心我的人，願與你們分享這份榮耀。





中文部分





摘要

本論文提出一個針對雙波段紅外線頻譜(Dual-band IR Spectrogram)，應用改良的盲源分離演算法(Blind Source Separation Algorithm)進行分析，來判定追蹤乳癌的長期化療之成效。本文採用雙波段紅外線頻譜的原始資料(RAW Data)為輸入檔，進行演算法分析，並針對盲源分離演算法的特性，在過程中運用資料分析進行改良，最後將該演算法加載至本實驗室開發的醫療控管處理器設計進行整合，提供我們一個更便利的管道來體現雙波段紅外線頻譜分析應用於醫療的成果。在標定率方面，該改良演算法設計比其他演算法平均高 15%，在癌細胞判斷的正確率方面，該演算法設計比其他演算法平均高 10%。

關鍵字：雙波段紅外線影像，盲源分離演算法。





目 錄

第一章	簡介.....	伍
第二章	背景.....	柒
第三章	盲源分離演算法.....	玖
第四章	模擬實驗結果.....	壹拾壹
第五章	結論.....	壹拾參





第一章

簡介

根據科學研究顯示，溫度高於攝氏-273.15 度的物體，皆會輻射出紅外線。我們可將人體輻射的紅外線訊號紀錄下來，進行分析並追蹤病徵，例如乳癌診斷系統。一般來說，乳癌診斷系統需要收集兩種不同波段的紅外線影像，才能有效地分析出病徵所在。由於現代技術之演進，現行只需要用一台紅外線攝影機做攝影就能取得雙波段紅外線頻譜(Dual-band IR Spectrogram)。

紅外線波段介於光波長0.75微米至1000.00微米的區間，紅外線主要有三種波段：近波段紅外線(0.75微米-1.50微米)、中波段(1.50微米-4.00微米)、和遠波段(4.00

微米-1000.00微米)。除此之外，中波段以及遠波段在人體中為自體輻射的訊號，故不需要照射額外的光源，因此對人體有無痛無害的特性，可進行低負擔的被動式醫療檢測。用該特性來取得人體的雙波段紅外線影像頻譜具有加強影像校正的功能，可幫助在進行盲源分離演算法(Blind Source Separation Algorithm)之前，增加其正確性。

盲源分離演算法的種類相當多樣化，例如針對不同病徵進行分類、針對不同溫度進行分類等等，因此我們的目標就是針對雙波段紅外線頻譜的訊號，進行盲源分離演算法之改良，以求達到更好的醫療檢測效果。





第二章

背景

紅外線影像的研究最早從十九世紀初開始，爾後越來越多的學者投入研究，紅外線分成數種波段：近波段紅外線、中波段紅外線、和遠波段紅外線。只要有熱度的物體都會輻射出紅外線，這意味著在我們身處的環境中，處處都有紅外線的存在，而在這裡我們會針對不同紅外線的特性，去評估哪些波段的紅外線頻譜適合做訊號分析。

在雙波段紅外線頻譜的訊號分析當中，盲源分離演算法扮演一個相當重要的核心角色。因此我們會以學術界發表過的一些盲源分離演算法為雛型，進而運行分析比較，並在此基礎下運用其他學術理論進一步進行整合與優化。經由使用資料分析的理論，如機器學習(Machine Learning)，可達到提升整體實驗流程當中實驗數據的可信度以及判斷的正確率。





第三章

盲源分離演算法

原始資料(RAW Data)為使用 SC4000 FLIR 攝影機所拍攝出來的，經過解碼之後，得到雙波段紅外線頻譜，其規格為 256 乘以 320，共 81920 個像素，每個像素為 14 位元。每次拍攝就會有兩張頻譜，分別包含 MIR 訊號及 LIR 訊號，在此提出三個方法來進行評估。

第一個方法為 Relative Temperature Algorithm，利用相對溫度的分布，進行直接觀察法，將相對溫度異常高的位置進行標定，該方法為最直觀的方法。

第二個方法為 Neighbor-based BSS Algorithm，利用觀察每個像素的兩筆資訊(分別為 MIR 訊號及 LIR 訊號) 其構成之向量與鄰近向量所形成之夾角的變化情形進行標定。一般來說，良性組織的周邊性質皆相當接近，故夾角變化會趨近於零，另一方面，惡性組織的周邊性質變化較大，故夾角變動幅度也比較劇烈且不規則。藉此特性我們可以在雙波段紅外線頻譜中標定出惡性組織的可疑區域，而 Neighbor-based BSS 必須仰賴兩筆資訊的頻譜對位，才能發揮其最大的成效。

第三個方法為我們所提出針對雙波段紅外線頻譜使用 Augmented Regression Model 的改良 Neighbor-based BSS Algorithm，利用惡性組織因微血管不正常增生而具較高溫的特性，先藉由 MIR 和 LIR 的數據變化，選出相對可疑的區域，再利用迴歸分析的方法來屏除雜訊以及加強數據的精確度。而 Augmented Regression Model 必須對 MIR 訊號及 LIR 訊號做初步的分類，故分類方法的設計將會主導 Augmented Regression Model 最後的成效。





第四章 模擬和實驗結果

雖然 Neighbor-based BSS Algorithm 可以成功標定出惡性組織的可疑區域，但該設計方法除了要求優良的影像對位設計，且要求鄰近檢測範圍必須大於等於惡性組織的分布範圍，以避免大範圍的惡性組織其內部可能會有大範圍的性質過於類似以致夾角變化程度不夠明顯，另外該演算法的訊息能量是利用夾角總和去估算，無法確切知道相對位置何處夾角變化較大，故標定位置相對不準。

我們所提出的 Improved Neighbor-based BSS Algorithm 為針對上述方法的缺點進行改良之演算法。此方法運用惡性組織在 MIR 以及 LIR 有不同的分布特性，成功體現雙波段紅外線頻譜的優勢並且成功屏除雜訊並且加強標定可疑區域的效果，而值得注意的是，Augmented Regression Model 在初步的分類會影響最後的精確性。

最後我們會在初始條件都相同的情況下，將各種演算法進行比較，並且各別評估其標定率、判斷正確率、和判斷錯誤率，然後提出未來可以改善的方向。





第五章

結論

在盲源分離演算法的基礎下，本論文提出成功運用雙波段紅外線頻譜的特性，在乳癌偵測中標定可疑區域的一個改良演算法。該改良演算法利用資料分析的理論來進行改良，與其他現有的演算法進行比較時，具有達到減少雜訊以及加強可疑區域標記之檢測效果。在標定率方面，該改良演算法設計比其他演算法平均高 15%，在癌細胞判斷的正確率方面，該演算法設計比其他演算法平均高 10%。未來在加載至本實驗室開發的醫療控管處理器設計時也提供另一種雙波段紅外線頻譜的分析方法，以及可因應不同的應用層面來優化電路設計的可能性。





英文部分





ABSTRACT

This work presents an application of Blind Source Separation (BSS) Algorithms on Dual-band IR Spectrogram for breast cancer detection, which is used to trace the effect of long-term chemotherapy for breast-cancer patients. We take Dual-band IR Spectrogram's RAW Data as an input to the BSS algorithms. Also, we plan to integrate this analytical algorithm into the back-end processor of our designed Dual-band IR Sensor and Readout Circuit Platform. This work will provide a more convenient medical application of our Improved Neighbor-based BSS algorithm on Dual-band IR Spectrogram for breast cancer detection. For Demarcating Degree, our Improved Neighbor-based BSS algorithm is approximately 15% better than other algorithms. For Correctness Rate, our improved algorithm approximately increases 10% compared with other algorithms.

Keywords: Dual-band IR Spectrogram, Blind Source Separation Algorithm.





TABLE OF CONTENTS

ABSTRACT	i
TABLE OF CONTENTS	iii
LIST OF FIGURES	v
LIST OF TABLES	vii
CHAPTER 1 INTRODUCTION	1
1.1 Motivation	1
1.2 Thesis Organization	2
CHAPTER 2 BACKGROUND	3
2.1 Dual-band IR Spectrogram	3
2.2 Blind Source Separation Algorithms	5
2.2.1 Single Pixel Blind Source Separation Algorithm.....	5
2.2.2 Neighbor-based Blind Source Separation Algorithm	7
2.3 Machine Learning Models	8
2.3.1 Supervised Learning and Unsupervised Learning	8
2.3.2 Linear Regression Model and Non-linear Regression Model.....	9
2.4 Breast Cancer Detection	11
2.4.1 Thermal Phenomena of Symptom	11
2.4.2 Critical Temperature and Gradient Temperature.....	12
2.4.3 Risk Level of Breast Cancer	13
CHAPTER 3 BLIND SOURCE SEPERATION ALGORITHMS.....	15
3.1 Single Pixel Blind Separation Algorithm	15
3.1.1 Principle of Single Pixel BSS	15
3.1.2 Application of Single Pixel BSS.....	16
3.1.3 Challenge in Single Pixel Blind Separation Algorithm	16

3.2	Neighbor-based Blind Separation Algorithm	17
3.2.1	Principle of Neighbor-based BSS	17
3.2.2	Application of Neighbor-based BSS.....	20
3.2.3	Challenge in Neighbor-based Blind Separation Algorithm	21
3.3	Machine Learning Analysis Algorithms	21
3.3.1	Linear Regression Model.....	21
3.3.2	Augmented Regression Model.....	22
3.4	Experiment Procedures	23
3.4.1	Specification of SC4000 FLIR Camera	23
3.4.2	Dual-band IR Platform.....	25
CHAPTER 4	SIMULATION AND EXPERIMENT RESULTS	29
4.1	RAW Data Pre-Processing.....	29
4.1.1	Histogram Equalization	29
4.1.2	Binary Occupied Histogram Projection.....	30
4.1.3	Result of Compression.....	30
4.2	Experimental Results.....	31
4.2.1	Algorithms used in our Experiments	31
4.2.2	Demarcating Degree	37
4.2.3	Correctness Rate and Error Rate.....	39
CHAPTER 5	CONCLUSIONS.....	43
REFERENCE	45





LIST OF FIGURES

Fig. 2-1.	Infrared Ray in the Electromagnetic Spectrum.....	4
Fig. 2-2.	Plank Radiation Law	6
Fig. 2-3.	Linear Mixture Model of Single Pixel BSS Algorithm	6
Fig. 2-4.	Experiment Process of Machine Learning Model	9
Fig. 2-5.	Examples of Linear and Non-linear Regression Models	10
Fig. 2-6.	Experiment Process of Relative Temperature Algorithm	12
Fig. 2-7.	Detective Process of Breast Cancer	14
Fig. 3-1.	Experiment Process of Single Pixel BSS Algorithm	17
Fig. 3-2.	Geometric Illustration of Unique Solution	18
Fig. 3-3.	Experiment Process of Neighbor-based BSS Algorithm	20
Fig. 3-4.	Distribution of Different Cells	21
Fig. 3-5.	Schematic Diagram of SC4000 FLIR Camera.....	23
Fig. 3-6.	Experiment Process of SC4000 FLIR Camera	23
Fig. 3-7.	Dual-band IR Platform.....	25
Fig. 3-8.	Chip Layout of Dual-band IR Readout Circuit.....	26
Fig. 4-1.	Diagram of Histogram Equalization	29
Fig. 4-2.	Diagram of Binary Occupied Histogram Projection.....	30
Fig. 4-3.	RAW Data Simulation of MIR Image and LIR Image	30
Fig. 4-4.	Pictures of MIR Image and LIR Image.....	31
Fig. 4-5.	Experiment Process of Improved Neighbor-based BSS Algorithm.....	32
Fig. 4-6.	Comparison of Detection Algorithms for C001_1.....	34
Fig. 4-7.	Comparison of Detection Algorithms for C001_2.....	34
Fig. 4-8.	Comparison of Detection Algorithms for C001_3.....	34
Fig. 4-9.	Comparison of Detection Algorithms for C002_1.....	35
Fig. 4-10.	Comparison of Detection Algorithms for C002_2.....	35

Fig. 4-11. Comparison of Detection Algorithms for C002_3.....	35
Fig. 4-12. Comparison of Detection Algorithms for C003_1.....	36
Fig. 4-13. Comparison of Detection Algorithms for C003_2.....	36
Fig. 4-14. Comparison of Detection Algorithms for C003_3.....	36
Fig. 4-15. Statistical Chart of C001	38
Fig. 4-16. Statistical Chart of C002.....	38
Fig. 4-17. Statistical Chart of C003.....	38
Fig. 4-18. Distribution of Correctness Rate and Error Rate	41





LIST OF TABLES

Table 2-1. Types of IR Spectrogram and their Characteristics.	3
Table 2-2. Types of Infrared Ray and their Characteristics.	5
Table 2-3. Formula of Single Pixel BSS Algorithm.	7
Table 2-4. Formula of Neighbor-based BSS Algorithm.	7
Table 2-5. Thermal Phenomena of Symptoms.	11
Table 2-6. Qualities of Thermal Phenomena.	11
Table 2-7. Risk Level of Breast Cancer.	13
Table 3-1. Specification of SC4000 FLIR Camera.	24
Table 3-2. Specification of Dual-band IR Readout Circuit.	27
Table 4-1. Comparison of Algorithms.	33
Table 4-2. Comparison of Patients.	37
Table 4-3. Correctness Rate and Error Rate.	40





CHAPTER 1

INTRODUCTION

This chapter gives an overview of our work. Section 1.1 will briefly introduce the motivation and background on Dual-band IR Spectrogram. Section 1.2 will summarize the organization of this Thesis.

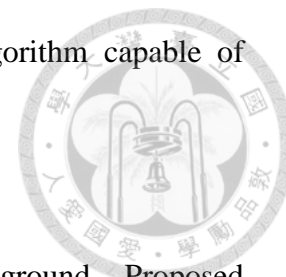
1.1 Motivation

According to scientific research, it shows that an object will radiate infrared rays if the temperature is higher than -273.15 degrees Celsius. We can record these signals of infrared rays from human bodies and design an analysis procedure to do symptom tracking, an example of such application is the diagnosis of breast cancer. Generally, diagnosis of breast cancer needs infrared sensors which can detect two different bands to get a Dual-band IR Spectrogram. Thanks to modern technology, we can use a Forward Looking Infrared (FLIR) Camera to get such Dual-band IR Spectrogram.

The range of infrared bands is from 0.75 micrometers to 1000.00 micrometers. These infrared rays have been divided into three main bands: near infrared rays, middle infrared rays, and long infrared rays. The Dual-band IR Spectrogram that we obtained in this work was combined with middle infrared rays and long infrared rays because they are autologous radiation from human bodies. Their characteristics are harmless and painless because we do not need to irradiate bodies additionally. The function of Dual-band IR Spectrogram could strengthen the correctness of images analyzed by using a Blind Source Separation Algorithm.

Many Blind Source Separation Algorithms have been proposed. And we can apply such algorithm according to different symptoms or different temperatures. Our

core goal is that we try to find a Blind Source Separation Algorithm capable of performing Proportional Analysis of Dual-band IR Spectrogram.



1.2 Thesis Organization

This Thesis contains five chapters: Introduction, Background, Proposed Algorithm, Simulation and Experiment Results, and Conclusions.

In Chapter 1, we present a brief introduction on advantages of applying Dual-band IR Spectrogram to cancer detection. On the other hand, we design an improved Blind Source Separation Algorithm as the core of our experimental architecture.

In Chapter 2, we depict sufficient background on Dual-band IR Spectrogram, and introduce some widely-used Blind Source Separation (BSS) Algorithms and Machine Learning Models. In the last section of this chapter, we will make a perfect integration between Blind Source Separation Algorithm and Machine Learning Model to deal with Dual-band IR Spectrogram.

In Chapter 3, we introduce a pre-process of RAW Data from FLIR. These RAW Data consist of MIR images and LIR images. Then we use our Improved Blind Source Separation Algorithm to detect the areas of images whether they display the locations of abnormal cells or normal cells. Additionally, we evaluate our Improved BSS Algorithm with some other BSS algorithms and compare the obtained results with the Clinical Trial's ones. By this way, we can observe the effect of long-term chemotherapy.

In Chapter 4, we show some procedures adopted in our experiments which include a suitable compression of Dual-band IR Spectrogram, and use demarcating degree, the correct rate, and the error rate to evaluate the effect of long-term chemotherapy of different BSS Algorithms.

In Chapter 5, we conclude this Thesis and suggest some research directions of Dual-band IR Spectrogram for the positive detection in the future.



CHAPTER 2

BACKGROUND

In this chapter, we will present some background information on Dual-band IR Spectrogram in Section 2.1, some Blind Source Separation Algorithms in Section 2.2, Machine Learning Models in Section 2.3, and Application of the above-mentioned algorithms to Breast Cancer Detection in Section 2.4. We will compare different kinds of methods and integrate some of them into our algorithm design.

2.1 Dual-band IR Spectrogram

Methods to analyze Passive Medical Images are introduced in this section. The types of IR Spectrogram commonly used are either Single Spectrum Infrared (SS-IR) Ray or Double Spectrum Infrared (DS-IR) Ray, which generate a Thermogram and a Dual-band IR Spectrogram respectively. The characteristics of two types of IR Spectrogram are listed in Table 2-1 and we will evaluate each of them and explain why Dual-band IR Spectrogram is more suitable for cancer detection.

Table 2-1. Types of IR Spectrogram and their Characteristics.

Type	Sensor	Position	Resolution	Cost	Result
Single Spectrum Infrared Ray	1	Worse	Low	Low	Thermo-gram
Double Spectrum Infrared Ray	2	Better	High	High	Dual-band IR Spectrogram

There is an obvious difference between Thermogram and Dual-band IR Spectrogram. Thermogram is generated in a low temperature environment during the

period of exposure. Unfortunately, kinds of unpredictable factors which influence the correctness of the detection are also present. On the other hand, Dual-band IR Spectrogram directly catches human information and strengthens the correctness using DS-IR Spectral Ratio, System Impulse Response, and Linear Approximation. Therefore, Dual-band IR Spectrogram is more suitable for cancer detection.

Figure 2-1 shows the location of Infrared Ray in Electromagnetic Spectrum. Infrared Ray is invisible but different ranges of Infrared Ray contain different human information for specific symptoms.

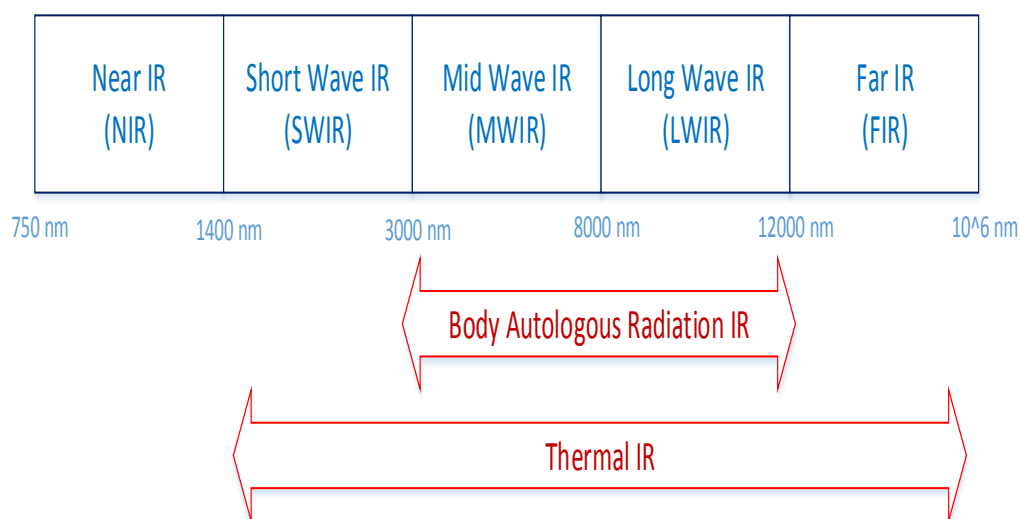


Fig. 2-1. Infrared Ray in the Electromagnetic Spectrum.

Given Double Spectrum Infrared Ray, one should decide which type of infrared ray is more suitable for breast cancer detection. The characteristics of infrared rays as listed in Table 2-2 show that middle infrared rays (MIR) and long infrared rays (LIR) are a better choice.

In Passive Medical Image Technology, Non-invading Detection is the most vital goal. Note that additional irradiation is a burden on human body. On the other hand, we do not need to irradiate bodies if Autologous Radiation of MIR and LIR are used

to obtain human information. These characteristics show that using MIR and LIR for cancer detection is more harmless and painless.

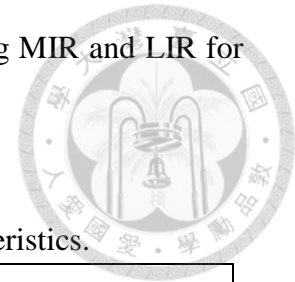


Table 2-2. Types of Infrared Ray and their Characteristics.

Type	Range (μm)	Invading Detection	Source
Near Infrared Ray (NIR)	0.75-1.50	Yes	External Light
Middle Infrared Ray (MIR)	1.50-4.00	No	Autologous Radiation
Long Infrared Ray (LIR)	4.00-1000.00	No	Autologous Radiation

2.2 Blind Source Separation Algorithms

Various sub-signals are obtained from the sources of a mixing signal during the processing of Dual-band IR Spectrogram. By designing an ideal Blind Source Separation Algorithm, one can extract particular signals to get an ideal Proportional Analysis called DS-IR Spectral Ratio. The core concepts of two kinds of Blind Source Separation Algorithms will be introduced in this section.

2.2.1 Single Pixel Blind Source Separation Algorithm

The core concept of Single Pixel Blind Source Separation (BSS) Algorithm, as described in [1], [2], [3] and [4], is the ratio which is composed of probability called s_a and s_b , where s_a represents the probability of abnormal cells and s_b represents the probability of normal cells. Intuitive assumption of this algorithm is $s_a + s_b = 1$. This intuitive assumption will help to build the formula used in the Single Pixel BSS Algorithm.

By detecting MIR (3-5 μm) and LIR (8-12 μm) from human body, we can get two kinds of thermo-vectors respectively from cancer cells and normal cells. The fact that cancer cells and normal cells are different in Plank radiation will help us to construct a Linear Mixture Model, because cancer cells have abnormal proliferation from microvascular. Fig. 2-2 shows Plank Radiation Law and variables which are $a(M)$, $a(L)$, $b(M)$, and $b(L)$ used in the Linear Mixture Model of Single Pixel Blind Source Separation Algorithm as shown in Fig. 2-3. The definitions of variables

and formulas used in the Linear Mixture Model of Single Pixel BSS Algorithm are listed in Table 2-3.

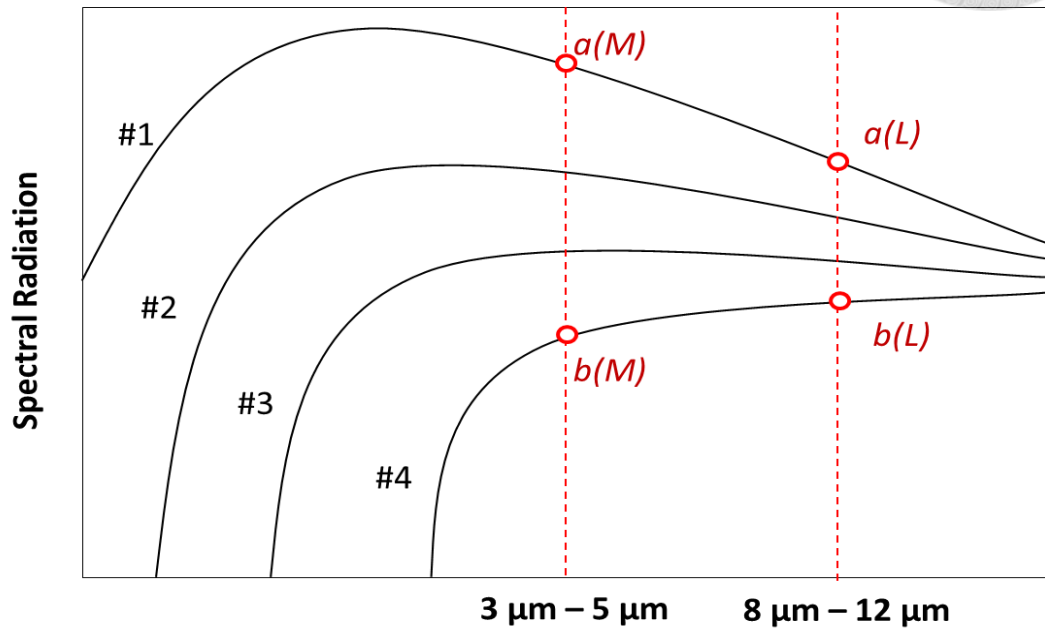
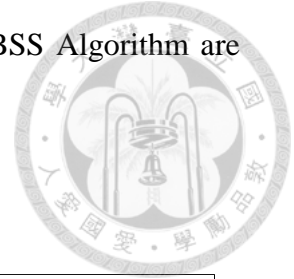


Fig. 2-2. Plank Radiation Law.

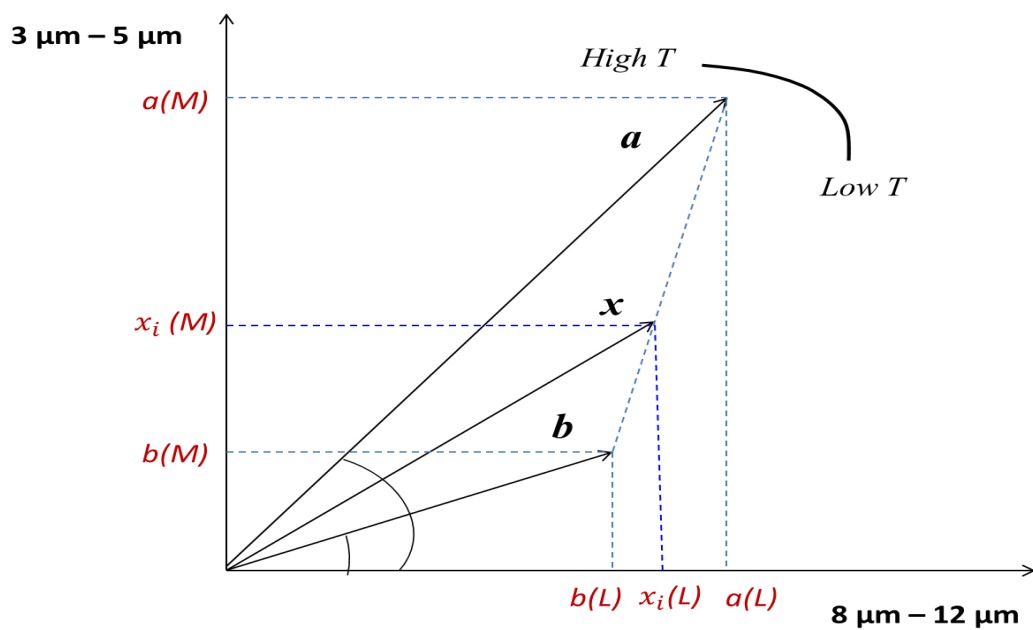


Fig. 2-3. Linear Mixture Model of Single Pixel BSS Algorithm.

Table 2-3. Formula of Single Pixel BSS Algorithm.

Algebra	Definition	Formula
$a(M) , a(L)$	Thermo-Vector of abnormal cells	$\begin{bmatrix} x_i(L) \\ x_i(M) \end{bmatrix} = \begin{bmatrix} a(L) & b(L) \\ a(M) & b(M) \end{bmatrix} \begin{bmatrix} x_i(S_a) \\ x_i(S_b) \end{bmatrix}$ $\begin{bmatrix} x_i(S_a) \\ x_i(S_b) \end{bmatrix} = \text{inv} \left(\begin{bmatrix} a(L) & b(L) \\ a(M) & b(M) \end{bmatrix} \right) \begin{bmatrix} x_i(L) \\ x_i(M) \end{bmatrix}$
$b(M) , b(L)$	Thermo-Vector of normal cells	
$x_i(M) , x_i(L)$	Thermo-Vector of test target	
$x_i(S_a) , x_i(S_b)$	Ratio of abnormal cells and normal cells	

2.2.2 Neighbor-based Blind Source Separation Algorithm

Another BSS algorithm applied in Dual-band IR Spectrogram is a Neighbor-based Blind Source Separation Algorithm. Consider each of the pixels in X_i which is $X_i = (x_i(M), x_i(L))$ as a vector where each pixel contains two signals emitted from MIR and LIR. We can compare them with the neighbor vectors of a control pixel called X_c which is $X_c = (x_c(M), x_c(L))$ and observe the variation of angles between pixels. Normal cells have similar quality and the variation of angles would be slight. On the other hand, abnormal (cancer) cells have an instable quality and the variation of angles would be violent. The details of Neighbor-based Blind Source Separation Algorithm are listed in Table 2-4.

Table 2-4. Formula of Neighbor-based BSS Algorithm.

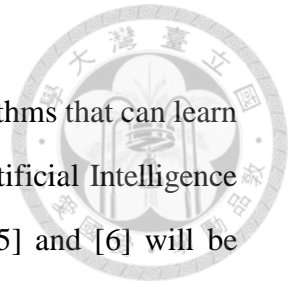
Algebra	Definition	Formula
X_c	Vector of Control Point	$E0 = \sum_{i=1}^n \cos^{-1} \frac{X_c X_i}{\ X_c\ \ X_i\ }$
X_i	Vector of Neighbor Points	

2.3 Machine Learning Models

Machine learning explores the study and construction of algorithms that can learn from and make predictions on data, and is an important topic in Artificial Intelligence Technology. Some machine learning techniques as described in [5] and [6] will be introduced in this section. When the data to process is huge, the computer can usually use same and duplicate operations to reduce overall efficiency of an algorithm. Applying Machine Learning, we can let computer to learn just like a living person. That is, the computer can get learning experience at each operation by making records such that these records can help the computer to work effectively. Testing data as described in [7], [8], [9] and [10] are input to our machine learning models and simulation of algorithms was done by using MATLAB Tools.

2.3.1 Supervised Learning and Unsupervised Learning

In classification, inputs are divided into two or more classes, and the learner must produce a model that assigns unseen inputs to one or more (multi-label classification) of these classes. After several operations are executed, we can put an Old Data into a particular Class according to its Feature. Above experience will help to construct a Model and such process is called Supervised Learning. On the other hand, we can also try to put a New Data into a particular Class according to its Feature. We can directly do this process with a similar Model as above and such process is called Unsupervised Learning. Experiment Process of Machine Learning Model is shown in Fig. 2-4.



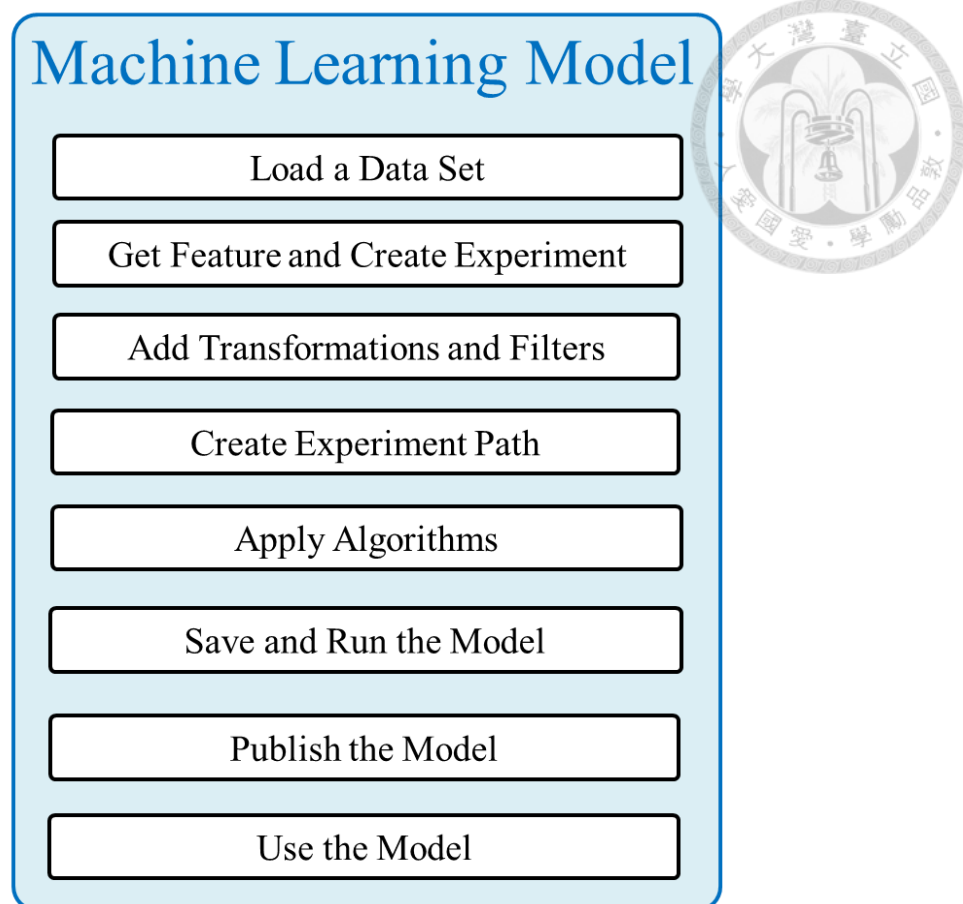


Fig. 2-4. Experiment Process of Machine Learning Model.

2.3.2 Linear Regression Model and Non-linear Regression Model

Regression analysis is a kind of supervised learning techniques, widely used for prediction and forecasting. Therefore, Regression Model can be applied to our Blind Source Separation Algorithms to make the calculating results more precise as well. Examples of using Linear and Non-linear Regression Models to classify Class 1 and Class 2 are shown in Fig. 2-5. By using such regression classification techniques, we can achieve our goal to separate original data into two different classes. As to the precision of the regression classification results, Non-linear Regression Model is better than Linear Regression Model because it uses a curve to make the regression instead of a straight line. We choose a Non-linear Regression Model called Augmented Regression Model to integrate into our algorithm design.

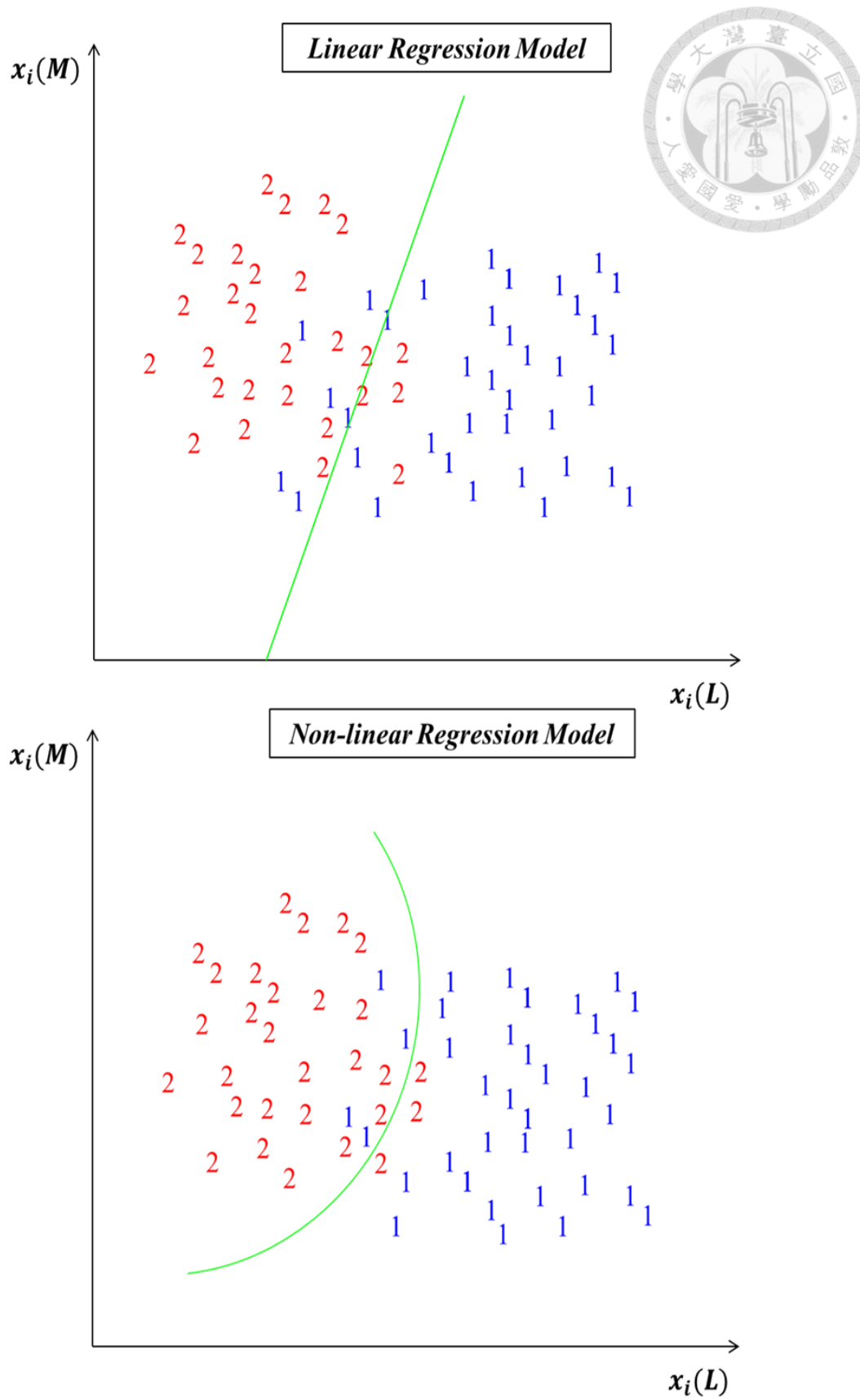


Fig. 2-5. Example of Linear and Non-linear Regression Model.

2.4 Application of Breast Cancer Detection

Due to the limit of atmosphere, wavelengths of MIR and LIR are set to be $3\mu\text{m}$ - $5\mu\text{m}$ for MIR and $8\mu\text{m}$ - $12\mu\text{m}$ for LIR in Dual-band IR Spectrogram. Local high temperature is most common phenomena in thermal images if some parts of the body are abnormal. Such characteristics of thermal phenomena can be used to judge which part of human body is abnormal.

2.4.1 Thermal Phenomena of Symptom

Different characteristics of thermal phenomena corresponding to different symptoms are listed in Table 2-5. On the other hand, these characteristics of thermal phenomena have different qualities of Clinical Test as well as listed in Table 2-6. Different thermal phenomena help us to evaluate the initial observation before we apply further algorithm of Dual-band IR Spectrogram.

Table 2-5. Thermal Phenomena of Symptoms.

Symptom	Thermal phenomena
Normal	Asymmetric Temperature is up to 0.5 degrees Celsius.
Tissue Dysplasia	Differences of Local High Temperature and Local Low Temperature are 1.0 degrees Celsius.
Pathological Changes	Local High Temperature is up to 1.5 degrees Celsius.
Cyst	Local Low Temperature is coated by Local High Temperature
Infection	Central Temperature is up to 5.0 degrees Celsius.
Tumor	Symmetric Temperature is up to 5.0 degrees Celsius.

Table 2-6. Qualities of Thermal Phenomena.

Thermal phenomena	Quality
Local High Temperature	Severity, Extensiveness, Distribution
Local Low Temperature	Severity, Extensiveness, Distribution
Asymmetric Temperature	Severity, Extensiveness, Symmetry
Temperature Gradient	Concentration, Dispersion

2.4.2 Critical Temperature and Gradient Temperature

Thermal temperature of an image can be calculated by a 14-bit gray level value of $X_i(x_i(M), x_i(L))$ as listed in Equation (2-1). By directly observing relative temperature, we get basic results of suspicious areas. Experiment process of Relative Temperature Algorithm is shown in Fig. 2-6.

$$T_i(t_i(M), t_i(L)) = X_i(x_i(M), x_i(L)) * \frac{T_{max} - T_{min}}{16383} - T_{env} \quad (2-1)$$

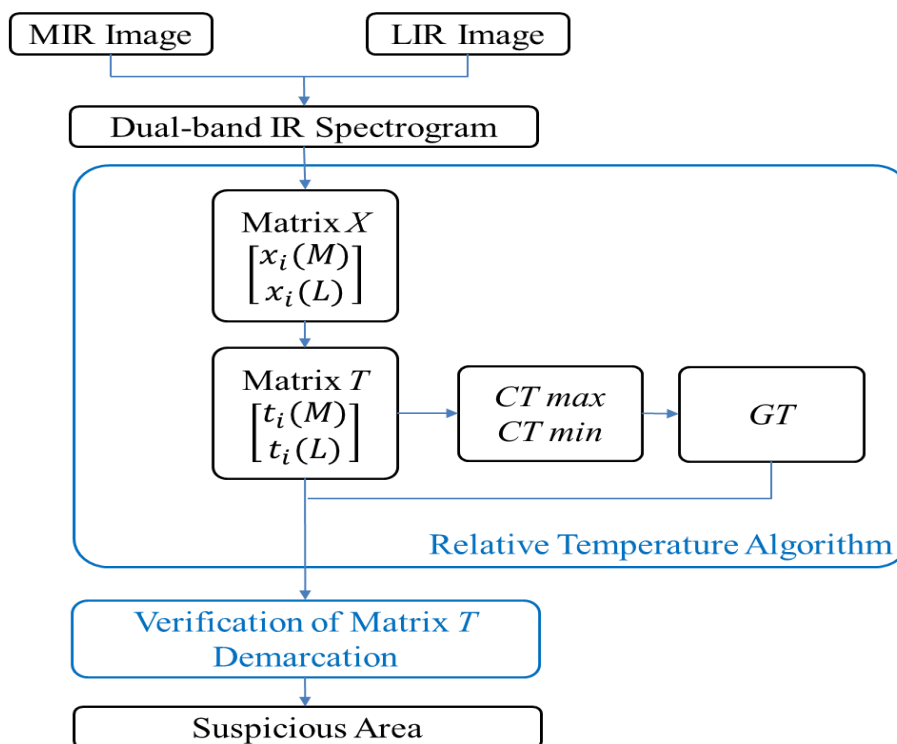


Fig. 2-6. Experiment Process of Relative Temperature Algorithm.

Given above relative temperature $T_i(t_i(M), t_i(L))$, we can calculate Critical Temperature using Equation (2-2) and Equation (2-3), where Maximum Critical Temperature ($CT\ max$) is eighty percent of the total heat, and Minimum Critical Temperature ($CT\ min$) is twenty percent of the total heat.

$$CT\ max = \sum T_i(t_i(M), t_i(L)) * 80\% \quad (2-2)$$

$$CT\ min = \sum T_i(t_i(M), t_i(L)) * 20\% \quad (2-3)$$

Gradient Temperature (GT) is calculated by Equation (2-4). Gradient Temperature helps to effectively observe the distribution of thermal phenomena in IR images. We will divide into A equal portions from $CTmax$ to $Tmax$, where A is the size of the image. Each portion differs from ΔT .

$$GT = \frac{\sum_{i=1 \text{ to } A} (CTmax + 0.5 * \Delta T * i) * (AREAi)}{CTmax - Tmax} \quad (2-4)$$

2.4.3 Risk Level of Breast Cancer

Breast cancer which is approximately 33% of all tumors in female bodies becomes the most common tumor. Breast cancer is the main cause of death of women whose ages are from 40 years old to 45 years old. There is a patient who is dead of breast cancer every 11 minutes. The risk level of breast cancer is listed in Table 2-7. As a result, building a compact detective process of breast cancer is an urgent topic.

Table 2-7. Risk Level of Breast Cancer.

Level	Risk	Treatment	Survival Rate (10-Year)
TH1	Very low	Annual Inspections	95%
TH2	Low	Symmetric Varices, Annual Inspections	88%
TH3	Middle	Semi-annual Inspections	66%
TH4	High	We need positive treatments to find the cause.	36%
TH5	Very High	Regular Inspections, Chemotherapy, Surgery	< 10%

The latest standard detective processes of breast cancer are shown in Fig. 2-7. Main detective processes consist of Nuclear Magnetic Resonance (NMR), X-ray and Ultrasound. NMR uses radio frequency as energy to make resonance of hydrogen atoms. According to Faraday Law, magnetic signals from resonance can be transformed to become current signals. We can get NMR images by current signals after we use Fourier Transform. Since the cost of NMR is huge, we often use NMR in final confirmation instead of periodic detection. The process of X-ray images is to

observe its X-ray decline. Patients need to be injected drugs and irradiated X-ray which become a burden to patients if they need periodic detection. The process of Ultrasound is to observe its sound reflection. Ultrasound images can be used to observe depth of tumors, size of tumors, and timely status of tumors. Since the precision of Ultrasound images depend on experiences of operators, it is difficult to build a standard process of Ultrasound detection. As a result, Ultrasound serves as an auxiliary tool in cancer

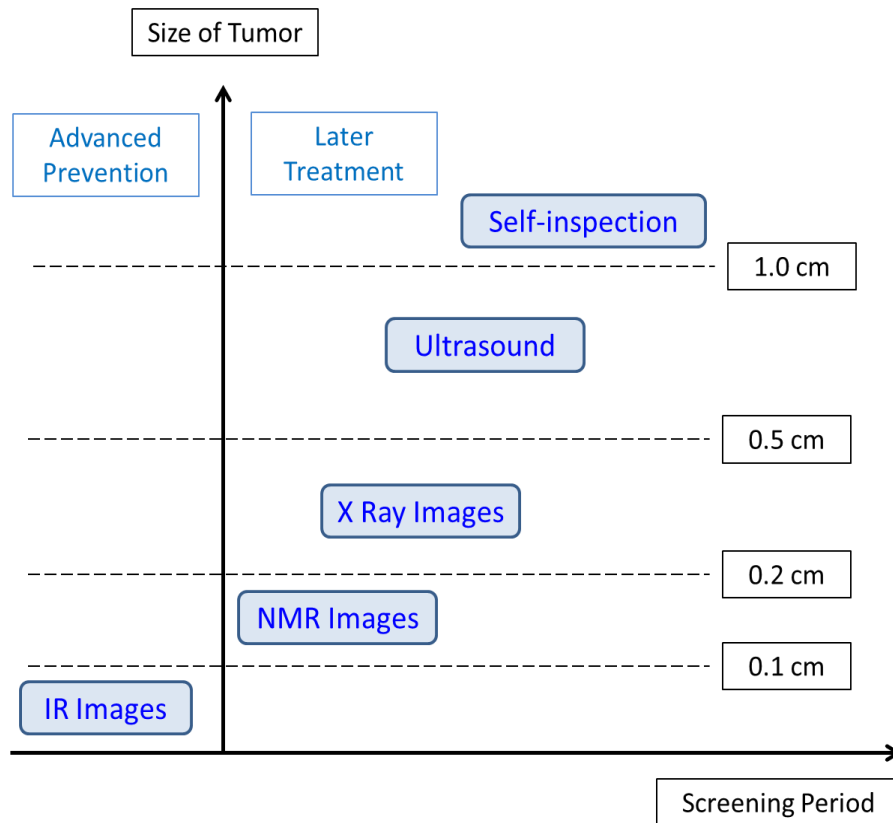


Fig. 2-7. Detective Process of Breast Cancer.



CHAPTER 3

BLIND SOURCE SEPERATION ALGORITHMS

In this chapter, we will present the experiment process of Dual-band IR Spectrogram on Single Pixel Blind Separation Algorithm in Section 3.1, Neighbor-based Blind Source Separation Algorithm in Section 3.2, and Machine Learning in Section 3.3. Then, we will introduce the IR system with which our algorithm will be integrated in Section 3.4.

3.1 Single Pixel Blind Source Separation Algorithm

In this section, we apply the proposed Single Pixel Blind Source Separation (BSS) Algorithm proposed in [1] to the two-band infrared spectrogram which contains a data vector at each individual pixel. This method is a basic BSS algorithm for abnormal tissue detection because two mixtures are needed to recover two unknown sources of normal tissue and abnormal tissue.

3.1.1 Principle of Single Pixel BSS

From the Plank Radiation Law figure as shown in Fig. 2.2, where four curves correspond to four different temperatures. For the high temperature curve #1, we define $a(M)$ and $a(L)$ which are respectively detectable by an MIR and an LIR sensors. For the low temperature curve #4, we define $b(M)$ and $b(L)$ which are respectively detectable by an MIR and an LIR sensors. We can find that as the temperature increases, the radiation increasing of a shorter wavelength band is faster the one of a longer wavelength band, such as the middle wavelength band and the long wavelength band ($a(M) - b(M) > a(L) - b(L)$). This important feature forms the basis of the linear mixture model.

According to the Linear Mixture Model, a linear mixing process operates on the two radiation vectors in matrix $A = [a \ b]$, where vector $a = [a(L) \ a(M)]^T =$

$[\cos \theta \quad \sin \theta]^T$ and vector $b = [b(L) \quad b(M)]^T = [\cos \varphi \quad \sin \varphi]^T$. The above two radiative angles θ and φ play important roles in determining the temperature of cells which include abnormal cells and normal cells. The measured data vector is a linear combination of the radiation vectors $X_i = x_i(s_a) * a + x_i(s_b) * b$ which include $x_i(M) = x_i(s_a) * a(M) + x_i(s_b) * b(M)$ and $x_i(L) = x_i(s_a) * a(L) + x_i(s_b) * b(L)$. Therefore, the following Equation (3-1) and Equation (3-2) form the main core of a Single Pixel Blind Source Separation (BSS) Algorithm.

$$a = \begin{bmatrix} a(L) \\ a(M) \end{bmatrix} = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, \quad b = \begin{bmatrix} b(L) \\ b(M) \end{bmatrix} = \begin{bmatrix} \cos \varphi \\ \sin \varphi \end{bmatrix} \quad (3-1)$$

$$X = AS, \quad \begin{bmatrix} x_i(L) \\ x_i(M) \end{bmatrix} = \begin{bmatrix} a(L) & b(L) \\ a(M) & b(M) \end{bmatrix} \begin{bmatrix} x_i(s_a) \\ x_i(s_b) \end{bmatrix} \quad (3-2)$$

3.1.2 Application of Single Pixel BSS

First, RAW Data obtained from FLIR SC4000 Camera contain two images called MIR image and LIR image. The size of each image is 256 x 320. Every pixel is 14-Bit and every pixel has an MIR value and an LIR value. Before we finish the alignment of images, we assign the abnormal pixel to be vector a and the normal pixel to be vector b . These two vectors are then combined into a mixing matrix A . Let the test pixel be matrix X and matrix S can be found by conversions of matrices. Fig. 3-1 shows the procedure of applying Single Pixel Blind Source Separation Algorithm to Dual-band IR Spectrogram. Finally, the accuracy of detection is evaluated by comparing the result of matrix S with the data obtained in Clinical Test.

3.1.3 Challenge in Single Pixel BSS Algorithm

On the other hand, if an iterative learning process is used in the Single Pixel BSS Algorithm to find a stable Matrix S , more noises will be generated. It is thus better to use a Neighbor-based BSS Algorithm which is an extension of the Single Pixel BSS Algorithm to find Matrix S . Neighbor-based BSS Algorithm finds Matrix S by using information energy which was ignored in Single Pixel BSS Algorithm. Details of Neighbor-based BSS Algorithm will be given in next section.

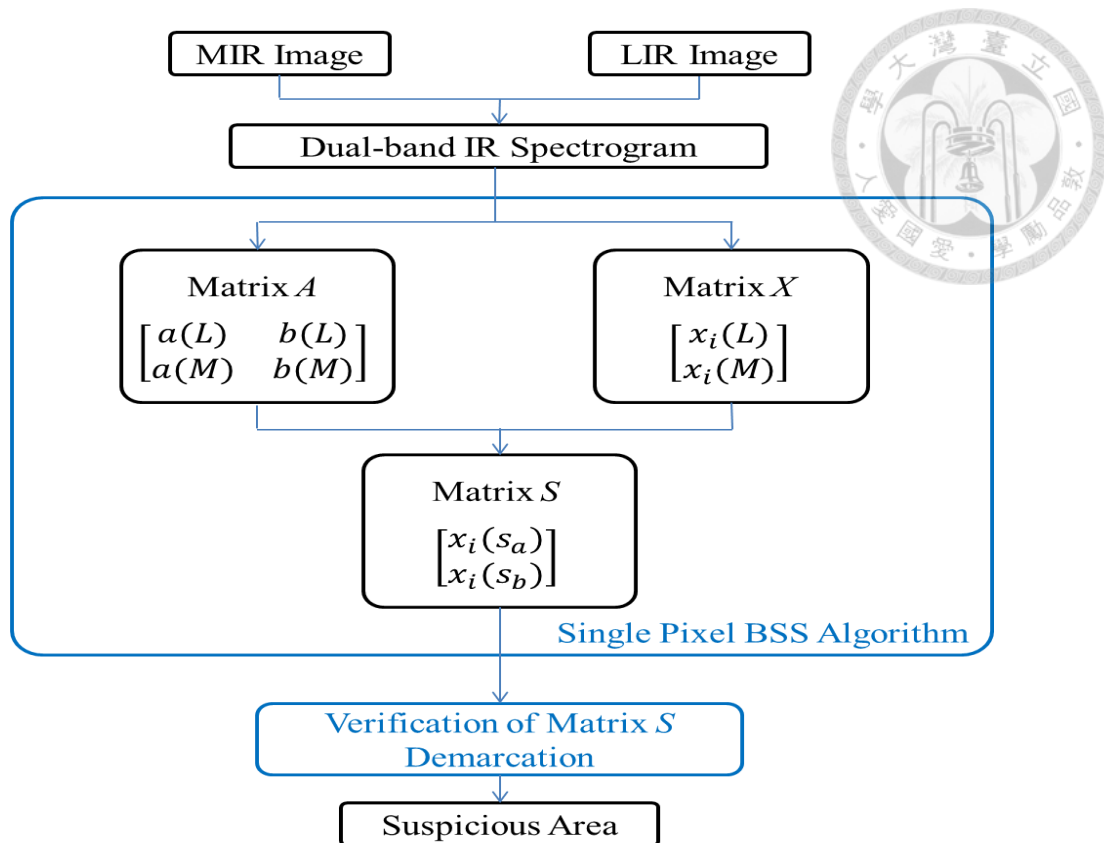


Fig. 3-1. Experiment Process of Single Pixel BSS Algorithm.

3.2 Neighbor-based Blind Source Separation Algorithm

In this section, an in-depth analysis on the uniqueness of solution obtained by the Neighbor-based BSS Algorithm is given by using geometric illustration, because Single Pixel BSS Algorithm is using an algebraic method and Neighbor-based BSS Algorithm is using a geometric method.

3.2.1 Principle of Neighbor-based BSS

Since the two sources $x_i(s_a)$ and $x_i(s_b)$ are not independent, we can eliminate one parameter by expressing $x_i(s_b)$ in terms of $x_i(s_a)$. Because of $x_i(s_a) + x_i(s_b) = 1$, the principle of transformation between Equation (3-3) and Equation (3-4) is shown as follows.

$$S = -x_i(s_a) * \ln(x_i(s_a)) - (1 - x_i(s_a)) * \ln(1 - x_i(s_a)) \quad (3-3)$$

$$E = E0 + (\mu2 - \mu1) * x_i(s_a) + \mu1 * w1 * x + \mu2 * w2 * x - \mu2 \quad (3-4)$$

Note that the information energy E is a straight line with a slope of $(\mu_2 - \mu_1)$ which is decided by Curve S . Curve S is defined for $x_i(s_a)$. When we find a tangent between straight line E and Curve S , we define $E = E^*$ and $x_i(s_a) = x_i(s_a)^*$ as the optimal solutions obtained in Fig. 3-2

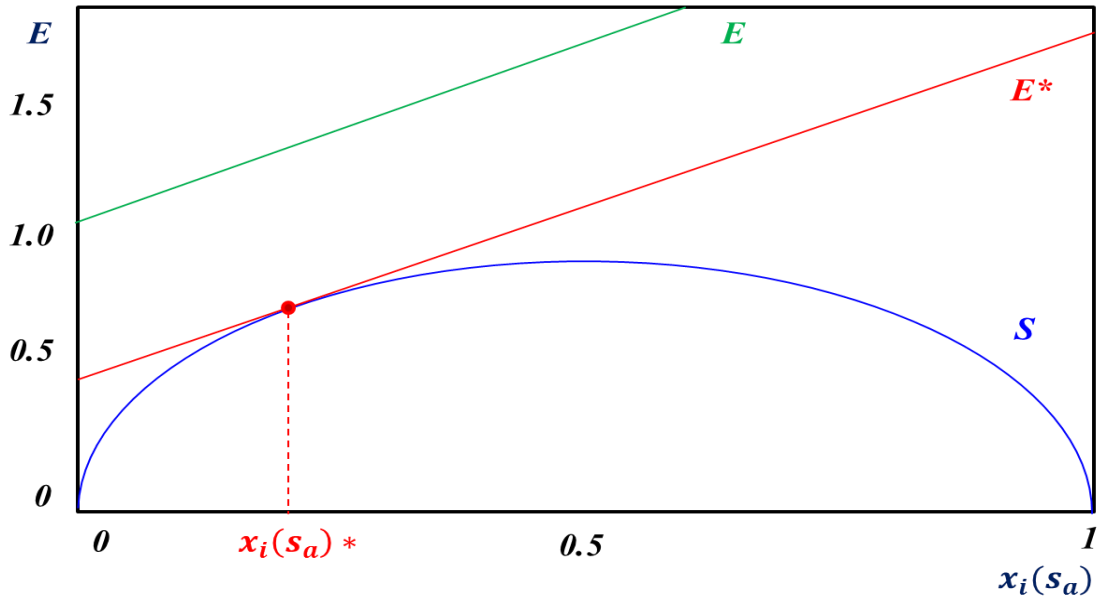


Fig. 3-2. Geometric illustration of unique solution.

The algorithm uses iterative learning rules to find an optimal solution. Note that the straight line with the vertical axis is given by $(E_0 + \mu_1 * w_1 * x + \mu_2 * w_2 * x - \mu_2)$. The term $(\mu_1 * w_1 * x + \mu_2 * w_2 * x - \mu_2)$ can be ignored because its influence is slight and we have now Equation (3-5).

$$E = E_0 + (\mu_2 - \mu_1) * x_i(s_a) \quad (3-5)$$

The interception of the energy line is uniquely determined by E_0 . In this case, if E_0 is determined, Matrix S can be deterministically obtained without the need of an iterative learning process. The mean value E_0 is estimated from the surrounding image of a point target. Our guess is that the E_0 of a certain pixel is a function of observation, which can be estimated by the majority of tissue everywhere from the neighborhood of an unknown location point target pixel. From the geometric point of

view, each point is a vector in a multidimensional space. If there is no target present in a neighborhood, then all data vectors point to the same direction of the background signature and the angles will be zero. Suppose that there is a subpixel target, one of the data vectors will depart from others and the angles are not zero any more. Obviously, the larger the fractional abundance of target, the further away this corresponding pixel departs from others, and the larger the angles. Based on these observations, we propose to use the angle between data vectors as the measure of information energy $E0$.

To estimate a suitable $E0$, we need to derive the relationship between $E0$ and $x_i(s_a)$ in Equation (3-6).

$$E0 = -\ln(1 - x_i(s_a)) \quad (3-6)$$

Let $(X_1, X_2 \dots X_n)$ denote the data within a neighborhood, where each column vector represents a data point and n is the size of neighborhood, and the control pixel is represented by X_c . The information energy is calculated by the summation of the angles between the control pixel and its neighbors in Equation (3-7).

$$E0 = \sum_{i=1}^n \cos^{-1} \frac{X_c X_i}{\|X_c\| \|X_i\|} \quad (3-7)$$

The information energy $E0$ plays an important role because it will influence $s1$ greatly. Every pixel is a vector containing an MIR signal and an LIR signal. Therefore, every pixel has an angle with its neighborhood pixels. $E0$ can then be built by the angles of neighborhood pixels. To estimate $E0$, we could observe the change of angles. A normal cell has similar quality and its variation of angle will be slight. On the other hand, an abnormal cell has instable quality and its variation of angle will be violent.

3.2.2 Application of Neighbor-based BSS

First, RAW Data input from the FLIR SC4000 Camera contain two images captured by the MIR and LIR camera sensors. The size of each image is 256 x 320, each pixel is 14-Bit containing an MIR value and an LIR value. Before the alignment of images, we build a vector for every pixel with the MIR value and LIR value. To observe the change of angle, we first set a control point called X_c and calculate with its n neighborhood points X_i , where $i = 1$ to n . Then, we mark a suspicious area if its change of angle is violent. Fig. 3-3 shows the experiment procedure of Neighbor-based Blind Source Separation Algorithm. Finally, the accuracy of detection can be evaluated by comparing the area with a large change of angle with the data obtained in Clinical Test.

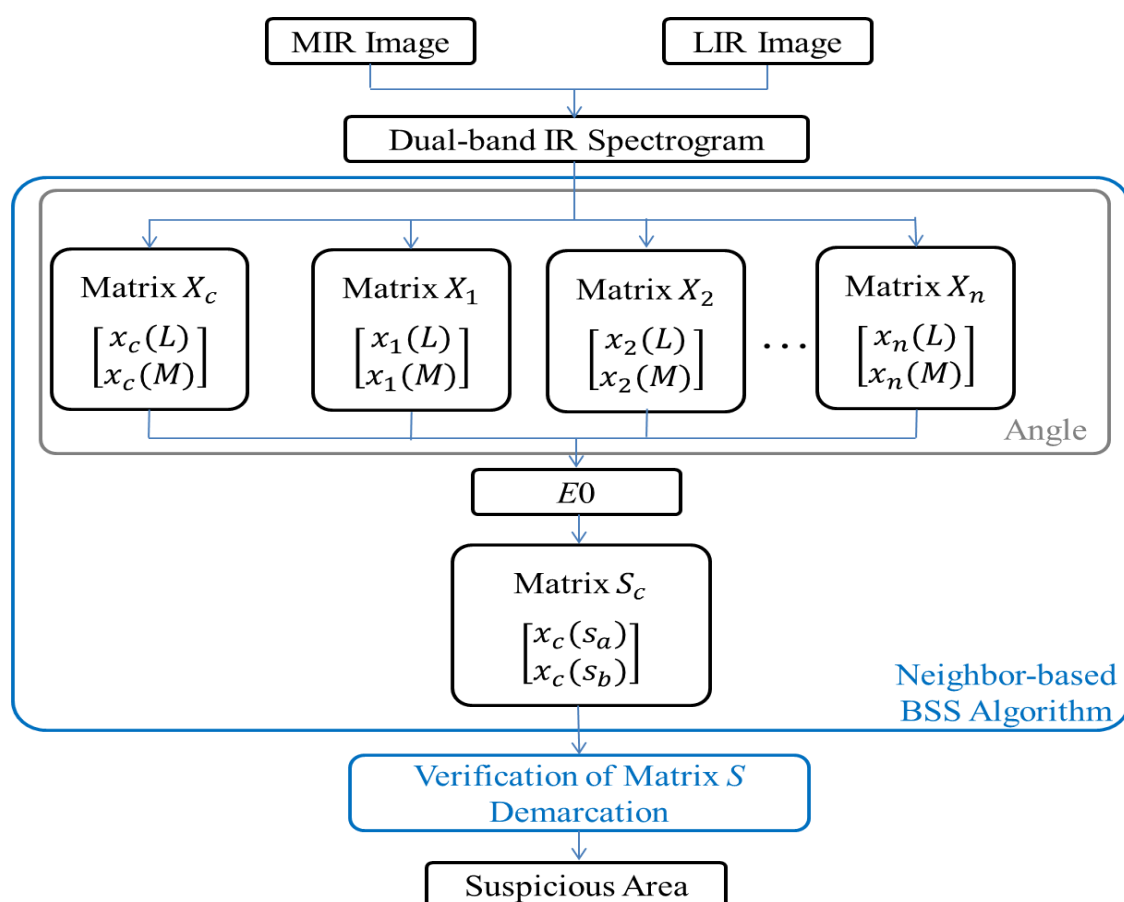


Fig. 3-3. Experiment Process of Neighbor-based BSS Algorithm.

3.2.3 Challenge in Neighbor-based BSS Algorithm

Before we apply Neighbor-based BSS Algorithm, we design a suitable method to align images obtained from the MIR and LIR cameras to avoid errors of mismatched location during the period of filming.

Distribution of abnormal cells and normal cells is shown in Fig. 3-4. Neighbor-based BSS Algorithm gets Matrix S by computing the sum of neighbor angles. When we compute the sum of neighbor angles directly, we might lose some details because the distribution of angles surrounding control point is unknown.

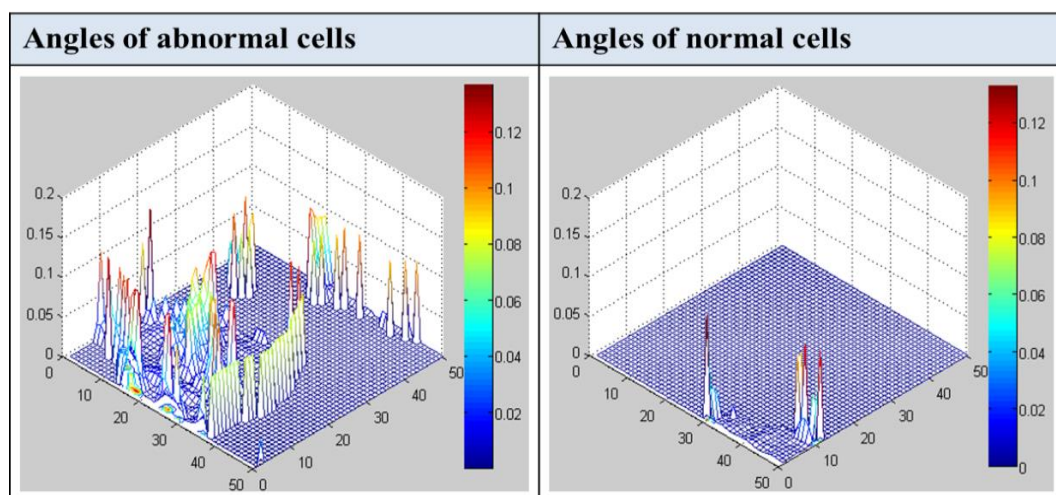


Fig. 3-4. Distribution of Different Cells.

3.3 Machine Learning Analysis Algorithms

A regression model can be used to distinguish abnormal tissue from normal tissue. Here, we define abnormal pixels as Class 1 pixels which scatter densely and normal pixels which scatter discretely as Class 2 in a suspicious area.

3.3.1 Linear Regression Model

In Supervised Learning, a data called X_i form the elements of Matrix X_L with coordinates of $X_i(M)$ and $X_i(L)$. All X_i have an intuitive class called $y(i)$ which form the elements of Matrix y . We use β_0 , β_1 and β_2 in Matrix β_L to draw the

regression line Y_{L_i} as described in Equation (3-9), Equation (3-10), and Equation (3-11). Finally, Matrix β_L can be obtained from Equation (3-8).

$$\beta_L = (X_L^T X_L)^{-1} X_L^T y, \beta_A = (X_A^T X_A)^{-1} X_A^T y \quad (3-8)$$

$$Y_{L_i} = \beta_0 + \beta_1 x_i(M) + \beta_2 x_i(L) \quad (3-9)$$

$$\beta_L = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, y = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(n) \end{bmatrix} \quad (3-10)$$

$$X_L = \begin{bmatrix} 1 & x_1(M) & x_1(L) \\ 1 & x_2(M) & x_2(L) \\ \vdots & \vdots & \vdots \\ 1 & x_n(M) & x_n(L) \end{bmatrix} \quad (3-11)$$

3.3.2 Augmented Regression Model

In Supervised Learning, the effect of Linear Regression Model is worse than Augmented Regression Model, if Class 1 and Class 2 do not scatter distinguishingly. In addition to $x_i(M)$ and $x_i(L)$, we need to add three variables called $x_i(M)x_i(L)$, $x_i(M)^2$ and $x_i(L)^2$. All X_i still have a class called $y(i)$ which are elements of Matrix y . We use β_0 , β_1 , β_2 , β_3 , β_4 and β_5 in Matrix β_A to draw the regression curve Y_{A_i} as described in Equation (3-12), Equation (3-13), and Equation (3-14). Finally, Matrix β_A Can be obtained from Equation (3-8).

$$Y_{A_i} = \beta_0 + \beta_1 x_i(M) + \beta_2 x_i(L) + \beta_3 x_i(M)x_i(L) + \beta_4 x_i(M)^2 + \beta_5 x_i(L)^2 \quad (3-12)$$

$$\beta_A = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}, y = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} \quad (3-13)$$

$$X_A = \begin{bmatrix} 1 & x_1(M) & x_1(L) & x_1(M)x_1(L) & x_1(M)^2 & x_1(L)^2 \\ 1 & x_2(M) & x_2(L) & x_2(M)x_2(L) & x_2(M)^2 & x_2(L)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n(M) & x_n(L) & x_n(M)x_n(L) & x_n(M)^2 & x_n(L)^2 \end{bmatrix} \quad (3-14)$$

3.4 Experiment Procedures

In this section, we introduce the specification of SC4000 FLIR Camera used to get RAW Data of MIR Images and LIR Images. In the future, we will apply our Improved Neighbor-based BSS Algorithm in the back-end processor of our developed Dual-band IR Sensor and Readout Circuit platform.

3.4.1 Specification of SC4000 FLIR Camera

We briefly introduce the FLIR Camera according to what was described in [9]. FLIR is a famous company which has developed many kinds of IR cameras for generating infrared ray spectrum. SC4000 FLIR Camera is used to capture RAW Data of Dual-band IR Spectrogram. A picture of SC4000 FLIR Camera is shown in Fig. 3-5. The Specification of SC4000 FLIR Camera is listed in Table 3-1.

The role of SC4000 FLIR Camera in our Experiment Procedure is shown in Fig. 3-6. Blind Source Separation Algorithm is run after Dual-band IR Spectrogram is generated by the SC4000 FLIR Camera. In the future, we will apply our improved Neighbor-based BSS Algorithm in our developed Dual-band IR Platform.



Fig. 3-5. Picture of SC4000 FLIR Camera.

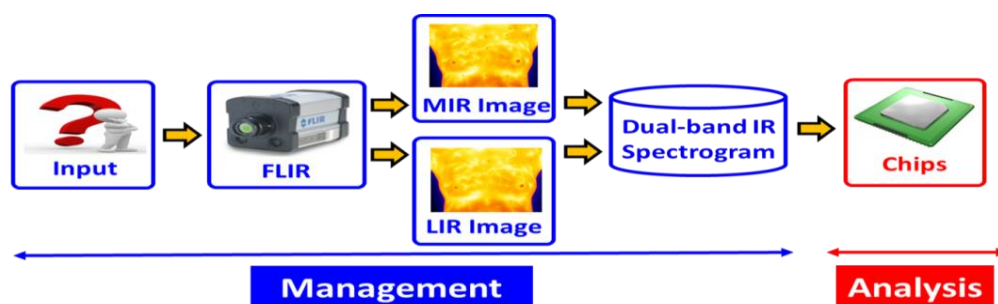


Fig. 3-6. Experiment Process of SC4000 FLIR Camera.

Table 3-1. Specification of SC4000 FLIR Camera.

	SC4000 MW IR	SC4000 LW IR
<i>Detector Specification</i>		
Detector	InSb	GaAs, QWIP
Spectral Range	3.0- μm ~5.0- μm	8.0- μm ~9.2- μm
Broadband Option	1.5- μm ~5.0- μm	NA
Resolution	320(H) * 256(V)	320(H) * 256(V)
Pixel Pitch	30- μm * 30- μm	30- μm * 30- μm
<i>Electronics & Data Rate</i>		
Integration Type	Snapshot	Snapshot
Integration Time	5- μs	5- μs
Read-out Mode	Asynchronous Integrate while read Asynchronous Integrate then read	Asynchronous Integrate while read Asynchronous Integrate then read
Dynamic Range	14-Bit	14-Bit
Data Rate	50-MHz	50-MHz
Full Frame Rate	0.1-Hz~432-Hz	0.1-Hz~432-Hz
Sub-windowing	Yes	Yes
Minimum Window Size	2 * 64	2 * 64
Super-framing	Yes	Yes
Preset Sequencing	Yes	Yes
<i>Performance Specification</i>		
NEI / NITD	< 18-mK	< 35-mK
Well Capacity	18-M Electrons / 4.5-M Electrons	18-M Electrons / 4.5-M Electrons
Operability	>99.95% Typically	>99.95% Typically

3.4.2 Dual-band IR Platform

To develop an algorithm to process Dual-band IR Spectrogram, one should have basic trainings on research ethics. The author took related courses offered by the Institutional Review Board/Ethics Committee (IRB/EC) which was described in [11] to ensure that any conduct of medical related research is appropriate and legal.

In the future, we will use Raspberry Pi which was described in [12] to be a bridge between Improved Neighbor-based BSS Algorithm and Dual-band IR Platform. According to characteristics of infrared ray which was described in [13], we apply our Improved Neighbor-based BSS Algorithm in a Dual-band IR Platform which is developed by our Lab., which includes a designed Dual-band IR sensor, a Dual-band IR Readout Circuit, an Analog-to-Digital Converter (ADC), and a Back-end ARM Processor. The whole Dual-band IR Platform is shown in Fig. 3-7

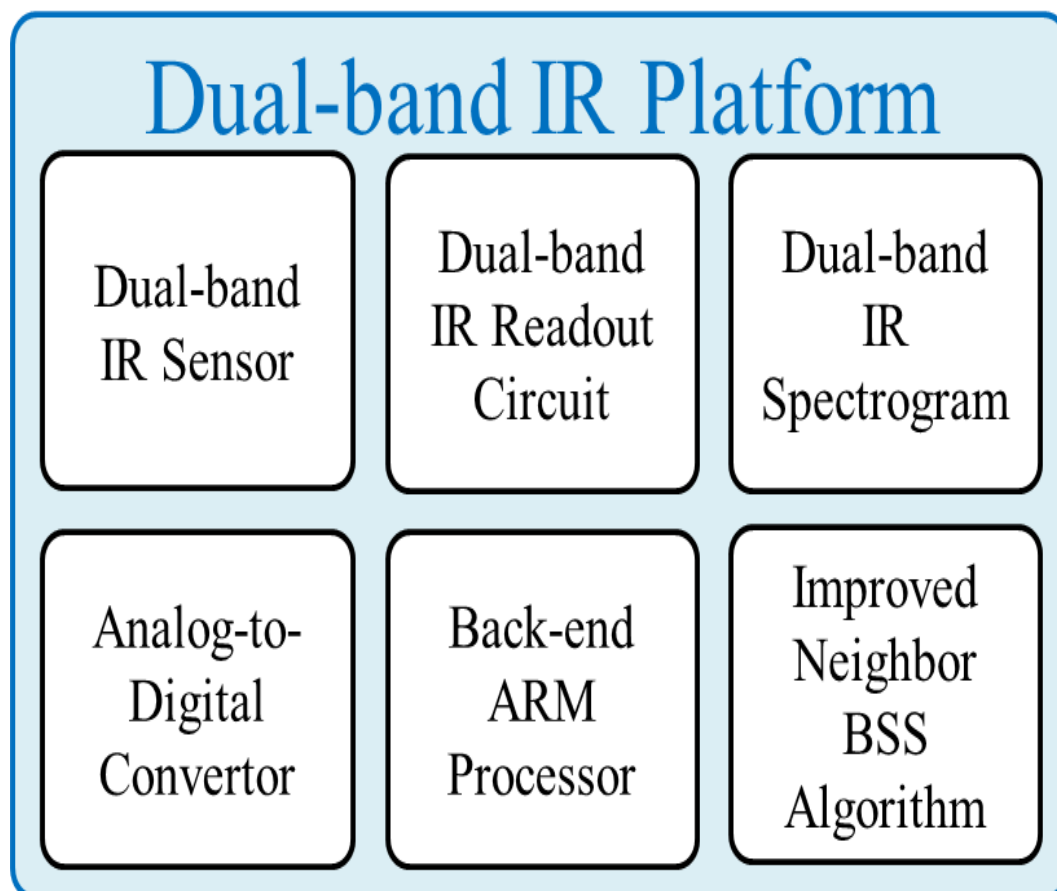


Fig. 3-7. Dual-band IR Platform.

The Dual-band IR Readout Circuit developed in our laboratory with background suppression function was fabricated in 90nm technology. Example of chip layout view, which was described in [14], is shown in Fig. 3-8 and the implementation specifications are described in Table 3-2.

The proposed Readout chip is designed for switching between Buffer Direct Injection (BDI) and Direct Injection (DI) with a pixel pitch $80\mu\text{m} \times 50\mu\text{m}$ and an operational frequency of 0.3MHz. Because of the two preamplifiers and selectable capacitor, the readout chip can support up to four input current ranges. The injection efficiency of BDI is 99.99% and DI is over 80%. The most importance parameter in a readout circuit design is its linearity of input to output, and our chip linearity reaches over 99%. The chip achieves a frequency of 0.3MHz per-pixel and with $176.84\mu\text{W}$ power consumption (without output buffer). Furthermore, our readout chip uses a switching mechanism to provide negative and positive bias voltages to our dual-dual IR sensors, such that we can detect the intensity of either MIR or LIR radiation. By getting signals of MIR and LIR from the Readout circuit, we can then use our Improved Neighbor-based BSS Algorithm to demarcate suspicious areas in accordance with Verification of Dual-band IR Spectral Ratio.

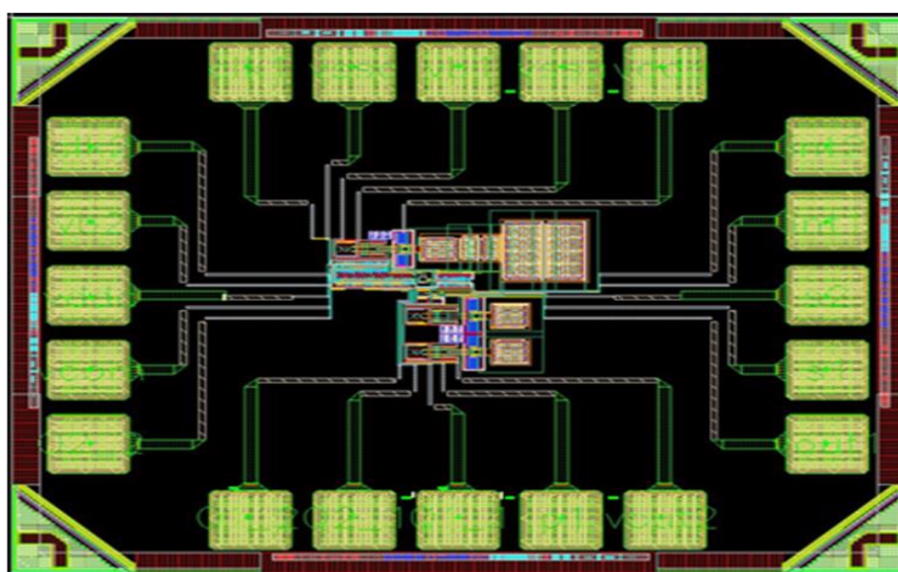


Fig. 3-8. Chip Layout of Dual-Band IR Readout Circuit.

Table 3-2. Specification of Dual-Band IR Readout Circuit.

Performance Parameter		Post Simulated Results	
Technology		TSMC 90-nm 1P9M CMOS	
Power Supply		1.2V	
Power Dissipation (without output buffer)		176.84 μ W	
Total Power Dissipation		535.89 μ W	
Pixel Pitch		80 μ m \times 50 μ m (2 pixels)	
Chip Size		631 μ m \times 545 μ m	
Integration Capacitance		250fF and 2pF	
Types of Bias		Negative & Positive	
Input Current	BDI	250fF	70nA ~ 180nA
		2pF	0.5 μ A ~ 1.2 μ A
	DI	250fF	102nA ~ 210nA
		2pF	0.9 μ A ~ 1.6 μ A
Injection Efficiency	BDI		99.99%
	DI		82%
Output Voltage Swing		0.5V ~ 1.05V	
Readout Speed		0.3MHz/one pixel	
Background Suppression Level		50nA ~ 250nA	
Linearity		>99%	





CHAPTER 4

SIMULATION AND EXPERIMENT RESULTS

In this chapter, we will introduce methods to pre-process the RAW Data in Section 4.1. We will show the Experiment Results obtained by our improved BSS algorithms in Section 4.2, and evaluate their advantages and disadvantages.

4.1 RAW Data Pre-Processing

To illustrate infrared images which are invisible, we have to compress RAW Data from 14-bit to 8-bit, as described in [15]. During the compression, some details are lost. As a result, a method capable to keep major details and abandon minor details after the compression is required. After compression, RAW Data Coding has to be verified whether it is correct or not because compression of RAW Data has new distribution of values.

4.1.1 Histogram Equalization

Histogram Equalization is the most common method used in compression of Dual-band IR Spectrogram. At most 16384 kinds of values exist in a 14-bit data. If RAW Data only has p kinds of values in 14-bit where $p < 16384$, one can make p be a denominator in place of 16384. With the proportional distribution of p , one can keep major details and abandon minor details in 8-bit after the compression. The diagram of Histogram Equalization is shown in Fig. 4-1.

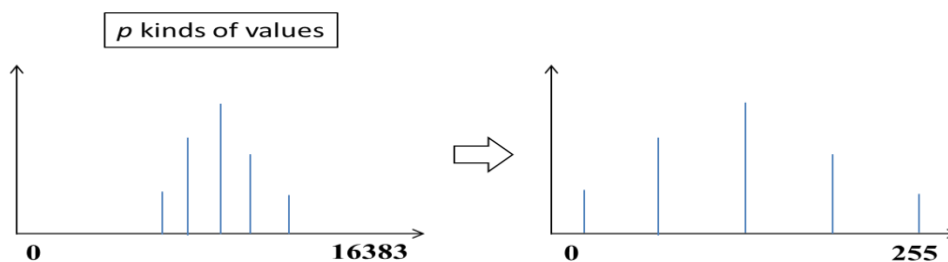


Fig. 4-1. Diagram of Histogram Equalization.

4.1.2 Binary Occupied Histogram Projection

Binary Occupied Histogram Projection is often used in industry. To find the p kinds of values which form q entries in 14-bit, we sort q entries to occupy the range of an 8-bit data with a proportional distribution of q . In this way, we can keep major details and abandon minor details in 8-bit after the compression. The diagram of Binary Occupied Histogram Projection with $p=256$ and $q=16384$ is shown in Fig. 4-2.

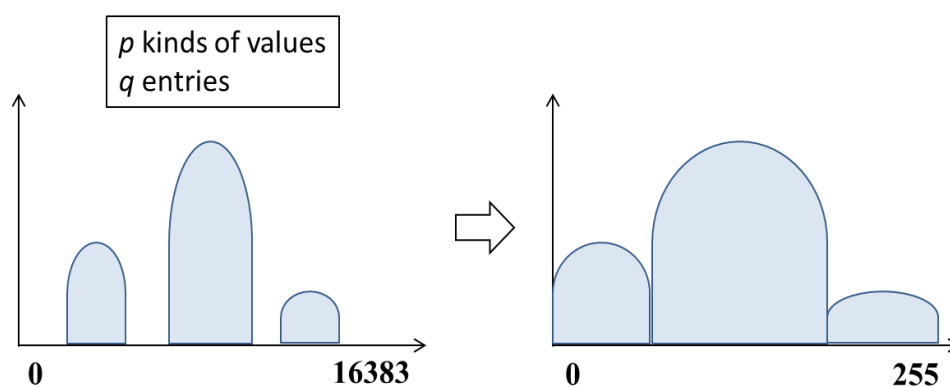


Fig. 4-2. Diagram of Binary Occupied Histogram Projection.

4.1.3 Results of Compression

The original images which include MIR image and LIR image are shown in Fig. 4-3, where we transform a Binary data into a Decimal data. Distribution of MIR is approximately 10000-12000. Distribution of LIR is approximately 9000-9500.

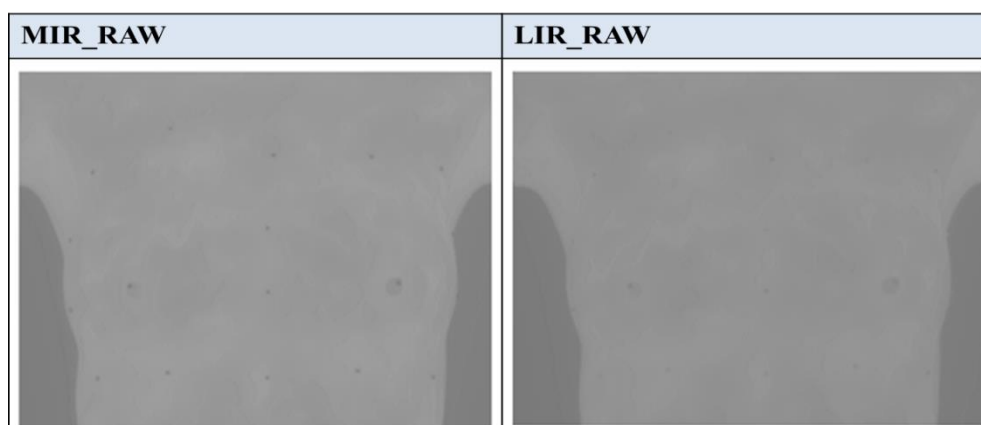


Fig. 4-3. RAW Data Pre-Processing of MIR Image and LIR Image.

4.2 Experiment Results

In this section, we use the clinical test data from patients who took long-term chemotherapy as our raw data. The photos taken by an SC4000 FLIR Camera, for the patients who receive chemotherapy each time, are Dual-band IR Spectrogram. First, pictures of MIR Image and LIR Image, obtained from the RAW Data by using Gamma Correction and Temperature Color, are shown in Fig. 4-4.

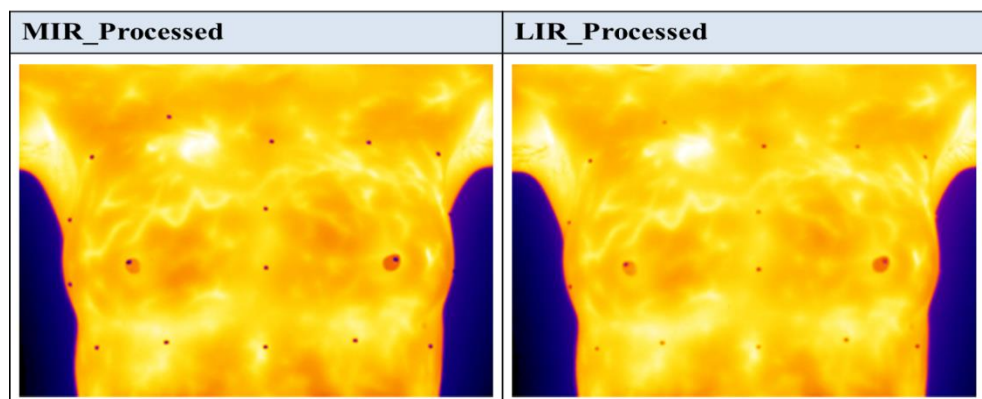


Fig. 4-4. Pictures of MIR Image and LIR Image.

4.2.1. Algorithms used in our Experiments

First, we use Relative Temperature Algorithm directly to demarcate pixels whose value of relative temperature (Matrix T) is abnormal. Though Single Pixel BSS Algorithm can find Matrix S by matrix operations, but it needs golden vectors such as $a(M)$, $a(L)$, $b(M)$, and $b(L)$, which values are difficult to define because of different environmental conditions. Moreover, if we use an iterative learning process to find a stable Matrix S , we will get more noises by using Single Pixel BSS Algorithm. Therefore, we just use a Neighbor-based BSS Algorithm which is an extension of the Single Pixel BSS Algorithm to find Matrix S as our second BSS algorithm. The disadvantage of using a Neighbor-based BSS Algorithm is that we are not able to know where change of angle is violent because we simply calculate the sum of neighborhood angles in that algorithm. Therefore, in our Improved Neighbor-based BSS Algorithm, we use an augmented regression model to confirm which part of

angles is certainly violent. The effects of above Relative Temperature Algorithm, Neighbor-based Blind Source Separation Algorithm, and Improved Neighbor-based Blind Source Separation Algorithm will be evaluated and the comparison of Algorithms is listed in Table 4-1. Without using the sum of neighborhood angles to derive Matrix S , our Improved Neighbor-based BSS Algorithm, which observes the change of angles directly and evaluates whether the distribution of angles is violent or not, might lose some details. The experiment procedure of the Improved Neighbor-based BSS Algorithm is shown in Fig. 4-5.

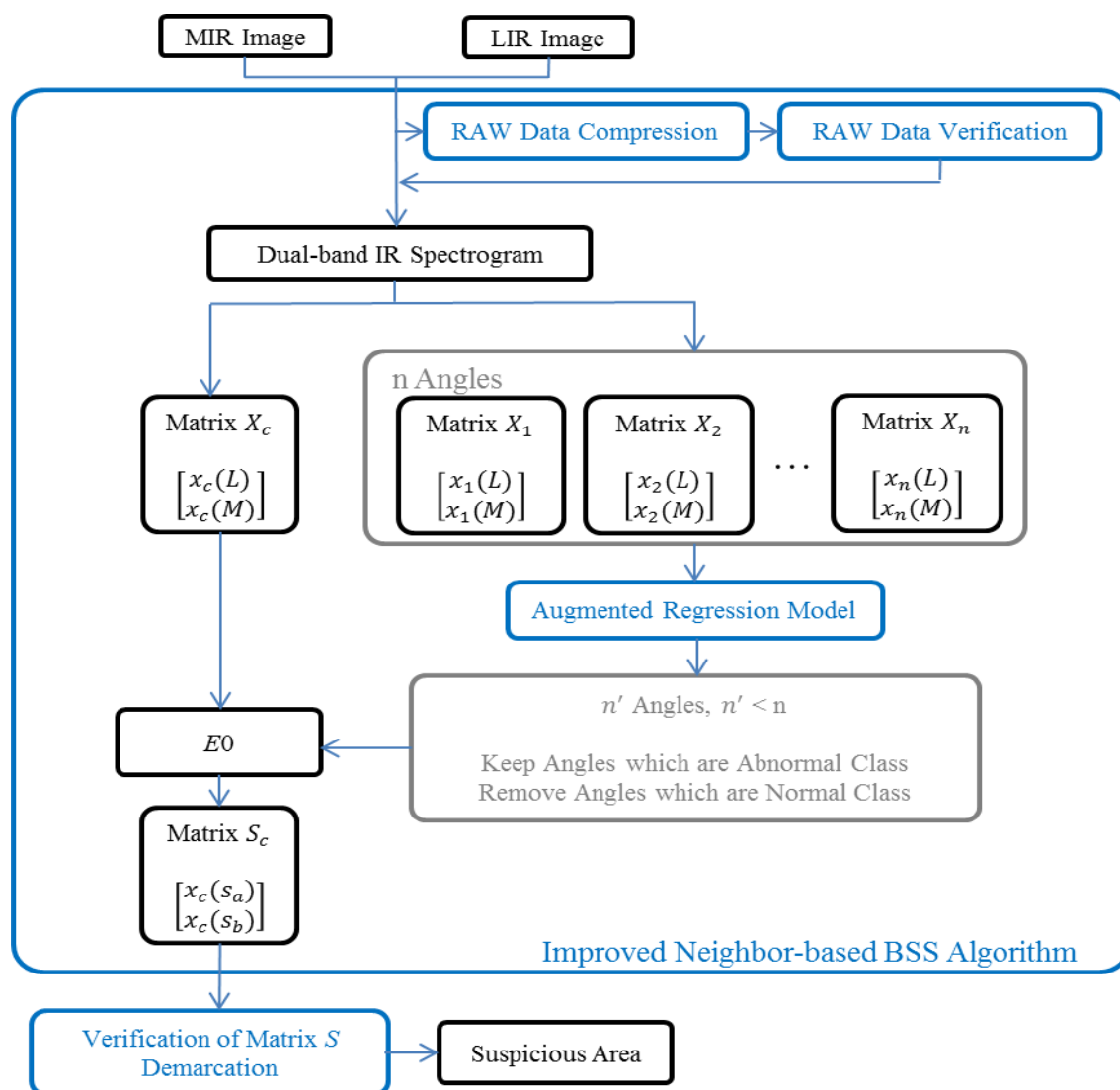


Fig. 4-5. Experiment Process of Improved Neighbor-based BSS Algorithm.

Three sets of clinical tests data are input to the above Relative Temperature Algorithm, Neighbor-based BSS Algorithm, and Improved Neighbor-based BSS Algorithm. We have in total 50 RAW Data from patients who accepted periodic chemotherapy and these data are classified into three main groups called as C001, C002 and C003. These groups of Dual-band IR Spectrograms show three representative treatment results after periodic chemotherapy. The comparisons of results obtained from the above three BSS algorithms on the three different sets of test data are respectively shown in Fig. 4-6, Fig. 4-7, Fig. 4-8, Fig. 4-9, Fig. 4-10, Fig. 4-11, Fig. 4-12, Fig. 4-13, and Fig. 4-14.

Table 4-1. Comparison of Algorithms

	Accordance	Advantage	Disadvantage
Relative Temperature Algorithm	Matrix T $\begin{bmatrix} t_i(M) \\ t_i(L) \end{bmatrix}$	Direct Observation	Less Precise
Single Pixel BSS Algorithm	Matrix S $\begin{bmatrix} x_i(s_a) \\ x_i(s_b) \end{bmatrix}$	Easy Simulation Matrix Operation	Less Precise Need Golden Vector
Neighbor-based BSS Algorithm	Matrix S $\begin{bmatrix} x_i(s_a) \\ x_i(s_b) \end{bmatrix}$	Extension of SP BSS More precise than SP BSS	We do not know which part of angle is more violent due to the sum of angles.
Improved Neighbor-based BSS Algorithm	Matrix S $\begin{bmatrix} x_i(s_a) \\ x_i(s_b) \end{bmatrix}$	Extension of N BSS More precise than N BSS	We need additional time of computation when we use machine learning model.

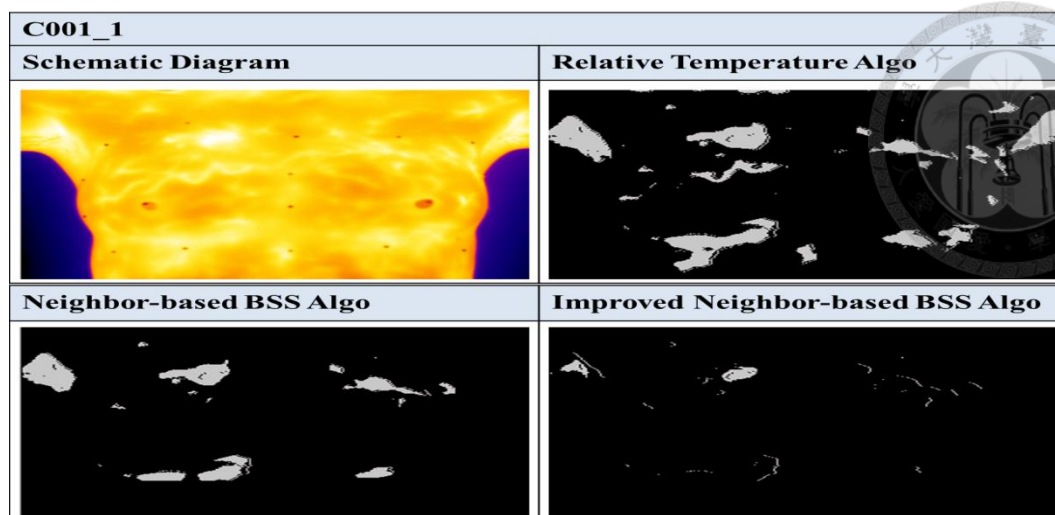


Fig. 4-6. Comparison of Detection Algorithms for C001_1.

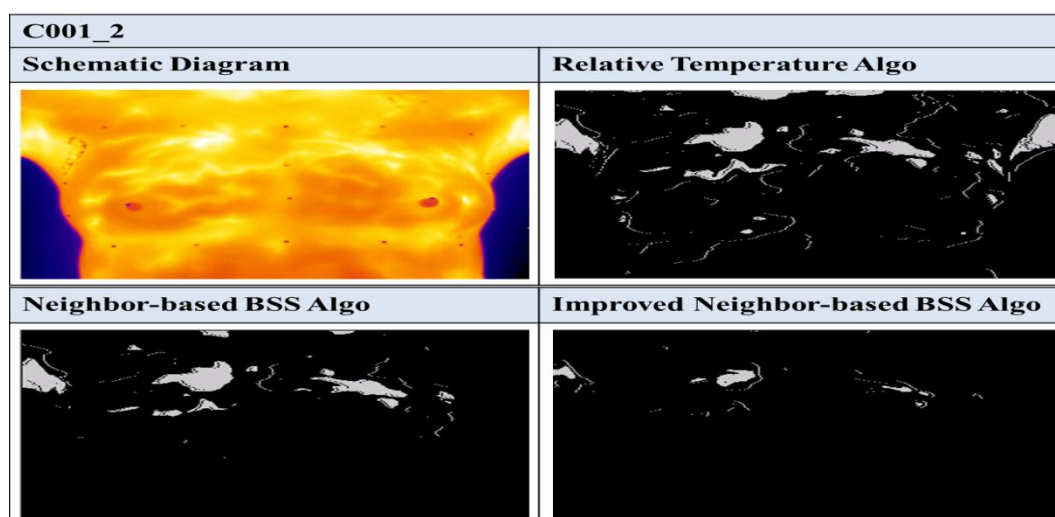


Fig. 4-7. Comparison of Detection Algorithms for C001_2.

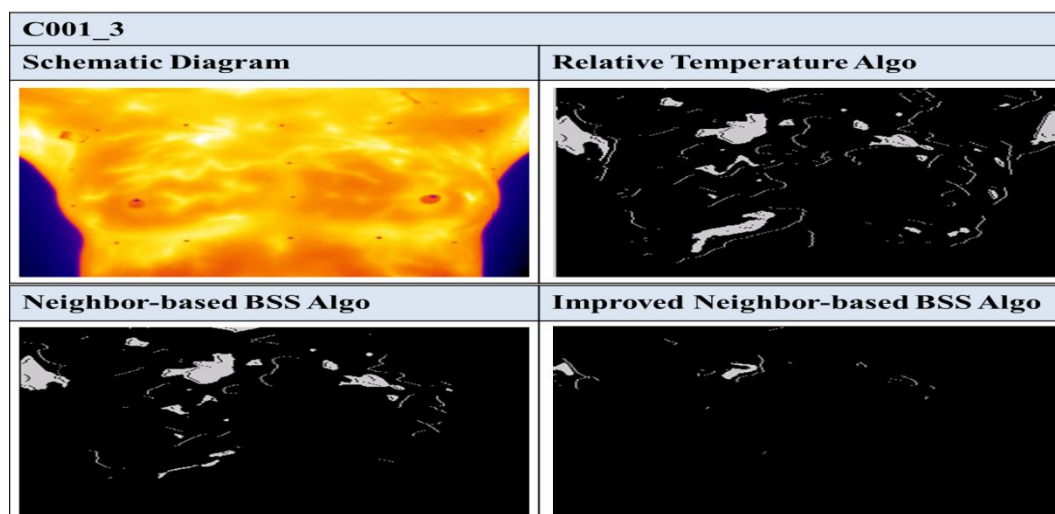


Fig. 4-8. Comparison of Detection Algorithms for C001_3.

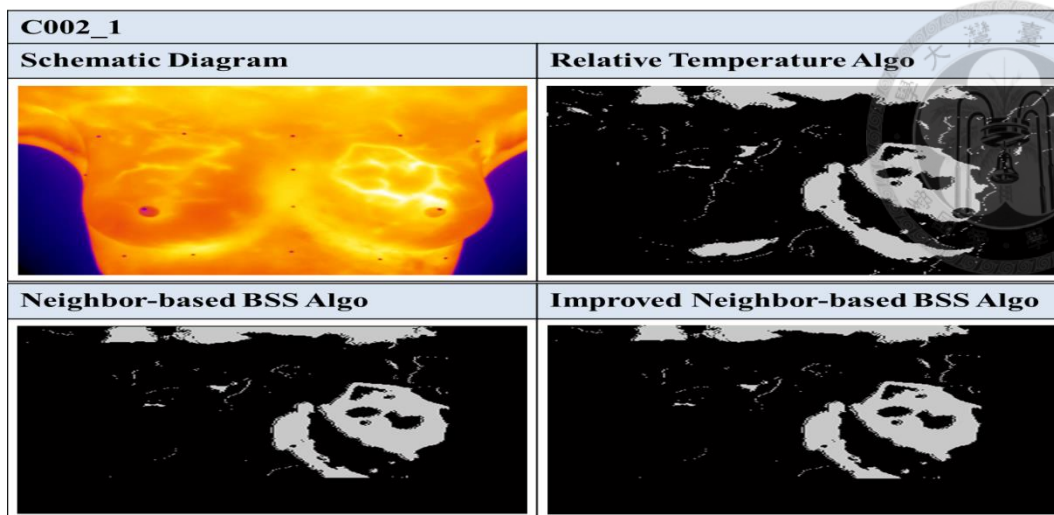


Fig. 4-9. Comparison of Detection Algorithms for C002_1.

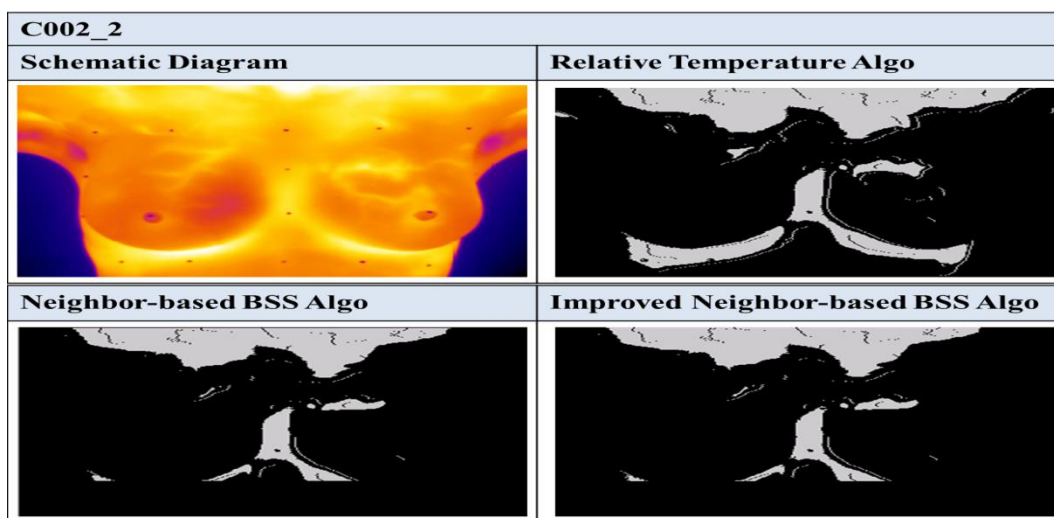


Fig. 4-10. Comparison of Detection Algorithms for C002_2.

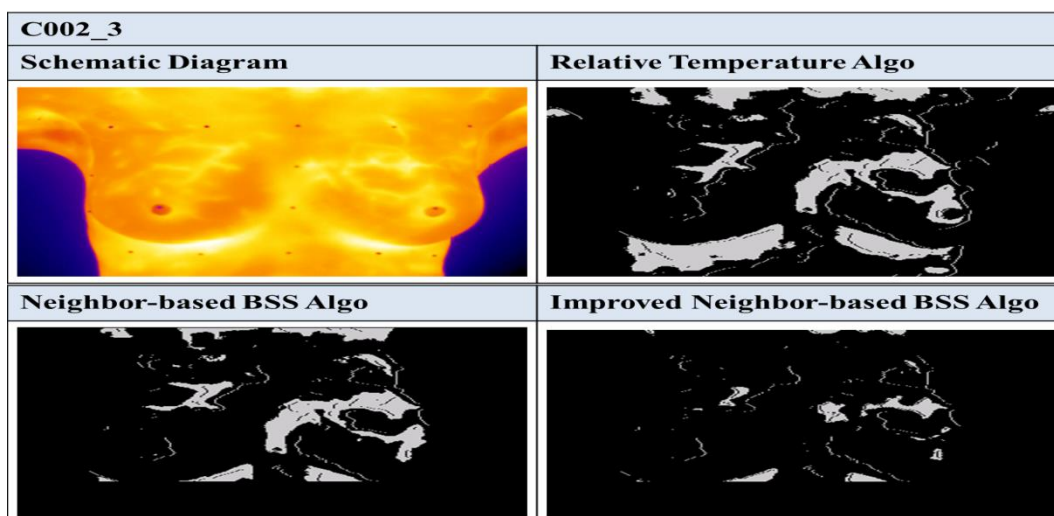


Fig. 4-11. Comparison of Detection Algorithms for C002_3.

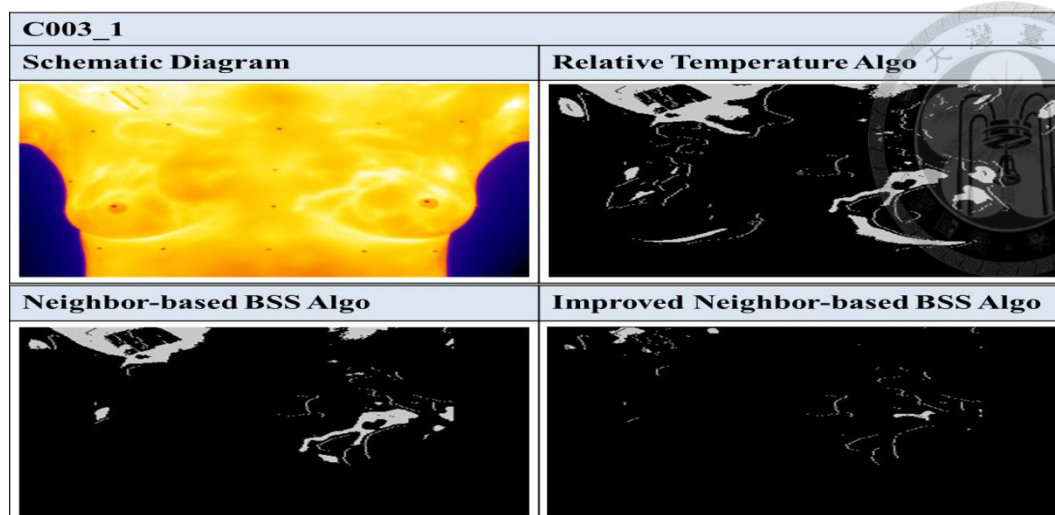


Fig. 4-12. Comparison of Detection Algorithms for C003_1.

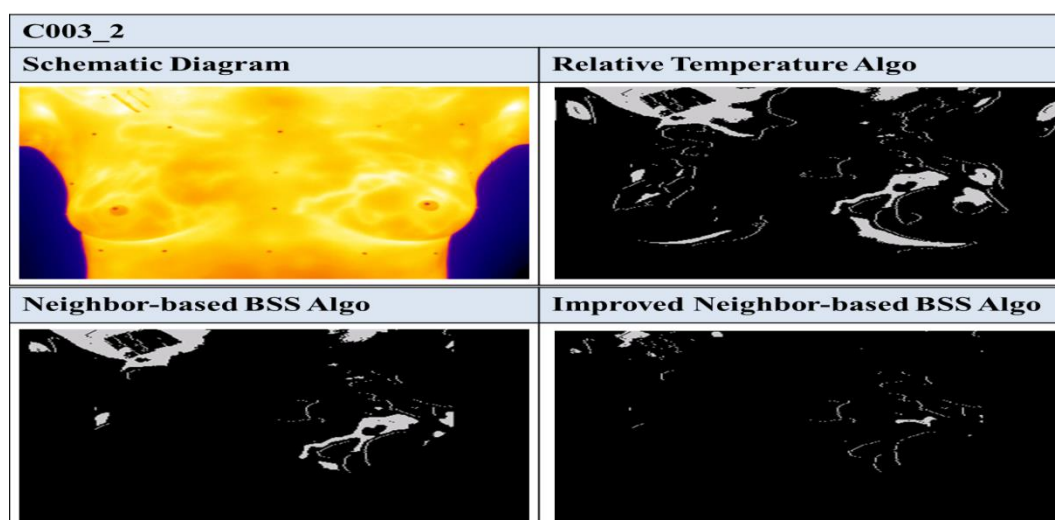


Fig. 4-13. Comparison of Detection Algorithms for C003_2.

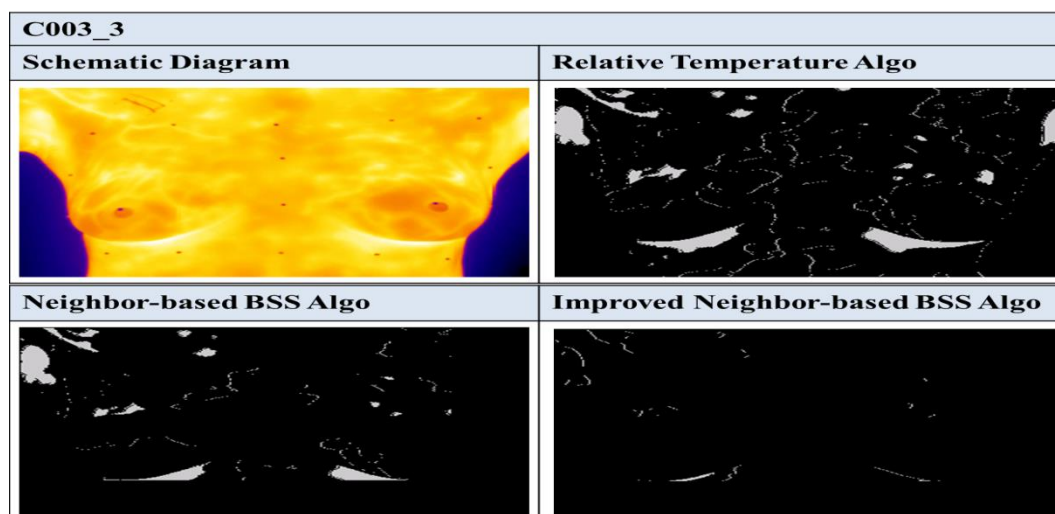
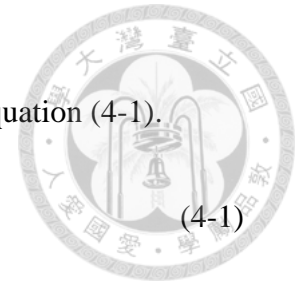


Fig. 4-14. Comparison of Detection Algorithms for C003_3.

4.2.2. Demarcating Degree

For comparison, we use a demarcating degree as defined in Equation (4-1).

$$\text{Demarcating Degree} = \frac{\text{Marked Pixels in areas of tumors}}{\text{All Marked Pixels}} \quad (4-1)$$



To observe the treatment results of these patients, we build statistical charts to record the demarcating degrees of suspicious areas, as shown in Fig. 4-15, Fig. 4-16, and Fig. 4-17. By observing these charts, we can find that the demarcating degree of Improved Neighbor-based Blind Source Separation Algorithm is better than others by its improved suspicious areas and removed noises in all cases. Fig. 4-15 illustrates a situation that the tumor of Patient_C001 neither grows nor shrinks, so we can get stable demarcating degrees. Fig. 4-16 illustrates a situation that the tumor of Patient_C002 is violent. Demarcating degrees decrease after periodic chemotherapy in the middle. However, Demarcating degrees increase again because of the tumor recurrence, which shows the worse effect of long-term chemotherapy. Fig. 4-17 illustrates a situation that the tumor of Patient_C003 shrinks slowly because the effect of long-term chemotherapy is very successful. We can find that demarcating degrees are very low because the tumor of Patient_C003 even disappears in the end. Table 4-2 summarizes the comparison of patients who accepted long-term chemotherapy.

Table 4-2. Comparison of Patients.

Group	Demarcating Degree	Chemotherapy Period	Disease Rehabilitation	Disease Recurrence
C001	Stable	6-Month	Middle	No
C002	Decreasing in the middle Increasing again	6-Month	Low	Yes
C003	Decreasing	6-Month	High	Yes

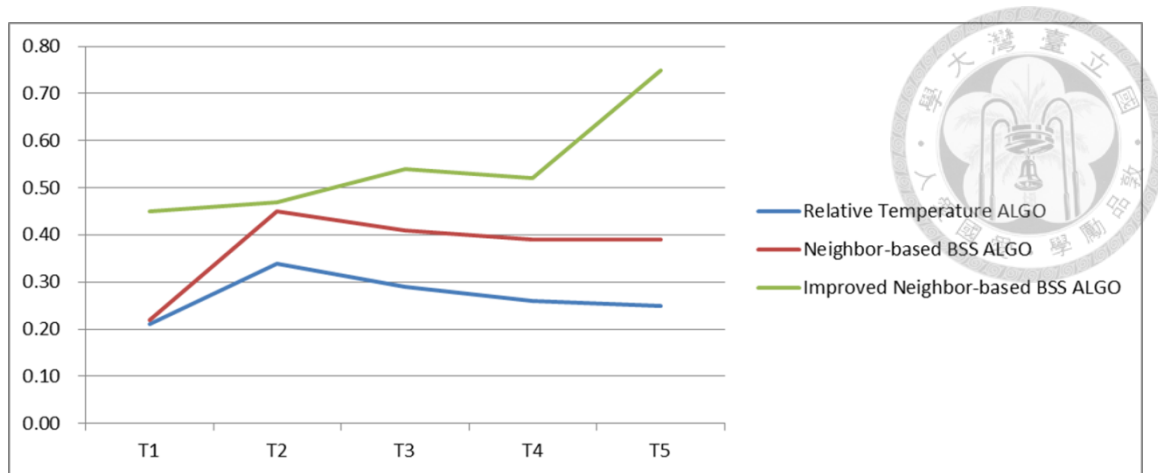


Fig. 4-15. Statistical Chart of C001.

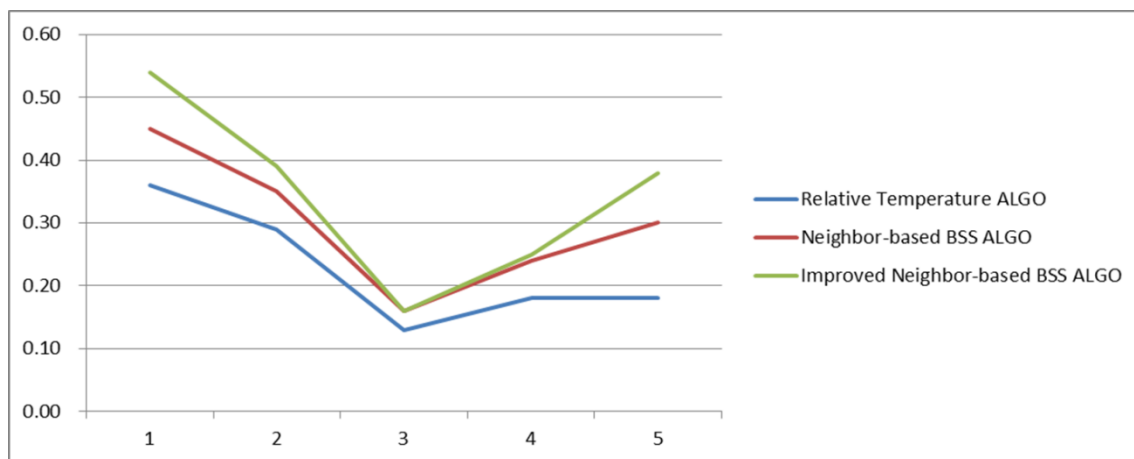


Fig. 4-16. Statistical Chart of C002.

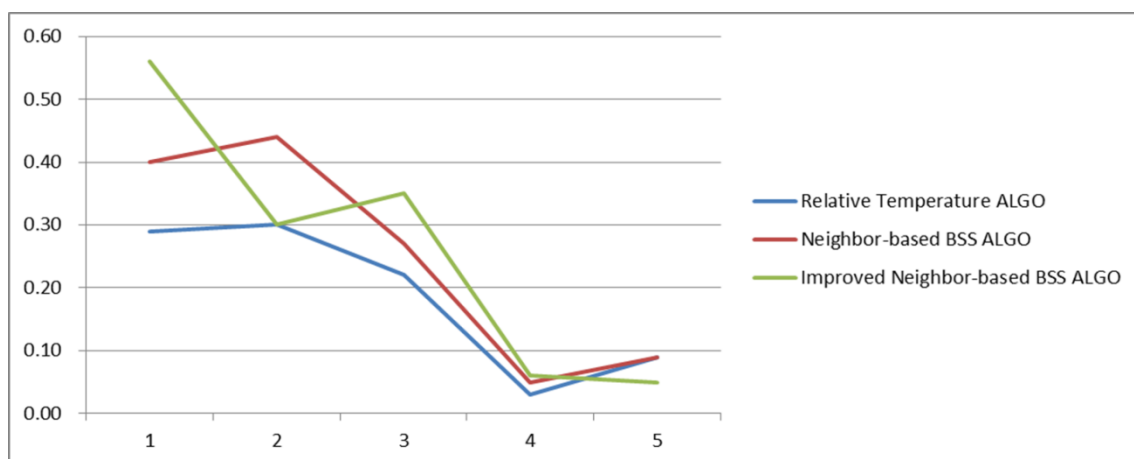
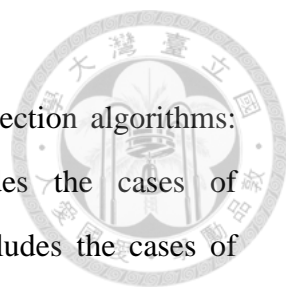


Fig. 4-17. Statistical Chart of C003.

4.2.3. Correctness Rate and Error Rate



We define two other metrics to evaluate the above three detection algorithms: correctness rate and error rate. Correctness situation includes the cases of Truth-Positive (TP) and False-Negative (FN). Error situation includes the cases of Truth-Negative (TN) and False-Positive (FP). Case_TP means that we see abnormal pixels as abnormal pixels. Case_FN means that we see normal pixels as normal pixels. Case_TN means that we see abnormal pixels as normal pixels. Case_FP means that we see normal pixels as abnormal pixels. There are total 81920 pixels in each image. Table 4-3 lists the average times of above 3 cases which are C001, C002 and C003. And the Correctness Rate and Error Rates can then be derived. RT_ALGO means Relative Temperature Algorithm. N_ALGO means Neighbor-based BSS Algorithm. IN_ALGO means Improved Neighbor-based BSS Algorithm. We can find that the Correctness Rate and Error Rate of our Improved Neighbor-based BSS Algorithm are better than other original algorithms. Otherwise, Fig.4-18 shows the distribution of Correctness Rate and Error Rate. We can find that our improved algorithm approximately increases 10% compared with other original algorithms.

Although people usually use regions instead of pixels to demarcate suspicious areas in Clinical Trial, we are convenient to calculate Correctness Rate and Error Rate by using pixels in our algorithm design. Our core goal is confirming that our algorithm design can demarcate suspicious areas by using Dual-band IR Spectrogram in Non-invasive and passive detection. To achieve this core goal, we can build suitable standard by using Demarcating Degree, Correctness Rate and Error Rate.

Table 4-3. Correctness Rate and Error Rate.

		Correctness Cases			Error Cases		
		TP	FN	Rate	TN	FP	Rate
C001	RT_ALGO	278	74344	91.09%	0	7298	8.91%
	N_ALGO	278	78302	95.92%	0	3340	4.08%
	IN_ALGO	278	81284	99.56%	0	358	0.44%
C002	RT_ALGO	2365	65823	83.24%	0	13732	16.76%
	N_ALGO	2365	70578	89.04%	0	8977	10.96%
	IN_ALGO	2365	73313	92.38%	0	6242	7.62%
C003	RT_ALGO	662	74406	91.64%	0	6852	8.36%
	N_ALGO	662	77636	95.58%	0	3622	4.42%
	IN_ALGO	662	80860	99.51%	0	398	0.49%

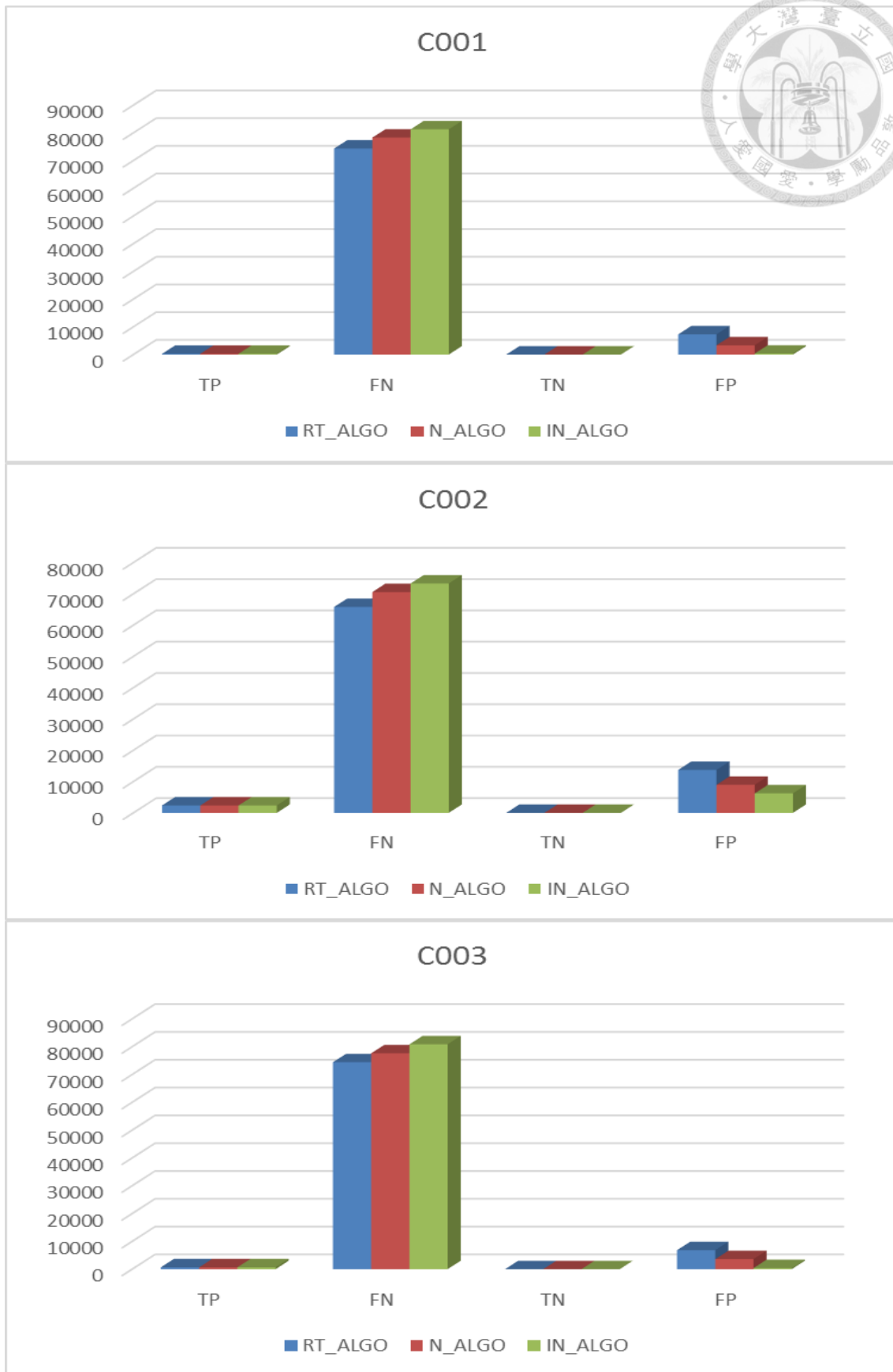


Fig. 4-18. Distribution of Correctness Rate and Error Rate





CHAPTER 5

CONCLUSION

In this work, we applied Blind Source Separation Algorithms to the Dual-band IR Spectrogram to find suspicious area of cancer cells and generated better results in breast cancer detection. In addition to using the basic characteristics of Blind Source Separation Algorithms, we also applied techniques in data analysis to improve the algorithm design.

Experiment results showed that noise reduction and improvement of marked suspicious areas are found using our Improved Neighbor-based BSS Algorithm. For Demarcating Degree, our designed Improved Neighbor-based BSS algorithm is approximately 15% better than other original algorithms. For Correctness Rate, our improved algorithm approximately increases 10% compared with other original algorithms. We will try to apply some other analytic methods to analyze Dual-band IR Spectrogram in the future.





REFERENCE

- [1] Szu, Harold, L. Miao, and H. Qi, "Thermodynamic Free-energy Minimization for Unsupervised Fusion of Dual-color Infrared Breast Images," *Defense and Security Symposium. International Society for Optics and Photonics*, pp.6-11, April 2006.
- [2] H.-Y. Hsieh, "Evaluation and Implementation of Dual-Spectrum IR Spectrogram Diagnostic System on Breast Cancer Detection," Master Thesis, National Taiwan University, July 2008.
- [3] C.-Y. Lee, "Evaluation of Dual-spectrum IR Spectrogram System on Invasive Ductal Carcinoma (IDC) Breast Cancer," *Biomedical Engineering: Applications, Basis and Communications*, pp.427-433, June, 2011.
- [4] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A Blind Source Separation Technique Using Second-order Statistics," *IEEE Transactions on Signal Processing*, vol.45, no.2, pp.434-444, February 1997.
- [5] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2011.
- [6] A.-C. Rencher, *Multivariate Statistical Inference and Applications*, Wiley-Interscience, 1998.
- [7] http://web.ntpu.edu.tw/~ccw/statmath/M_group_1.pdf
- [8] http://web.ntpu.edu.tw/~ccw/statmath/M_group_2.pdf
- [9] http://web.ntpu.edu.tw/~ccw/statmath/M_group_3.pdf
- [10] http://web.ntpu.edu.tw/~ccw/statmath/M_group_4.pdf
- [11] <http://www.fda.gov/RegulatoryInformation/Guidances/ucm126420.htm>
- [12] <https://wssllab.cd.github.io/blog/2013/01/31/how-to-setup-raspberry-pi/>

- [13] W. Herschel, "Experiments on the Refrangibility of the Invisible Rays of the Sun," *Philosophical Transactions of the Royal Society of London*, vol. 90, pp. 284-292, 1800.
- [14] S.-Y. Yin, "Design of a Dual-band Quantum-well IR Spectrogram Readout Circuit," Master Thesis, National Taiwan University, July 2014.
- [15] C.-C. Wang, "A Practical Infrared Image Contrast Enhancement Method," Master Thesis, National Chiao Tung University, June 2004.

