6-24-2015

Communication-theoretic Approach for Skin Cancer Detection using Dynamic Thermal Imaging

Sebastian Godoy

Sebastian Eugenio Godoy

Candidate

Electrical and Computer Engineering

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Majeed M. Hayat , Chairperson

Sanjay Krishna

Vincent Calhoun

R. Steven Padilla

Communication-theoretic Approach for Skin Cancer Detection using Dynamic Thermal Imaging

by

Sebastián Eugenio Godoy

B.S., Engineering, University of Concepción, Chile, 2006
M.S., Engineering, University of Concepción, Chile, 2009

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Engineering

The University of New Mexico

Albuquerque, New Mexico

May, 2015

Dedication

To my greatest creation ever, Cristóbal.

Acknowledgments

First of all, thank you Cristóbal for all the times you saw me tired and told me: "Never give up, daddy!" Thank you for the hugs, kisses and your infinite energy that motivated me through the years. Thank you my beloved wife, Elizabeth, for joining me on this huge adventure and for your endless support through the years.

Thanks also to my parents and sister in Chile and to my sister in Norway for their endless support. The distance was not an issue for you to emotionally support me through the years.

I would like to thank my advisor, Prof. Majeed M. Hayat and my co-advisor, Prof. Sanjay Krishna for their friendly and professional guidance. Your encouragement and kindness taught me more than how to do science. I really feel honored to have had the opportunity of work under your guidance. Without your constant encouragement it would not have been possible to successfully complete my Ph.D.. Thank you also to my committee members, Prof. Vincent Calhoun and Dr. R. Steven Padilla for their encouragement and kindness.

Special thanks to Dr. Biliana Paskaleva from Sandia National Labs and Dr. Woo-Yong Jang from AFRL for their infinite help at the first stages of my doctorate studies. Also special thanks to Dr. David Ramirez and Dr. Stephen Myers for their help in collecting patient data, performing image registration and managing the biopsy database.

I would like to also thank my colleagues and friends, David, Stephen, Mahshid, Javad, Biliana, Eric and Greg for all your help, advise and friendship. The coffee breaks and the talks about life helped me in a way you cannot imagine. I also thank all other group members from ECE and CHTM research groups and the Skinfrared, LLC staff. I will remember all the good memories we had together.

Finally, I would like to thank The University of Concepción, Chile (UdeC), the Comisión Nacional de Ciencia y Tecnología (CONICYT) and the Cancer Nanoscience and Microsystems Training Center (CNTC) for their finantial support. This work was partially supported by the Department of Energy NNSA (NA-22) award number DE-NA0002494.

Communication-theoretic Approach for Skin Cancer Detection using Dynamic Thermal Imaging

by

Sebastián Eugenio Godoy

B.S., Engineering, University of Concepción, Chile, 2006
M.S., Engineering, University of Concepción, Chile, 2009

Ph.D., Engineering, University of New Mexico, 2015

Abstract

Skin cancer is the most common cancer in the United States with over 3.5M annual cases. Statistics from the Americans Cancer Society indicate that 20% of the American population will develop this disease during their lifetime. Presently, visual inspection by a dermatologist has good sensitivity (>90%) but poor specificity (<10%), especially for melanoma conditions, which is the most dangerous type of skin cancer with a five-year survival rate between 16–62%.

Over the past few decades, several studies have evaluated the use of infrared imaging to diagnose skin cancer. Here we use dynamic thermal imaging (DTI) to demonstrate a rapid, accurate and non-invasive imaging and processing technique to diagnose melanoma and non-melanoma skin cancer lesions. In DTI, the suspicious lesion is cooled down and the thermal recovery of the skin is monitored with an infrared camera. The proposed algorithm exploits the intrinsic order present in the time evolution of the thermal recoveries of the skin of human subjects to diagnose

the malignancy and it achieves outstanding performance for discriminating between benign and malignant skin lesions.

In this dissertation we propose a stochastic parametric representation of the thermal recovery curve, which is extracted from a heat equation. The statistics of the random parameters associated with the proposed stochastic model are estimated from measured thermal recovery curves of subjects with known condition. The stochastic model is, in turn, utilized to derive an analytical autocorrelation function (ACF) of the stochastic recovery curves. The analytical ACF is utilized in the context of continuous-time detection theory in order to define an optimal statistical decision rule such that the sensitivity of the algorithm is guaranteed to be at a maximum for every prescribed false-alarm probability. The proposed algorithm was tested in a pilot study including 140 human subjects and we have demonstrated sensitivity in excess of 99% for a prescribed false-alarm probability of 1% (specificity in excess of 99%) for detection of skin cancer. To the best of our knowledge, this is the highest reported accuracy for any non-invasive skin cancer diagnosis method.

The proposed algorithm is studied in details for different patient permutations demonstrating robustness in maximizing the probability of detecting those subjects with malignant condition. Moreover, the proposed method is further generalized to include thermal recovery curves of the tissue that surrounds the suspicious lesion as a local reference. Such a local reference permits the compensation of any possible anomalous behavior in the lesion thermal recovery, which, in turn, improves both the theoretical and empirical performance of the method.

As a final contribution, we develop a novel edge-detection algorithm—specifically targeted for multispectral (MS) and hyperspectral (HS) imagery—which performs l edge detection based solely on spectral (color) information. More precisely, this algorithm fuses the process of detecting edges through ratios of pixels with critical information resulting from spectral classification of the very image whose edges are

to be identified. This algorithm is tested in multicolor (spectral) imagery achieving superior results as compared with other alternatives. The edge-detection algorithm is subsequently utilized in the skin-cancer detection context to define the lesion boundary from a visible color image by exploiting the color contrast between the pigmented tissue and the surrounding skin. With this automated lesion selection, we develop a method to extract spatial features equivalent to those utilized by the dermatologists in diagnosing malignant conditions. These spatial features are fused with the temporal features, obtained from the thermal-recovery method, to yield a spatio-temporal method for skin-cancer detection.

While providing a rigorous mathematical foundation for the viability of the dynamic thermal recovery approach for skin-cancer detection, the research completed in this dissertation also provides the first reliable, accurate and non-invasive diagnosis method for preliminary skin-cancer detection. This dissertation, therefore, paves the way for future clinical studies to produce new skin-cancer diagnosis practices that minimize the need for unnecessary biopsies without sacrificing reliability.

Contents

List of Figures

List of Tables

Glossary

AHI	Airborne Hyperspectral Imager
BCC	Basal-cell carcinoma
DP	Detection probability
DTI	Dynamic thermal imaging
DWELL	Quantum dot-in-a-well
FAP	False-alarm probability
FPA	Focal-plane array
HS	Hyper-spectral
IR	Infrared
KL	Karhunen-Loève
MM	Malignant melanoma
MS	Multispectral
RKHS	Reproducing-kernel Hilbert spaces
SCC	Squamous-cell carcinoma

TRC	Thermal recovery curve

Chapter 1

Introduction

1.1 Motivation

Skin cancer is the most common form of cancer in the United States with over 3.5 million cases of skin cancer reported annually [1]. There is a higher incidence of skin cancer than the combined occurrence of breast, prostate, lung and colon cancers [2]. Melanoma, which accounts for an estimated 4% of skin cancer cases, is responsible for approximately 75% of all deaths from skin cancer. The total deaths in the United States due to melanomas and other types of skin cancer are estimated to be more than 12,000 for 2014 [1]. Currently, the detection of melanoma relies on a subjective ABCDE (*Asymmetry*, *Border*, *Color*, *Diameter* and *Evolution*, Fig. 1.1) test performed visually by dermatologists, general practitioners (GP) or primary care physicians (PCP) [3]. However, the ABCDE test provides a qualitative guideline and it requires a trained specialist to actually distinguish malignant lesions from benign nevi. Moreover, the ABCDE approach has a relatively high false-alarm probability (0.35–0.44, i.e., a specificity in the range 56% to 65%) and moderate detection probability (0.47–0.89) [3, 4, 5]. Since a false negative (i.e., a patient with malig-

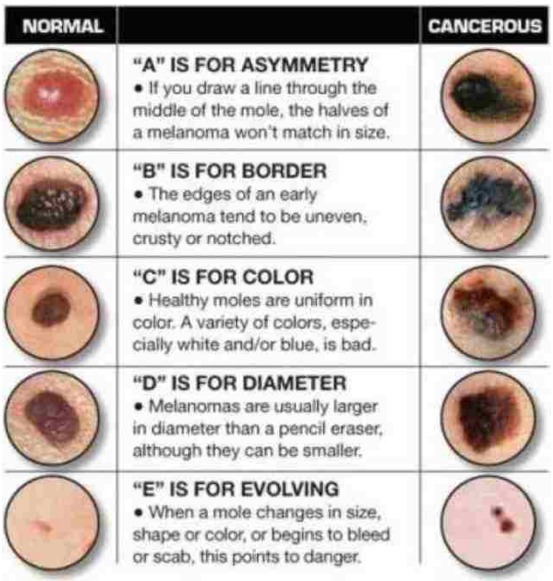


Figure 1.1: The ABCDE visual test performed as a first screening mechanism for the detection of melanoma lesions.

nant condition that is declared to have benign condition) could lead to metastasis (spreading to other parts of the body) and death, excisional biopsies are routinely performed even on lesions that are non-cancerous. It was estimated that the number of biopsies undertaken in nine geographical areas of the US between 1986 and 2001 is close to 60 for every melanoma detected [6].

As in any type of cancer, if malignant lesions can be diagnosed and excised early in their evolution, patients have a survival rate of almost 100%. Once the cancer advances into the deeper layers of skin, the risk of metastasis increases and cancer cells grow promptly into the subcutaneous layers, invading lymphatic and blood-vessels resulting in a serious and possibly lethal clinical problems [1]. Early diagnosis is therefore essential to increase the survival rate of skin cancer patients [7]. One of the critical barriers in early skin-cancer detection is the lack of a non-invasive techniques that can detect the cancer at an early stage with high detection probability (i.e., the probability of correctly detecting a malignant lesion) and low false-alarm probability

(i.e., the probability of declaring a benign lesion as malignant). In this dissertation, we aim to develop a fast, accurate and noninvasive technique that by means of infrared imaging and optimal decision theory can achieve a maximum detection probability for a constrained false-alarm probability.

1.2 Prior art

1.2.1 Non-invasive techniques for skin cancer detection

Since biopsies are intrusive and can be painful, different non-invasive techniques are being researched in order to minimize the number of excess biopsies performed [8, 9]. Some of these techniques include multispectral (MS) imaging [10, 11, 12], digital dermatoscopy and videodermatoscopy (sequential digital dermatoscopy) [13, 14], reflectance-mode confocal microscopy [15], ultrasound [16, 17], laser Doppler perfusion imaging [18], and optical coherence tomography (OCT) [19, 20], to name a few. Approximately 25% of dermatologists use a dermatoscope, which consists of a magnifier, a light source, a transparent plate, and a liquid medium. The dermatoscope minimizes the light reflection of external sources, such that the suspected lesions can be inspected without the obstruction normally caused by the normal reflection of light on the human skin [21]. Its adoption has been limited due to several factors, including the extensive training requirements and a steep learning curve for marginal improvements in sensitivity and specificity [5]. MelaFind™, a non-invasive device approved by the Food and Drug Administration (FDA), employs MS imaging in the visible and near-infrared range as well as computer vision algorithms to evaluate clinically atypical pigmented skin lesions and it classifies them based upon their level of 3D morphological disorganization [22]. The product presents a high-level of detection probability (>95%) [23], but high false-alarm probability

(>90%) [24]. MoleMate is a hand-held device that detects melanoma based on images of the epidermal and dermal melanin and vasculature, and the collagen content of the lesion [25]. A recent, large clinical trial found no evidence that MoleMate improved appropriateness of referral by PCPs [26]. Aura is a Raman spectroscopy system that scans suspicious lesions and produces a risk score related to the probability that the lesion is malignant, but the device is very expensive and has high false-alarm probability [27]. Vivosight Multi-Beam System™, another FDA-approved device, utilizes OCT to achieve a detection probability between 0.79–0.94 and a false-alarm probability between 0.04–0.15 for non-melanoma skin cancer lesions [28]. The problem they present is that the suspicious lesion must be probed several times to achieve such an accuracy, which makes the acquisition time prohibitively high.

1.2.2 The role of infrared imaging in the detection of skin cancer

In order to address the aforementioned problems, different groups have investigated the utilization of infrared (IR) thermal imaging to detect skin cancer over the past few decades [29, 30]. The reason that makes infrared imaging attractive is the fact that it constitutes a functional and non-invasive imaging method, providing information on the normal and abnormal physiologic response of the nervous and vascular systems, as well as the local metabolic rate and inflammatory processes that ultimately appear as differences in the skin temperature [31]. The first attempts of utilizing IR imaging for the detection of skin cancer were prevalently static, with point-and-shoot imaging and subjective IR image evaluation leading the experts to believe that IR imaging was useless for medical applications [32].

The attention later was turned to the dynamic IR imaging (DTI) approach where the tissue of a suspicious lesion is cooled (or warmed) in order to observe the thermal

recovery of the tissue with an IR camera. DTI offers advantages in clinical applications, for which the duration of the measurement and the ease of use are critical, as it is much less dependent on ambient temperatures and conditions [30]. The DTI technique was pioneered by Anbar [33], who described how changes in human skin temperature convey valuable physiological and pathophysiological information that can lead to a diagnosis. Buzug *et al.* [34] studied the thermal recovery curve (TRC) of basal-cell carcinomas (BCC) lesions after applying an external stimulus to cool down the lesion and the surrounding tissue. Their work established that the TRC of a BCC lesion differs from the TRC of the surrounding tissue, which is assumed to be benign. Similarly, Centigul and Herman [35, 36, 37] studied the TRC of malignant melanoma (MM) lesions, and they observed the same phenomena, i.e., that the TRC of a benign lesions differs from the TRC of malignant lesions, and such a difference contain useful information that has the potential to non-invasively differentiate benign from malignant lesions.

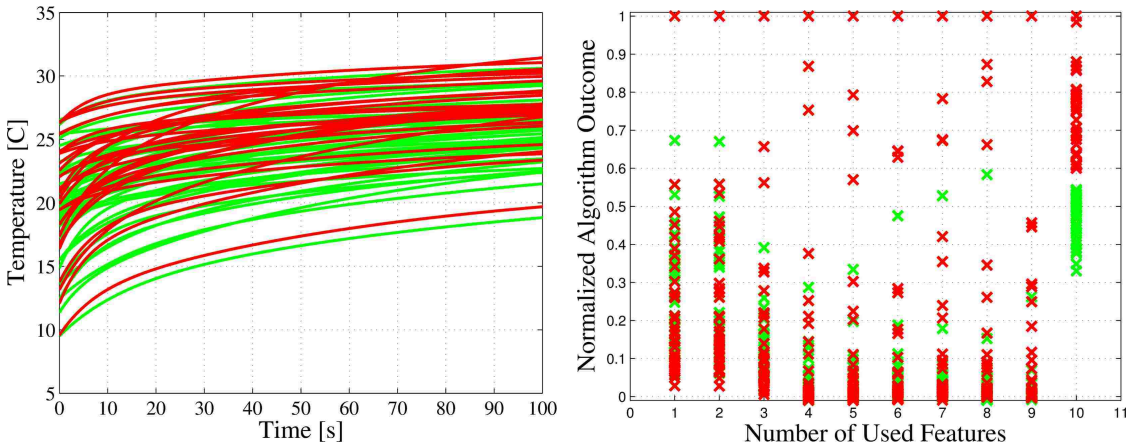
Nevertheless, these methods only partially extracted the information present in the temporal evolution of the recovery process. More specifically, the existing DTI techniques have neglected the temporal statistical features inherent in the thermal recovery process. To fully extract the vital information present in the recovery process, which will enable us to make a reliable inference on the malignancy of lesion, two problems must be solved. First, the recovery process must be viewed as a random function of time, or a stochastic process, and its temporal statistical properties, such as its temporal correlations, must be mathematically characterized. Second, such complete statistical understanding of the thermal recovery process must, in turn, be utilized in a statistical-inference framework that yields the optimal decision rule for classifying a lesion as malignant or benign. Both of these problems are formulated and solved rigorously in this dissertation.

1.3 Contributions of this dissertation

The main thrust of this dissertation is to develop and test an optimal hypothesis-testing approach to classify skin cancer lesions based on the TRC of the lesions obtained through DTI.

The first contribution of this dissertation is the development of a physics-based stochastic mathematical model that describe the time evolution of the TRCs. Modeling the local temperature of the tissue by a heat equation and defining the appropriate boundary conditions, a mathematical structure of the TRCs was obtained. The lack of precise knowledge of skin parameters is addressed by considering the unknown parameters as random variables. The distribution and correlation of these parameters is estimated and utilized in the algorithm development to define an analytical autocorrelation function that describes the stochastic model.

The second contribution is the proposed solution of the signal-against-signal hypothesis-testing problem by means of Grenander's approach [38]. This approach is studied in detail using different scenarios to train the hypothesis-tester. Through our method, the continuous-time TRCs are mapped onto a feature space, where each feature is the projection of the TRC onto the eigenfunctions of the autocorrelation function associated with each hypothesis. These projections, known as the Karhunen-Loève coefficients, are combined into a single number, termed the test-statistics, that can be compared to a threshold to declare the malignancy of the lesion. Example of continuous-time TRCs are depicted in Fig. 1.2(a), where the green and red curves represent TRCs of patients with benign and malignant condition, respectively. While these curves suggest some modest level of discernibility between the two conditions, no clear separation can be visually established. The test-statistics, which compactly contains the temporal statistical information present in these curves at different desired levels of granularity, is represented in Fig. 1.2(b).



(a) (b)

Figure 1.2: Input and output examples for the proposed algorithm for skin cancer detection: (a) Continuous-time TRCs, which are not clearly separable and (b) normalized test-statistics, which become clearly separable when ten features are used in the computation of the test-statistics.

For each patient with unknown condition, the corresponding test-statistics is computed, and the value of such test-statistic is then compared to an optimally selected threshold. The result of the comparison indicates to which hypothesis the patient's TRC belongs to. Our method was tested in a pilot study including 140 human subjects and we have demonstrated sensitivity in excess of 99% for a prescribed false-alarm probability below 1% for detection of skin cancer. To the best of our knowledge, this is the highest reported accuracy for any non-invasive device for detection of skin cancer. We present the theoretical receiver operating characteristic (ROC) curve for the proposed method in Fig. 1.3. The ROC curve, is a graphical plot that illustrates the performance of our method as its discrimination threshold is varied. It can be noted that as we increase the level of granularity in the computation of the test-statistics the theoretical performance of the detector is increased

The third contribution of this dissertation is the study of the hypothesis-testing problem by utilizing a multiple-signal generalization of the first method: It includes

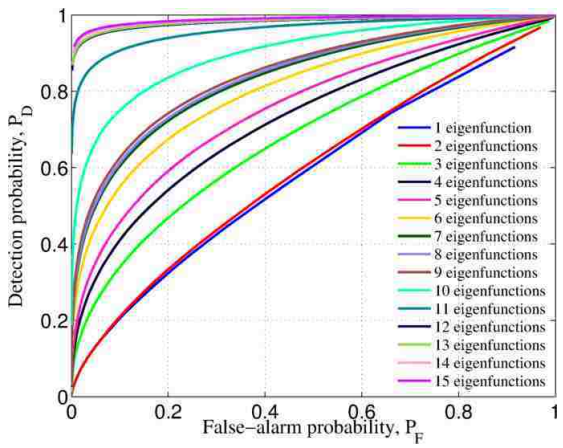


Figure 1.3: Theoretical receiver operating characteristic (ROC) curve for the proposed method as the number of eigenfunctions is increased in computing the test-statistics.

one TRC signal from inside the lesion and another from the surrounding tissue. This approach allows the detection of abnormal features in the TRCs by self-referencing, that is, comparing the responses of suspected tissue (lesion TRC) and reference tissue (surrounding TRC) across all the patient dataset. It was observed that the inclusion of reference TRC helped the hypothesis tester to achieve the same performance of the previous approach requiring less features extracted from the TRCs.

The fourth contribution of this dissertation is a novel method to fuse material/tissue classification with an edge detection algorithm in order to improve the performance of the later. More precisely, we extend the Spectral Ratio Contrast (SRC) Algorithm (presented by our group) [42, 43] by utilizing spectral classification to further enhance the detection of edges that are solely due to material (not intensity) changes. We term this extension the Adaptive Spectral Ratio Contrast (ASRC) edge detection algorithm since it adaptively changes the SRC algorithm sensitivity to edges (at each pixel) by considering the material-classification results of the neighboring pixels. The SRC algorithm detects edges that arise from both intensity and spectral changes while the ASRC algorithm detects edges based on spectral (color)

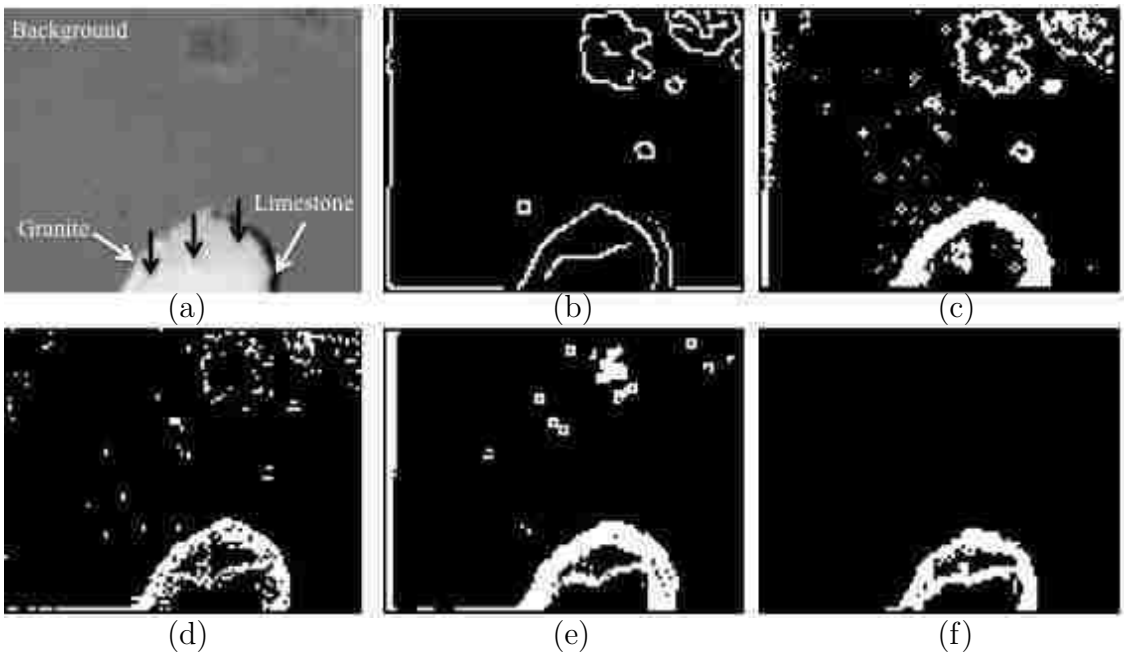


Figure 1.4: Comparison between the edge maps obtained by the ASRC edge detector and four benchmark edge detectors. (a) DWELL FPA image comprising background, and granite and limestone rocks. The (invisible) isoluminant edge exists between granite and limestone rocks and it is marked by the tip of the black arrows; (b) Canny edge detector [39] applied over the same band of the image presented in (a); (c) Multicolor gradient (MCG) [40] edge map; (d) HySPADE edge map [41]; (e) the SRC edge map; and (f) the ASRC edge map.

changes only. A comparison between four different multicolor edge detectors and the ASRC algorithm is depicted in Fig. 1.4. In these experiments, we have utilized a 320×256 quantum dots-in-a-well (DWELL) focal-plane array (FPA). The DWELL sensor was designed and fabricated at the Center for High Technology Materials at the University of New Mexico [44, 45].

The fifth and final contribution of this dissertation is the utilization of the ASRC algorithm to perform automatic boundary identification of the suspected lesions by exploiting the color contrast between the mole and the surrounding tissue. Two example of the automatic boundary detection are shown in Fig. 1.5. After defining

Figure 1.5: Examples of two lesion boundaries automatically detected by means of the ASRC algorithm. These boundaries are used to define spatial properties of the lesions in a similar way the dermatologist perform the ABCDE test.

the boundary, we define and extract spatial properties of the lesions in a similar fashion as the dermatologist do it using the ABCDE test. Moreover, we propose and test a methodology to fuse these spatial features (Asymmetry, Border irregularity, Color change and Diameter) with the TRC (time Evolution) in a single hypothesis-testing problem.

In summary, while providing a rigorous mathematical foundation for the viability of the dynamic thermal recovery approach for skin-cancer detection, the research completed in this dissertation also provides the first reliable, accurate and non-invasive diagnosis method for preliminary skin-cancer detection. This dissertation, therefore, paves the way for future clinical studies to produce new skin-cancer diagnosis practices that minimize the need for unnecessary biopsies without sacrificing reliability.

1.4 Organization of the dissertation

This dissertation is organized as follows. In Chapter 2 we present the skin-cancer detection algorithm formulation and the main assumptions. The performance and robustness of the method is detailed in Chapter 3. In Chapter 4 we generalize the algorithm to a multiple-signal setting, which has the added feature of self (within-

patient)-referencing. In Chapter 5 we introduce the ASRC edge-detection algorithm and analyze its performance using spectral data. Later we utilize the ASRC to automatically detect the lesion boundaries. The automatic lesion selection is, in turn, used to define spatial features of the patient lesions. The generalized spatio-temporal approach is presented at the end of Chapter 5. Finally, in Chapter 6. we summarize this dissertation and propose future reseasrch directions.

1.5 Products of the dissertation

A comprehensive list of publications and patent applications generated during the development of this dissertation is presented next.

1.5.1 Patent applications

1. Majeed M. Hayat, Sanjay Krishna, **Sebastián E. Godoy** and David Ramirez, "Method for Skin Cancer Detection Based on the Application of Statistical Decision Theory to Dynamic Infrared Image Sequences," provisional patent application, October 20, 2014.

2. Majeed M. Hayat, Sanjay Krishna and **Sebastián E. Godoy**, *"Method to Fuse Material Classification with Spatio-Spectral Edge Detection in Spectral Imagery,"* utility patent application, July 17, 2014.

3. Sanjay Krishna, Pavan Mutil, **Sebastián E. Godoy** and Julia Hautmann, *"Methods for Accurate and Simultaneous Measurement of Multiple Properties of Spray Plume,"* utility patent application, November 8, 2014

4. **Sebastián E. Godoy** and Sanchita Krishna, *"Dynamic Infrared Imaging for Skin Cancer Screening and Diagnosis,"* provisional patent application, July 1,

2014

5. Sanjay Krishna, Sanchita Krishna, Majeed M. Hayat, Pradeep Sen, Maziar Yaesoubi, **Sebastián E. Godoy** and Ajit V. Barve, *"A Method and System for Feature Extraction and Decision Making from Series of Images,"* utility patent application, March 14, 2013

1.5.2 Journal manuscripts

1. **Sebastián E. Godoy**, Majeed M. Hayat, David A. Ramirez, Stephen A. Myers, R. Steven Padilla and Sanjay Krishna, "Dynamic Thermal Imaging Accurately and Non-Invasively Detects Skin Cancer," to be submitted.

2. **Sebastián E. Godoy**, David A. Ramirez, Stephen A. Myers, Greg v. Winckel, Sanchita Krishna, Marianne Berwick, R. Steven Padilla, Pradeep Sen and Sanjay Krishna, "Dynamic Infrared Imaging for Skin Cancer Screening," Infrared Physics and Technology, in press, September 2014

3. Zhaobing Tian, **Sebastián E. Godoy**, H. S. Kim, Ted Schuler-Sandy, John A. Montoya and Sanjay Krishna, "High operating temperature interband cascade focal plane arrays," Appl. Phys. Lett. vol. 105, no. 5, pp. 051109, 051109-5, Aug 2014

4. Stephen Myers, **Sebastián E. Godoy**, David Ramirez, Greg Von Winckel, Sanchita Krishna, Marianne Berwick, R. Steven Padilla and Sanjay Krishna, "Dynamic thermal infrared imaging non-invasively detects skin cancer with high sensitivity and specificity," Journal of Investigative Dermatology, vol. 134, p. 102, Nature Publishing Group, May 1, 2014

5. Biliana Paskaleva, **Sebastián E. Godoy**, Woo-Yong Jang, Steven C. Bender, Sanjay Krishna and Majeed M. Hayat, "Model-based Edge Detector for

Spectral Imagery using Sparse Spatio-spectral Masks," IEEE Transactions on Image Processing, vol. 23, no. 5, pp. 2315–2327, May 2014

6. Cesar San Martin, Carlos Deocares, **Sebastián E. Godoy**, Pablo Meza and Daniella Bonilla, "Wavelet-FFT Filter Applied to Non Uniformity Correction in Infrared Imaging System," Lecture Notes in Computer Science, Volume 7441/2012, pp. 355–363, 2012

7. Ajit V. Barve, Saumya Sengupta, Jun Oh Kim, John Montoya, Brianna Klein, Mohammed A. Shirazi, Marziyeh Zamiri, Yagya D. Sharma, Sourav Adhikary, **Sebastián E. Godoy**, Woo-Yong Jang, Glauco R. C. Fiorante and Sanjay Krishna, "Barrier Selection Rules for Quantum Dots-in-a-Well Infrared Photodetector," IEEE Journal of Quantum Electronics, Vol. 48, No. 10, 2012

8. Woo-Yong Jang, Majeed M. Hayat, **Sebastián E. Godoy**, Payman Zarkesh-Ha, Steven C. Bender and Sanjay Krishna, "Data-compressive paradigm for multispectral sensing using tunable DWELL mid-infrared detectors," Optics Express, Vol. 19, Issue 20, pp. 19454–19472, 2011

1.5.3 Conference manuscripts

1. R. M. Clark, B. Coffman, D. A. Ramirez, **Sebastián E. Godoy**, S. A. Myers, T. McGregor, S. Krishna, P. G. McGuire and T. R. Howdieshell, "Accelerated Myocutaneous Revascularization Following Graded-Ischemia in db/db Mice," Submitted to the 10th Annual Academic Surgical Congress, to be held February 3 to 5, 2015, Encore, Las Vegas, NV

2. Zhaobing Tiang, Ted Schuler-Sandy, **Sebastián E. Godoy**, C. Kadlec, H. S. Kim and Sanjay Krishna, "High Operating Temperature InAs/GaSb Type-II Superlattices," 18th International Conference on Molecular Beam Epitaxy (MBE), September 7–12, 2014, Flagstaff, AZ

3. Javad Ghasemi, Payman Zarkesh-Ha, Sanjay Krishna, **Sebastián E. Godoy**, and Majeed M. Hayat, "A Novel Readout Circuit for On-sensor Multispectral Classification," IEEE 57th International Midwest Symposium on Circuits and Systems (MWSCAS 2014), August 3–6, 2014, College Station, TX

4. **Sebastián E. Godoy**, David Ramirez, Stephen Myers, Greg von Winckle, Sanchita Krishna and Sanjay Krishna, "Dynamic Infrared Imaging for Skin Cancer," Quantum Structured Infrared Photodetector International Conference, QSIP 2014 (June 29 – July 3, 2014)

5. Zhaobing Tiang, Ted Schuler-Sandy, **Sebastián E. Godoy**, C. Kadlec, H. S. Kim and Sanjay Krishna, "Interband Cascade Infrared Photodetectors and their Focal Plane Arrays," Quantum Structured Infrared Photodetector International Conference, QSIP 2014 (June 29 - July 3, 2014)

6. Zhaobing Tiang, **Sebastián E. Godoy**, H. S. Kim, Ted Schuler-Sandy, J. Montoya and Sanjay Krishna, "Mid-wave infrared interband cascade photodetector and focal plane arrays," Proc. SPIE 9070, Infrared Technology and Applications XL, 90701K (June 26, 2014)

7. Zhaobing Tiang, Ted Schuler-Sandy, **Sebastián E. Godoy**, H. S. Kim, J. Montoya and Sanjay Krishna, "Mid-wave infrared interband cascade photodetector and focal plane arrays," IEEE Photonics Conference (IPC), pp. 598–599, 8–12 September 2013, Bellevue, WA

8. Julia Hautmann, **Sebastián E. Godoy**, Patricia Marshik, Ramesh Chand, Jason McConville, Sanjay Krishna, Sanchita Krishna, and Pavan Muttil, "Metered Dose Inhalers: Varying the time between multiple actuations could influence the emitted dose," 2013 American College of Clinical Pharmacy (ACCP) Annual Meeting, Albuquerque, NM. October 13-16, 2013

9. Zhaobing Tiang, Ted Schuler-Sandy, **Sebastián E. Godoy**, H. S. Kim and Sanjay Krishna, "High-operating-temperature MWIR detectors using type II superlattices," Proc. SPIE 8867, Infrared Remote Sensing and Instrumentation XXI, 88670S (September 19, 2013)

10. Z. Tiang, T. Schuler-Sandy, **S. E. Godoy**, H. S. Kim, J. Montoya, S. Myers, B. Klein, E. Plis and S. Krishna, "Quantum-engineered mid-infrared type-II InAs/GaSb superlattice photodetectors for high temperature operations," Proc. SPIE 8704, Infrared Technology and Applications XXXIX, 87041T (June 18, 2013)

11. Julia Hautmann, **Sebastián E. Godoy**, P. Marshik, R. Chand, J. McConville, Sanchita Krishna, Sanjay Krishna, and Pavan Muttil, "Effect of Time Between Actuation on the Dose Variability for Three Metered Dose Inhalers," RDD Europe (2013), vol. 2, pp.429-434

12. **Sebastián E. Godoy**, Majeed M. Hayat, Woo-Yong Jang and Sanjay Krishna, "Classifier-enhanced algorithm for compressive spatio-spectral edge detection," IEEE Photonics Conference (IPC), 23-27 September 2012, San Francisco, CA

13. Woo-Yong Jang, Majeed Hayat, **Sebastián E. Godoy** and Payman Zarkesh-Ha, "Compressive Multispectral Sensing Algorithm with Tunable Quantum Dots-in-a-Well Infrared Photodetectors," IEEE Photonics Conference (IPC), p. 147-148, 9-13 October 2011

Chapter 2

Skin-cancer detection using dynamic infrared imaging

2.1 Introduction

Here we report a method for statistical inference, which uses the technique of dynamic thermal imaging (DTI) and it demonstrates a rapid, accurate and non-invasive imaging system for detection of skin cancer. DTI is a technique in which a thermal stimulus is applied to the suspected lesion and the thermal recovery is captured as function of time using an infrared (IR) camera [33]. Even though several groups have reported that the thermal recovery of a skin-cancer lesion and the surrounding healthy skin is different [34, 36, 46, 30], these methods only partially extracted the information present in the temporal evolution of the recovery process. More specifically, the existing DTI techniques have neglected the temporal statistical features inherent in the thermal recovery process. To fully extract the vital information present in the recovery process, which will enable us to make a reliable inference on the malignancy of lesion, two problems must be solved. Firstly, the recovery

process must be viewed as a random function of time and its temporal statistical properties, such as its temporal correlations, must be mathematically characterized. Secondly, such complete statistical understanding of the thermal recovery process must, in turn, be utilized in a statistical-inference framework that yields the optimal decision rule for classifying a lesion as malignant or benign. Both of these problems are formulated and solved in the present work.

Let us explain the concept of DTI and the patient data acquisition before developing the algorithm.

2.2 Dynamic thermal imaging

Thermal infrared imaging is a non-invasive, non-contact sensing method that captures the emitted electromagnetic radiation in the infrared region of the spectrum (i.e., 3–14μm) from an object. Thermal imaging provides a quantitative estimate of the spatial and temporal temperature profile of the target object. Since the surface temperature distribution directly depends on the subsurface properties [47], thermal infrared imaging can be used to unveil those subsurface properties by monitoring changes in the spatial and/or temporal profiles of the surface temperature.

In general, infrared imaging can be performed either passively (static thermal imaging) or actively (dynamic thermal imaging, DTI). On one hand, static thermal imaging involves the monitoring of the target temperature without applying any external stimulus to the surface of such a target. On the other hand, DTI consists of monitoring the target temperature after an initial temperature stimulus is applied to the target (such as deliberate heating or cooling). It was observed that static infrared imaging was not suitable for dermatological applications due to the poor detection and high false-alarm probabilities [32]. As a contrast, DTI have demonstrated to have high potential for this particular application [29, 33, 48]. Moreover, other

groups [34, 37] have explored the utilization of DTI for the detection of skin cancer, but, as was aforementioned, they have neglected the temporal statistical features inherent in the thermal recovery processes.

2.2.1 Image acquisition hardware

We perform DTI with three components. The first component is a cooling unit that is used to impart the temperature stimulus to the lesion and the surrounding skin tissue. Two different cooling unit were used in our study. The first one was a a Ranque-Hilsch vortex tube that generates an oil-free, moisture-free, ultra-quiet air flow. It was later replaced by a commercially available air-conditioning (AC) unit due to its portability. It was observed that by properly modifying the time the cooling unit was applied to the skin, the imparted temperature was almost the same for both units.

The second component is an infrared marker, which is used for correction of involuntary movement of the subject (i.e., image registration); the IR evolved from a canvas paper marker to a square piece of plastic with a square opening in the middle. Since the only purpose of the marker is to aid in the registration of the infrared sequence of frames, changing the material of the marker did not change the acquisition protocol.

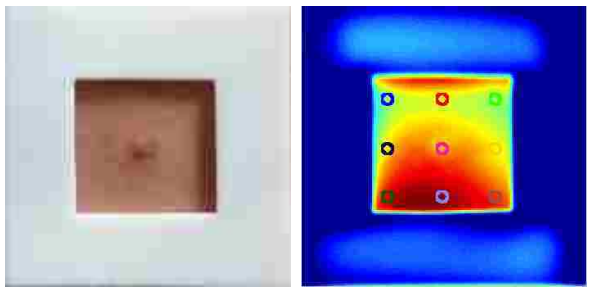
The third component includes the imagers. The first imager is a commercial visible still camera that is used to capture a reference image before the DTI acquisition commences. The second and most important imager is a longwave infrared (LWIR) camera that is used to capture a sequence of frames of the thermal recovery of the skin after the cool stimulus is applied. The LWIR camera consists of a 320×256 focal-plane array (FPA) of quantum-well infrared photodetectors (QWIP) operating at 60K. The noise equivalent temperature difference (NEDT) of the FPA is 20mK

and the QWIP camera is fitted with a 50mm, f/2 LWIR lens, yielding an approximate spatial resolution of 300 microns per pixel. The QWIP camera was chosen for our study because it has higher array uniformity, lower NETD and high spatial resolution as compared with other IR camera technologies [29, 49].

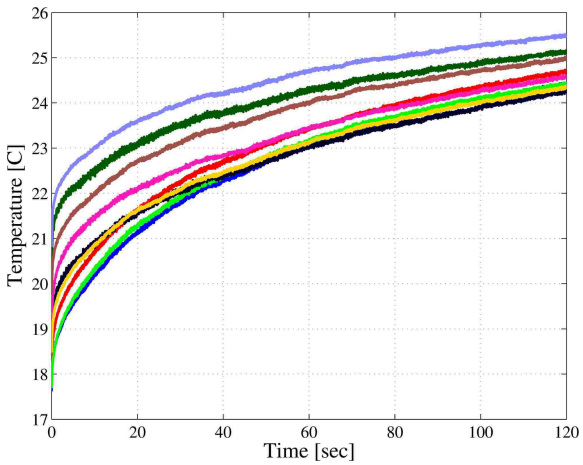
2.2.2 Imaging procedure

After informed consent, each subject was escorted to a designated room in the UNM Dermatology Clinic to perform the imaging procedure. The temperature of the room was controlled to be between 20^{o}C to 22^{o}C to make sure that all the patients were exposed to the same temperature before applying the cooling stimulus to the area of interest. At the beginning of the procedure, the square registration marker was placed around the lesion with the lesion centered in the opening, as shown in Fig. 2.1(a). A visible image of the lesion was then taken with the digital camera for reference. After collection of the visible image, a 15 second infrared image sequence of the marked area was collected to serve as a baseline. Later, the subjects skin within the marker opening was cooled for the 15 or 110 seconds, depending of the cooling unit used. After cooling, the exposed area was allowed to warm up naturally to ambient temperature. During the warm-up phase, thermal images of the skin were captured for a total of 2 minutes at a rate of 60 frames per second with the QWIP camera. All the thermal images were recorded using an uncompressed 14-bit format. The total time required to complete the imaging procedure was less than five minutes.

If the subject was scheduled for a biopsy, the biopsy was performed following the data collection by the attending dermatologist and sent to pathology for diagnosis. The biopsy results were delivered to us within the two weeks following the imaging procedure. Some patients were clinically diagnosed with a benign condition by the staff, and, therefore, no biopsy was performed. These patients are considered as



(a) (b)



(c)

Figure 2.1: Example of a patient dataset: (a) example of one square plastic marker used in the data acquisition step; (b) first frame of the infrared sequence, note that the visible and this frame are spatially aligned; and (c) the thermal recovery curves (TRCs) for the labeled pixels in (b).

control patients and are included in the set of benign patients.

2.2.3 Image registration

Since involuntary movements of the patients cannot be avoided, image registration must be performed over the infrared sequence of images. Moreover, to correctly reference the lesion location within the IR sequence (i.e., the mole, which not necessarily can be spotted in the IR sequence), the visible picture must also be spatially aligned

with the IR sequence. Therefore, the visible image is considered as an additional frame for the purposes of the registration process.

The registration is conducted as follows. First, we use the Harris corner-detector algorithm [50] to automatically detect the four corners of the plastic marker amongst the entire sequence of frames. Second, by assuming rigid movement of the scene, we estimate an affine transformation matrix that maps such a movement between the corners of consecutive frames (one matrix is estimated for each pair of consecutive frames) [51]. Third, we utilize the inverse of each transformation matrix to align each frame with respect to the first frame of the sequence [52]. After registration, both the visible image and the entire IR sequence are spatially aligned, generating a three-dimensional (3D) array real numbers that we term the *patient dataset*. Figure 2.1(b) depicts the first IR frame after the cooling was removed of the same example case presented in Fig. 2.1(a); note that both the visible and the first IR frame are spatially aligned. The thermal recovery curves (TRCs) of the labeled pixels are shown in Fig. 2.1(b), where it can be noted that there is some non-uniformity in the cooling process that make these TRCs to start at different initial temperature.

Most of the image registration of the patient dataset utilized in this dissertation was performed by Dr. David Ramirez.

2.2.4 Camera calibration

In order to have a temperature measurement of the skin surface as accurate as possible, the QWIP camera must be radiometrically calibrated. As in any FPA, the camera suffers of the non-uniform response of its detectors (a problem known in the literature as non-uniformity) and it is compensated by means of non-uniformity correction (NUC) tables performed and stored during the factory calibration process.

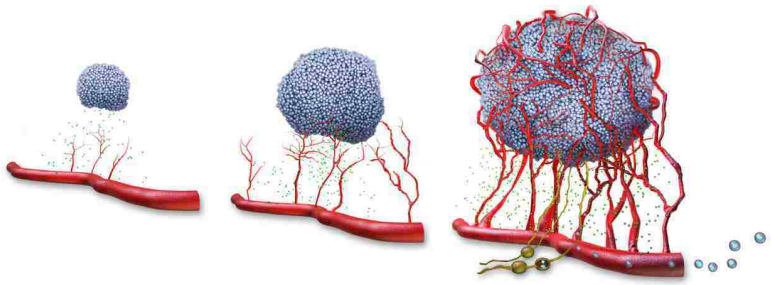
The radiometric calibration is achieved by means of the two-point calibration

technique [53]. This calibration is performed by placing, in the field-of-view (FOV) of the camera, a uniform-intensity calibration device such as a black-body source at two distinct and known temperatures [54]. The gain and the bias of each detector are then calibrated across the array so that all detectors produce a radiometrically accurate and uniform readout at the two reference temperatures. The reference temperatures where chosen to be within the normal temperature of the skin, i.e., 25^oC and 40^oC. Examples of thermal recovery curves after the temperature calibration was performed were already shown in Fig. 2.1(c).

2.3 Postulation of the detection problem

The physical principle that the diagnosis will be based upon is the following. Skin cancers, like all solid malignant tumors, require a blood supply in order to grow larger than a few millimeters in diameter [55, 56]. Tumors induce the growth of new capillary blood vessels (a process called angiogenesis) by producing specific angiogenesis-promoting growth factors. The so-called *precancerous* lesions of the skin, including atypical moles, are already angiogenic, as indicated by their higher density of capillaries than surrounding normal skin. New blood vessel growth continues through the progression from precancerous skin lesions to full-blown skin cancer as depicted in Fig. 2.2, [57]. The presence of new blood vessels and the increased blood supply should somewhat change the thermal response of the tumor cells when a stimulus is applied.

Under this scenario, we can assume that the patient condition is hidden within TRCs of suspicious lesions. Moreover, we assume that the malignancy of a lesion can be inferred only by monitoring the tissue of the mole. Thus, let us denote by $S_j(t),\ j = 0, 1$ the average TRCs (across all the pixels within the mole) of patients with benign and malignant condition, respectively. We account the intrinsic ran-



(a) (b) (c)

Figure 2.2: Tumor angiogenesis in cancer at different stages: (a) The tumor release growth factors that activate the growing cells generating blood vessel sprouts. (b) The blood vessels feed the tumor that growths thanks to cell proliferation. (c) The tumor becomes vascularized and it starts to metastasize through the blood stream.

domness of these signals by assuming that, under each hypotheses, these signals are parameterized by different random vectors, i.e., $S_j(t) \equiv S_j(t;\boldsymbol{\Theta}_j)$, $j = 0, 1$. As a consequence, the binary detection problem of detecting if a measured TRC, $Y(t)$, is either from benign tissue (null hypothesis, H_0) or malignant tissue (alternative hypothesis, H_1) is formulated by

$$H_0 \ : Y(t) = S_0(t;\boldsymbol{\Theta}_0) \ , \quad t \in [0, T] \tag{2.1a}$$

$$H_1 \ : Y(t) = S_1(t;\boldsymbol{\Theta}_1) \ , \quad t \in [0, T] \tag{2.1b}$$

where T represents the TRCs acquisition time. The alternative hypothesis is assumed to include all the conditions classified as non-benign, including malignant melanomas (MM), basal-cell carcinoma (BCC) and squamous-cell carcinoma (SCC) cases.

The continuous-time signals $S_j(t;\boldsymbol{\Theta}_j)$, $j = 0, 1$ are modeled as stochastic processes due to the stochastic nature of the problem, the unknown quantities and the patient-specific parameters that cannot be determined a priori, as discussed next.

2.4 Physics-based stochastic model for the skin thermal recovery

Before developing the algorithm for mapping each TRC to a decision on its condition, we must develop a physics-based, probabilistic mathematical model for the TRCs.

At this stage, the lesion boundary is defined by visual inspection over the mole region in the visible image. Since the area that defines the mole may contain malignant and benign tissue we need to look at the aggregated effect of the lesion TRC. The most natural and simple way to do such an aggregation is by computing the average TRC over all the pixels within the region that defines the mole. In what follows, we use the term TRC to refer to the average TRC of a lesion unless the opposite it is specifically stated.

2.4.1 Heat equation

The mathematical model should incorporate detailed physical and statistical information about the TRC signals; however, it must be simple enough to ensure a feasible solution with the available information. Moreover, the model must be derived based on the physics that generates the TRCs. First, we are using an infrared camera to measure thermal evolution through the acquisition of the infrared radiation emitted by the skin; thus, we can assume that the physics of the problem is governed by a heat equation, which describes the spatio-temporal thermal response of the human skin. Second, it is believed that the veins in the fat-dermis layer and the blood-vessels within the dermis layer (depicted in Fig. 2.3) are the most relevant source of heat acquired by the infrared camera. Moreover there is no evidence that we can resolve, using an infrared camera, subcutaneous thermal processes beyond such an interface.

Third, there is evidence suggesting that the *in vivo* emissivity of the skin is near

unity [58, 59, 60]. Therefore, it can be assumed that the skin mimics a perfect black-body source, meaning that all the subcutaneous thermal processes are somewhat integrated to the surface of the skin. As a consequence, the temperature of the skin captures the cumulative effect of all the subcutaneous thermal processes, which allow us to define an *effective* diffusion constant, $\mathcal{D}$, that consolidates all the skin thermal parameters (i.e., the *effective* tissue density, ρ, the *effective* specific heat of the tissue, C and the *effective* thermal conductivity, k).

The model only will be affected by the variations on the depth of the lesion, x, due to the averaging of TRCs previously discussed. Therefore, the temperature of the skin sample is assumed to be modeled by the one-dimensional heat equation,

$$\rho C \frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2} \ , \quad \begin{matrix} t \in [0, T] \\ x \in [0, H] \end{matrix} \ ,$$

where T represents the acquisition time and H the bottom of the spatial domain. (To clarify, $x = 0$ represents the skin surface.) The effective diffusion constant, $\mathcal{D}$, is defined by $k/\rho C$.

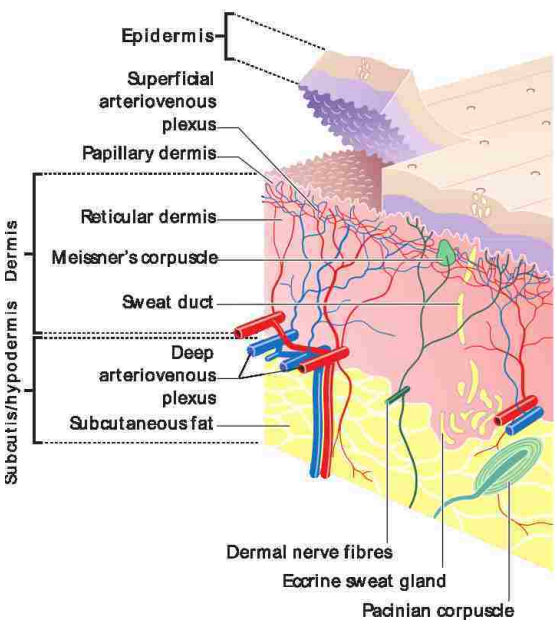


Figure 2.3: Pictorial representation of the skin layers, where the network of veins located in the fat-dermis interface can be identified. It is believed that these veins and the blood-vessels within the dermis layer are the most relevant source of heat acquired by the infrared camera, which by thermal diffusion reach the skin surface.

Since we have stated that we cannot resolve beyond the fat-dermis interface because of the veins (that carry blood at a temperature determined by the core of the human body), we set the *bottom* of our spatial domain to be at fat-dermis interface; this means H is given by the sum of the thicknesses of the epidermis and dermis layers. Since this bottom is in contact with the veins of the fat-dermis interface and zones with a high density of blood vessels, which assure a constant temperature, termed the blood or core temperature, T_B, for all time. This defines the first boundary condition.

As explained in Section 2.2.2, in our study we perform the initial cooling of the lesion and acquire the subsequent thermal recovery. Right after the cooling is stopped the surface of the skin is exposed to the natural convection with the air. Therefore, the heat flux at the surface, $q'' = -ku_x(0,t)$, must be equal to the product of the convective heat transfer coefficient, h_∞, and the temperature difference between the skin surface, $u(0,t)$, and the temperature of the air, T_A.

In summary, the model for the temperature is given by

$$\frac{\partial u}{\partial t} = \mathcal{D}\frac{\partial^2 u}{\partial x^2} \ , \quad \begin{matrix} t \in [0,T] \\ x \in [0,H] \end{matrix} \tag{2.2a}$$

with boundary conditions

$$q'' = -ku_x(0,t) = h_\infty \left(u(0,t) - T_A\right), \tag{2.2b}$$

$$u(H,t) = T_B \ , \tag{2.2c}$$

$$u_x(H,t) = 0 \ , \tag{2.2d}$$

where $\mathcal{D} = k/\rho C$ is the effective diffusion coefficient. For convenience, we impose the additional condition (2.2d). At this stage, we have not specified the initial condition function $u_0(x)$ because it is actually unknown. (We will discuss how we address this later.) Next, we present one alternative to solve (2.2).

2.4.2 Closed-form analytical solution

The temperature model is solved by dividing the solution into the steady-state solution, $u_{ss}(x)$, and the transient solution, $v(x,t)$, i.e., $u(x,t) = u_{ss}(x) + v(x,t)$.

Steady-state solution

The steady-state solution can be easily found by setting $u_t(x,t) = 0$ in (2.2a). Therefore we need to solve $u_{xx} = 0$, which have solution of the form $u_{ss}(x) = Ax + B$. From (2.2c) we have that $u_{ss}(H) = AH + B = T_B$, or $B = T_B - AH$. Replacing this into (2.2b) we have

$$\begin{aligned} -ku_x(0,t) &= h_\infty \left(u(0,t) - T_A\right) \\ \therefore -kA &= h_\infty(B - T_A) = h_\infty(T_B - AH - T_A) \\ \therefore (h_\infty H - k)A &= h_\infty(T_B - T_A) \\ \therefore A &= \frac{\frac{h_\infty}{k}(T_B - T_A)}{\frac{h_\infty H}{k} - 1} \end{aligned}$$

and $B = T_B - AH = \frac{T_B - \frac{h_\infty H}{k} T_A}{1 - \frac{h_\infty H}{k}}$. Then, the steady-state solution is

$$u_{ss}(x) = \left(\frac{\frac{h_\infty}{k}(T_B - T_A)}{\frac{h_\infty H}{k} - 1}\right) x + \left(\frac{T_B - \frac{h_\infty H}{k} T_A}{1 - \frac{h_\infty H}{k}}\right), \quad x \in [0, H] \ . \tag{2.3}$$

Transient solution

The transient solution, $v(x,t)$, also satisfy (2.2a) but the boundary conditions are different because of the steady-state solution $u_{ss}(x)$. Substituting $u(x,t) = u_{ss}(x) + v(x,t)$ in (2.2c) we have $u(H,t) = u_{ss}(H) + v(H,t) = T_B$, but $u_{ss}(H) = T_B$, so

$v(H,t) = 0$. Doing the same for (2.2b) yields

$$\begin{aligned} -ku_x(0,t) &= h_\infty \left(u(0,t) - T_A\right) \\ -k(u_{ss_x}(0) + v_x(0,t)) &= h_\infty \left(u_{ss}(0) + v(0,t) - T_A\right) \\ -ku_{ss_x}(0) - kv_x(0,t) &= h_\infty \left(u_{ss}(0) - T_A\right) + h_\infty v(0,t) \\ h_\infty v(0,t) &= -kv_x(0,t) \ , \end{aligned}$$

where the last equality comes from the fact that $u_{ss}(x)$ was defined to satisfy (2.2b), thus $h_\infty \left(u_{ss}(0) - T_A\right) + ku_{ss_x}(0) = 0$. We now impose the condition that the heat-flux from the skin into the blood stream is zero for the transient response only, i.e., $v_x(H,t) = 0$. In summary, the problem to solve for the transient response, is governed by

$$\frac{\partial v}{\partial t} = \mathcal{D}\frac{\partial^2 v}{\partial x^2} \ , \quad \begin{array}{l} t \in [0,T] \\ x \in [0,H] \end{array} \ , \tag{2.4a}$$

with boundary conditions

$$h_\infty v(0,t) + kv_x(0,t) = 0 \ , \tag{2.4b}$$

$$v(H,t) = 0 \ , \tag{2.4c}$$

$$v_x(H,t) = 0 \ . \tag{2.4d}$$

The solution is obtained by following the separation of variables approach, i.e., we assume that the transient solution can be defined as the product of two functions: $v(x,t) = X(x)T(t)$, transforming the transient equation (2.4a) to $X\dot{T} = \mathcal{D}X''T$, or $\frac{X''}{X} = \frac{\dot{T}}{\mathcal{D}T} = \alpha$, where the coefficient α is known as the separation constant. These equation can be written as

$$X''(x) - \alpha X(x) = 0 \ , \quad x \in [0,H] \ , \tag{2.5a}$$

$$\dot{T}(t) - \alpha\mathcal{D}T(t) = 0 \ , \quad t \in [0,T] \ , \tag{2.5b}$$

where the boundary conditions (2.4b)-(2.4d) for $X(x) \neq 0$, $x \in [0, H]$ and $T(t) \neq 0$, $t \in [0, T]$ are recast as

$$X'(0) + \frac{h_\infty}{k} X(0) = 0 \ , \tag{2.6a}$$

$$X(H) = 0 \ , \tag{2.6b}$$

$$X'(H) = 0 \ . \tag{2.6c}$$

We solve (2.5a) first and find the solutions associated with the separation constant α. In particular, there are three cases of interest:

(i) $\boldsymbol{\alpha = 0}$. Here (2.5a) becomes $X'' = 0$, which have solutions of the form $X(x) = Cx + D$. Using this solution in (2.6c) we have that $X'(H) = D = 0$ and from (2.6b) we have that $X(H) = AH = 0$. Hence, $X(x) = 0$, which means that there are no eigenfunctions associated with $\alpha = 0$.

(ii) $\boldsymbol{\alpha = \mu^2 > 0}$. Here (2.5a) becomes $X'' - \mu^2 X = 0$, which has a general solution of the form $X(x) = Ce^{\mu x} + De^{-\mu x}$. By applying the BCs (2.6b) and (2.6c) to this solution, we have $X(H) = Ce^{\mu H} + De^{-\mu H} = 0$ and $X'(H) = C\mu e^{\mu H} - D\mu e^{-\mu H} = 0$. Now, $\mu X(H) + X'(H) = 2C\mu e^{\mu H} = 0$. Since $\mu \neq 0$, the only alternative is to have $C = 0$. Following the same rationale for $\mu X(H) - X'(H)$, we find that $D = 0$ as well. Therefore, there are no eigenfunctions for $\alpha > 0$.

(iii) $\boldsymbol{\alpha = -\mu^2 < 0}$. Here the equation becomes $Z'' + \mu^2 Z = 0$, which has a general solution of the form

$$X(x) = C \cos \mu x + D \sin \mu x \ .$$

Since $X'(x) = -C\mu \sin \mu x + D\mu \cos \mu x$, then $X'(0) = D\mu$. Thus, (2.6a) becomes $X'(0) + \frac{h_\infty}{k} X(0) = D\mu + \frac{h_\infty}{k} C = 0$, or $C = -\frac{k}{h_\infty} \mu D$. Replacing in (2.6b) we have

$$\begin{aligned} X(H) = C \cos \mu H + D \sin \mu H &= 0 \\ -\frac{k}{h_\infty} \mu D \cos \mu H + D \sin \mu H &= 0 \\ \left(\sin \mu H - \frac{k}{h_\infty} \mu \cos \mu H \right) D &= 0 \ . \end{aligned}$$

To ensure non-trivial solution, we impose $D \neq 0$, therefore we must find μ such that $\sin \mu H - \frac{k}{h_\infty} \mu \cos \mu H = 0$, or

$$\tan \mu H = \frac{k}{h_\infty} \mu \ . \tag{2.7}$$

Given the periodicity of (2.7), we will have infinite solutions $0 < \mu_1 < \mu_2 < \dots$ with the property

$$(n-1)\frac{\pi}{H} < \mu_n < (2n-1)\frac{\pi}{2H} \ ,$$

where μ_n is the n-th solution to the equation. Since the RHS in (2.7) have positive slope, then

$$\lim_{n \to \infty} \mu_n = (2n-1)\frac{\pi}{2H} \ .$$

Figure 2.4(a) shows an example of the solutions when parameter values from the literature [36] are used. Once the values μ_n are computed, we will have solutions of the form

$$X_n(x) = C_n \cos \mu_n x + D_n \sin \mu_n x \ , \quad \begin{matrix} x \in [0, H] \\ n = 1, 2, \dots \end{matrix} \ ,$$

known in the literature as the eigenfunctions associated with the negatives eigenvalues $\alpha_n = -\mu_n^2$. Now for the nth eigenvalue, the time ODE (2.5b) becomes

$$\dot{T}_n + \mathcal{D}\mu_n^2 T_n = 0 \ , \quad t \in [0, T] \ ,$$

which have solutions of the form $T_n(t) = e^{-\mathcal{D}\mu_n^2 t}$, where $\mathcal{D} = k/\rho C$. By multiplying these solutions together we have the n-th normal mode solution (nth eigenfunction)

$$v_n(x,t) = X_n(x)T_n(t) = \big(C_n \cos\mu_n x + D_n \sin\mu_n x\big)e^{-\mathcal{D}\mu_n^2 t} \ , \quad \begin{array}{l} x \in [0,H] \\ t \in [0,T] \\ n = 1,2,\ldots \end{array} \ .$$

We know that if $\{v_n\}$ is a sequence of solutions of the heat equation which satisfy the specific boundary conditions, then any linear combination of these solutions will also be a solution. Thus, the superposition of normal modes gives the general solution for the transient function,

$$v(x,t) = \sum_{n=1}^{\infty}\big(C_n \cos\mu_n x + D_n \sin\mu_n x\big)e^{-\mathcal{D}\mu_n^2 t} \ , \quad \begin{array}{l} x \in [0,H] \\ t \in [0,T] \end{array} \ , \tag{2.8}$$

where the coefficients C_n and D_n are determined by the generalized Fourier series expansion of the initial condition function, $u_0(x)$, as explained later. By composing the temperature model using (2.3) and (2.8) we obtain the analytical solution of the heat-equation problem

$$\begin{aligned} u(x,t) &= \left(\frac{\frac{h_\infty}{k}(T_B - T_A)}{\frac{h_\infty H}{k} - 1}\right)x + \left(\frac{T_B - \frac{h_\infty H}{k}T_A}{1 - \frac{h_\infty}{k}H}\right) \\ &+ \sum_{n=1}^{\infty}\big(C_n \cos\mu_n x + D_n \sin\mu_n x\big)e^{-\mathcal{D}\mu_n^2 t} \ , \end{aligned} \tag{2.9}$$

where $x \in [0,H]$ and $t \in [0,T]$. If the initial condition function $u_0(x)$ is known, we have that $u_0(x) = u(x,0) = u_{ss}(x) + \sum_{n=1}^{\infty}\big(C_n \cos\mu_n x + D_n \sin\mu_n x\big)$, or

$$u_0^{ss}(x) = \sum_{n=1}^{\infty}\big(C_n \cos\mu_n x + D_n \sin\mu_n x\big) \ , \tag{2.10}$$

where $u_0^{ss}(x) = u_0(x) - u_{ss}(x)$. From (2.10) we see that knowing the initial condition, the coefficients C_n and D_n correspond to the coefficients of the generalized Fourier

series expansion with basis functions $\{\cos \mu_n z\}$ and $\{\sin \mu_n z\}$, where $\{\mu_n\}$ are the solutions of (2.7). If we can determine a finite number of the coefficients C_n and D_n required to model the TRCs, say N, then the TRCs' mathematical model, at the skin surface will be governed by the structure

$$u(0,t) = B + \sum_{n=1}^{N} C_n e^{-\mathcal{D}\mu_n^2 t} \ , \tag{2.11}$$

where $B = (T_B - \frac{h_\infty H}{k} T_A)/(1 - \frac{h_\infty H}{k})$ as defined in (2.3). Next we utilize test functions as initial condition function $u_0(x)$ in order to study the decay of the C_n coefficients and determine a proper value for N.

2.4.3 Initial-condition analysis

Given the lack of knowledge regarding the initial temperature profile for the subcutaneous layers of the skin, there is no way for us to know *a priori* the initial condition function, $u_0(x)$. Some approximations have been proposed in the literature [61], but in general, the level of uniformity of the subcutaneous layers assumed is too unrealistic to make this alternatives suitable for all the cases.

Here we used four test function including a linear function, a quadratic and cubic polynomial functions as well as an exponential radial basis function constructed from the normal cumulative distribution function (CDF) implemented in MatLab. These four test functions are depicted in Fig. 2.4(b), where the required parameters (e.g., skin depth, blood temperature, etc) were obtained from the literature [36]. We numerically compute the coefficients C_n and D_n using the following rationale. Let us consider the K equidistant sample points $\boldsymbol{x} = [x_0 \ x_2 \ \cdots \ x_K]^T$ of the skin subsurface, being $x_0 = 0$ and $x_K = H$. Assume that we have numerically solved for the first N solutions of (2.7), i.e., we have $\boldsymbol{\mu} = [\mu_1 \ \mu_2 \ \cdots \ \mu_N]^T$, then (2.10) can be numerically recast as

$$\boldsymbol{h} = \boldsymbol{\Phi b} \ ,$$

where $\boldsymbol{h}$ is the sampled test function minus the sampled steady-state solution, $\boldsymbol{\Phi}$ is the basis function matrix of the generalized Fourier series expansion and $\boldsymbol{b}$ is the array of coefficients to be determined. In symbols,

$$\boldsymbol{h} = \begin{bmatrix} u_0(x_1) - u_{ss}(x_1) \\ u_0(x_2) - u_{ss}(x_2) \\ \vdots \\ u_0(x_K) - u_{ss}(x_K) \end{bmatrix}, \quad \boldsymbol{\Phi} = \left[\cos \boldsymbol{x}\boldsymbol{\mu}^T \sin \boldsymbol{x}\boldsymbol{\mu}^T\right], \quad \text{and, } \boldsymbol{b} = \begin{bmatrix} C_1 \\ \vdots \\ C_N \\ D_1 \\ \vdots \\ D_N \end{bmatrix},$$

which can be easily solved in Matlab. The first six coefficients for each test function depicted in Fig. 2.4(b) were computed. The results of the coefficients C_n, $n = 1, \ldots, 6$, are shown in Fig. 2.4(c)-2.4(f) for the four test functions. The actual value of the coefficients were normalized with respect to the maximum coefficient for each test function in order to facilitate their comparison.

The most relevant information that can be extracted from this analysis is that for all the test functions utilized here, two coefficients seem to be sufficient in the model because they decay rapidly as we increase N. This situation is particularly true for the case of cubic and exponential functions, which, based on Wilson and Spence's work [61], present the most feasible functions as initial condition. As a consequence, from now on we shall work under the assumption that the solution of the temperature model (2.11) can be approximated by the constant B plus two exponential functions, i.e., $u(0,t) = B + C_1 e^{-\mathcal{D}\mu_1^2 t} + C_2 e^{-\mathcal{D}\mu_2^2 t}$.

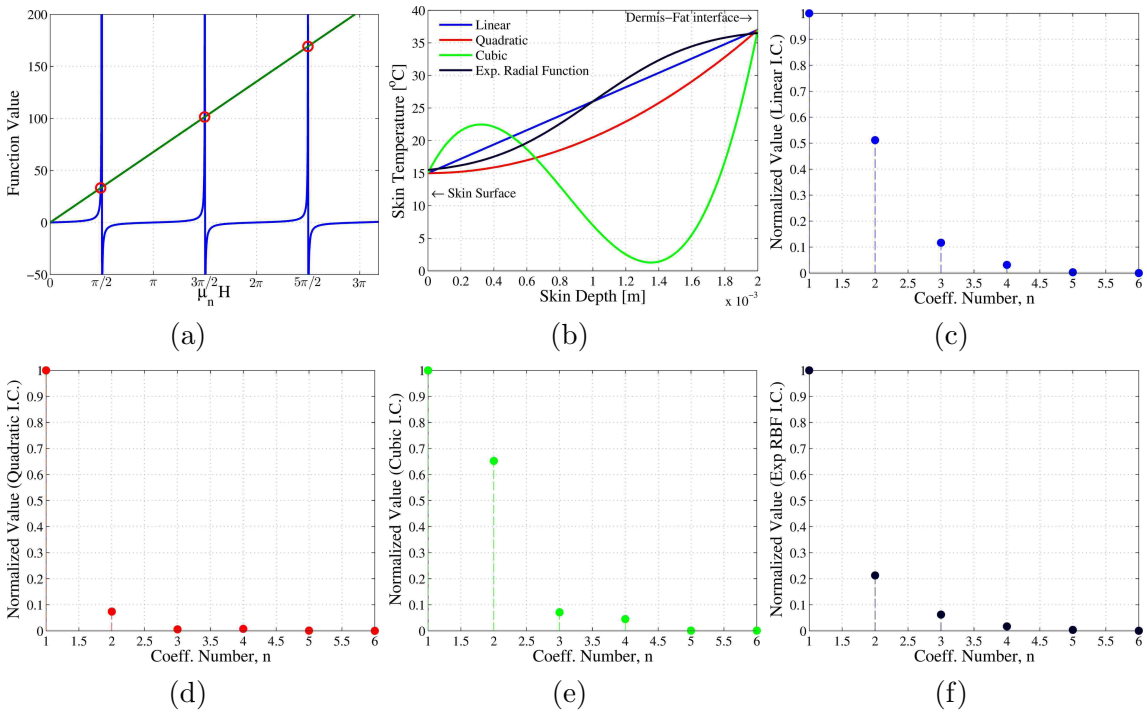


Figure 2.4: Numerical study of parts of the model analytical solution, using literature parameters for the required quantities. (a) Example solutions, μ_n for the non-linear equation (2.7). (b) Test functions used in the initial condition function analysis where for different initial condition test functions, the decay of the coefficients C_n is studied using literature parameters. (c)-(f) Coefficient decay for linear, quadratic, cubic and exponencial test functions, respectively

2.4.4 Stochastic model for the thermal recovery of human skin

By solving the proposed temperature model with proper boundary conditions we have determined that the solution at the surface of the skin can be approximated by a constant and two exponential functions. Nevertheless, when we solved the temperature model we assumed the knowledge of thermal and physical parameters of the skin that are either unknown or vary from patient-to-patient. For example, the effective diffusion coefficient, $\mathcal{D}$, appears in the solution and it is defined as the

ratio $k/\rho C$, where k is the effective thermal conductivity of the skin, ρ is the effective tissue density and C is the effective specific heat of the tissue. It is clear that the value of these parameters cannot be described in a deterministic fashion. A similar problem appears in (2.7), which requires, for example, the precise knowledge of the depth of the lesion to compute the values of μ_1 and μ_2 for the temperature model. We address these inherent uncertainties by considering each parameter within the model as a random variable. Therefore, the stochastic-process model for the TRCs is given by the temporal structure given by the solution of a heat equation and it is parameterized by a set of five random variables. Thus, the TRCs are modeled by the stochastic parameters given by $S(t;\boldsymbol{\Theta}) = \theta_1 + \theta_2 e^{-\theta_3 t} + \theta_4 e^{-\theta_5 t}$, where the entries of the random vector $\boldsymbol{\Theta}$ are correlated random variables. Since the random vector $\boldsymbol{\Theta}$ directly depends of the nature of the lesion under study, we assume that each hypothesis in the detection problem (2.1b) have its particular random vector, namely $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_1$. This implies that the binary hypothesis problem (2.1b) can be recast as

$$H_0 \ : Y(t) = S(t;\boldsymbol{\Theta}_0) = \theta_{0,1} + \theta_{0,2} e^{-\theta_{0,3} t} + \theta_{0,4} e^{-\theta_{0,5} t} \ , \quad t \in [0,T] \tag{2.12a}$$

$$H_1 \ : Y(t) = S(t;\boldsymbol{\Theta}_1) = \theta_{1,1} + \theta_{1,2} e^{-\theta_{1,3} t} + \theta_{1,4} e^{-\theta_{1,5} t} \ , \quad t \in [0,T] \ . \tag{2.12b}$$

The distribution of the random vectors $\boldsymbol{\Theta}_j = [\theta_{j,1} \ \cdots \ \theta_{j,5}]$, $j = 0, 1$ must be determined from patient data with known condition. For our purposes, the ground truth is taken as the biopsy result and/or the clinical diagnosis determined by the UNM's dermatology specialists at the moment of the patient data acquisition.

Random parameters distribution and their correlation

We assume that for each patient with known condition, each pixel within the selected lesion gives a realization of the random variables of the model. Each realization is

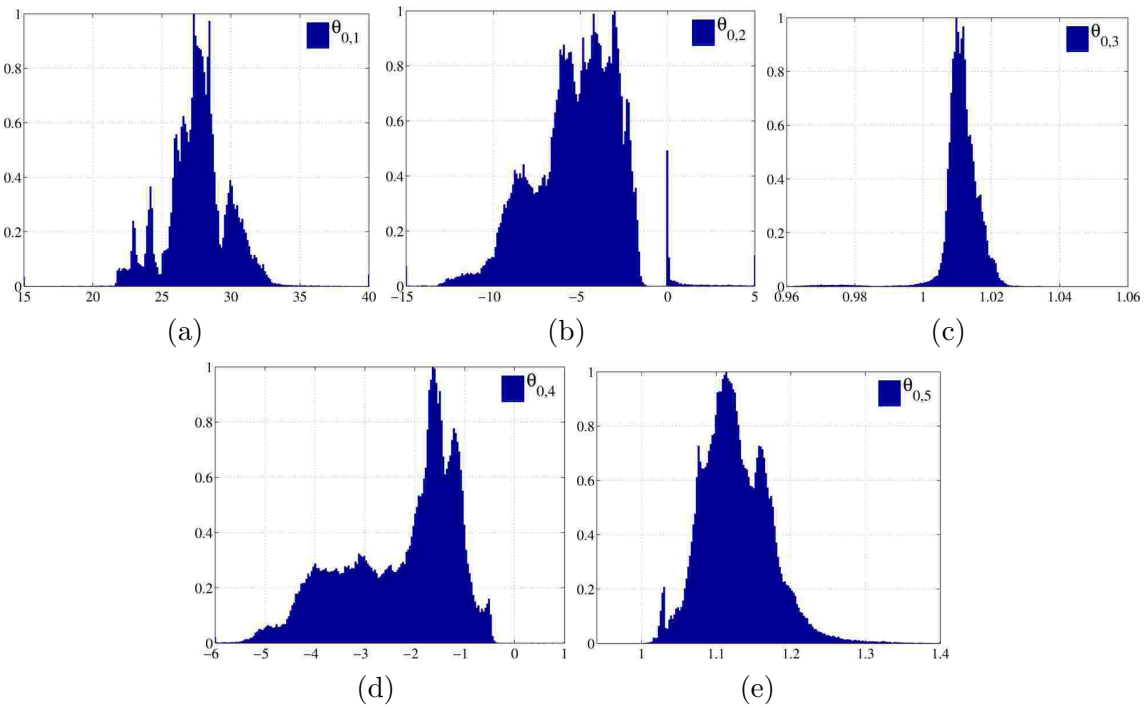


Figure 2.5: Histograms of the realizations of each of the random variables within the TRC model for the null hypothesis, H_0.

obtained by doing non-linear fitting of the the measured TRCs of each pixel with the dual exponential model. In turn, we can utilize the histograms of these realizations to estimate the marginal distribution of each random variable, under each one of the hypothesis, if required. For the purposes of illustration we obtained the histograms for the realizations of 140 patients with known condition, as determined by their biopsy results. (More details about this patient data will be given later in Chapter 3.)

As an example, we obtained the histograms of the realizations for a dataset of 140 patients, which includes 58 malignant cases and 82 benign cases (more details will be given later). The resulting histograms for the null and alternative hypothesis are depicted in Fig. 2.5 and Fig. 2.6, respectively. The correlation between the coefficients is not shown in the presented figures, but are given in Table 2.1 for $\mathbf{\Theta}_0$ and in tableTable 2.2 for $\mathbf{\Theta}_1$.

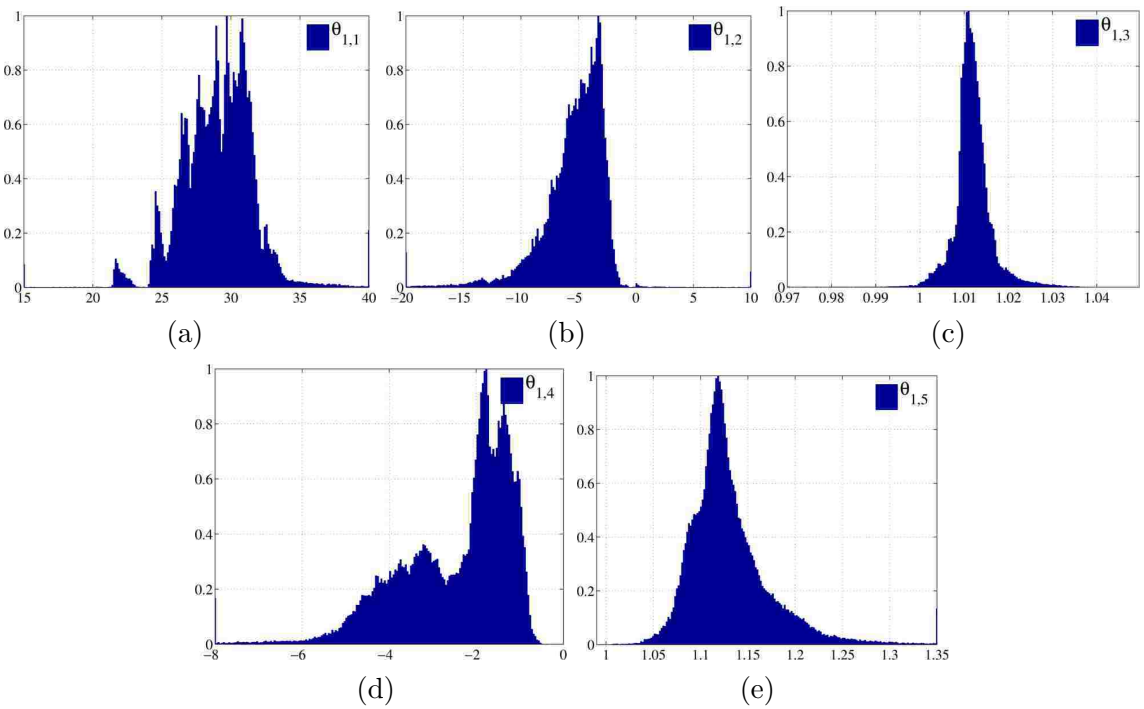


Figure 2.6: Histograms of the realizations of each of the random variables within the TRC model for the alternative hypothesis, H_1.

Even though the correlation between some parameters is not strong, we have decided to consider all the parameters to be correlated. Such a correlation is partially accounted in the autocorrelation functions, which we computed in the following section by using the structure of the stochastic model. Let us first describe the procedure we use to solve the detection problem by means of Grenander's approach.

Table 2.1: Correlation coefficient between the random parameters $\theta_{0,\ell}$, for $\ell = 1, 2, \ldots, 5$, which define the stochastic signal for the hypothesis H_0

	$\theta_{0,1}$	$\theta_{0,2}$	$\theta_{0,3}$	$\theta_{0,4}$	$\theta_{0,5}$
$\theta_{0,1}$	1.00	-0.98	0.01	0.01	-0.01
$\theta_{0,2}$		1.00	-0.03	0.08	0.02
$\theta_{0,3}$			1.00	-0.10	0.64
$\theta_{0,4}$				1.00	-0.12
$\theta_{0,5}$					1.00

Table 2.2: Correlation coefficient between the random parameters $\theta_{1,\ell}$, for $\ell = 1, 2, \ldots, 5$, which define the stochastic signal for the hypothesis H_1

	$\theta_{0,1}$	$\theta_{0,2}$	$\theta_{0,3}$	$\theta_{0,4}$	$\theta_{0,5}$
$\theta_{0,1}$	1.00	-1.00	-0.02	-0.01	-0.01
$\theta_{0,2}$		1.00	0.01	0.03	0.001
$\theta_{0,3}$			1.00	-0.06	0.53
$\theta_{0,4}$				1.00	-0.07
$\theta_{0,5}$					1.00

2.5 Solution of the detection (statistical-inference) problem

2.5.1 Background

With the stochastic model for the random processes under each hypothesis at hand, we proceed to solve the detection problem of classifying the TRC of a suspicious lesion to be either benign or malignant. The main difference of our detection problem and the continuous-time problems one can find in the literature is the presence of a stochastic signal under both hypotheses. Hence, each hypothesis have an autocorrelation function (determined by means of the corresponding stochastic signals).

Three main approaches to address the problem with two autocorrelation functions have been proposed in the literature. The first approach, termed the Grenander's approach [38], utilizes an orthogonal series expansion (more precisely, the Karhunen-Loève expansion) to transform the continuous-time random processes into statistically-equivalent discrete sequences, which compactly contain all the statistical information from the continuous-time random processes. This sequence of coefficients is later utilized to construct an optimal detector under a certain prescribed rule.

The second approach, termed the reproducing-kernel Hilbert space (RKHS) approach [62], defines the likelihood ratio test in terms of norms of the observed random

processes, which may have different definition on each RKHS. (In simple terms, a RKHS as a Hilbert space of functions with all evaluation functionals bounded and linear that is *generated* by one autocorrelation function.) For example for the unequal covariance problem the test-statistics can be defined by [63]

$$Z = \|X\|^2_{R_0} - \|X\|^2_{R_1} \ ,$$

where $\|\cdot\|^2_{R_j}$ represents the square of the norm defined under the RKHS associated with the autocorrelation R_j. The success of this approach is directly linked to the definition of the norms of the data vectors in the RKHS under each hypothesis, which are hard or even impossible to obtain for complex autocorrelation functions. The closed-form solution for these norms have been obtained for very specific structures of the ACFs and their extension to complex structures is not trivial. See, for example [62, 63, 64] for further details. As such, given the complexity of our autocorrelation functions (which we will introduce in the next section) this approach is not applicable to our specific problem.

The third approach is termed the whitening-filter approach and it was proposed by Van Trees [65]. The idea is to define a filter, $h_w(t,u)$, such that the autocorrelation function of the null-hypothesis is a white process. In symbols, if we denote by $S^*(t;\boldsymbol{\Theta}_0)$ the output of this filter when $S(t;\boldsymbol{\Theta}_0)$ is the input, we choose $h_w(t,u)$ such that $\mathsf{R}^*_0(t,u) = \mathsf{E}\left[S^*(t;\boldsymbol{\Theta}_0)S^*(u;\boldsymbol{\Theta}_0)\right] = \delta(t-u)$. Once this filter is defined, the equivalent detection problem becomes a classical detection problem, because H^*_0 has white ACF and H^*_1 has ACF given by $\mathsf{R}^*_1(t,u) = \mathsf{E}\left[S^*(t;\boldsymbol{\Theta}_1)S^*(u;\boldsymbol{\Theta}_1)\right]$. It is clear that this approach requires the computation of integral convolutions to use the whitening filter. The main difficulty of this approach is that the optimality of the method is directly linked to the inverse of the whitening filter; this means that solving this problem requires to solve inverse integral equations, which can be difficult if not impossible to solve because this is, in general, an *ill-posed* problem. This problem is similar in nature to the solution of integral inverse problems in

image processing which are ill-posed and are directly affected by the presence of noise. (See, for example, [66] for an introductory survey of these type of problems.) As a consequence, this alternative is not feasible to implement and, it if is, then there is no guarantee that the filter is reversible, which will make this approach to be non-optimal.

In this work, we solve the detection problem by following Grenander's approach because its optimality is guaranteed if one work the solution under the correct assumptions. Let us explain this approach with more details and the main assumptions we followed that make this alternative the optimal solution for our problem.

2.5.2 Grenander's approach

As it was explained above, Grenander's approach [38] utilizes an orthogonal series expansions of the random processes under each hypothesis known as the Karhunen-Loève (KL) expansion. After this expansion, the continuous-time random processes are represented by a discrete-sequence of coefficients termed the KL coefficients. These coefficients, in turn, compactly contain all the statistical information extracted from the stochastic TRCs and can be utilized in an optimal statistical decision theory framework to address the detection problem.

In simple terms, the steps required to solve the problem are under this approach are: (1) Obtain the autocorrelation function (ACF) that characterize the stochastic processes under each hypothesis; (2) Utilize Mercer's theorem to characterize the autocorrelation functions by its corresponding eigenvalue-eigenfunction pairs[1]; (3) Apply the KL expansion and Grenander's theorem to construct the likelihood ratio of the problem; and (4) Utilize optimal statistical inference (optimum decision theory)

[1]For simplicity, in what follows, we term the eigenvalue-eigenfunction pairs as the eigenpairs.

to declare the malignancy of a lesion by means of the likelihood ratio test.

In what follows, we propose and develop each one of these steps in detail.

2.6 Autocorrelation functions

2.6.1 Analytical form of the autocorrelation function

By definition, the ACF of a zero-mean stochastic process $S(t;\boldsymbol{\Theta}_j)$ is defined by $\mathsf{R}_j(t,u) = \mathsf{E}\left[S(t;\boldsymbol{\Theta}_j)S(u;\boldsymbol{\Theta}_j)\right]$, where $\mathsf{E}\left[\cdot\right]$ denotes the mathematical expectation. Expanding this expression, the jth ACF is defined by

$$\begin{aligned}
\mathsf{R}_j(t,u) &= \mathsf{E}\left[S(t;\boldsymbol{\Theta}_j)S(u;\boldsymbol{\Theta}_j)\right] \\
&= \mathsf{E}\left[\left(\theta_{j,1}+\theta_{j,2}e^{-\theta_{j,3}t}+\theta_{j,4}e^{-\theta_{j,5}t}\right)\left(\theta_{j,1}+\theta_{j,2}e^{-\theta_{j,3}u}+\theta_{j,4}e^{-\theta_{j,5}u}\right)\right] \\
&= \mathsf{E}\left[{\theta_{j,1}}^2\right]+\mathsf{E}\left[\theta_{j,1}\theta_{j,2}e^{-\theta_{j,3}u}\right]+\mathsf{E}\left[\theta_{j,1}\theta_{j,4}e^{-\theta_{j,5}u}\right] \\
&\quad +\mathsf{E}\left[\theta_{j,1}\theta_{j,2}e^{-\theta_{j,3}t}\right]+\mathsf{E}\left[{\theta_{j,2}}^2e^{-\theta_{j,3}(t+u)}\right]+\mathsf{E}\left[\theta_{j,2}\theta_{j,4}e^{-\theta_{j,3}t}e^{-\theta_{j,5}u}\right] \\
&\quad +\mathsf{E}\left[\theta_{j,1}\theta_{j,4}e^{-\theta_{j,5}t}\right]+\mathsf{E}\left[\theta_{j,2}\theta_{j,4}e^{-\theta_{j,5}t}e^{-\theta_{j,3}u}\right] \\
&\quad +\mathsf{E}\left[\theta_{j,4}^2e^{-\theta_{j,5}(t+u)}\right]\ , \qquad (2.13)
\end{aligned}$$

for $(t,u)\in[0,T]$. From (2.13) we observe that we require to know the expectation of the product of two, three and four random variables. The procedure to address these products when computing the ACF for our modeled stochastic processes is explained in detail next.

Expectation of the product of random variables

Let us determine the expectation of the product of four random variables X, Y, U and V, i.e., we want to determine an analytical expression for $\mathsf{E}\left[XYUV\right]$ from

pairwise relations between these random variables. It is known that $\mathsf{cov}\,(X,Y) = \mathsf{E}\,[XY] - \mathsf{E}\,[X]\,\mathsf{E}\,[Y]$, thus the expectation for the product of two random variables is

$$\mathsf{E}\,[XY] = \mathsf{E}\,[X]\,\mathsf{E}\,[Y] + \mathsf{cov}\,(X,Y)\ . \tag{2.14}$$

Let us define the random variables $\Delta X = X - \mathsf{E}\,[X]$, $\Delta Y = Y - \mathsf{E}\,[Y]$, $\Delta U = U - \mathsf{E}\,[U]$ and $\Delta V = V - \mathsf{E}\,[V]$. In terms of these zero-mean random variables, the product XY can be also expressed by $XY = (\Delta X + \mathsf{E}\,[X])(\Delta Y + \mathsf{E}\,[Y]) = \mathsf{E}\,[X]\,\mathsf{E}\,[Y] + \mathsf{E}\,[X]\,\Delta Y + \mathsf{E}\,[Y]\,\Delta X + \Delta X \Delta Y$. Subtracting (2.14) from this last expression we have

$$XY - \mathsf{E}\,[XY] = \mathsf{E}\,[X]\,\Delta Y + \mathsf{E}\,[Y]\,\Delta X + \Delta X \Delta Y - \mathsf{cov}\,(X,Y)\ . \tag{2.15}$$

Similarly, for U and V we have

$$UV - \mathsf{E}\,[UV] = \mathsf{E}\,[U]\,\Delta V + \mathsf{E}\,[V]\,\Delta U + \Delta U \Delta V - \mathsf{cov}\,(U,V)\ . \tag{2.16}$$

If we multiply (2.15) and (2.16) and take expectation we obtain

$$\begin{aligned}
\mathsf{cov}\,(XY,UV) &= \mathsf{E}\,[X]\,\mathsf{E}\,[U]\,\mathsf{cov}\,(Y,V) + \mathsf{E}\,[X]\,\mathsf{E}\,[V]\,\mathsf{cov}\,(Y,U) + \mathsf{E}\,[Y]\,\mathsf{E}\,[U]\,\mathsf{cov}\,(X,V) \\
&+ \mathsf{E}\,[Y]\,\mathsf{E}\,[V]\,\mathsf{cov}\,(X,U) + \mathsf{E}\,[\Delta X \Delta Y \Delta U \Delta V] + \mathsf{E}\,[X]\,\mathsf{E}\,[\Delta Y \Delta U \Delta V] \\
&+ \mathsf{E}\,[Y]\,\mathsf{E}\,[\Delta X \Delta U \Delta V] + \mathsf{E}\,[U]\,\mathsf{E}\,[\Delta X \Delta Y \Delta V] + \mathsf{E}\,[V]\,\mathsf{E}\,[\Delta X \Delta Y \Delta U] \\
&- \mathsf{cov}\,(X,Y)\,\mathsf{cov}\,(U,V)\,.
\end{aligned} \tag{2.17}$$

One case of interest is when $Y = 1$ in (2.17), which means that $\mathsf{E}\,[Y] = 1$, $\Delta Y = 0$ and $\mathsf{cov}\,(Y,\cdot) = 0$; therefore we have

$$\mathsf{cov}\,(X,UV) = \mathsf{E}\,[U]\,\mathsf{cov}\,(X,V) + \mathsf{E}\,[V]\,\mathsf{cov}\,(X,U) + \mathsf{E}\,[\Delta X \Delta U \Delta V]\ . \tag{2.18}$$

Now, we use the derived relations to define the product of three and four random

variables. The expectation of the product of three random variables will be

$$\begin{aligned}
\mathsf{E}[XUV] &= \mathsf{E}[X(UV)] = \mathsf{E}[XZ] \overset{(2.14)}{=} \mathsf{E}[X]\,\mathsf{E}[Z] + \mathsf{cov}(X,Z) \\
&= \mathsf{E}[X]\,\mathsf{E}[UV] + \mathsf{cov}(X,UV) \\
&\overset{(2.14)}{=} \mathsf{E}[X]\,(\mathsf{E}[U]\,\mathsf{E}[V] + \mathsf{cov}(U,V)) + \mathsf{cov}(X,UV) \\
&\overset{(2.18)}{=} \mathsf{E}[X]\,\mathsf{E}[U]\,\mathsf{E}[V] + \mathsf{E}[X]\,\mathsf{cov}(U,V) + \mathsf{E}[U]\,\mathsf{cov}(X,V) \\
&\quad + \mathsf{E}[V]\,\mathsf{cov}(X,U) + \mathsf{E}[\Delta X \Delta U \Delta V] \ . \qquad (2.19)
\end{aligned}$$

For the product of four random variables we have

$$\begin{aligned}
\mathsf{E}[XYUV] &= \mathsf{E}[(XY)(UV)] \overset{(2.14)}{=} \mathsf{cov}(XY,UV) + \mathsf{E}[XY]\,\mathsf{E}[UV] \\
&\overset{(2.14)}{=} \mathsf{cov}(XY,UV) + (\mathsf{cov}(X,Y) + \mathsf{E}[X]\,\mathsf{E}[Y])(\mathsf{cov}(U,V) + \mathsf{E}[U]\,\mathsf{E}[V]) \\
&\overset{(2.17)}{=} \mathsf{E}[X]\,\mathsf{E}[U]\,\mathsf{cov}(Y,V) + \mathsf{E}[X]\,\mathsf{E}[V]\,\mathsf{cov}(Y,U) + \mathsf{E}[Y]\,\mathsf{E}[U]\,\mathsf{cov}(X,V) \\
&\quad + \mathsf{E}[\Delta X \Delta Y \Delta U \Delta V] \\
&\quad + \mathsf{E}[Y]\,\mathsf{E}[V]\,\mathsf{cov}(X,U) + \mathsf{E}[X]\,\mathsf{E}[\Delta Y \Delta U \Delta V] + \mathsf{E}[Y]\,\mathsf{E}[\Delta X \Delta U \Delta V] \\
&\quad + \mathsf{E}[U]\,\mathsf{E}[\Delta X \Delta Y \Delta V] + \mathsf{E}[V]\,\mathsf{E}[\Delta X \Delta Y \Delta U] + \mathsf{E}[U]\,\mathsf{E}[V]\,\mathsf{cov}(X,Y) \\
&\quad + \mathsf{E}[X]\,\mathsf{E}[Y]\,\mathsf{cov}(U,V) + \mathsf{E}[X]\,\mathsf{E}[Y]\,\mathsf{E}[U]\,\mathsf{E}[V] \ , \qquad (2.20)
\end{aligned}$$

which under the assumption that the four random variables are Gaussian (2.20) can be reduced to

$$\begin{aligned}
\mathsf{E}[XYUV] &= \mathsf{E}[X]\,\mathsf{E}[Y]\,\mathsf{cov}(U,V) + \mathsf{E}[X]\,\mathsf{E}[U]\,\mathsf{cov}(Y,V) + \mathsf{E}[X]\,\mathsf{E}[V]\,\mathsf{cov}(Y,U) \\
&\quad + \mathsf{E}[Y]\,\mathsf{E}[U]\,\mathsf{cov}(X,V) + \mathsf{E}[Y]\,\mathsf{E}[V]\,\mathsf{cov}(X,U) + \mathsf{E}[U]\,\mathsf{E}[V]\,\mathsf{cov}(X,Y) \\
&\quad + \mathsf{cov}(X,Y)\,\mathsf{cov}(U,V) + \mathsf{cov}(X,U)\,\mathsf{cov}(Y,V) + \mathsf{cov}(X,V)\,\mathsf{cov}(Y,U) \\
&\quad + \mathsf{E}[X]\,\mathsf{E}[Y]\,\mathsf{E}[U]\,\mathsf{E}[V] \ . \qquad (2.21)
\end{aligned}$$

From (2.14), (2.19) and (2.20) one can obtain the expressions for each term within the expectation formula without any assumption regarding the distribution of these

random variables. If we assume that the parameters depicted in Fig. 2.5 and Fig. 2.6 are Gaussian, then one can use the simplified version of the expectation of the product given by (2.21). The first observation that worth mentioning is that after the random parameters for the stochastic model are characterized, the computation of the analytical ACF only requires the computation of the expectation and pairwise covariances of those parameters.

An new assumption is introduced here. We assume that the correlation function between benign and malignant patients can be neglected when compared to the autocorrelation function of each hypothesis. In symbols, we assume that

$$\mathsf{R}_{k,\ell}(t,u) \triangleq \mathsf{E}\left[S(t;\boldsymbol{\Theta}_k)S(u;\boldsymbol{\Theta}_\ell)\right] = 0 \tag{2.22}$$

for $k \neq \ell$. This assumption states that patients with benign and malignant conditions are uncorrelated. As a consequence, the jth hypothesis is statistically characterized in full by its corresponding ACF, R_j, defined in (2.13).

Autocorrelation function

We estimate the ACF by applying the formulas for the expectation of the product of random variables to each expectation in (2.13) as follows.

The first term in (2.13) is $\mathsf{E}\left[{\theta_{j,1}}^2\right]$, which can be easily estimated from the patients with known diagnosis. The second term is

$$\begin{aligned}
\mathsf{E}\left[\theta_{j,1}\theta_{j,2}e^{-\theta_{j,3}u}\right] &= \mathsf{E}\left[XUV\right] = \mathsf{E}\left[X\right]\mathsf{E}\left[U\right]\mathsf{E}\left[V\right] + \mathsf{E}\left[X\right]\mathsf{cov}\left(U,V\right) \\
&\quad + \mathsf{E}\left[U\right]\mathsf{cov}\left(X,V\right) + \mathsf{E}\left[V\right]\mathsf{cov}\left(X,U\right) \\
&= \mathsf{E}\left[\theta_{j,1}\right]\mathsf{E}\left[\theta_{j,2}\right]\mathsf{E}\left[e^{-\theta_{j,3}u}\right] \\
&\quad + \mathsf{E}\left[\theta_{j,1}\right]\mathsf{cov}\left(\theta_{j,2}, e^{-\theta_{j,3}u}\right) \\
&\quad + \mathsf{E}\left[\theta_{j,2}\right]\mathsf{cov}\left(\theta_{j,1}, e^{-\theta_{j,3}u}\right) \\
&\quad + \mathsf{E}\left[e^{-\theta_{j,3}u}\right]\mathsf{cov}\left(\theta_{j,1}, \theta_{j,2}\right) \ .
\end{aligned} \tag{2.23}$$

Where the expectation of the exponential function of a random variable is given by

$$\mathsf{E}\left[e^{-\theta_{j,3}u}\right] = \exp\left[-u\mu_{j,3} + u^2\frac{\sigma_{j,3}^2}{2}\right] , \tag{2.24}$$

with $\mu_{j,3} = \mathsf{E}\left[\theta_{j,3}\right]$ and $\sigma_{j,3}^2 = \mathsf{var}\left[\theta_{j,3}\right]$. Under the assumption of Gaussian parameters, the random variable $\theta_{j,k}, e^{-\theta_{j,3}u}$ can be approximated to as a random variable that follows a Normal-Lognormal distribution [67, 68, 69].

The random variable Z is said to follow a Normal-Lognormal (NLN) distribution if $Z = Xe^Y$, where the two random variables X and Y are jointly Gaussian random variables, i.e., [67]

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \rho_{XY}\sigma_X\sigma_Y \\ \rho_{XY}\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}\right) ,$$

where ρ_{XY} is the correlation coefficient between X and Y. From Chen's work we know that [67]

$$\mathsf{cov}\left(Z\right) = \mathsf{cov}\left(Xe^Y\right) = \rho\sigma_X\sigma_Y \exp\left[\mu_Y + \frac{\sigma_Y^2}{2}\right]$$

where ρ is the correlation coefficient between X and $\ln(Y)$ Thus, the pairwise covariances in (2.23) can be approximated by [67]

$$\mathsf{cov}\left(\theta_{j,k}, e^{-\theta_{j,3}u}\right) = \rho_{k,3}\sigma_{j,k}u\sigma_{j,3} \exp\left[-u\mu_{j,3} + u^2\frac{\sigma_{j,3}^2}{2}\right]$$

where $\rho_{k,3}$ is the estimated correlation coefficient between $\theta_{j,k}$ and $\ln(\theta_{j,3})$.

Replacing everything back into (2.23), the second term of the ACF is

$$\begin{aligned} \mathsf{E}\left[\theta_{j,1}\theta_{j,2}e^{-\theta_{j,3}u}\right] &= \mu_{j,1}\mu_{j,2} \exp\left[-u\mu_{j,3} + u^2\frac{\sigma_{j,3}^2}{2}\right] \\ &\quad + \mu_{j,1}\rho_{2,3}\sigma_{j,2}u\sigma_{j,3} \exp\left[-u\mu_{j,3} + u^2\frac{\sigma_{j,3}^2}{2}\right] \\ &\quad + \mu_{j,2}\rho_{1,3}\sigma_{j,1}u\sigma_{j,3} \exp\left[-u\mu_{j,3} + u^2\frac{\sigma_{j,3}^2}{2}\right] \\ &\quad + \mathsf{cov}\left(\theta_{j,1}, \theta_{j,2}\right) \exp\left[-u\mu_{j,3} + u^2\frac{\sigma_{j,3}^2}{2}\right] , \end{aligned} \tag{2.25}$$

where $\mathsf{cov}\left(\theta_{j,1}, \theta_{j,2}\right)$ can be estimated from the data.

Using the same procedure, the third term can be expressed as

$$\begin{aligned}
\mathsf{E}\left[\theta_{j,1}\theta_{j,4}e^{-\theta_{j,5}u}\right] &= \mu_{j,1}\mu_{j,4}\exp\left[-u\mu_{j,5}+u^2\frac{\sigma_{j,5}^2}{2}\right] \\
&\quad +\mu_{j,1}\rho_{4,5}\sigma_{j,4}u\sigma_{j,5}\exp\left[-u\mu_{j,5}+u^2\frac{\sigma_{j,5}^2}{2}\right] \\
&\quad +\mu_{j,4}\rho_{1,5}\sigma_{j,1}u\sigma_{j,5}\exp\left[-u\mu_{j,5}+u^2\frac{\sigma_{j,5}^2}{2}\right] \\
&\quad +\mathsf{cov}\left(\theta_{j,1},\theta_{j,4}\right)\exp\left[-u\mu_{j,5}+u^2\frac{\sigma_{j,5}^2}{2}\right] . \qquad (2.26)
\end{aligned}$$

Similarly, the fourth term in (2.13) is

$$\begin{aligned}
\mathsf{E}\left[\theta_{j,1}\theta_{j,2}e^{-\theta_{j,3}t}\right] &= \mu_{j,1}\mu_{j,2}\exp\left[-t\mu_{j,3}+t^2\frac{\sigma_{j,3}^2}{2}\right] \\
&\quad +\mu_{j,1}\rho_{2,3}\sigma_{j,2}t\sigma_{j,3}\exp\left[-t\mu_{j,3}+t^2\frac{\sigma_{j,3}^2}{2}\right] \\
&\quad +\mu_{j,2}\rho_{1,3}\sigma_{j,1}t\sigma_{j,3}\exp\left[-t\mu_{j,3}+t^2\frac{\sigma_{j,3}^2}{2}\right] \\
&\quad +\mathsf{cov}\left(\theta_{j,1},\theta_{j,2}\right)\exp\left[-t\mu_{j,3}+t^2\frac{\sigma_{j,3}^2}{2}\right] . \qquad (2.27)
\end{aligned}$$

The fifth term in (2.13) is

$$\begin{aligned}
\mathsf{E}\left[{\theta_{j,2}}^2e^{-\theta_{j,3}(t+u)}\right] &= \mu_{j,2}^2\exp\left[-(t+u)\mu_{j,3}+(t+u)^2\frac{\sigma_{j,3}^2}{2}\right] \\
&\quad +2\mu_{j,2}\rho_{2,3}\sigma_{j,2}(t+u)\sigma_{j,3}\exp\left[-(t+u)\mu_{j,3}+(t+u)^2\frac{\sigma_{j,3}^2}{2}\right] \\
&\quad +\mathsf{cov}\left(\theta_{j,2}^2\right)\exp\left[-(t+u)\mu_{j,3}+(t+u)^2\frac{\sigma_{j,3}^2}{2}\right] . \qquad (2.28)
\end{aligned}$$

The seventh term in (2.13) is

$$\begin{aligned}
\mathsf{E}\left[\theta_{j,1}\theta_{j,4}e^{-\theta_{j,5}t}\right] &= \mu_{j,1}\mu_{j,4}\exp\left[-t\mu_{j,5}+t^2\frac{\sigma_{j,5}^2}{2}\right] \\
&\quad +\mu_{j,1}\rho_{4,5}\sigma_{j,4}t\sigma_{j,5}\exp\left[-t\mu_{j,5}+t^2\frac{\sigma_{j,5}^2}{2}\right] \\
&\quad +\mu_{j,4}\rho_{1,5}\sigma_{j,1}t\sigma_{j,5}\exp\left[-t\mu_{j,5}+t^2\frac{\sigma_{j,5}^2}{2}\right] \\
&\quad +\mathsf{cov}\left(\theta_{j,1},\theta_{j,4}\right)\exp\left[-t\mu_{j,5}+t^2\frac{\sigma_{j,5}^2}{2}\right]. \qquad (2.29)
\end{aligned}$$

The ninth term in (2.13) is

$$\begin{aligned}
\mathsf{E}\left[{\theta_{j,4}}^2e^{-\theta_{j,5}(t+u)}\right] &= \mu_{j,4}^2\exp\left[-(t+u)\mu_{j,5}+(t+u)^2\frac{\sigma_{j,5}^2}{2}\right] \\
&\quad +2\mu_{j,4}\rho_{4,5}\sigma_{j,4}(t+u)\sigma_{j,5}\exp\left[-(t+u)\mu_{j,5}+(t+u)^2\frac{\sigma_{j,5}^2}{2}\right] \\
&\quad +\mathsf{cov}\left(\theta_{j,4}^2\right)\exp\left[-(t+u)\mu_{j,5}+(t+u)^2\frac{\sigma_{j,5}^2}{2}\right]. \qquad (2.30)
\end{aligned}$$

Now we turn our attention to the case of four random variables within the ex-

pectation. The sixth term in (2.13) is

$$
\begin{aligned}
\mathsf{E}\left[\theta_{j,2}e^{-\theta_{j,3}t}\theta_{j,4}e^{-\theta_{j,5}u}\right] &= \mathsf{E}\left[XYUV\right] \\
&= \mathsf{E}\left[X\right]\mathsf{E}\left[Y\right]\mathsf{cov}\left(U,V\right) + \mathsf{E}\left[X\right]\mathsf{E}\left[U\right]\mathsf{cov}\left(Y,V\right) \\
&\quad +\mathsf{E}\left[X\right]\mathsf{E}\left[V\right]\mathsf{cov}\left(Y,U\right) + \mathsf{E}\left[Y\right]\mathsf{E}\left[U\right]\mathsf{cov}\left(X,V\right) \\
&\quad +\mathsf{E}\left[Y\right]\mathsf{E}\left[V\right]\mathsf{cov}\left(X,U\right) + \mathsf{E}\left[U\right]\mathsf{E}\left[V\right]\mathsf{cov}\left(X,Y\right) \\
&\quad +\mathsf{cov}\left(X,Y\right)\mathsf{cov}\left(U,V\right) + \mathsf{cov}\left(X,U\right)\mathsf{cov}\left(Y,V\right) \\
&\quad +\mathsf{cov}\left(X,V\right)\mathsf{cov}\left(Y,U\right) + \mathsf{E}\left[X\right]\mathsf{E}\left[Y\right]\mathsf{E}\left[U\right]\mathsf{E}\left[V\right] \\
&= \mu_{j,2}\exp\left[-t\mu_{j,3} + t^2\frac{\sigma_{j,3}^2}{2}\right]\mathsf{cov}\left(\theta_{j,4}e^{-\theta_{j,5}u}\right) \\
&\quad +\mu_{j,2}\mu_{j,4}\mathsf{cov}\left(e^{-\theta_{j,3}t}, e^{-\theta_{j,5}u}\right) \\
&\quad +\mu_{j,2}\exp\left[-u\mu_{j,5} + u^2\frac{\sigma_{j,5}^2}{2}\right]\mathsf{cov}\left(e^{-\theta_{j,3}t}, \theta_{j,4}\right) \\
&\quad +\mu_{j,4}\exp\left[-t\mu_{j,3} + t^2\frac{\sigma_{j,3}^2}{2}\right]\mathsf{cov}\left(\theta_{j,2}, e^{-\theta_{j,5}u}\right) \\
&\quad +\exp\left[-t\mu_{j,3} + t^2\frac{\sigma_{j,3}^2}{2}\right]\exp\left[-u\mu_{j,5} + u^2\frac{\sigma_{j,5}^2}{2}\right]\mathsf{cov}\left(\theta_{j,2}, \theta_{j,4}\right) \\
&\quad +\mu_{j,4}\exp\left[-u\mu_{j,5} + u^2\frac{\sigma_{j,5}^2}{2}\right]\mathsf{cov}\left(\theta_{j,2}e^{-\theta_{j,3}t}\right) \\
&\quad +\mathsf{cov}\left(\theta_{j,2}, e^{-\theta_{j,3}t}\right)\mathsf{cov}\left(\theta_{j,4}, e^{-\theta_{j,5}u}\right) \\
&\quad +\mathsf{cov}\left(\theta_{j,2}, \theta_{j,4}\right)\mathsf{cov}\left(e^{-\theta_{j,3}t}, e^{-\theta_{j,5}u}\right) \\
&\quad +\mathsf{cov}\left(\theta_{j,2}, e^{-\theta_{j,5}u}\right)\mathsf{cov}\left(e^{-\theta_{j,3}t}, \theta_{j,4}\right) \\
&\quad +\mu_{j,2}\exp\left[-t\mu_{j,3} + t^2\frac{\sigma_{j,3}^2}{2}\right]\mu_{j,4}\exp\left[-u\mu_{j,5} + u^2\frac{\sigma_{j,5}^2}{2}\right] .
\end{aligned}
$$

Here, all the terms are already defined in the previous paragraphs except for the term $\mathsf{cov}\left(e^{-\theta_{j,3}t}, e^{-\theta_{j,5}u}\right)$, which we define next. Under the assumption of Gaussian random variables, with $\theta_{j,k} \sim \mathcal{N}(\mu_{j,k}, \sigma_{j,k}^2)$, $k = 3, 5$, the random variables $e^{\theta_{j,k}t}$ are log-Normal random variables and the covariance between them is given by the

formulae

$$\begin{aligned}\mathsf{cov}\left(e^{-\theta_{j,3}t}, e^{-\theta_{j,5}u}\right) &= \mathsf{E}\left[e^{-\theta_{j,3}t}\right]\mathsf{E}\left[e^{-\theta_{j,5}u}\right]\left(e^{\mathsf{cov}(\theta_{j,3},\theta_{j,5})}-1\right) \\ &= \exp\left[-t\mu_{j,3}+t^2\frac{\sigma_{j,3}^2}{2}\right]\exp\left[-u\mu_{j,5}+u^2\frac{\sigma_{j,5}^2}{2}\right]\left(e^{\mathsf{cov}(\theta_{j,3},\theta_{j,5})}-1\right),\end{aligned} \tag{2.31}$$

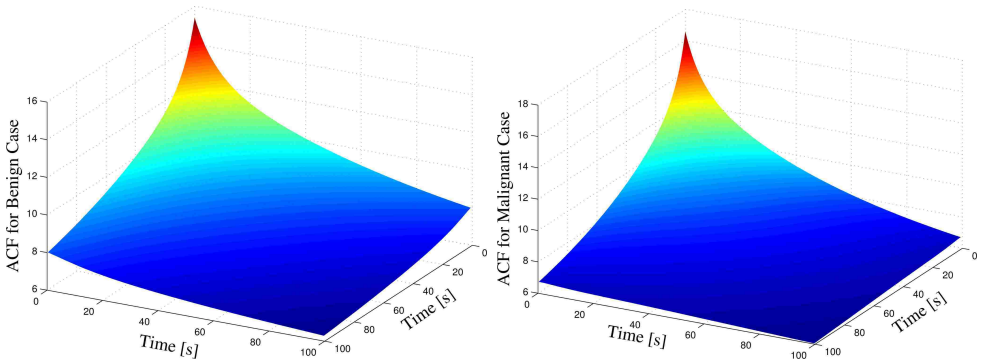
where $\mathsf{cov}\left(\theta_{j,3},\theta_{j,5}\right)$ can be estimated from the data. Replacing this expression back in (2.31) we obtain the expression to compute the sixth term of (2.13). The eigth term is computed exactly in the same way, making the appropriate changes in the variables as done for the other terms. Using all the described assumptions and derived expressions we can define the analytical ACF by adding up (2.25)-(2.31) The resulting ACF from the parameters of patients with known benign and malignant conditions are depicted in Fig. 2.7(a) and Fig. 2.7(b), respectively. Please recall that benign conditions are considered to define the null hypothesis (H_0) and that the malignant conditions (including MM, BCC and SCC conditions) are considered to define the alternative hypothesis (H_1). In both cases, the acquisition time was set to be 100 seconds in order to avoid small differences that affected the image registration over the actual 120 seconds used as the acquisition time.

2.6.2 Mercer's theorem

Let us assume that the stochastic signals under each hypothesis are zero-mean second-order random processes. Therefore, the jth ACF ($j = 0, 1$) can be expanded by the absolutely convergent series (Mercer's theorem) [70]

$$\mathsf{R}_j(t,u) = \sum_{k=1}^{\infty}\lambda_{j,k}\phi_{j,k}(t)\phi_{j,k}(u)\ , \quad (t,u)\in[0,T]^2\ , \tag{2.32}$$

where $\{\lambda_{j,k}\}_{k=1}^{\infty}$ and $\{\phi_{j,k}\}_{k=1}^{\infty}$ are the eigenvalues and the corresponding orthonormal eigenfunctions of the jth ACF, $\mathsf{R_j}$. The eigenvalues and eigenfunctions are the



(a) (b)

Figure 2.7: (a) Autocorrelation function for the null-hypothesis (H_0) estimated from patient data with known benign condition. (b) Autocorrelation function for the alternative-hypothesis (H_1) estimated from patient data with known malignant condition.

solutions of the integral equation

$$\lambda_{j,k}\phi_{j,k}(t) = \int_0^T \mathsf{R}_j(t,u)\phi_{j,k}(u)\ du\ , \quad t \in [0,T]\ , \tag{2.33}$$

with $\int_0^T \phi_{j,k}(t)\phi_{j,\ell}(t) = \delta_{k,\ell}$, where $\delta_{k,\ell}$ is the Kronecker delta. This equation is known as a Fredholm integral equation of the second kind. The expansion (2.32) is known as the Mercer's theorem and it is the key enabling theorem to solve our problem. Details on how we solve (2.33) are given later in Subsection 2.6.4, whereas other alternatives to solve this equation numerically can be found, for example in [71].

2.6.3 Karhunen-Loève expansion of the thermal recovery curves

The two sets of eigenfunctions, namely $\{\phi_{0,k}\}_{k=1}^{\infty}$ and $\{\phi_{1,k}\}_{k=1}^{\infty}$, are two complete sets because the corresponding ACFs, R_0 and R_1, respectively, are symmetric and positive definite [70, 72]. The completeness of these two sets allow us to represent any process with any of these sets (in the mean-square sense). We represent each

signal with its corresponding set, namely, for the jth hypothesis

$$S_j(t;\boldsymbol{\Theta}_j) = \sum_{k=1}^{\infty} S_{j,k}\phi_{j,k}(t) \ , \quad t \in [0,T] \ , \tag{2.34}$$

where the expansion coefficients $S_{j,k}$ are known as the Karhunen-Loève (KL) coefficients associated with the stochastic process $S_j(t;\boldsymbol{\Theta}_j)$. The KL coefficients are computed as the projections of each process on its respective basis functions, namely

$$S_{j,k} = \int_0^T \phi_{j,k}(t) S_j(t;\boldsymbol{\Theta}_j)\ dt \ , \quad \begin{matrix} k = 1,2,\ldots \\ j = 0,1 \end{matrix} \ . \tag{2.35}$$

One of the main properties related with the KL expansion for random processes is that the expansion coefficients have zero mean and are uncorrelated. More details can be found elsewhere [70] and are included here for completeness. The mean of the coefficients is $\mathsf{E}\left[S_{j,k}\right] = \mathsf{E}\left[\int_0^T \phi_{j,k}(t)S_j(t;\boldsymbol{\Theta}_j)\ dt\right] = \int_0^T \phi_{j,k}(t)\mathsf{E}\left[S_j(t;\boldsymbol{\Theta}_j)\right]\ dt = 0$, owing the assumption the our random processes have zero mean. The correlation of these coefficients is on the other hand,

$$\begin{aligned} \mathsf{cov}\left(S_{j,k}, S_{j,\ell}\right) &= \mathsf{E}\left[S_{j,k}S_{j,\ell}\right] = \mathsf{E}\left[\int_0^T \phi_{j,k}(t)S_j(t;\boldsymbol{\Theta}_j)\ dt \int_0^T \phi_{j,\ell}(u)S_j(u;\boldsymbol{\Theta}_j)\ du\right] \\ &= \int_0^T \phi_{j,k}(t) \int_0^T \mathsf{E}\left[S_j(t;\boldsymbol{\Theta}_j)S_j(u;\boldsymbol{\Theta}_j)\right]\phi_{j,\ell}(u)\ du\ dt \\ &= \int_0^T \phi_{j,k}(t) \int_0^T \mathsf{R}_j(t,u)\phi_{j,\ell}(u)\ du\ dt = \lambda_\ell \delta_{k,\ell} \ . \end{aligned} \tag{2.36}$$

The KL expansion enables us to conveniently decouple randomness (compactly contained in the KL coefficients, $S_{j,k}$) and time-variations (embodied in the sequence of eigenfunctions, $\phi_{j,k}(t)$) for the TRCs under each hypothesis. As such, the KL expansion enables us to equivalently view the continuous-time TRC stochastic signals $S_j(t;\boldsymbol{\Theta}_j)$ for $j = 0, 1$, as sequences of uncorrelated random variables, namely the KL coefficients $S_{j,k}$, for $j = 0, 1$ and $k = 1, 2, \ldots$. These two sequences of KL coefficients in effect constitute the set of statistical features that fully describe the TRC for each patient under each hypothesis. With such statistical equivalence between a TRC and

its KL sequences, we can employ optimal-inference theory to announce the hypothesis based on the KL coefficients of an observed TRC. The established procedure for doing so is to construct the likelihood ratio associated with the equivalent KL representation, as detailed in Section 2.7. Next we describe how we have obtained the eigenfunctions and eigenvalues for both hypotheses.

2.6.4 Numerical solution of the Fredholm integral equation of the second kind

For this study, we solve (2.33) numerically following the *Nystrom method* which is explained below [73]. The method requires the choice of some approximation for the integral by means of a quadrature rule,

$$\int_0^T y(u)\, du \approx \sum_{\ell=0}^{N} w_\ell y(u_\ell) = \boldsymbol{w}^T \boldsymbol{y} \ , \tag{2.37}$$

where $\boldsymbol{w} = [w_1 \ w_2 \ \cdots \ w_N]^T$ and $\boldsymbol{y} = [y(u_1) \ y(u_2) \ \cdots \ y(u_N)]^T$, with the points $\{u_\ell\}$ as samples points in time within a grid and the weight $\{w_\ell\}$ are those from the quadrature approximation. For example, under an uniform grid of N points the trapezoidal rule for numerical integration is

$$\int_0^T y(u)\, du \approx \frac{T}{2N}\left[y(u_0) + 2y(u_1) + 2y(u_1) + \cdots + 2y(u_{N-1}) + y(u_N)\right] \ ,$$

thus the vector of weights can be $\boldsymbol{w} = \frac{T}{2N}[1 \ 2 \ 2 \ \cdots \ 2 \ 1]$. Now, applying the quadrature rule (2.37) to the equation (2.33) we get

$$\lambda_{j,k}\phi_{j,k}(t) \approx \sum_{\ell=0}^{N} w_\ell \mathsf{R}_j(t, u_\ell)\phi_{j,k}(u_\ell) \ ,$$

which can be evaluated at the quadrature points to get

$$\lambda_{j,k}\phi_{j,k}(t_i) \approx \sum_{\ell=0}^{N} w_\ell \mathsf{R}_j(t_i, u_\ell)\phi_{j,k}(u_\ell) \ , \quad i = 0, 1, 2, \ldots, N \ .$$

The last equation is a linear system of N equations with N unknown, which in vector form becomes

$$\lambda_{j,k}\boldsymbol{\Phi}_{j,k} = \boldsymbol{R}_j^*\boldsymbol{\Phi}_{j,k} \ , \tag{2.38}$$

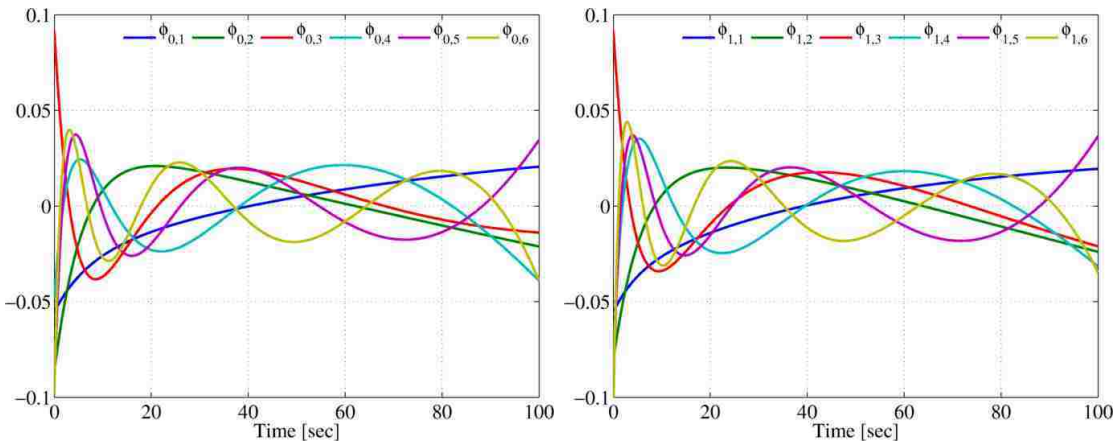
where

$$\boldsymbol{\Phi}_{j,k} \triangleq \begin{bmatrix} \phi_{j,k}(t_0) \\ \phi_{j,k}(t_1) \\ \vdots \\ \phi_{j,k}(t_N) \end{bmatrix} \ , \tag{2.39}$$

and

$$\boldsymbol{R}_j^* \triangleq \begin{bmatrix} \mathsf{R}_j(t_0,u_0)w_0 & \mathsf{R}_j(t_0,u_1)w_1 & \cdots & \mathsf{R}_j(t_0,u_N)w_N \\ \mathsf{R}_j(t_1,u_0)w_0 & \mathsf{R}_j(t_1,u_1)w_1 & \cdots & \mathsf{R}_j(t_1,u_N)w_N \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{R}_j(t_N,u_0)w_0 & \mathsf{R}_j(t_N,u_1)w_1 & \cdots & \mathsf{R}_j(t_N,u_N)w_N \end{bmatrix} . \tag{2.40}$$

Note that since the weights $\{w_\ell\}$ are not equal for most quadrature rules, the matrix $\boldsymbol{R}_j^*$ in (2.38) is, in general, not symmetric; indeed, if we define $\boldsymbol{W} = diag(\boldsymbol{w})$, then $\boldsymbol{R}_j^* = \boldsymbol{R}_j\boldsymbol{W}$, where the entries of the matrix $\boldsymbol{R}_j$ are $R_j(n,m) = \mathsf{R}_j(t_n,u_m)$. It is important to mention that this method will, in general, give us N eigenvalues (recall that N is the number of quadrature points used). It is stated by Press *et al.* [73] that for square-integrable kernels, as in our case, this method provides a good approximations to the N eigenvalues of the integral equation (2.33). For other alternatives to numerically solve (2.33) see, for example, Delves and Mohamed [74] or Chen *et al.* [71]. We solve (2.38) with MatLab's `eig` function for the same patient dataset utilized to compute the ACFs presented in Fig. 2.7. The resulting, six most important eigenfunctions (provided that we have sorted the eigenvalues in descending order) for the null and alternative hypotheses are depicted in Fig. 2.8(a) and Fig. 2.8(b), respectively.



(a) (b)

Figure 2.8: First six, most important eigenfunctions for both hypotheses, (a) null hypothesis and (b) alternative hypothesis. The importance of each eigenfunction is determined by the value of its corresponding eigenvalue.

In theory, each autocorrelation function can be expanded by its eigenvalues and eigenfunctions which form a complete orthonormal set [72]. Since each ACF create a complete basis function set, both sets can be used to compute the KL coefficients. Nevertheless, since the ACFs are *estimated* from a small number of patient data by computing the means and pairwise covariances between the model parameters they are not "rich enough" to allow us to estimate a big number of eigenvalues. As such, there is a practical limit in the number of eigenpairs that can be reliably extracted from the estimated ACFs. (The variability of the eigenvalue-eigenfunction pairs is studied in detail in Section 3.4.1.) This practical limit in the number of eigenvalues force us to limit the KL expansions to the most important coefficients as discussed later in the construction of the likelihood ratio.

Table 2.3: Eigenvalues sorted by their magnitude, corresponding to those eigenfunctions depicted in Fig. 2.8

k	1	2	3	4	5	6
$\lambda_{0,k}$	955.47	38.89	2.24	0.55	0.02	7.47×10^{-4}
$\lambda_{1,k}$	1609.30	44.69	5.70	1.22	0.03	2.25×10^{-3}

2.7 Likelihood ratio

The likelihood ratio is an explicit function of all the KL coefficients and all the eigenvalues and it is constructed as follows. With the statistical equivalence between a TRC and its KL sequences, we can recast the detection problem (2.12) as [75]

$$H_0 \ : \ S_{0,k} \ , \quad k = 1, 2, \ldots \tag{2.41a}$$

$$H_1 \ : \ S_{1,k} \ , \quad k = 1, 2, \ldots \tag{2.41b}$$

The main difficulty at this stage is the application of Grenander's theorem as we require to have the probability density function of the KL coefficients under each hypothesis and for each $k = 1, 2, \ldots$. As explained by Poor [70], with the exception of some particular cases, it is extremely difficult, if not impossible, to find even the marginal density of one of the KL coefficient. Thus without further assumptions, one cannot proceed to find likelihood ratios in this way. An important exception to this difficulty is when the random processes $S_j(t; \boldsymbol{\Theta}_j)$ are Gaussian. A process, say $S_j(t; \boldsymbol{\Theta}_j)$, is a Gaussian random process if every linear functional of $S_j(t; \boldsymbol{\Theta}_j)$ is a Gaussian random variable [72]. Namely, if $S_j = \int_0^T g(u) S_j(u; \boldsymbol{\Theta}_j) \, du$, where $g(u)$ is any function such that $\mathsf{E}\left[S_j^2\right] < \infty$ and S_j is a Gaussian random variable, then the process $S_j(t; \boldsymbol{\Theta}_j)$ is Gaussian. Since integration is a linear operation, if we assume that $S_j(t; \boldsymbol{\Theta}_j)$ is a Gaussian random process, then the KL coefficients will be also Gaussian and the density under each hypothesis can be easily specified by the mean and the covariances of the KL coefficients, $\mathsf{E}\left[S_{j,k}\right] = 0$ and $\mathsf{cov}\left(S_{j,k} S_{j,\ell}\right) = \lambda_k \delta_{k,\ell}$, respectively. Since we are assuming the processes are Gaussian and the KL coefficients are uncorrelated, they also are independent random variables, with $S_{j,k} \sim \mathcal{N}\left(0, \lambda_{j,k}\right)$; thus, the original continuous-time detection problem becomes the discrete (but infinite) detection problem between two Gaussian distributions with different (diagonal) covariance matrices; hence, for the observation process $Y(t)$, we have the

discrete detection problem

$$H_0 \ : Y_{0,k} \sim \mathcal{N}(0, \lambda_{0,k}) \ , \quad k = 1, 2, \ldots \tag{2.42a}$$

$$H_1 \ : Y_{1,k} \sim \mathcal{N}(0, \lambda_{1,k}) \ , \quad k = 1, 2, \ldots \tag{2.42b}$$

where the KL coefficients of the observation $Y(t)$, namely $Y_{j,k}$, $j = 0, 1$ are the expansion coefficients of $Y(t)$ under the jth hypothesis, namely $Y_{j,k} = \int_0^T \phi_{j,k}(t) Y(t) \ dt$, $j = 0, 1$. This discrete detection problem have a likelihood ratio defined by

$$\begin{aligned} L(Y) \ &\triangleq \ \frac{p_1(Y)}{p_0(Y)} = \frac{\prod_{k=1}^{\infty} \frac{1}{\sqrt{2\pi\lambda_{1,k}}} \exp\left[-\frac{1}{2}\frac{Y_{1,k}^2}{\lambda_{1,k}}\right]}{\prod_{k=1}^{\infty} \frac{1}{\sqrt{2\pi\lambda_{0,k}}} \exp\left[-\frac{1}{2}\frac{Y_{0,k}^2}{\lambda_{0,k}}\right]} \\ &= \ \prod_{k=1}^{\infty} \left(\frac{\lambda_{0,k}}{\lambda_{1,k}}\right)^{1/2} \exp\left[\frac{1}{2}\sum_{k=1}^{\infty}\left(\frac{Y_{0,k}^2}{\lambda_{0,k}} - \frac{Y_{1,k}^2}{\lambda_{1,k}}\right)\right] \end{aligned} \tag{2.43}$$

where Y denotes $\{Y_k\}_{k=1}^{\infty}$, the vector that contains all the KL coefficients. By inspection, one can see that the first restriction required to ensure the convergence of (2.43) is that $\lambda_{1,k} > \lambda_{0,k}$ for $k = 1, 2, \ldots$, and the convergence of the second term is ensured if

$$\sum_{k=1}^{\infty}\left(\frac{Y_{0,k}^2}{\lambda_{0,k}} - \frac{Y_{1,k}^2}{\lambda_{1,k}}\right) < \infty \ ,$$

because the logarithm function is monotonic. The convergence in mean-square of each term within the summation can be proven by following the same procedure as in Poor [70] (pp. 305-306) by letting $\hat{X}_k^2 = Y_{j,k}^2$ and $\lambda_k = \lambda_{j,k}$, for $j = 0, 1$ in Equation (VI.D.20), and will not be shown here.

2.7.1 Test statistic for the detection problem

The dependence of the likelihood ratio on the KL coefficients and the eigenvalues is further simplified to produce the test-statistic, and the latter is compared to a single pre-specified threshold and the result of the comparison is used to declare the

hypothesis associated with the TRC, as studied next for a Neyman-Pearson (NP) decision rule. Let us denote the optimal threshold determined by the NP decision rule by η.

The test statistics is obtained as usual by separating the terms that depend on the KL coefficients of the observation process and letting the remaining terms be absorbed by the threshold. We take logarithm in (2.43) we obtain

$$\begin{aligned} \log L(Y) &= \log\left[\prod_{k=1}^{\infty}\left(\frac{\lambda_{0,k}}{\lambda_{1,k}}\right)^{1/2}\right] + \frac{1}{2}\sum_{k=1}^{\infty}\left(\frac{Y_{0,k}^2}{\lambda_{0,k}} - \frac{Y_{1,k}^2}{\lambda_{1,k}}\right) \\ &= \frac{1}{2}\sum_{k=1}^{\infty}\log\left[\frac{\lambda_{0,k}}{\lambda_{1,k}}\right] + \frac{1}{2}\sum_{k=1}^{\infty}\left(\frac{Y_{0,k}^2}{\lambda_{0,k}} - \frac{Y_{1,k}^2}{\lambda_{1,k}}\right) \end{aligned} \tag{2.44}$$

Now, the test statistics, Z is defined by

$$Z = \sum_{k=1}^{\infty}\left(\frac{Y_{0,k}^2}{\lambda_{0,k}} - \frac{Y_{1,k}^2}{\lambda_{1,k}}\right) \gtrless \eta \ . \tag{2.45}$$

where the threshold η must be determined under an optimal prescribed decision rule.

As was discussed before, there is a practical limit in the number of eigenpairs one can reliably extract from the estimated ACFs. The KL expansion offers the optimality-under-truncation property, i.e., the mean-square error resulting from a finite representation of the process is minimized [72]. Such a property allow us to still optimally represent our processes and our resulting test-statistics, when the most important eigenpairs are used (i.e., those corresponding to the eigenvalues with the highest value). As a consequence, we define the truncated version of the test-statistics as

$$Z^{(K)} = \sum_{k=1}^{K}\left(\frac{Y_{0,k}^2}{\lambda_{0,k}} - \frac{Y_{1,k}^2}{\lambda_{1,k}}\right) \ , \tag{2.46}$$

where the superscript $^{(K)}$ means that the first K KL coefficients and eigenvalues of each hypothesis where used to define the test statistics. The truncated version

of the test statistics is used later to implement the algorithm. Next, by using the theoretical test statistics we define our optimal decision rule.

2.8 Neyman-Pearson decision rule

In this section, we describe how to optimally define the threshold, η such that detection probability is maximized for a fix, prescribed false-alarm probability. The procedure to do so, is to first obtain the distribution of the test-statistics under each hypothesis. Later, the optimal decision rule is obtained as a function of this distributions as described in the following sections.

2.8.1 Neyman-Pearson lemma

Let us state the Neyman-Pearson (NP) lemma to set the baseline for the upcoming discussion. Assume that we observe a random variable distributed according to one of two distributions, namely $H_j \ : \ Y \sim p_j$, for $j = 0, 1$. Consider the likelihood ratio test

$$L(Y) \triangleq \frac{p_1(Y)}{p_0(Y)} \underset{H_0}{\overset{H_1}{\gtrless}} \tau \ ,$$

with $\tau > 0$ chosen so that the false alarm probability $P_F = \Pr(L(Y) > \tau | H_0) \leq \alpha$. In simple words, the false-alarm probability is the probability of declaring H_1 given that H_0 is true. The detection probability (the probability of correctly declaring H_1 when H_1 is true) is defined by $P_D = \Pr(L(Y) > \tau | H_1)$. The NP lemma states that does not exist another test with $P_F \leq \alpha$ such that the P_D is higher that the one achieved by the likelihood ratio test (LRT). That is, the LRT is the most powerful test with probability of false-alarm less than or equal to α. The proof of this lemma can be found, for example, in [70] and will be omitted here. It is clear that the definitions of P_F and P_D require us to know the distribution of either the likelihood

ratio or the test-statistics, which is what we address next.

2.8.2 Distribution of the test-statistics under each hypothesis

If H_0 is true, we know that the incoming signal, $Y(t)$, will be equal to $S_0(t;\boldsymbol{\Theta}_0)$. Thus, the KL coefficients in (2.45) are

$$\begin{aligned} Y_{0,k} &= \int_0^T \phi_{0,k}(t)Y(t)\ dt = \int_0^T \phi_{0,k}(t)S_0(t;\boldsymbol{\Theta}_0)\ dt \\ &= \int_0^T \phi_{0,k}(t)\left[\sum_{\ell=1}^{\infty} S_{0,\ell}\phi_{0,\ell}(t)\right]\ dt = \sum_{\ell=1}^{\infty} S_{0,\ell}\int_0^T \phi_{0,k}(t)\phi_{0,\ell}(t)\ dt \\ &= \sum_{\ell=1}^{\infty} S_{0,\ell}\delta_{k,\ell} = S_{0,k}\ , \end{aligned} \tag{2.47}$$

where we have used the KL expansion of $S_0(t;\boldsymbol{\Theta}_0)$ and the orthonormality of the eigenfunction set $\{\phi_{0,k}\}_{k=1}^{\infty}$. The other KL coefficient is

$$\begin{aligned} Y_{1,k} &= \int_0^T \phi_{1,k}(t)Y(t)\ dt = \int_0^T \phi_{1,k}(t)S_0(t;\boldsymbol{\Theta}_0)\ dt \\ &= \int_0^T \phi_{1,k}(t)\left[\sum_{\ell=1}^{\infty} S_{0,\ell}\phi_{0,\ell}(t)\right]\ dt = \sum_{\ell=1}^{\infty} S_{0,\ell}\int_0^T \phi_{1,k}(t)\phi_{0,\ell}(t)\ dt \\ &= \sum_{\ell=1}^{\infty} S_{0,\ell}\delta_{k,\ell} = S_{0,k}\ , \end{aligned} \tag{2.48}$$

where we have assumed that the sets $\{\phi_{0,k}\}_{k=1}^{\infty}$ and $\{\phi_{1,k}\}_{k=1}^{\infty}$ are cross-orthogonal. Numerically, we have observed this cross-orthogonality for the first 12 eigenfunctions, which are precisely the number of reliable eigenfunctions one can estimate from our data. (We will explore this details carefully in Chapter 3.) The cross-orthogonality (numerically computed from the computed eigenfunctions) can be seen in Fig. 2.9 for a particular combination of patients used as training.

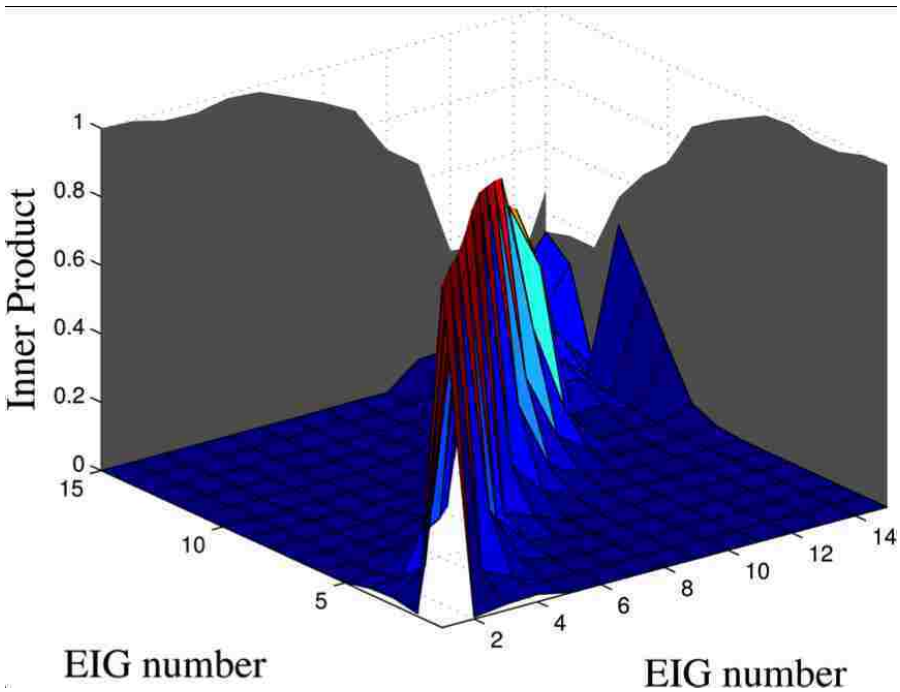


Figure 2.9: Study of the cross-orthogonality between the sets $\{\phi_{0,k}\}_{k=1}^{\infty}$ and $\{\phi_{1,k}\}_{k=1}^{\infty}$ computed numerically for the estimated eigenfunctions

Under this assumptions, therefore, the KL coefficients under H_0 are $Y_{0,k} = Y_{1,k} = S_{0,k}$, $k = 1, 2, \ldots$. As such, the test statistics under H_0, denoted by Z_0, is given by

$$\begin{aligned} Z_0 &= \sum_{k=1}^{\infty} \left(\frac{Y_{0,k}^2}{\lambda_{0,k}} - \frac{Y_{1,k}^2}{\lambda_{1,k}} \right) = \sum_{k=1}^{\infty} \left(\frac{S_{0,k}^2}{\lambda_{0,k}} - \frac{S_{0,k}^2}{\lambda_{1,k}} \right) \\ &= \sum_{k=1}^{\infty} \left(1 - \frac{\lambda_{0,k}}{\lambda_{1,k}} \right) \frac{S_{0,k}^2}{\lambda_{0,k}} \, . \end{aligned} \tag{2.49}$$

Given that we have assumed the processes to be Gaussian, and, as a consequence the KL coefficients are also Gaussian, then the random variables $S_{0,k}^2/\lambda_{0,k}$, for $k = 1, 2, \ldots$ are χ^2-distributed because the variance of the kth KL coefficient $S_{0,k}$ is precisely $\lambda_{0,k}$. Moreover, as was previously discussed, $\lambda_{1,k} > \lambda_{0,k}$, for $k = 1, 2, \ldots$ in order to ensure the convergence of the likelihood ratio; thus, the weights in the summation are positive. If we denote the χ^2-distributed random variables by X_k,

then (2.49) can be recast as

$$Z_0 = \frac{1}{2}\sum_{k=1}^{\infty} a_k X_k \ , \tag{2.50}$$

where $a_k = 2(1 - \lambda_{0,k}/\lambda_{1,k}) > 0$ are the coefficients of a linear combination of χ^2-distributed random variables. (The pdf for this linear combination is discussed later.) Similarly, if H_1 is true, we have that the KL coefficients are

$$\begin{aligned} Y_{0,k} &= \int_0^T \phi_{0,k}(t)Y(t)\ dt = \int_0^T \phi_{0,k}(t)S_1(t;\boldsymbol{\Theta}_0)\ dt \\ &= \int_0^T \phi_{0,k}(t)\left[\sum_{\ell=1}^{\infty} S_{1,\ell}\phi_{1,\ell}(t)\right]\ dt = \sum_{\ell=1}^{\infty} S_{1,\ell}\int_0^T \phi_{0,k}(t)\phi_{1,\ell}(t)\ dt \\ &= \sum_{\ell=1}^{\infty} S_{1,\ell}\delta_{k,\ell} = S_{1,k}\ , \end{aligned}$$

where again we assumed cross-orthogonality of the eigenfunctions, and

$$\begin{aligned} Y_{1,k} &= \int_0^T \phi_{1,k}(t)Y(t)\ dt = \int_0^T \phi_{1,k}(t)S_1(t;\boldsymbol{\Theta}_0)\ dt \\ &= \int_0^T \phi_{1,k}(t)\left[\sum_{\ell=1}^{\infty} S_{1,\ell}\phi_{1,\ell}(t)\right]\ dt = \sum_{\ell=1}^{\infty} S_{1,\ell}\int_0^T \phi_{1,k}(t)\phi_{1,\ell}(t)\ dt \\ &= \sum_{\ell=1}^{\infty} S_{1,\ell}\delta_{k,\ell} = S_{1,k}\ , \end{aligned}$$

thus, the test-statistics under H_1 is given by

$$Z_1 = \sum_{k=1}^{\infty}\left(\frac{S_{1,k}^2}{\lambda_{0,k}} - \frac{S_{1,k}^2}{\lambda_{1,k}}\right) = \sum_{k=1}^{\infty}\left(\frac{\lambda_{1,k}}{\lambda_{0,k}} - 1\right)\frac{S_{1,k}^2}{\lambda_{1,k}} = \frac{1}{2}\sum_{k=1}^{\infty} b_k X_k \ , \tag{2.51}$$

where $b_k = 2(\lambda_{1,k}/\lambda_{0,k} - 1) > 0$ are the coefficients of another linear combination of χ^2-distributed random variables. In summary, the test-statistics under each hypothesis is a linear combination of χ^2-distributed random variables, with different set of positive coefficients. For a finite number of KL coefficients (2.50) and (2.51) are quadratic forms of the Gaussian random variables, $S_{0,k}$ and $S_{1,k}$, $k = 1, 2, \ldots, K$.

(Recall that the KL coefficients are uncorrelated and Gaussian, therefore they are also independent.) The distribution of quadratic forms of i.i.d. Gaussian random variables with positive coefficients was studied by Pachares [76] and his main results are summarized in the Appendix A. Pachares' main result is that a finite linear combination of independent χ^2-distributed random variables with weights $\boldsymbol{a} = [a_1 \ a_2 \ \cdots \ a_K]$,

$$S_r = \frac{1}{2}\left[a_1\chi^2_{m_1} + a_2\chi^2_{m_2} + \cdots + a_K\chi^2_{m_K}\right] ,$$

has a cumulative distribution function (CDF) given by

$$G(\tau;\boldsymbol{a}) = \Pr(S_r \leq \tau) = \sqrt{\frac{\tau^M}{a_1^{m_1}\cdots a_r^{m_K}}}\sum_{k=0}^{\infty}\frac{(-\tau)^k}{k!}\frac{\mathsf{E}\left[S_r^*\right]_k}{\Gamma(M/2+k+1)} , \qquad (2.52)$$

where $M = \sum_i m_i$, $S_r^* = \sum_i a_i^{-1}\chi^2_{m_i}$ and $\mathsf{E}\left[S_r^*\right]_k$ is the kth moment of S_r^* (see (A.1) in Appendix A for more details). We use this CDF in the next section to define the optimum decision rule based on the Neyman-Pearson lemma.

2.8.3 Optimal decision rule

Now we turn back into the NP lemma. At this stage we truncate the number of KL coefficients and see how the performance, i.e., P_F and P_D change with respect to the number of these coefficients. We denote by K the number of eigenvalue-eigenfunction pairs used in the next computations. The false-alarm probability is given by

$$P_F \triangleq \Pr(L(Y) > \tau|H_0) = \Pr(Z_0 > \eta) = 1 - \Pr(Z_0 \leq \eta) \overset{(2.50)}{=} 1 - G(\eta;\boldsymbol{a}) , \quad (2.53)$$

where $\boldsymbol{a} = [a_1 \ a_2 \ \cdots \ a_K]$ are the coefficients from the test statistics, with $a_k = 2(1-\lambda_{0,k}/\lambda_{1,k})$, as defined in (2.49). Similarly, the detection probability is given by

$$P_D \triangleq \Pr(L(Y) > \tau|H_1) = 1 - \Pr(Z_1 \leq \eta) \overset{(2.51)}{=} 1 - G(\eta;\boldsymbol{b}) , \qquad (2.54)$$

where $\boldsymbol{b} = [b_1 \ b_2 \ \cdots \ b_K]$, with $b_k = 2(\lambda_{1,k}/\lambda_{0,k} - 1)$. Now, for a prescribed level of false-alarm, say α, the NP lemma tell us that the optimal threshold, η_0, will be given

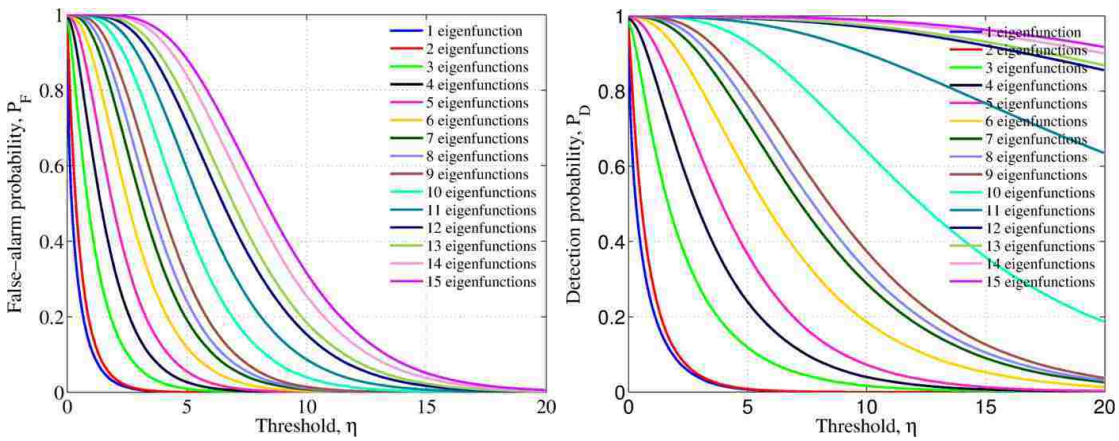


Figure 2.10: False-alarm and detection probabilities parameterized by the threshold value, η, for different number of eigenfunctions used in the construction of the test-statistics (2.46)

by

$$\eta_0 = G^{-1}(1-\alpha;\boldsymbol{a}) \ , \tag{2.55}$$

where G^{-1} represent the inverse of the CDF function (2.52). When using this optimum threshold, by the NP lemma, the detection probability $P_D = 1 - G(\eta_0;\boldsymbol{b})$ is maximum amongst all the other possible test one may design. The complexity of implementing (2.52) makes the implementation of its inverse function an almost impossible task. Therefore, we numerically solve the equivalence by parameterizing the false-alarm and detection probabilities by the threshold, η. Figure 2.10 depicts the how the false-alarm and detection probabilities are parameterized by η for different number of used eigenfunctions. (Recall that the number of eigenfunctions determine the number of weights $\{a_k\}$ and $\{b_k\}$ that we use in the function (2.52).)

Once the desired false-alarm probability is specified, say $P_F = \alpha$, the parameterized optimum threshold, η_0 can be obtained from Fig. 2.10 and (2.55). Such an optimum threshold is later used to classify patient data by comparing the test-statistic of a patient with unknown diagnosis: if the test statistic exceeds the optimum threshold

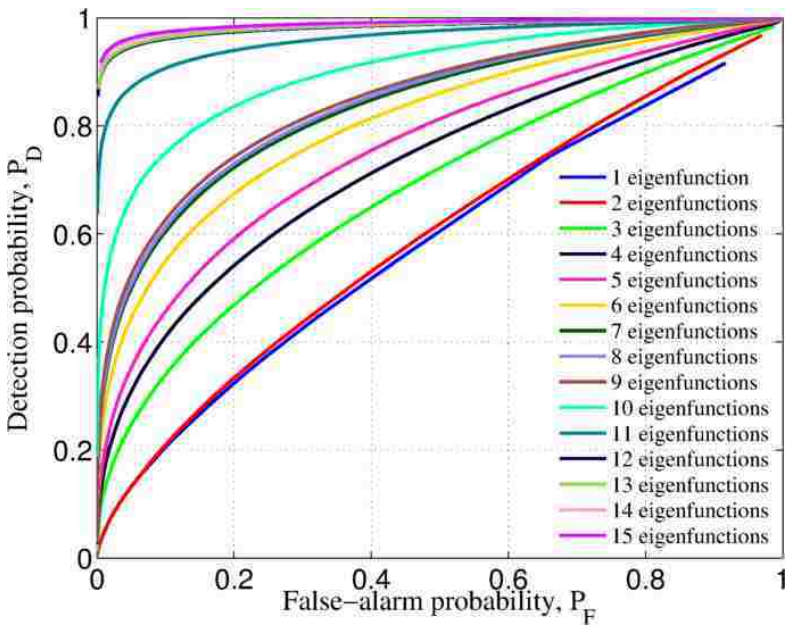


Figure 2.11: The theoretical receiver-operating characteristic (ROC) curve graphically shows the expected performance of the detector as we increase the number of eigenvalue-eigenfunction pairs. The bigger the number of the pairs utilized to construct the test-statistics, the more statistical features utilized and the better the performance of the algorithm

then is classified as malignant. More details are given in the following section.

With the parameterized false-alarm and detection probabilities, one can construct the so-called receiver-operating characteristic (ROC) curve, a standard measure of the decision-rule performance that depicts the direct relationship between the theoretical false-alarm probability and the corresponding (theoretical) detection probability. The ROC curve of a perfect classifier correspond to a line that achieves 100% detection (sensitivity) for any value of the false-alarm probability (i.e., $P_D = 1.0$ for $P_F \in (0, 1)$). We show the ROC curve corresponding to the parameterized probabilities in Fig. 2.11, where it can be noted that, as expected, as we include more eigenvalue-eigenfunction pairs, more features are extracted from the TRC and the theoretical performance is improved, plateauing at a level, say at the Kth pair. The

Table 2.4: Computed area under the ROC curve for the presented ROC curves. The plateau of the performance is observed to occur after 12 eigenfunctions

Number of pairs	AUC
1	0.50
2	0.56
3	0.67
4	0.72
5	0.76
6	0.81
7	0.83
8	0.84
9	0.85
10	0.90
11	0.96
12	0.98
13	0.98
14	0.98
15	0.98

plateau can be easily observed by computing the area under the ROC curve (AUC), which is a metric of the accuracy of the test (a perfect classifier will have AUC equal to unity). A rule-of-thumb to evaluate the performance of a classifier by the AUC is that an *excellent classifier* will have an AUC in the range 0.9-1.0, a good classifier have an AUC 0.80-0.89 and a fair classifier will present an AUC in the range 0.60-0.79. A classifier with an AUC below 0.59 is considered poor or simply worthless. The resulting AUC for the presented ROC curves are shown in Table 2.4 where it can be seen that the performance of the classifier plateau occurs after 12 eigenvalue-eigenfunction pairs ($K = 12$) are utilized, and that for $K > 10$ the detector can be classified to have an *excellent* performance. We explore the variability of the theoretical performance in detail in Chapter 3. Next we explain how the eigenvalue-eigenfunction pairs and the optimum threshold can be utilized to classify the patient data.

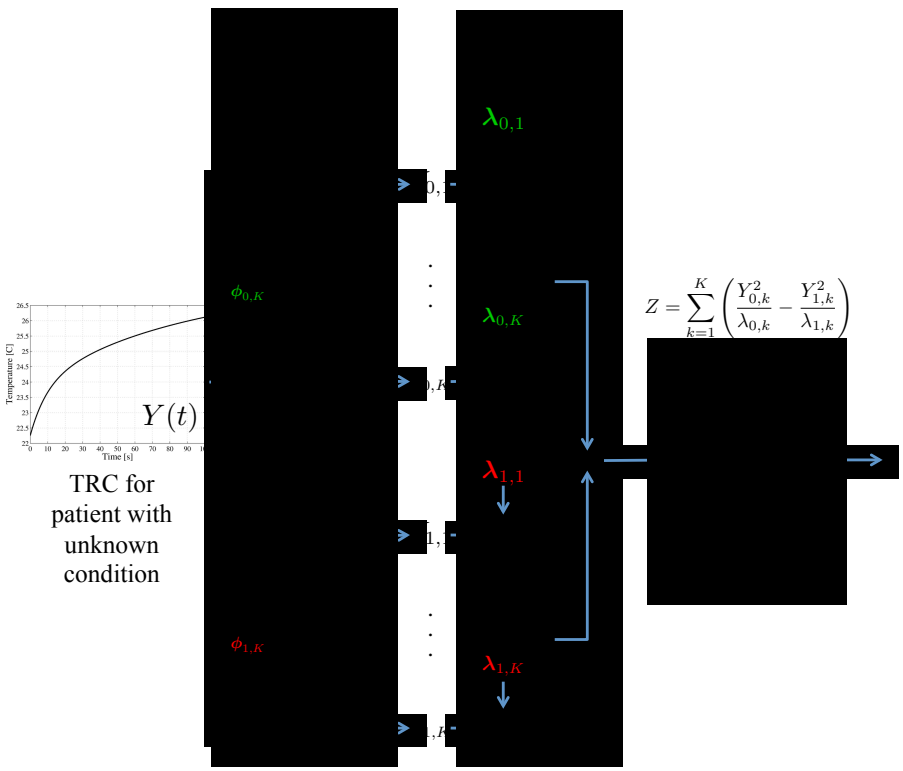


Figure 2.12: Block diagram of the detection stage of the proposed algorithm. The KL coefficients are computed by using the eigenfunctions of each hypothesis. These coefficients and the eigenvalues are used to compute the patient's test-statistics, which is later compared with the optimum threshold to declare the malignancy

2.8.4 Summary of the steps followed to detect malignancy

Let us explain in detail how a patient with unknown diagnosis can be diagnosed with the proposed approach. Let the average TRC of this patient be denoted by $Y(t)$. At this stage, we assume that the autocorrelation functions were correctly estimated from patient data with known condition (i.e., training patient data), and that the eigenvalues-eigenfunction pairs were also obtained and sorted based on the value of the eigenvalues. We denote the number of the stored (and sorted) pairs under each hypothesis by K. We also assume that the optimum threshold, η_0 was already defined by means of the NP decision rule as described in the previous section.

We first utilize the eigenfunctions to compute the KL coefficients under each hypothesis, i.e., $Y_{j,k} = \int_0^T \phi_{j,k}(t)Y(t)$, for $j = 0, 1$ and $k = 1, 2, \ldots, K$. This computation is depicted graphically in Fig. 2.12 (left dashed block). It is worth to reiterate that the set of KL coefficients represent the most important statistical features that completely describe the statistical content within the TRC $Y(t)$[2]. We next utilize the KL coefficients and the corresponding eigenvalues to construct the test-statistics as defined in (2.46). Such a calculation is depicted in Fig. 2.12 (middle dashed block). The test-statistics is later compared with the optimum threshold, η_0. This last step is named the decision stage and it is depicted in Fig. 2.12 (right dashed block). By the extraction of the KL coefficients and the thresholding of the test statistics by the optimally defined threshold, the detection problem is guaranteed to ensure the maximum achievable detection probability, and, as such, it presents an upper bound in the expected performance of the detector.

2.9 Concluding remarks

In this chapter we have introduced the concept of dynamic thermal imaging, the acquisition hardware utilized during the actual patient data acquisition and the preprocessing applied to the infrared sequence before commencing the development of the algorithm.

We proposed a physics-inspired stochastic mathematical model that describes the stochastic thermal recovery of the patient tissue after a cool stimuli is applied. Using this stochastic model for the thermal random processes monitored by the infrared camera, we have we postulated the detection problem we aim to solve. Moreover, we have proposed a way to determine the analytical expression of an autocorrelation

[2]As such, they can be used to feed any *out-of-the-shelf* classifier to detect the malignancy of a suspicious lesion. This idea is expanded in the last chapter in the section of proposed future work.

function for stochastic processes with a parameterized model.

By numerically extracting the eigenvalues and eigenfunctions of each autocorrelation function, and by using Grenander's theorem and the Karhunen-Loève expansion we have presented the NP optimal decision rule to maximize the number of malignant cases that are correctly classified while constraining the number of false positives that are declared.

In the next chapter, we explore the performance of the designed algorithm under different operation scenarios of the developed algorithm.

Chapter 3

Performance and robustness analysis for the method

3.1 Introduction

In this Chapter we demonstrate the efficacy and delimit the scope and robustness of the skin cancer detection algorithm. To this end, we focus our attention in three different scenarios to evaluate the algorithm performance. First, we perform the separability analysis of the patient data. This is made by training the algorithm with a set of patients and then testing the performance by means of the empirical ROC curve of the very same patients used in the training. Second, we study the effect in the performance by changing the training size (i.e., the number of patients used in training) and by using different permutations of patients as training/testing. Third, we study how selecting the lesion boundary affects the performance of the algorithm.

In all of the previous settings we evaluate the variability in the theoretical and empirical performance. The theoretical performance is measured by the theoretical

false-alarm and detection probabilities, as well as the ROC curves and the corresponding AUCs. The metrics we utilize to evaluate the empirical performance of the algorithm are the empirical false-alarm and detection probabilities, which we define next.

3.2 Empirical performance metrics

For the purpose of this study, our gold standard is the biopsy result performed at the UNM Dermatology Clinic. With the biopsy result and the result of our algorithm we classify the patients as:

1. Those that are malignant (by biopsy) and are declared as malignant by the algorithm are known as the true positives (TP);

2. Those that are malignant and are declared as benign by the algorithm are the false negatives (FN);

3. Those that are benign and are declared as benign are the true negatives (TN); and

4. Those that are benign and are declared as malignant are the false positives (FP).

With these four quantities, we compute the empirical detection probability as the ratio between the true positives and all the malignant patients, i.e., $P_{D,e} = TP/(TP+FN)$, where the subscript $_e$ denotes the empirical nature of the definition. Similarly, the empirical false-alarm probability is computed as the ratio of the true negatives and all the benign cases, $P_{F,e} = TN/(TN + FP)$. The empirical ROC curve is obtained by plotting the empirical detection probability with respect to the empirical

false-alarm probabilities. We obtain the empirical AUC by numerically integrating (with the trapezoidal quadrature rule) the empirical ROC curves.

3.3 Separability analysis of the method

We begin our study of the algorithm performance by performing the separability analysis, which is the study of the the discernibility between the two hypothesis by testing the algorithm with the very same patient data that was used to train such an algorithm. Here we utilize all the 140 patient data with known condition used to describe the algorithm in the previous chapter.

Let us detail some information regarding the actual patient dataset acquired at the UNM Dermatology Clinic by the author.

3.3.1 Patient data cohort

A cohort study with 140 subjects is performed to investigate the proposed approach. Fifty eight percent of the subjects were male and, from the biopsy result, out of the 140 subjects 82 had benign condition and 58 had malignant condition. Out of those 58 subjects with malignant condition, 6 were diagnosed with malignant-melanomas (MM), 42 with basal-cell carcinoma (BCC) and 10 with squamous-cell carcinoma (SCC).

The majority of the infrared sequences utilized in this dissertation were acquired by the author at the UNM Dermatology Clinic in a period of 2 years (2012-2014).

3.3.2 Eigenvalue-eigenfunction pairs selection

In Chapter 2 we presented the ACFs computed from all the 140 subject data (see Fig. 2.7). Here we want to investigate the maximum number of reliable eigenvalue-eigenfunction pairs that can be extracted from these ACFs. In what follows we term an eigenvalue-eigenfunction pair as an *eigenpair*.

Provided that we sorted the first sixteen eigenvalues as presented in Table 3.1, in Fig. 3.1 we present the corresponding eigenfunctions for both hypotheses. It can

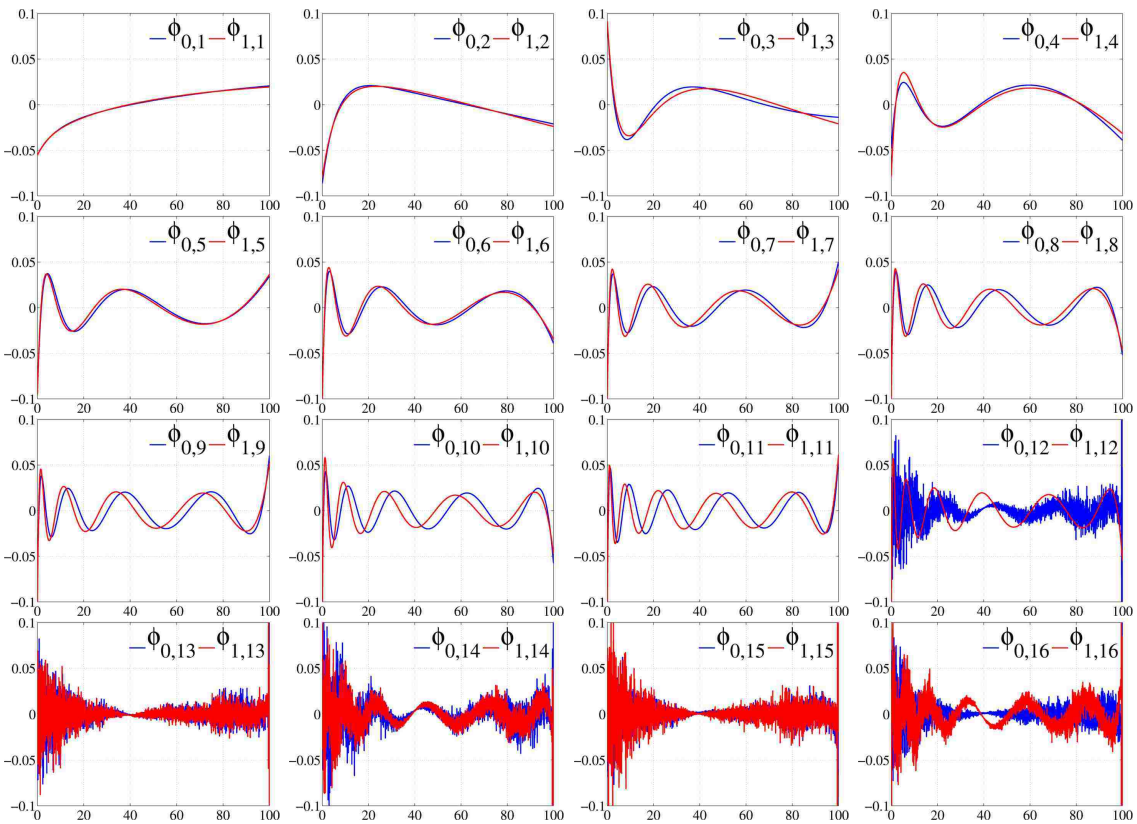


Figure 3.1: Eigenfunctions computed for each hypothesis sorted by the value of their corresponding eigenvalue. It can be noted that after the twelfth eigenpair the eigenfunctions account for noise, and, therefore, they cannot be considered as reliable anymore

Table 3.1: Eigenvalues corresponding to the eigenfunctions depicted in Fig. 3.1. These eigenvalues were computed using all 140 patients with known condition

k	$\lambda_{0,k}$	$\lambda_{1,k}$
1	955.47	1609.33
2	37.89	44.69
3	2.24	5.70
4	0.55	1.22
5	1.68×10^{-2}	3.45×10^{-2}
6	7.47×10^{-4}	2.25×10^{-3}
7	3.58×10^{-5}	8.97×10^{-5}
8	1.31×10^{-6}	1.99×10^{-6}
9	4.54×10^{-8}	7.52×10^{-8}
10	1.11×10^{-9}	6.39×10^{-9}
11	2.68×10^{-11}	6.38×10^{-10}
12	1.64×10^{-12}	3.94×10^{-11}
13	1.62×10^{-12}	2.85×10^{-12}
14	6.44×10^{-13}	2.64×10^{-12}
15	5.82×10^{-13}	1.50×10^{-12}
16	4.58×10^{-13}	1.20×10^{-12}

be noted from the presented plots that both eigenfunction sets are essentially the same under each hypothesis, but as we require more eigenfunctions from the ACFs a slight phase difference start to appear. This result is consistent with the numerical cross-orthogonality observed and presented in Fig. 2.9. Once the phase difference becomes notorious (e.g., for $K > 10$) the eigenfunctions loose their similarity, and, as a consequence, the cross-orthogonality is also lost.

We have discussed in the previous subsection that the size of the patient dataset under each hypothesis is small. The main problem related with this situation is that the statistical information obtained from the dataset is not "rich enough" to permit an appropriated estimation of the ACFs. This poor estimation of the ACFs, in turn, affects the reliability of the eigenvalue-eigenfunction pairs. As such, as can be observed in Fig. 2.9, after the twelfth eigenvalue, the eigenfunction start to be noisy. It is expected that after having more patient data available with sufficient

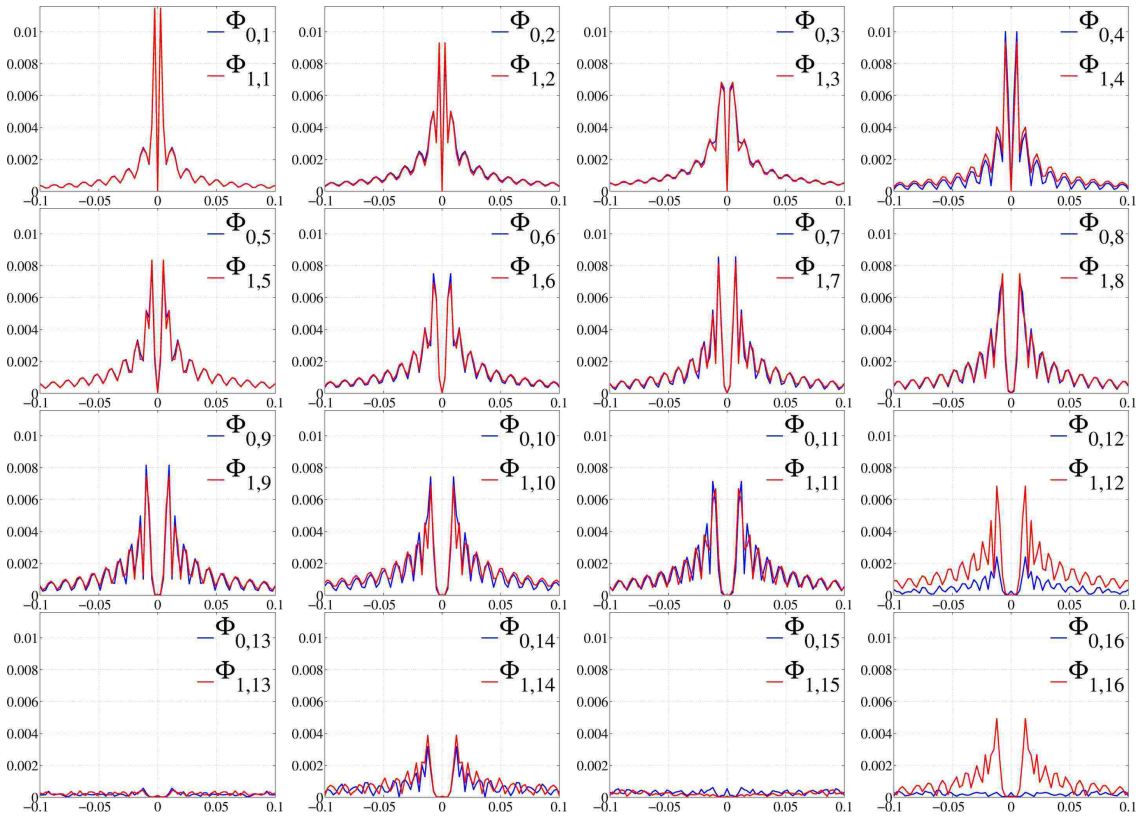
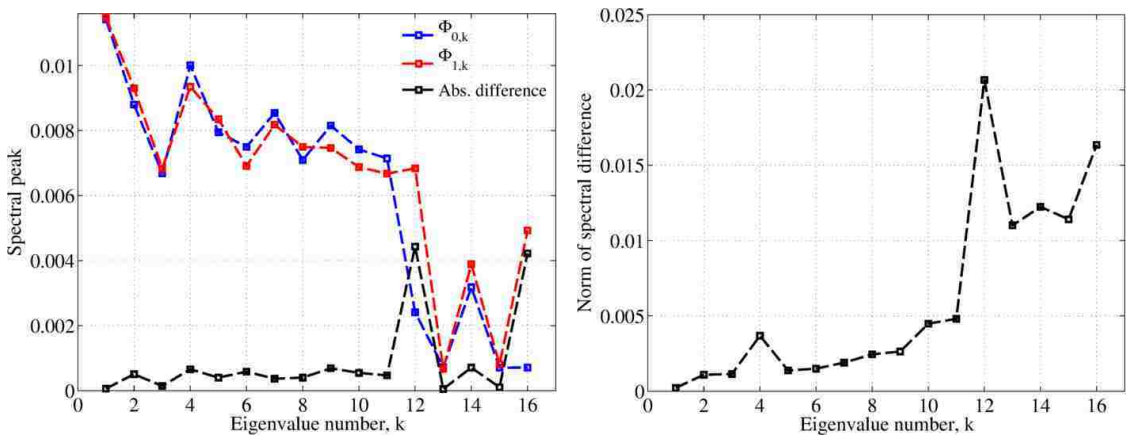


Figure 3.2: Fast-Fourier transform computed over the eigenfunction sets presented in Fig. 3.1. The scale is the same in the resulting plots to facilitate comparison

variability, this problem can be undertaken due to the increase statistical information. Probably, a different approach in estimating the ACF can also improve the number of reliable eigenfunctions. The later approach is suggested in the future work section in Chapter 6.

Since we have the limitation of noisy eigenfunctions after the Kth eigenvalue, we propose to find such a K by means of the Fourier transform of the estimated eigenfunctions. We computed the FFT of each eigenfunction, namely $\Phi_{j,k}(f) = \mathsf{FFT}\left[\phi_{j,k}(t)\right]$, for $j = 0, 1$ and $k = 1, \ldots, 16$. The resulting FFTs are shown in Fig. 3.2, where we used the same scale for all the spectra in order to facilitate the comparison. As the eigenvalue number is increased, the peak spectrum of each eigen-



(a) (b)

Figure 3.3: Proposed analysis alternatives to select the number of reliable eigenvalues based on the spectral content of the corresponding eigenfunctions

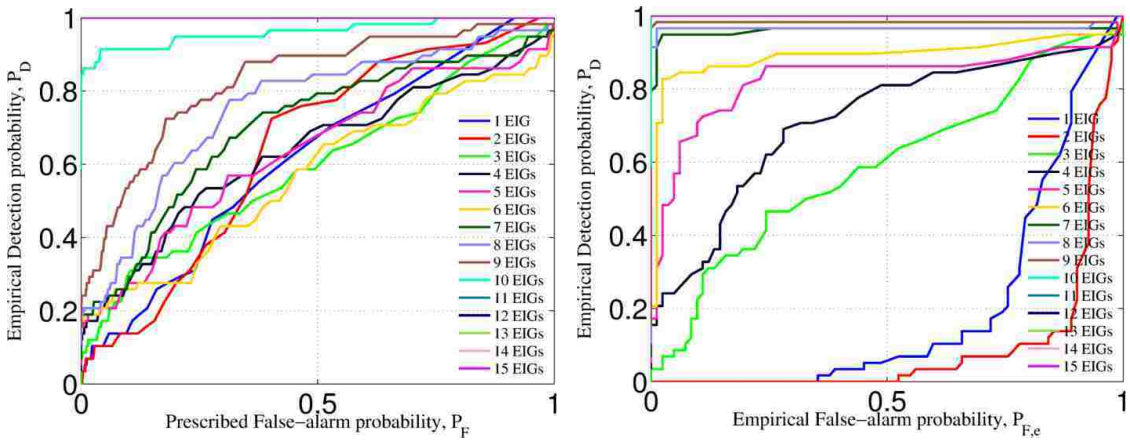
function is slightly shifted and reduced in amplitude, so one alternative to determine K is by looking into abrupt changes in the peak of the spectrum of the set of eigenfunctions. This alternative is shown in Fig. 3.3(a), where it can be noted that the peak spectra for the benign case (blue line) drops abruptly at $k = 12$, and that it is also detected by the absolute difference between the two peak spectra (black line). An alternative approach to detect the number of reliable eigenvalues is by exploiting the spectral similitude observed within each eigenfunction corresponding to the same eigenvalue number. As we estimate more eigenvalues for each hypothesis, the spectra of the corresponding eigenfunctions start to depart slightly. The breaking point of having a noisy eigenfunction can be easily detected as the difference of the spectral increases abruptly. This is because a noisy eigenfunction will have its spectral content dispersed among different frequencies instead of collocated around zero. Therefore, if we compute the norm of the difference between the absolute value of the spectral we can easily find the noisy eigenfunction by finding the maximum of such a norm, as shown in Fig. 3.3(b). It is clear that this approach will only work if one of the eigenfunction becomes noise but the other stays with low noise content.

In what follows, we select the noiseless eigenvalues and functions by using the former method, i.e., by comparing the absolute difference of the eigenfunction spectra. We now utilize this to perform the sensitivity analysis

3.3.3 Separability analysis

By limiting our attention to the first twelve eigenvalue-eigenfunction pairs, we computed the KL coefficients of all the 140 patients and constructed their respective test-statistics. As the specified false-alarm is changed, so does the optimum NP decision threshold, and, thus, the empirical detection probability. When doing so, we can generate the empirical ROC curve, which is the plot of the prescribed false-alarm probability versus the empirical detection probability as depicted in Fig. 3.4(a). When the *purely* empirical ROC curve is generated (that is, the empirical false-alarm vs. empirical detection probabilities) the resulting performance is a slightly degraded as shown in Fig. 3.4(b). Please note that the NP decision rule does not impose any direct restriction to the empirical false-alarm probability and the computed value is simply a direct consequence of the definition of the test-statistics and the threshold that optimizes the detection probability. The degradation of the empirical false-alarm probability as compared with the prescribed one is a direct consequence of the assumptions made in the algorithm development and will affect directly the empirical performance of the algorithm. Other decision rules may introduce constraints to the empirical false-alarm probability directly, or indirectly by means of, for example, minimizing the error probability, which is $P_E = P(H_1|H_0) + P(H_0|H_1)$ where we recognize the false-alarm probability in the first term.

In order to illustrate the result of the sensitivity analysis with more detail, we have generated Table 3.2 where we present the empirical detection probability $P_{D,e}$ for different levels of prescribed false-alarm probabilities (specifically, $P_F = 0.1$, $P_F = 0.05$ and $P_F = 0.01$). For comparison, we have also included the empirical detection



(a) (b)

Figure 3.4: Sensitivity analysis resulting ROC curves when using the (a) prescribed (theoretical) false-alarm probability versus the empirical detection probability, and (b) the empirical false-alarm and detection probabilities.

probability when the prescribed false-alarm is enforced over the empirical false-alarm probability, $P_{F,e}$. Clearly, the later achieve worse performance than the former, but

Table 3.2: Empirical detection probability, $P_{D,e}$, for different levels of prescribed FA, P_F, and empirical FA, $P_{F,e}$

Number of used eigenfunctions	Empirical detection probability for					
	Prescribed FA, P_F			Empirical FA, $P_{F,e}$		
	0.1	0.05	0.01	0.1	0.05	0.01
1	0.14	0.12	0.05	0.00	0.00	0.00
2	0.14	0.10	0.07	0.00	0.00	0.00
3	0.29	0.17	0.09	0.22	0.09	0.03
4	0.28	0.22	0.16	0.33	0.24	0.21
5	0.28	0.21	0.17	0.71	0.57	0.31
6	0.26	0.21	0.19	0.84	0.84	0.71
7	0.29	0.22	0.19	0.95	0.95	0.91
8	0.34	0.22	0.21	0.97	0.97	0.97
9	0.53	0.40	0.24	0.98	0.98	0.98
10	0.91	0.91	0.86	1.00	1.00	1.00
11	1.00	1.00	1.00	1.00	1.00	1.00
12	1.00	1.00	1.00	1.00	1.00	1.00

since it is the *real* performance of the algorithm it worth attention. Please note that perfect detection of all the malignant cases is achieved (i.e., $P_{D,e} = 1.0$) when ten eigenvalue-eigenfunction pairs or more are used to determine the test statistics. This detection is achieved for an empirical false-alarm probability of $P_{F,e} = 0.01$. In simple terms, this means that the proposed algorithm permits the correct detection of all 58 malignant lesions in the dataset, at the cost of missing 1 patient with benign condition out of the 82 benign cases present in the dataset. To the best of our knowledge, this is the best performance ever reported for a non-invasive technique in skin cancer detection.

3.4 Robustness of the algorithm to random permutations of the training set from the total set

In the the previous study we have utilized all subject data to train and then test the proposed algorithm. In this study we divide the patient data (140 subjects; 58 with malignant and 82 with benign condition) in training and testing sets following the following criteria:

1. Train with 110 subjects (55M and 55B) and test with 30 subjects;
2. Train with 100 subjects (50M and 50B) and testing with 40 subjects;
3. Train with 80 subjects (40M and 40B) and testing with 60 subjects;
4. Train with 60 subjects (30M and 30B) and testing with 60 subjects,

where in all the cases the set with malignant condition is forced to have at east a few MM, a few BCC and a few SCC cases; more precisely, we force the traning

set to have the same percentile distribution of lesions found in our dataset. For example our dataset has 6 MM out of the 58 subject with malignant diagnosis; thus we force our split to have a 10.34% of MM in the training set, which for the case of 60 subjects means that at least 3 MM must be included in the training. For the set of benign cases, we force to have at least one half of control samples (those with benign clinical diagnosis instead of formal biopsy result). These two conditions are included to ensure that the training and testing sets have an equivalent variability of cases under both hypotheses.

In addition, we have performed 200 different permutations of the patient data to define each one of the different four settings, allowing us to study the robustness of the presented algorithm under different operation conditions. In what follows we present the variability in the eigenvalue-eigenfunction pairs as well as the variability in the theoretical and empirical performance for all these settings.

3.4.1 Eigenvalue-eigenfunctions pairs

The distributions of the random parameters that define the stochastic signals change as we permute the patients from the training set. Such a change will change the estimated autocorrelation function, which, in turn, will alter the estimated eigenvalue-eigenfunction pairs. Therefore, we first investigate the variability of these pairs for each of the training/testing settings we described above. The minimum, maximum and mean value of each one of the first twelve eigenvalues for training sizes of 60, 80, 100 and 110 subjects are shown in Tables 3.3-3.6, respectively.

For the first five eigenvalues, the variabilities are small as compared with the variation for other eigenvalues, where there are orders of magnitude between the minimum and maximum value. Indeed, it can be noted from the presented tables that the variability of the eigenvalues becomes severe as we require more eigenvalues

Table 3.3: Eigenvalue variability over 200 permutations for a training size of 60 subjects

k	$\min \lambda_{0,k}$	$\max \lambda_{0,k}$	$\mathsf{E}[\lambda_{0,k}]$	$\min \lambda_{1,k}$	$\max \lambda_{1,k}$	$\mathsf{E}[\lambda_{1,k}]$
1	860.45	1365.75	1062.93	1235.09	2698.75	1897.37
2	12.93	41.64	28.59	31.41	68.45	52.13
3	1.40	3.49	2.36	3.99	20.42	11.59
4	0.14	1.02	0.38	0.71	2.74	1.79
5	1.67e-03	4.14e-02	1.32e-02	3.06e-02	1.20e-01	7.03e-02
6	3.25e-05	1.35e-03	2.94e-04	1.04e-03	8.82e-03	3.95e-03
7	5.78e-07	4.66e-05	7.92e-06	1.56e-05	3.58e-04	1.45e-04
8	3.39e-09	9.07e-07	1.54e-07	4.82e-07	6.90e-06	3.00e-06
9	2.20e-11	2.33e-08	2.56e-09	1.57e-08	4.09e-07	1.71e-07
10	7.58e-13	3.94e-10	2.71e-11	6.14e-10	2.13e-08	8.36e-09
11	5.62e-13	1.16e-11	1.39e-12	1.45e-11	3.06e-09	7.01e-10
12	4.69e-13	2.03e-12	9.66e-13	1.56e-12	1.47e-10	1.81e-11

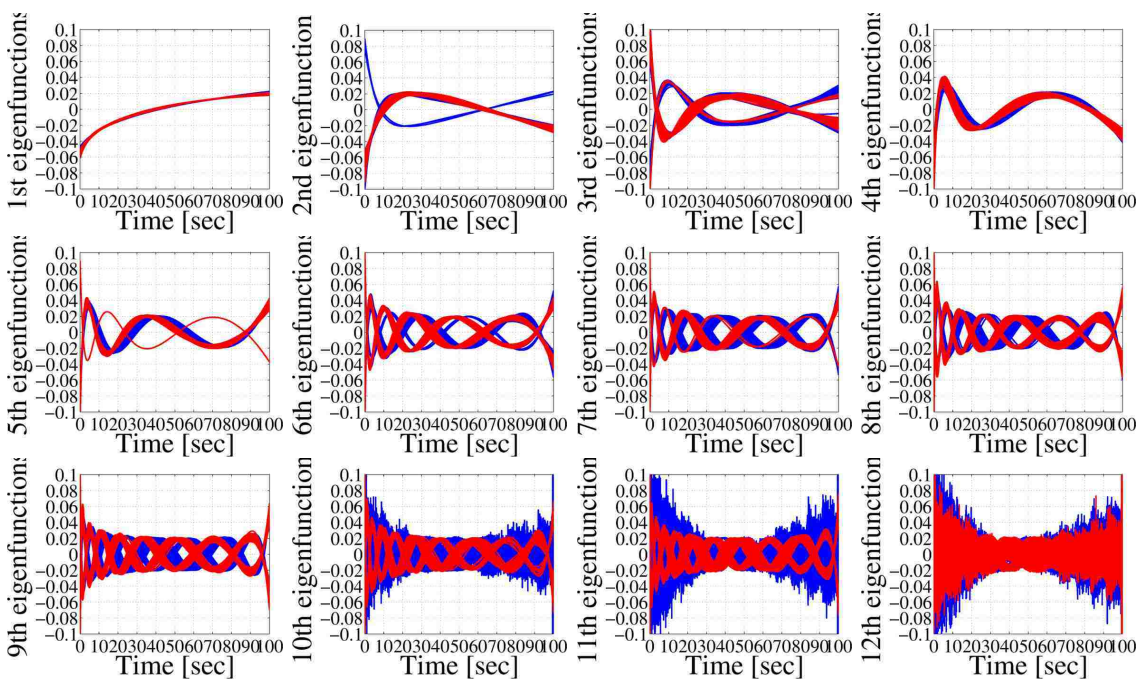


Figure 3.5: Variability of the eigenfunctions for the 200 permutations on the 60 training/80 testing setting. The blue and red lines represent the variability of $\phi_{0,k}$ and $\phi_{1,k}$, respectively, for $k = 1, 2, \ldots, 12$.

Table 3.4: Eigenvalue variability over 200 permutations for a training size of 80 subjects

k	$\min \lambda_{0,k}$	$\max \lambda_{0,k}$	$\mathsf{E}[\lambda_{0,k}]$	$\min \lambda_{1,k}$	$\max \lambda_{1,k}$	$\mathsf{E}[\lambda_{1,k}]$
1	890.84	1345.32	1103.11	1324.74	2331.21	1899.43
2	21.26	46.29	34.56	41.28	64.38	53.30
3	1.60	4.04	2.68	4.70	16.33	11.54
4	0.23	1.24	0.55	0.81	2.48	1.83
5	5.75e-03	4.26e-02	1.97e-02	3.50e-02	1.02e-01	7.35e-02
6	4.41e-05	1.50e-03	6.76e-04	1.35e-03	6.75e-03	4.40e-03
7	2.41e-06	6.40e-05	2.23e-05	3.40e-05	3.02e-04	1.73e-04
8	1.50e-08	1.85e-06	5.20e-07	1.03e-06	5.87e-06	3.65e-06
9	1.30e-10	5.81e-08	1.05e-08	7.16e-08	3.65e-07	2.29e-07
10	1.88e-12	3.75e-09	2.42e-10	2.59e-09	1.88e-08	1.17e-08
11	5.98e-13	5.29e-11	4.04e-12	6.72e-11	2.90e-09	1.24e-09
12	3.99e-13	2.80e-12	1.21e-12	2.36e-12	1.15e-10	3.94e-11

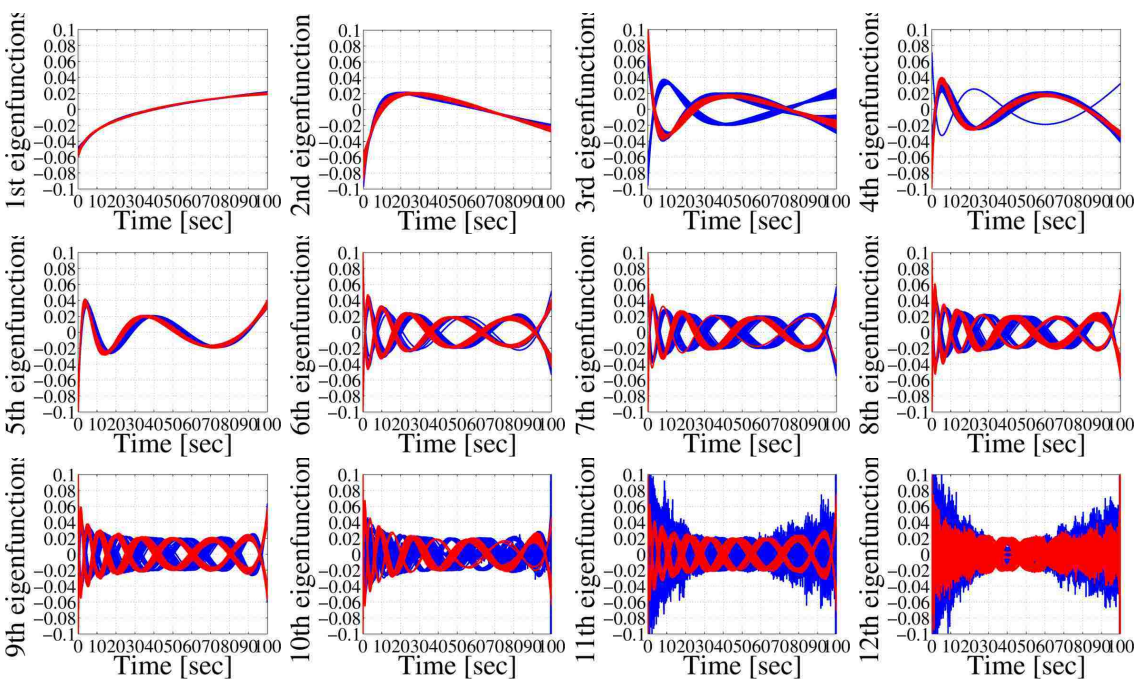


Figure 3.6: Variability of the eigenfunctions for the 200 permutations on the 80 training/60 testing setting. The blue and red lines represent the variability of $\phi_{0,k}$ and $\phi_{1,k}$, respectively, for $k = 1, 2, \ldots, 12$.

Table 3.5: Eigenvalue variability over 200 permutations for a training size of 100 subjects

k	$\min \lambda_{0,k}$	$\max \lambda_{0,k}$	$\mathsf{E}[\lambda_{0,k}]$	$\min \lambda_{1,k}$	$\max \lambda_{1,k}$	$\mathsf{E}[\lambda_{1,k}]$
1	893.77	1371.49	1113.82	1461.09	2040.84	1862.96
2	26.61	42.87	36.59	44.97	57.93	52.85
3	1.84	4.50	2.87	6.94	13.65	11.76
4	0.33	0.94	0.62	1.20	2.08	1.85
5	8.57e-03	3.83e-02	2.43e-02	5.68e-02	8.76e-02	7.53e-02
6	1.58e-04	1.52e-03	9.30e-04	2.18e-03	5.88e-03	4.89e-03
7	4.10e-06	6.08e-05	3.46e-05	5.67e-05	2.60e-04	2.05e-04
8	8.88e-08	1.78e-06	9.09e-07	1.48e-06	5.09e-06	4.11e-06
9	9.83e-10	5.72e-08	2.13e-08	9.72e-08	3.26e-07	2.55e-07
10	1.41e-11	3.86e-09	7.62e-10	4.21e-09	1.66e-08	1.30e-08
11	8.75e-13	6.42e-11	1.37e-11	3.33e-10	2.43e-09	1.67e-09
12	6.53e-13	2.87e-12	1.23e-12	1.40e-11	9.73e-11	6.48e-11

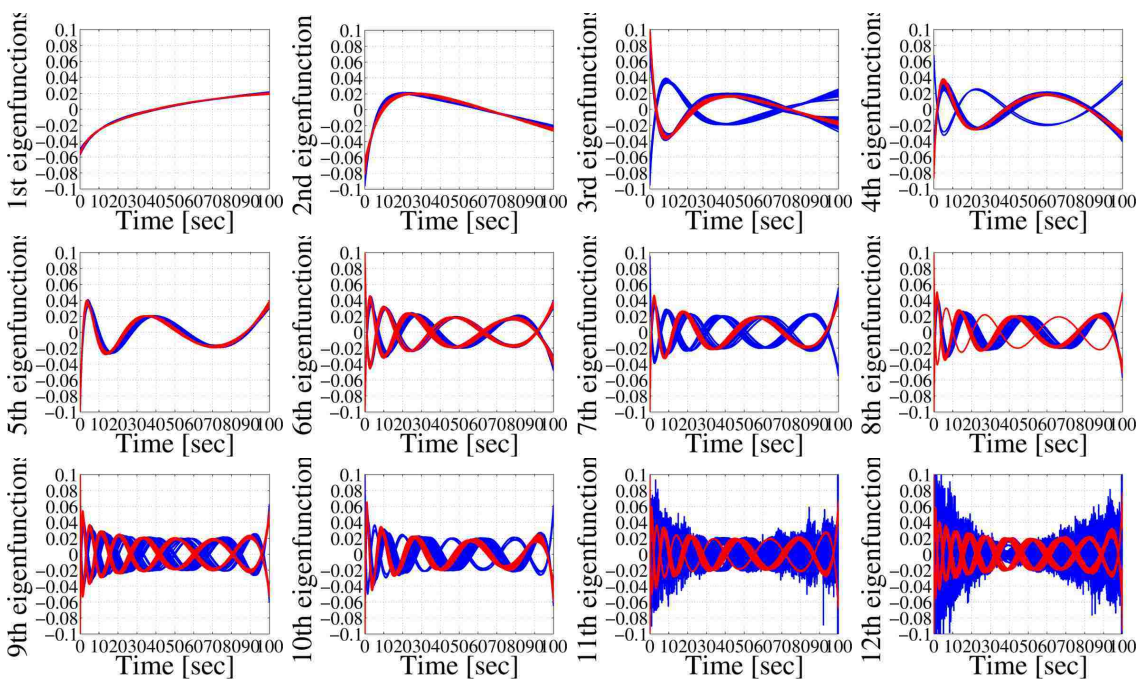


Figure 3.7: Variability of the eigenfunctions for the 200 permutations on the 100 training/40 testing setting. The blue and red lines represent the variability of $\phi_{0,k}$ and $\phi_{1,k}$, respectively, for $k = 1, 2, \ldots, 12$.

Table 3.6: Eigenvalue variability over 200 permutations for a training size of 110 subjects

k	$\min \lambda_{0,k}$	$\max \lambda_{0,k}$	$\mathsf{E}[\lambda_{0,k}]$	$\min \lambda_{1,k}$	$\max \lambda_{1,k}$	$\mathsf{E}[\lambda_{1,k}]$
1	907.07	1353.96	1112.31	1474.71	1937.50	1847.76
2	28.83	50.31	37.61	47.13	55.03	52.18
3	1.88	3.83	2.95	7.08	12.51	11.56
4	0.35	0.96	0.67	1.37	1.93	1.84
5	1.09e-02	3.60e-02	2.54e-02	5.96e-02	8.17e-02	7.69e-02
6	1.90e-04	1.53e-03	1.04e-03	2.34e-03	5.38e-03	4.93e-03
7	5.16e-06	6.03e-05	3.79e-05	5.38e-05	2.39e-04	2.12e-04
8	8.25e-08	1.76e-06	1.04e-06	2.69e-06	4.69e-06	4.29e-06
9	7.78e-10	5.64e-08	2.59e-08	1.65e-07	3.00e-07	2.72e-07
10	1.44e-11	3.92e-09	1.19e-09	8.16e-09	1.55e-08	1.41e-08
11	8.82e-13	6.45e-11	2.12e-11	5.64e-10	2.23e-09	1.91e-09
12	6.07e-13	3.44e-12	1.40e-12	1.61e-11	8.89e-11	7.21e-11

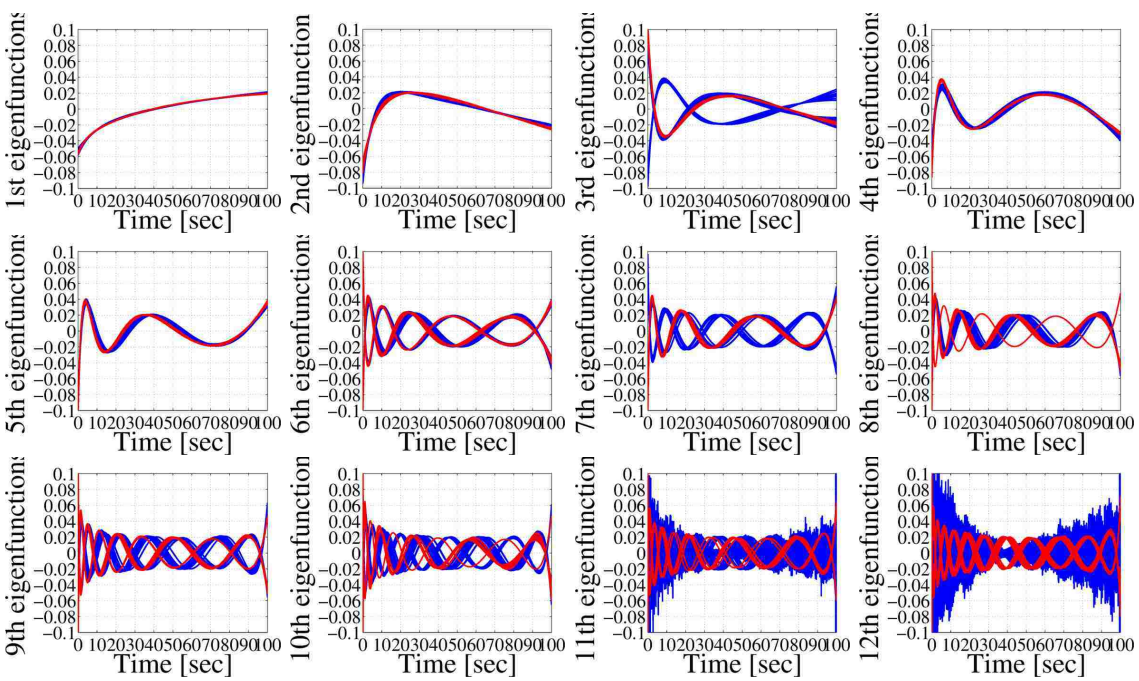


Figure 3.8: Variability of the eigenfunctions for the 200 permutations on the 110 training/30 testing setting. The blue and red lines represent the variability of $\phi_{0,k}$ and $\phi_{1,k}$, respectively, for $k = 1, 2, \ldots, 12$.

from the ACFs. We believe that this tremendous variability is mainly due to the estimated ACF with a limited number of patient data. We believe that the more patients become available, the better the ACF will be estimated and the less the variability of the eigenvalues should be.

A similar problem is presented with the eigenfunctions, which slightly vary for the first five eigenvalues but present big changes for higher orders. Such a change occurs to ensure that the corresponding eigenvalues are positive (recall that the autocorrelation function is symmetric and positive definite, and, as such, its eigenvalues must be real and positive). This variability is mainly due to the method we have developed to estimate the autocorrelation function, which relies in the numerical computation of the means and cross-covariances between the random parameters that characterize the TRC under each hypothesis. As we change the patients inside the training set, the sample mean and sample covariance of the parameters is dramatically changed due to the small number of patients we have to training. In fact, it can be noted from comparing Figs.3.5–3.8 that, as more patients are included into the training set, the estimated eigenfunctions become more reliable. Hence, we expect that as more patient data is acquired, the estimation of the eigenvalue-eigenfunction pairs will become more reliable and the performance of the algorithm will be superior. Next, we study the theoretical performance variability over the 200 permutation for each training size.

3.4.2 Theoretical performance

It was already shown in the previous section that as we utilize different permutations of patients and different number of patients to train the algorithm, there is considerable variations in the estimated eigenvalue-eigenfunction pairs. To asses how such a variability affects the theoretical performance, we utilize the estimated eigenvalue-eigenfunction pairs for each one of the permutations and computed the theoretical

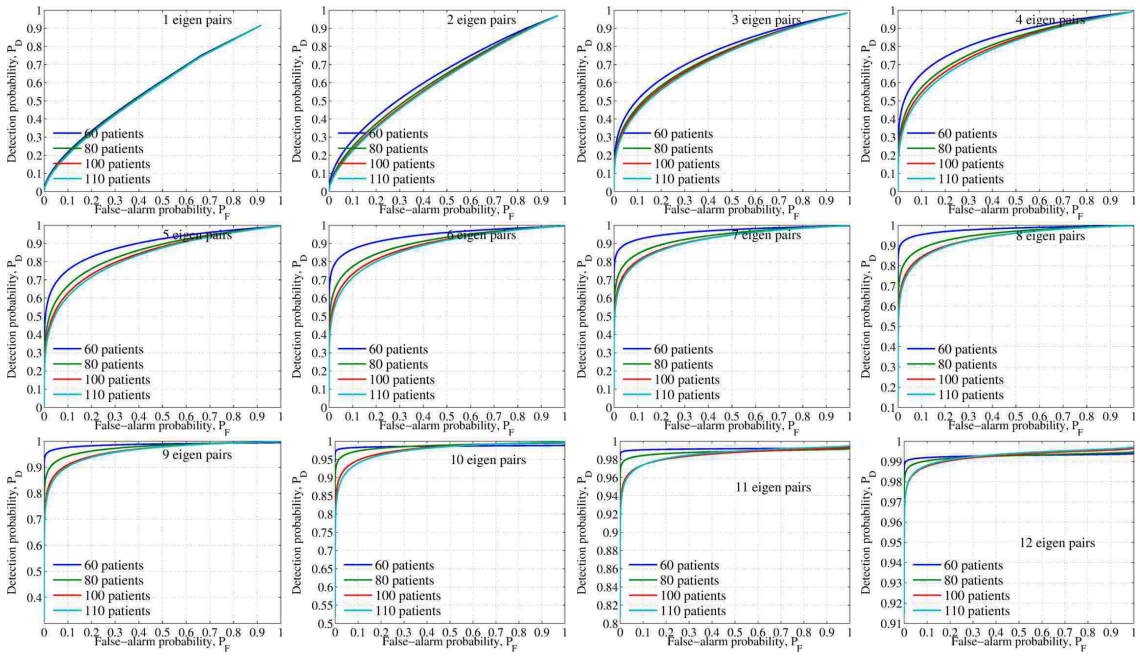


Figure 3.9: Mean ROC curves (over 200 different patient permutations) for different training sets with 60, 80, 100 and 110 patients.

false-alarm and detection probabilities.

The theoretical performance variability as a function of the size of the training set is assess by the mean ROC curves for different number of eigenvalue-eigenfunction pairs, as depicted in Fig. 3.9. Beside some small variabilities, the mean theoretical performance is consistent across the changes in the the training size, but the real difference can be observed when the area under the curve and its mean absolute error over the 200 permutations is computed. Hence, we quantify the variability in the theoretical performance by computing the mean AUC, for each training size, over the 200 permutations. The results are summarized in Table 3.7, where it can be clearly observed that the average performance when 60 patients are used in the training is better, but the mean absolute error (shown within the parenthesis) demonstrate that there is more variability in the performance due to the reduced number of patients. It can be noted from the same table that the theoretical performance slightly changes

Table 3.7: Variability of the AUC over 200 permutations measured by the mean AUC and the mean absolute error (given within the parenthesis) for different training sizes and different number of used eigenvalue-eigenfunction pairs.

Number of used eigenfunctions	Size of the training set			
	60 patients	80 patients	100 patients	110 patients
1	0.51 (± 0.06)	0.50 (± 0.06)	0.50 (± 0.01)	0.50 (± 0.01)
2	0.61 (± 0.11)	0.59 (± 0.07)	0.58 (± 0.01)	0.57 (± 0.01)
3	0.76 (± 0.11)	0.74 (± 0.10)	0.73 (± 0.02)	0.72 (± 0.02)
4	0.84 (± 0.11)	0.80 (± 0.10)	0.79 (± 0.02)	0.78 (± 0.02)
5	0.89 (± 0.14)	0.85 (± 0.13)	0.84 (± 0.03)	0.83 (± 0.02)
6	0.94 (± 0.15)	0.90 (± 0.16)	0.89 (± 0.03)	0.88 (± 0.03)
7	0.96 (± 0.16)	0.93 (± 0.17)	0.92 (± 0.03)	0.91 (± 0.02)
8	0.98 (± 0.12)	0.95 (± 0.16)	0.94 (± 0.03)	0.93 (± 0.02)
9	0.99 (± 0.04)	0.98 (± 0.11)	0.96 (± 0.02)	0.96 (± 0.02)
10	0.99 (± 0.03)	0.99 (± 0.05)	0.98 (± 0.01)	0.97 (± 0.01)
11	0.99 (± 0.02)	0.99 (± 0.03)	0.98 (± 0.01)	0.99 (± 0.01)
12	0.99 (± 0.02)	0.99 (± 0.01)	0.99 (± 0.00)	0.99 (± 0.00)

when we use 100 or 110 patients in the training set, which is consistent with the reduced variability in the eigenvalues and eigenfunctions that we discussed earlier.

Next, we utilize the same settings to evaluate the changes in the empirical performance as determined by the empirical detection probability.

3.4.3 Empirical performance

Using the same setting from the previous section, here we measure the variability in the empirical detection probability, computed for different levels of prescribed false-alarm probability. In what follows we present the tables with the results, which include the mean empirical detection probability (over all 200 permutations) and the corresponding mean absolute error within the parenthesis. The mean absolute error (MAE) for our purposes is defined as

$$MAE = \frac{1}{L}\sum_{i=1}^{L} |P_{D,e}^{(i)} - \bar{P}_{D,e}| \ , \tag{3.1}$$

Table 3.8: Mean and mean absolute error (given within the parenthesis) of the empirical detection probability for different levels of prescribed false-alarm probability when 80 patients are used in the training and 60 in the testing stage.

Number of used eigenfunctions	False-alarm probability		
	$P_F = 0.1$	$P_F = 0.05$	$P_F = 0.01$
1	0.00 (± 0.00)	0.00 (± 0.00)	0.00 (± 0.00)
2	0.02 (± 0.04)	0.02 (± 0.04)	0.02 (± 0.04)
3	0.42 (± 0.14)	0.42 (± 0.14)	0.42 (± 0.14)
4	0.46 (± 0.12)	0.46 (± 0.12)	0.46 (± 0.12)
5	0.80 (± 0.11)	0.80 (± 0.11)	0.80 (± 0.11)
6	0.91 (± 0.09)	0.91 (± 0.09)	0.91 (± 0.09)
7	0.94 (± 0.11)	0.94 (± 0.11)	0.94 (± 0.11)
8	0.81 (± 0.31)	0.81 (± 0.31)	0.81 (± 0.31)
9	0.86 (± 0.24)	0.86 (± 0.24)	0.86 (± 0.24)
10	0.85 (± 0.25)	0.85 (± 0.25)	0.85 (± 0.25)
11	0.90 (± 0.19)	0.90 (± 0.19)	0.90 (± 0.19)
12	0.86 (± 0.23)	0.86 (± 0.23)	0.86 (± 0.23)

where $\bar{P}_{D,e}$ represents the mean empirical detection probability, computed over the 200 permutations and $P^i_{D,e}$ is the ith permutation. Clearly here $L = 200$.

Table 3.9: Mean and mean absolute error (given within the parenthesis) of the empirical detection probability for different levels of prescribed false-alarm probability when 100 patients are used in the training and 40 in the testing stage.

Number of used eigenfunctions	False-alarm probability		
	$P_F = 0.1$	$P_F = 0.05$	$P_F = 0.01$
1	0.00 (± 0.00)	0.00 (± 0.00)	0.00 (± 0.00)
2	0.00 (± 0.00)	0.00 (± 0.00)	0.00 (± 0.00)
3	0.15 (± 0.09)	0.10 (± 0.08)	0.13 (± 0.07)
4	0.29 (± 0.12)	0.28 (± 0.12)	0.26 (± 0.10)
5	0.69 (± 0.16)	0.68 (± 0.16)	0.67 (± 0.16)
6	0.89 (± 0.09)	0.89 (± 0.09)	0.89 (± 0.09)
7	0.96 (± 0.08)	0.96 (± 0.08)	0.96 (± 0.08)
8	0.85 (± 0.25)	0.85 (± 0.25)	0.85 (± 0.25)
9	0.88 (± 0.22)	0.88 (± 0.22)	0.88 (± 0.22)
10	0.93 (± 0.14)	0.93 (± 0.14)	0.93 (± 0.14)
11	0.86 (± 0.24)	0.86 (± 0.24)	0.86 (± 0.24)
12	0.89 (± 0.20)	0.89 (± 0.20)	0.89 (± 0.20)

Table 3.10: Mean and mean absolute error (given within the parenthesis) of the empirical detection probability for different levels of prescribed false-alarm probability when 110 patients are used in the training and 30 in the testing stage.

Number of used eigenfunctions	False-alarm probability		
	$P_F = 0.1$	$P_F = 0.05$	$P_F = 0.01$
1	0.00 (± 0.00)	0.00 (± 0.00)	0.00 (± 0.00)
2	0.00 (± 0.00)	0.00 (± 0.00)	0.00 (± 0.00)
3	0.19 (± 0.19)	0.17 (± 0.19)	0.16 (± 0.18)
4	0.29 (± 0.21)	0.28 (± 0.20)	0.26 (± 0.20)
5	0.76 (± 0.21)	0.75 (± 0.21)	0.73 (± 0.21)
6	0.89 (± 0.15)	0.89 (± 0.15)	0.89 (± 0.15)
7	0.97 (± 0.06)	0.97 (± 0.06)	0.97 (± 0.06)
8	0.84 (± 0.26)	0.84 (± 0.26)	0.84 (± 0.26)
9	0.90 (± 0.18)	0.90 (± 0.18)	0.90 (± 0.18)
10	1.00 (± 0.00)	1.00 (± 0.00)	1.00 (± 0.00)
11	1.00 (± 0.00)	1.00 (± 0.00)	1.00 (± 0.00)
12	1.00 (± 0.00)	1.00 (± 0.00)	1.00 (± 0.00)

3.4.4 Additional results over a different set of patients

Here, we present the classification results of the proposed method over eleven new patient TRCs that were acquired at the UNM Dermatology Clinic while our method was developed. Out of those 11 patients, 5 of them were diagnosed with malignant condition. It is important to mention that none of these patients were used in the training of the algorithm and that their biopsy result is utilized here only to evaluate the method performance.

In the following table, we report the empirical detection probability averaged over all 200 permutations when 110 patients are used in training the algorithm. We focus our attention to the the range of 5–12 eingepairs because they should present the best performance as predicted by our previous results. In fact, once again, a perfect detection of the malignant lesions is achieved by the algorithm.

Table 3.11: Mean of the empirical detection probability for different levels of prescribed false-alarm probability when 110 patients are used in the training the eleven additional patients are used in the training.

Number of used eigenfunctions	False-alarm probability		
	$P_F = 0.1$	$P_F = 0.05$	$P_F = 0.01$
5	0.31	0.25	0.18
6	0.31	0.25	0.20
7	0.29	0.23	0.21
8	0.35	0.29	0.20
9	0.48	0.50	0.42
10	0.97	0.97	0.97
11	1.00	1.00	1.00
12	1.00	1.00	1.00

3.4.5 Discussion

It was demonstrated by changing the size of the training size that the more the patients that are used to train the algorithm, the better the reliability in estimating the eigenvalue-eigenfunction pairs and the variability in the theoretical performance is minimized as measured by the mean absolute error of the AUC.

The empirical performance achieved by the method when 100 or 80 patients are used to train the algorithm serves as another evidence that a reduce number of patients does not include all the statistical information required to estimate the autocorrelation functions. In the empirical performance evaluation, it seems that ten eigenfunctions is a magic number that ensures that all the malignant cases are detected when 110 patients are used to train the algorithm. Nevertheless, it seems that this behaviour is due to the lack of reliability in estimating beyond the tenth eigenfunction for the benign hypothesis, as was observed in Section 3.4.1 (see, for example, Fig. 3.8). Once the eigenfunctions of one hypothesis becomes noisy, the value of the projection coefficients (namely the KL coefficients) abruptly depart from the value of the coefficients of the other hypothesis, leading to an abrupt change in the test-statistics, which, may yield to the perfect classification of the lesions.

One important thing to mention is the resolution of the error for the presented empirical results. For example, when 100 patients are used in the training, only 8 malignant patients are left for testing the algorithm. That means that if one patient is misclassified as benign, the empirical detection probability will drop from 1.0 (8 out of 8 malignant cases detected) to 0.875 (7 out of 8 malignant cases detected). Under this scenario, if one patient is misclassified the empirical detection probability will match the theoretical detection probability.

It is expected that as we collect more patient data, the algorithm ACFs and the corresponding eigenpairs will reach a plateau, leading to an stable and reliable estimation of the performance of the proposed method. At this stage, our performance metrics are only estimates, making hard to predict an expected performance for a large dataset of patients.

3.5 Robustness of the method to variations in the selection of the lesion boundary

3.5.1 Background

As it was described in Chapter 2, the fist step required to compute the average TRC of a patient is to select the lesion boundary. In the presented studies, we did so by manually defining the boundary of the pigmented area from the visible color picture, which is spatially aligned with respect to the infrared sequence by the image registration procedure. In this section we investigate the robustness of the method to changes in such a selection. The rationale behind the study is given next. On one hand, if the boundary of the lesion is selected to be bigger that the boundary of the pigmented area, such that it contains pigmented and non-pigmented tissue, some of

the pixels used to compute the TRC will be from benign tissue. These benign pixels will clearly affect the average TRC and, as a consequence, the characterization of malignant lesions. On the other hand, if the boundary is selected to be small such that the pixels inside such a selection are purely of malignant tissue, one may expect to have a better characterization of the lesion, but at the tradeoff of having less pixels to compute the average TRC, such an average may not be totally representative of the lesion either.

Here, for each patient, we defined four different boundaries for the lesions to assess the variability in the following way. From the original boundary used in the previous studies we determined the centroid and the area (in pixels) of the region that defines the lesion boundary. By using the same centroid, we adjusted the radii of the region at four different values, chosen such that the area of the resulting regions are one quarter, one half, the double and four times the area of the original selection. These four regions define the four different boundaries for the presented study. In Fig. 3.10 we show by the red lines the four boundaries defined for one patient as an example. The black line on each picture is the original boundary selection made for this specific patient. As before we explore the variability of the eigenvalues and eigenfunctions, the variability in the theoretical and empirical detection probability.

3.5.2 Eigenvalue-eigenfunctions pairs

First, we want to study how the eigenvalues and eigenfunctions change as we have more or less surrounding tissue. Benign lesions and the surrounding tissue should present the same characteristics on their TRCs because both should have similar subcutaneous characteristics. (Authors have declared that the surrounding skin is actually benign tissue [34, 35].) As such, it is expected to observe some degree of insensitivity to the changes in the lesion selection for the eigenpairs of the benign

Figure 3.10: Example of the four areas selected for this study on an actual patient visible image. The selection of the regions (red line) are made such that the area of each selection is (from left to right) one quarter of the original selection, one half, the double and four times the area of the original selection. The original boundary selection is shown in black for each picture as a reference.

hypothesis. On the other hand, by including surrounding tissue to the malignant lesions, we are modifying their average TRC, so the eigenpairs of the malignant hypothesis should be more sensitive to the changes we are introducing here.

Let us focus our attention to Table 3.12 and Fig. 3.11 were we present the changes in the eigenvalues and eigenfunctions, respectively, for the benign hypothesis. As predicted by our rationale regarding the benign nature of the surrounding tissue, the eigenvalues for the benign hypothesis present minimum variability. For example, the maximum difference for the first, most important eigenvalue is $\max \Delta\lambda_{0,1} < 5\%$ for a 400% increment in the selected area (an equivalent increment in the selection area radii of 37 pixels). The same observation can be extended to the eigenfunctions of the benign hypothesis, Fig. 3.11, where one observe virtually no variability for the first 7 most important eigenfunctions. The first real change in the eigenfunction set occurs for the eight eigenfunction, and it is only a phase change, similarly to the one observed when changing the permutation of patients in the training performed in the previous stage. Such a small difference in the eigenvalue and eigenfunctions guarantees that the impact in the theoretical performance (from the perspective of the benign hypothesis) will be minimal for different selections of the lesion.

Table 3.12: Variability of the eigenvalues for the benign hypothesis when the lesion selection area is changed between one quarter until four times the original area utilized in the previous studies.

k	$\lambda_{0,k}$				
	Quarter	Half	Original area	Twofold	Fourfold
1	1128.6	1118.1	1103.4	1076.5	1049.5
2	44.5141	43.5454	41.4486	39.6900	38.4422
3	3.4695	3.3634	3.2924	3.1513	2.8122
4	0.6371	0.6132	0.5735	0.5247	0.4733
5	0.0351	0.0335	0.0321	0.0291	0.0253
6	0.0016	0.0015	0.0014	0.0012	0.0010
7	0.3316×10^{-4}	0.3020×10^{-4}	0.2716×10^{-4}	0.2593×10^{-4}	0.2611×10^{-4}
8	0.1279×10^{-5}	0.0938×10^{-5}	0.0782×10^{-5}	0.0818×10^{-5}	0.0815×10^{-5}
9	0.7397×10^{-7}	0.5502×10^{-7}	0.2422×10^{-7}	0.1238×10^{-7}	0.1×10^{-7}
10	0.8287×10^{-9}	0.7194×10^{-9}	0.5672×10^{-9}	0.2495×10^{-9}	0.1526×10^{-9}
11	0.7274×10^{-11}	0.6929×10^{-11}	0.7980×10^{-11}	0.9319×10^{-11}	0.3899×10^{-11}
12	0.1514×10^{-11}	0.1758×10^{-11}	0.1170×10^{-11}	0.1099×10^{-11}	0.1270×10^{-11}

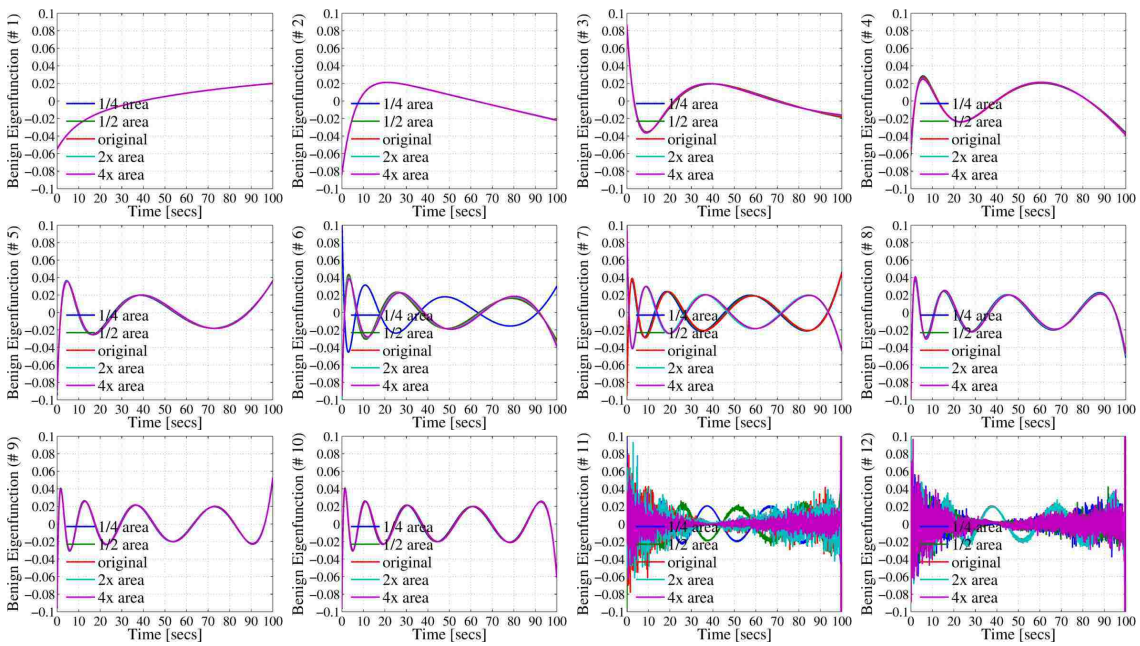


Figure 3.11: Comparison of the eigenfunctions of the benign hypothesis for the four different areas defined in this study.

Table 3.13: Variability of the eigenvalues for the malignant hypothesis when the lesion selection area is changed between one quarter until four times the original area utilized in the previous studies.

k	$\lambda_{1,k}$				
	Quarter	Half	Original area	Twofold	Fourfold
1	2045.2	1981.8	1904.7	1814.6	1740.8
2	54.4070	54.4183	52.9904	51.1407	47.7552
3	15.0219	13.8023	12.4111	10.2148	8.6727
4	2.3186	2.1633	1.8864	1.5662	1.4030
5	0.1205	0.1003	0.0793	0.0631	0.0549
6	0.0077	0.0065	0.0053	0.0038	0.0029
7	0.3666×10^{-3}	0.2916×10^{-3}	0.2356×10^{-3}	0.1708×10^{-3}	0.1313×10^{-3}
8	0.1155×10^{-4}	0.0638×10^{-4}	0.0448×10^{-4}	0.0369×10^{-4}	0.0341×10^{-4}
9	0.8122×10^{-6}	0.4618×10^{-6}	0.2881×10^{-6}	0.1843×10^{-6}	0.1339×10^{-6}
10	0.5395×10^{-7}	0.3232×10^{-7}	0.1395×10^{-7}	0.0688×10^{-7}	0.0606×10^{-7}
11	0.5234×10^{-8}	0.3890×10^{-8}	0.2206×10^{-8}	0.1090×10^{-8}	0.0895×10^{-8}
12	0.2027×10^{-9}	0.1303×10^{-9}	0.0888×10^{-9}	0.0578×10^{-9}	0.0415×10^{-9}

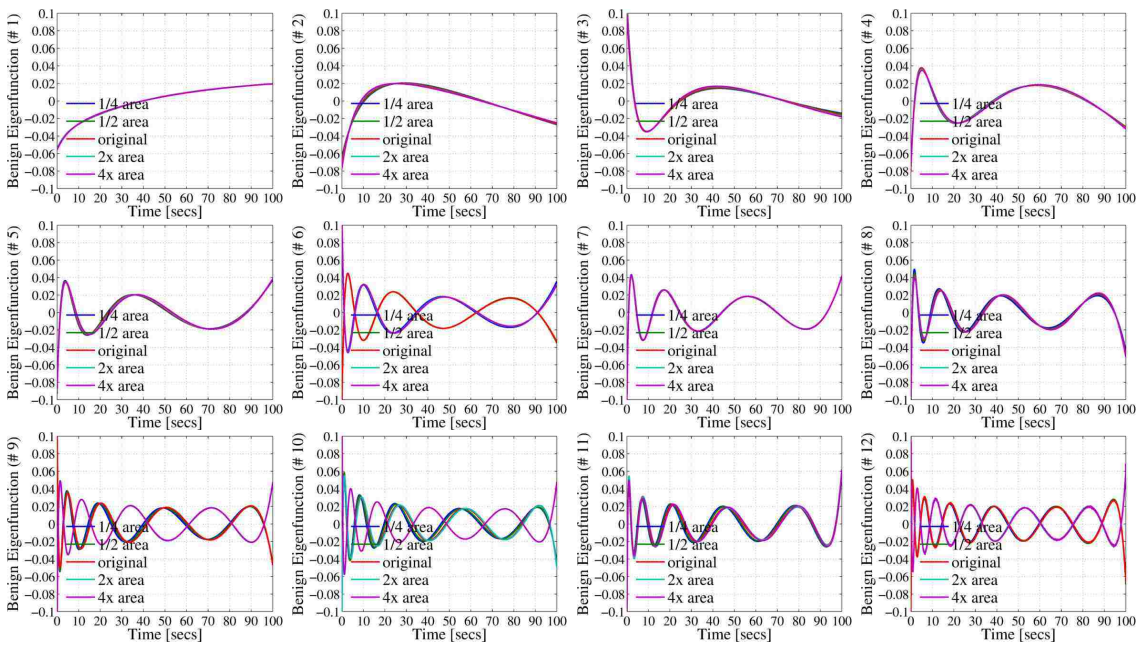


Figure 3.12: Comparison of the eigenfunctions of the malignant hypothesis for the four different areas defined in this study.

Now we turn our attention to the eigenvalues and eigenfunctions from the malignant hypothesis, shown un Table 3.13 and Fig. 3.12, respectively. The first observation one can made is that, for the same average change in the radii of the selection area (i.e., 37 pixels) the maximum variability for the first eigenvalue is $\max \Delta\lambda_{1,1} > 8\%$, almost the double of the benign case, but still small enough to have little effect over the eigenfunctions, as can be noted in Fig. 3.12, where, as for the benign hypothesis, the set is virtually the same except for changes in the phases, which can be easily accounted in the algorithm implementation.

The small variability observed in the eigenpairs under both hypotheses predict an small variability in the theoretical performance of the algorithm, which we study next.

3.5.3 Theoretical performance

We asses the changes in the theoretical performance by obtaining the ROC curves of the approach for each one of the four new regions. We specifically select a single permutation with 100 patients used in the training and 40 in the testing. The idea is, as it was stated before, to evaluate the changes that are only due to the changes in the selection of the lesion and not to changes in the training.

As before, we present the theoretical variability by show the obtained ROC curves (including the obtained with the original lesion selection, for this permutation, as a reference). Such ROC curves are shown in Fig. 3.13. As it was predicted by the study of the eigenvalues and eigenvectors performed in the previous section it can be noted that the obtained ROC curves are essentially equal for the first two rows presented in the figure. Small changes start to appear when more eigenpairs are used, observing an interesting change in the slope for the smaller regions, plateauing at lower detection probabilities as their counterparts with higher number of pixels (i.e.,

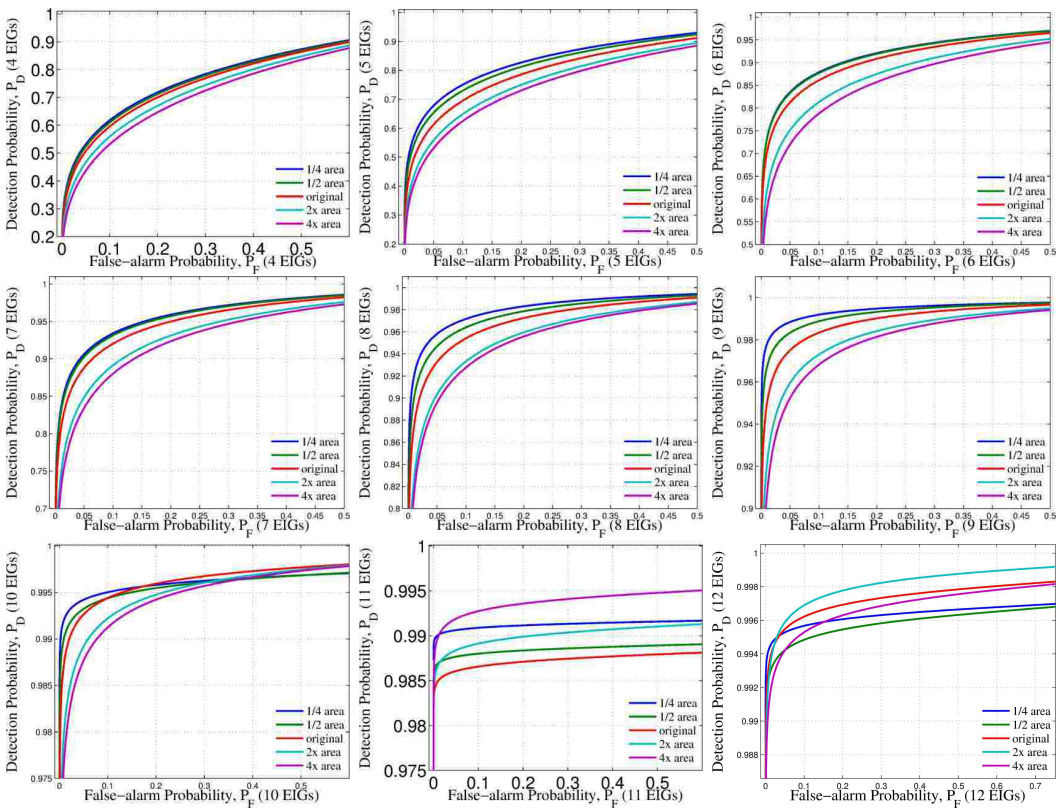


Figure 3.13: Theoretical performance variability as a function of the area utilized to define the boundary of the lesions when we use from four up to twelve eigenfunctions.

the selection with bigger areas). This observation can be justifies by the fact that having less pixels to compute the average TRCs yields to an inaccurate aggregation of the effects of the TRCs. Such a lower detection probability in the plateau is compensated by a faster rate in reaching such a plateau, which can be observed when the AUC is computed (see Table 3.14). It can be concluded, therefore, that since there is a small variability in the eigenpairs when we change the selection of the lesion, then the theoretical performance is virtually insensitive to those changes. The small changes in the AUC quantify such an insensitivity. Next we explore the variability of the empirical detection probability for the same permutation of patients utilized in this section.

Table 3.14: AUC of the theoretical ROC curves for the four different selection regions utilized. The original selected area is included as a reference

k	Quarter	Half	Original area	Twofold	Fourfold
1	0.51	0.51	0.51	0.50	0.50
2	0.58	0.58	0.57	0.57	0.56
3	0.77	0.77	0.76	0.74	0.73
4	0.82	0.82	0.81	0.80	0.78
5	0.89	0.88	0.87	0.85	0.83
6	0.94	0.94	0.94	0.92	0.91
7	0.97	0.97	0.96	0.95	0.95
8	0.98	0.98	0.98	0.97	0.97
9	0.99	0.99	0.99	0.98	0.98
10	0.99	0.99	0.99	0.99	0.99
11	0.99	0.98	0.99	0.99	0.99
12	0.99	0.99	0.99	0.99	0.99

3.5.4 Empirical performance

As before, we evaluate the empirical performance by means of the empirical detection probability for different levels of prescribed false-alarm probabilities of 0.1, 0.05 and 0.01 (i.e., a specificity of 90%, 95% and 99%, respectively). To focus our attention to nine, ten and eleven eigenfunctions to see how the region that defines the lesion affects the empirical performance with respect to the one achieved for this permutation with the original area.

The results are summarized in Table 3.15. It can be observed that the empirical performance for 10 and 11 eigenpairs remains intact. The unchanged performance for these number of features is expected from the observation made earlier regarding the small variability of the eigenpairs and the theoretical performance. If there is no change in these two stages, then KL coefficients for the patients and the optimum threshold will be virtually the same as the original setting. Nevertheless, it can be observed that when 9 eigenpairs are used, then there is a slight variability in the obtained performance. The original $P_{D,e} = 0.875$ achieved with the original lesion

Table 3.15: Empirical detection probability for different levels of prescribed false-alarm probability when 100 patients are used in the training and 40 in the testing stage as a function of the lesion selection

Area	Number of used eigenfunctions	False-alarm probability		
		$P_F = 0.1$	$P_F = 0.05$	$P_F = 0.01$
Quarter	9	0.625	0.25	0.25
	10	1.00	1.00	1.00
	11	1.00	1.00	1.00
Half	9	0.625	0.125	0.125
	10	1.00	1.00	1.00
	11	1.00	1.00	1.00
Original area	9	0.875	0.5	0.375
	10	1.00	1.00	1.00
	11	1.00	1.00	1.00
Twofold	9	0.625	0.375	0.25
	10	1.00	1.00	1.00
	11	1.00	1.00	1.00
Fourfold	9	0.625	0.125	0.125
	10	1.00	1.00	1.00
	11	1.00	1.00	1.00

(recall that this implies 7 true positive and one false negative) changes to the case of having 5 true positives and 3 false negatives. This dramatic drop in performance is due to the small number of patients utilized in testing the algorithm.

3.5.5 Discussion

It was observed that for one particular setting, the eigenvalues and eigenfunction variability was small for a relatively small variation in the selected area. It can be concluded, therefore, that the small variability in the eigenpairs as we change the selection of the lesion, makes the theoretical performance virtually insensitive, which yield to a robust almost invariant empirical performance when ten eigenpairs are used. (Recall that ten eigenpairs were recomended in the previous study as the optimal number of features per hypothesis to correctly classify all the malignant

lesions.)

As a final remark of the present study, we want to mention that even though the changes in the area look big in terms of number of pixels, the real changes (in terms of physical area) is significantly smaller. Indeed, the average change in radii is around 37 pixels, which in physical terms means an average change of about 1cm.

3.6 Concluding remarks

We first performed a separability analysis of the proposed algorithm and proposed a method to determine the maximum number of reliable eigenfunctions based on their spectra. It was observed that the eigenfunctions of both ACFs were virtually the same. Later, it was noted that as we reduce the number of patients, these two sets of eigenfunctions departed in amplitude and their cross orthogonality was lost. As such, there is a practical minimum number of patients that must be utilized to reliably define the eigenpairs for each hypothesis. The author believes that once this practical minimum is attained, both sets of eigenfunctions will be the same, and the problem will be solvable by selecting one of these sets. Unfortunately, there is no way to demonstrate this from a practical perspective without having more patient data.

It was observed in the separability analysis that the algorithm can correctly classify all the malignant cases for different levels of prescribed false-alarm probability as we introduce more statistical information: when using 10 eigenpairs per hypothesis the algorithm achieves $P_{D,e} = 1.0$. This means that the algorithm is capable of detecting all 58 malignant lesions. It is worth mentioning that the error resolution is low, because if we fail one of these 58 patients, then the empirical detection probability will drop to $P_{D,e} = 0.98$, which is the theoretical predicted value as shown in Fig. 2.11.

The proposed decision rule does not impose any restriction over the empirical false-alarm probability. Hence, any departure from the stated assumptions makes the empirical performance to be degraded as compared with the constrained, theoretical performance. A practical way to address this is by estimating the practical distributions and possible correlation of the KL coefficients under both hypothesses. This requires to have a sufficient number of patient, which, unfortunately, are not available at this stage. The empirical distribution of the KL coefficients can be used instead of the assumed Gaussianity. The modified likelihood ratio and corresponding test-statistics will present, for sure, a more realistic empirical performance of the presented in this dissertation. Moreover, this approach will set the actual performance and impact of this work f applied in future clinical studies. It is worth mentioning that other decision rules can be designed in order to evaluate the performance of the algorithm under different objectives. For example, one may directly impose restrictions over the false-alarm probability by, for example minimizing the error probability

At the end of this chapter it was observed that the performance of the algorithm is virtually insensitivity for small changes in the selected area of the lesion. For example, it was observed that the theorerical performance presented a maximum change in performance (measured by the AUC) of 5% for a maximum change of 30% in the radii of the selected region (measured by the number of pixels).

Chapter 4

Generalization of the method to self-referenced dual-signal setting

4.1 Introduction

It was demonstrated in the previous chapter that the skin cancer detection approach leads to the most accurate and robust results presented so far. Nevertheless the requirement of twenty KL coefficients per patient (ten KL coefficients per hypothesis) is excessive when the reliability of the estimated eigenvalue-eigenfunction pairs beyond seven eigenfunctions is not high.

In this chapter, we study the inclusion of a reference signal to the detection problem, obtained locally from the very same patient under study. More precisely, for each hypothesis, we define a self-reference signal from the tissue that surrounds the suspicious lesions. The hope is that the by self-referencing the patient's TRC, abnormal features from the lesion TRCs can be compensated out and, as a consequence, less KL coefficients are required to correctly detect the malignant lesions.

Including the reference signal from the surrounding tissue into the detection problem can be stated by the following hypothesis-testing problem

$$H_0 \ : \boldsymbol{Y}(t) = \boldsymbol{S}(t;\boldsymbol{\Theta}_0) = \begin{bmatrix} S^{(1)}(t;\boldsymbol{\Theta}_0^{(1)}) \\ S^{(2)}(t;\boldsymbol{\Theta}_0^{(2)}) \end{bmatrix} \ , \quad t \in [0,T] \ , \tag{4.1a}$$

$$H_1 \ : \boldsymbol{Y}(t) = \boldsymbol{S}(t;\boldsymbol{\Theta}_1) = \begin{bmatrix} S^{(1)}(t;\boldsymbol{\Theta}_1^{(1)}) \\ S^{(2)}(t;\boldsymbol{\Theta}_1^{(2)}) \end{bmatrix} \ , \quad t \in [0,T] \ , \tag{4.1b}$$

where the entries of the vector of random processes under the jth hypothesis, namely $S^{(1)}(t;\boldsymbol{\Theta}_j^{(1)})$ and $S^{(2)}(t;\boldsymbol{\Theta}_j^{(2)})$, are defined to be the lesion and the reference TRC, respectively, under the jth hypothesis. As before, we define the lesion TRC as the average TRC over all the pixels within the lesion boundary. Similarly, we define the reference TRC as the average TRC over all the pixels *outside* the lesion boundary. Examples of dual TRCs for two subjects with benign condition and two subjects with malignant condition are shown in Fig. 4.1. We begin the algorithm generalization by first describing how to represent vector random processes in terms of series expansions.

4.2 Representation of vector random processes

In the present section, we follow Van Tress [72] explanation of vector random processes (specifically Chapter 3) and it is included here for completeness.

Consider the vector containing N random process

$$\boldsymbol{X}(t) = \begin{bmatrix} X_1(t) \\ X_2(t) \\ \vdots \\ X_N(t) \end{bmatrix} \ , \tag{4.2}$$

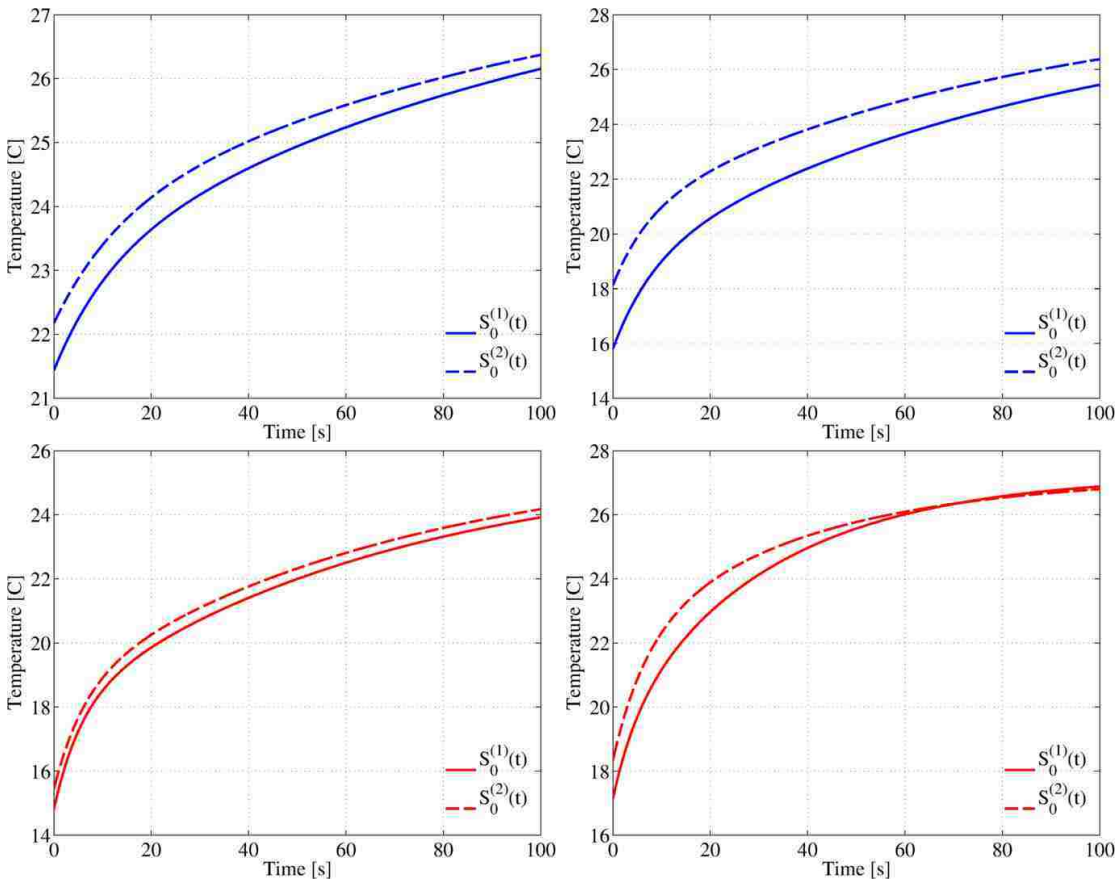


Figure 4.1: Examples of dual TRC from two subjects with benign condition (top row) and two patients with malignant condition (bottom row).

which can be generated, for example, by N sensors located at different spatial locations. We are interested to obtain a series expansion representation for this vector process such that the classical detection theory can be applied as before. The mean of the random process, $\boldsymbol{m}_X(t) = \mathsf{E}\left[\boldsymbol{X}(t)\right]$, is assumed to be zero. The autocorrelation

of this vector process is defined as the $N \times N$ matrix

$$\begin{aligned} \boldsymbol{R}(t,u) &\triangleq \mathsf{E}\left[\boldsymbol{X}(t)\boldsymbol{X}^T(u)\right] \\ &= \begin{bmatrix} \mathsf{E}\left[X_1(t)X_1(u)\right] & \mathsf{E}\left[X_1(t)X_2(u)\right] & \cdots & \mathsf{E}\left[X_1(t)X_N(u)\right] \\ \mathsf{E}\left[X_2(t)X_1(u)\right] & \mathsf{E}\left[X_2(t)X_2(u)\right] & \cdots & \mathsf{E}\left[X_2(t)X_N(u)\right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{E}\left[X_N(t)X_1(u)\right] & \mathsf{E}\left[X_N(t)X_2(u)\right] & \cdots & \mathsf{E}\left[X_N(t)X_N(u)\right] \end{bmatrix} \end{aligned} \tag{4.3}$$

The complexity of the vector processes can be passed to the eigenfunctions (i.e., represent the signals by vector eigenfunctions and scalar eigenvalues) or to the eigenvalues (i.e., represent the signals by scalar eigenfunctions and matrix eigenvalues). Van Tress [72] and Oya *et al.* [77] stated that the optimal way to address this problem is by means of the former, since it was proved by Kelly and Root [78] that the vector eigenfunction with scalar eigenvalue representation is a generalization of the KL expansion.

Hence, let us assume that we have a complete set of vector eigenfunctions $\{\boldsymbol{\Phi}_k\}_{k=1}^{\infty}$ to represent the vector random process, where each vector eigenfunction is defined by

$$\boldsymbol{\Phi}_k(t) = \begin{bmatrix} \phi_k^{(1)}(t) \\ \phi_k^{(2)}(t) \\ \vdots \\ \phi_k^{(N)}(t) \end{bmatrix} , \tag{4.4}$$

and the vector random process $\boldsymbol{X}(t)$ can be expanded over this vector eigenfunctions by

$$\boldsymbol{X}(t) = \sum_{k=1}^{\infty} X_k \boldsymbol{\Phi}_k(t) \ . \tag{4.5}$$

The corresponding expansion coefficients associated with this representation are

given by

$$X_k \triangleq \int_0^T \boldsymbol{\Phi}_k^T(t)\boldsymbol{X}(t)\ dt = \int_0^T \boldsymbol{X}^T(t)\boldsymbol{\Phi}_k(t)\ dt. \tag{4.6}$$

As before, we want the projection coefficients to be uncorrelated, i.e., $\mathsf{E}\left[X_k X_\ell\right] = \lambda_\ell \delta_{k,\ell}$, thus

$$\begin{aligned}
\lambda_\ell \delta_{k,\ell} &= \mathsf{E}\left[X_k X_\ell\right] = \mathsf{E}\left[\int_0^T \boldsymbol{\Phi}_k^T(t)\boldsymbol{X}(t)\ dt \int_0^T \boldsymbol{X}^T(u)\boldsymbol{\Phi}_\ell(u)\ du\right] \\
&= \int_0^T \int_0^T \boldsymbol{\Phi}_k^T(t)\ \mathsf{E}\left[\boldsymbol{X}(t)\boldsymbol{X}^T(u)\right]\ \boldsymbol{\Phi}_\ell(u)\ dt\ du \\
&= \int_0^T \boldsymbol{\Phi}_k^T(t)\left[\int_0^T \boldsymbol{R}(t,u)\boldsymbol{\Phi}_\ell(u)\ du\right]\ dt\ .
\end{aligned} \tag{4.7}$$

As in the scalar case, we will achieve the desired uncorrelated coefficients if the quantity within the brackets in (4.7) equals $\lambda_\ell \boldsymbol{\phi}_\ell(t)$, or

$$\int_0^T \boldsymbol{R}(t,u)\boldsymbol{\Phi}_\ell(u)\ du\ dt = \lambda_\ell \boldsymbol{\Phi}_\ell(t)\ , \tag{4.8}$$

which is the equivalent of the eigenvalue-eigenfunction integral equation for vector-valued processes. According to Van Tress [72] and Oya *et al.* [77] this representation satisfy Mercer's theorem; as a consequence

$$\boldsymbol{R}(t,u) = \mathsf{E}\left[\boldsymbol{X}(t)\boldsymbol{X}^T(u)\right] = \sum_{k=1}^{\infty} \lambda_k \boldsymbol{\Phi}_k(t)\boldsymbol{\Phi}_k^T(u)\ . \tag{4.9}$$

A useful property of this expansion is that the expansion coefficients are scalar, which is a property we exploit in the following section.

4.3 Dual-signal detection problem

4.3.1 Autocorrelation functions

In the same way that we solved the single-signal detection problem, in this approach we assume that each hypothesis have its own autocorrelation function, namely

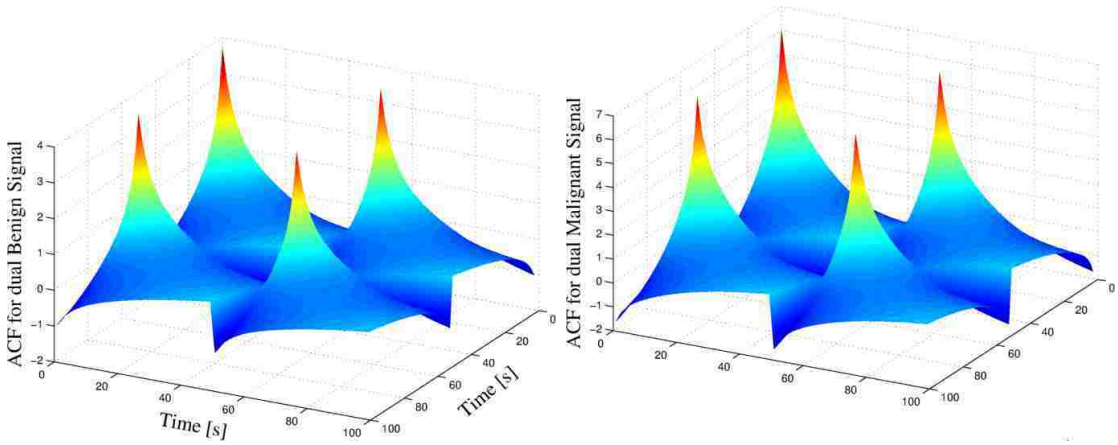


Figure 4.2: Matrix of autocorrelation functions for the benign (left) and the malignant (right) hypotheses. The presented functions were computed using all 140 patients with known diagnosis.

$\boldsymbol{R}_0(t,u)$ and $\boldsymbol{R}_1(t,u)$, with $\boldsymbol{R}_j(t,u) = \mathsf{E}\left[\boldsymbol{S}_j(t;\boldsymbol{\Theta}_j)\boldsymbol{S}_j^T(t;\boldsymbol{\Theta}_j)\right]$, for $j = 0, 1$. Hence, $\boldsymbol{R}_j(t,u)$, $j = 0, 1$ is 2×2 matrix of autocorrelation functions. Since each entry of this matrix is simply an scalar autocorrelation function, we can define the entire matrix in the same way we did for the single signal approach. The obtained autocorrelation functions for the vector representation of are depicted in Fig. 4.2 for the benign and the malignant hypotheses.

As before, each autocorrelation will have its own set of orthonormal eigenfunctions. The kth vector eigenfunction for the jth autocorrelation function, namely $\boldsymbol{\Phi}_{j,k}(t) = \left[\phi_j^{(1)}(t)\ \phi_j^{(2)}(t)\right]^T$, is determined by the eigenvalue-eigenfunction integral equation

$$\lambda_{j,k}\boldsymbol{\Phi}_{j,k}(t) = \int_0^T \boldsymbol{R}_j(t,u)\boldsymbol{\Phi}_{j,k}(u)\ du\ .$$

We solve this integral equation by the Nystrom method, which was introduced in Section 2.6.4 for the single TRC approach. The resulting first sixteen eigenfunctions for the benign and malignant hypothesis are depicted in Fig. 4.3 and Fig. 4.4, respectively. It can be noted, that the resulting vector eigenfunctions can only be reliably extracted from the autocorrelation functions until a certain number of eigenvalues.

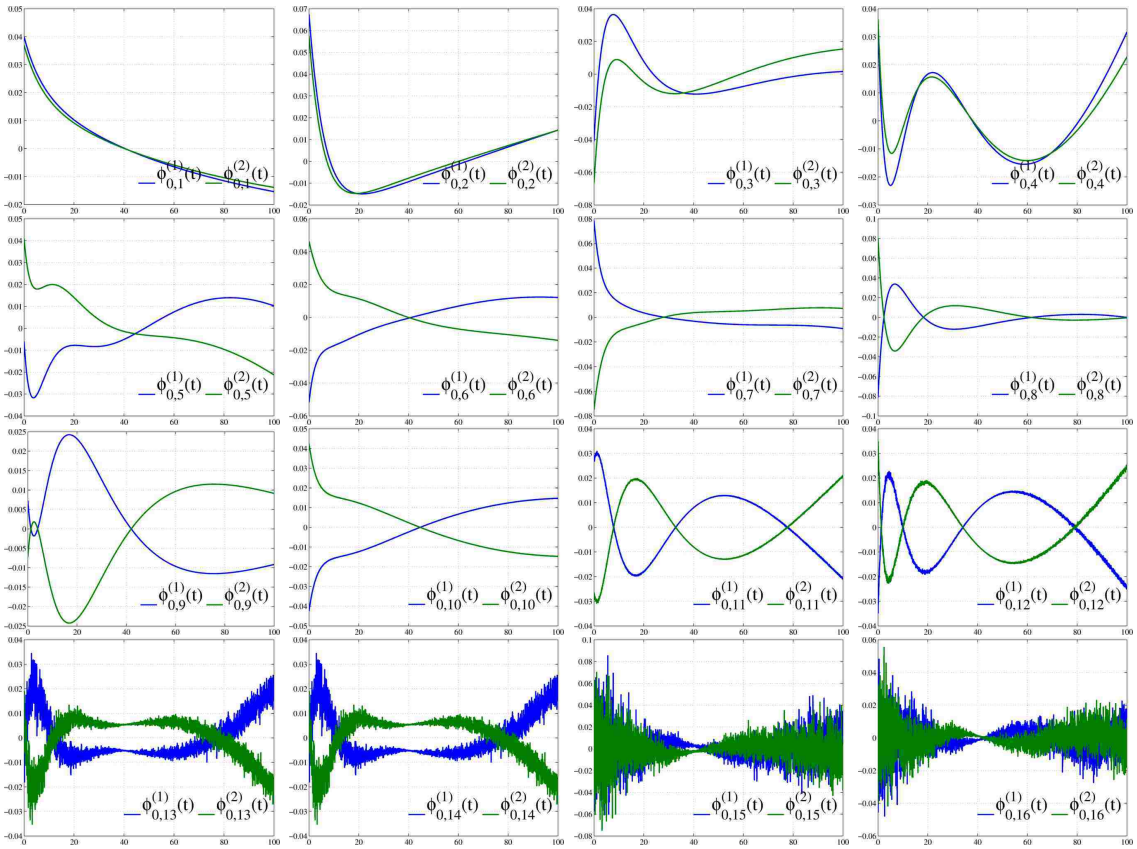


Figure 4.3: Benign vector eigenfunctions

As in the single TRC approach, twelve eigenfunctions is the upper limit. We will see later, that as we reduce the number of patients to train the algorithm, complex components start to appear in the eigenfunctions.

The resulting vector eigenfunctions allow us to represent the dual TRC under each hypothesis with scalar expansion coefficients by

$$\boldsymbol{S}(t;\boldsymbol{\Theta}_j) = \sum_{k=1}^{\infty} S_{j,k}\boldsymbol{\Phi}_k(t) \ . \tag{4.10}$$

Since this series representation for vector random processes can be considered as a generalization of the KL expansion [72, 77], then the scalar expansion coefficients contain all the statistical features of the dual TRC, in a similar way that the KL

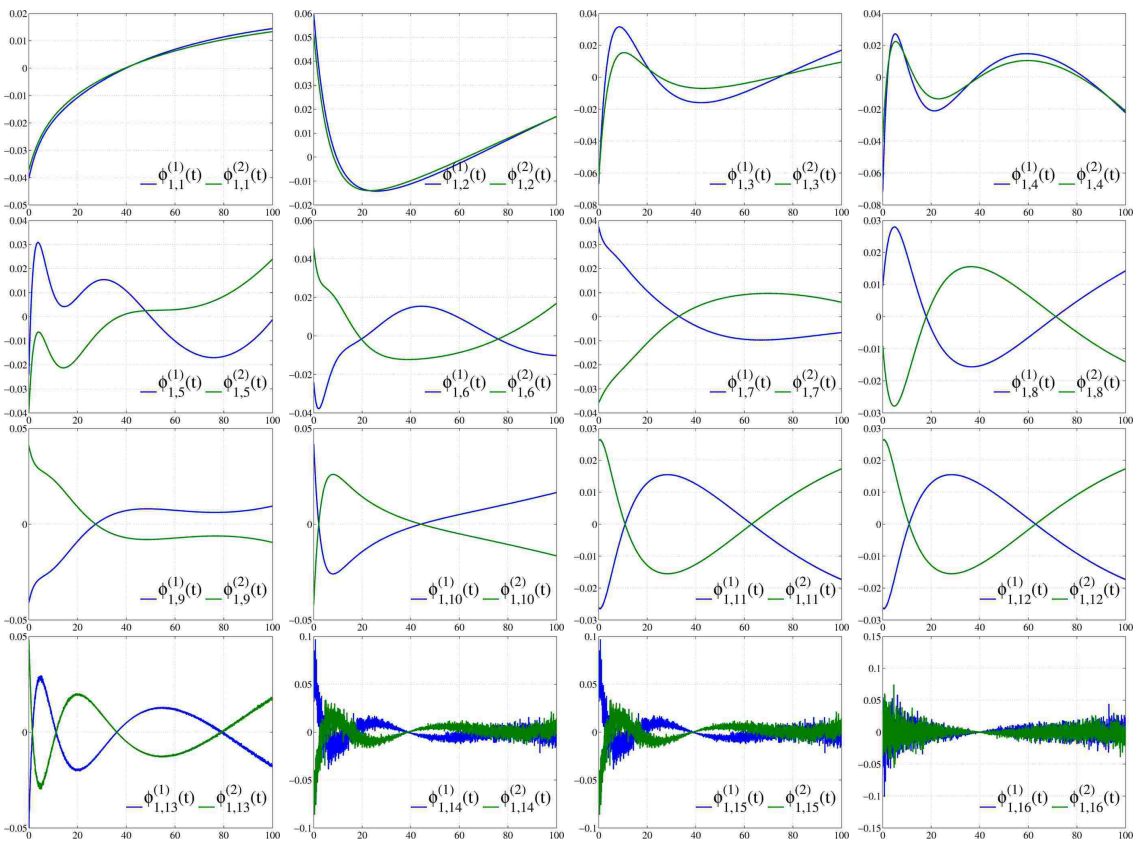


Figure 4.4: Malignant vector eigenfunctions

coefficients contained the statistical temporal features in the single TRC approach. As a consequence, we can also solve the dual TRC problem by solving the statistical equivalent problem of detecting the malignancy of the lesion based on the expansion coefficients. In symbols, the dual TRC detection problem can be recast as

$$H_0 \ : \ S_{0,k} \ , \quad k = 1, 2, \ldots \tag{4.11}$$

$$H_1 \ : \ S_{1,k} \ , \quad k = 1, 2, \ldots \tag{4.12}$$

where, as before, we assume that the expansion coefficients are Gaussian as a consequence of assuming that the TRCs are Gaussian random processes. The mean of

the expansion coefficients under the jth hypothesis is

$$\begin{aligned} \mathsf{E}\left[S_{j,k}\right] &= \mathsf{E}\left[\int_0^T \boldsymbol{\Phi}_k^T(t)\boldsymbol{S}(t;\boldsymbol{\Theta}_j)\ dt\right] \\ &= \int_0^T \boldsymbol{\Phi}_k^T(t)\mathsf{E}\left[\boldsymbol{S}(t;\boldsymbol{\Theta}_j)\right]\ dt = 0\ , \end{aligned} \tag{4.13}$$

by the assumption that the Gaussian random processes that define the entries of the dual TRC have zero mean. The covariance between these coefficients under the same hypothesis is

$$\begin{aligned} \mathsf{cov}\left(S_{j,k}, S_{j,\ell}\right) &= \mathsf{E}\left[S_{j,k}S_{j,\ell}\right] \\ &= \mathsf{E}\left[\int_0^T \boldsymbol{\Phi}_k^T(t)\boldsymbol{S}(t;\boldsymbol{\Theta}_j)\ dt \int_0^T \boldsymbol{S}^T(u;\boldsymbol{\Theta}_j)\boldsymbol{\Phi}_\ell(u)\ du\right] \\ &= \int_0^T\int_0^T \boldsymbol{\Phi}_k^T(t)\mathsf{E}\left[\boldsymbol{S}(t;\boldsymbol{\Theta}_j)\boldsymbol{S}^T(u;\boldsymbol{\Theta}_j)\right]\boldsymbol{\Phi}_\ell(u)\ du\ dt \\ &= \int_0^T \boldsymbol{\Phi}_k^T(t)\int_0^T \boldsymbol{R}_j(t,u)\boldsymbol{\Phi}_\ell(u)\ du\ dt \\ &= \int_0^T \boldsymbol{\Phi}_k^T(t)\lambda_{j,\ell}\boldsymbol{\Phi}_\ell(t)\ dt \\ &= \lambda_{j,\ell}\delta_{j,\ell}\ . \end{aligned}$$

Now, since the expansion coefficients are also Gaussian, they are independent. It is clear now, that under the series expansion utilized in this approach, we have transformed the original dual TRC problem onto a new problem of scalar and independent coefficients, that follow a Gaussian distribution with variance equal to the corresponding eigenvalues. This is exactly the same problem that we already solved for the single TRC approach, and therefore, the mathematical structure of the likelihood ratio and the distribution of the test-statistic are the same. Moreover, we utilize the same structure of the NP decision rule, because the resulting test-statistics is a linear combination of χ^2-distributed random variables (just as in the single-TRC approach) but with different eigenvalues, and, as a consequence, different coefficients for the CDFs defined in (2.52).

In the following sections, we evaluate the theoretical and empirical performance for the dual TRC approach and compare its performance with the single-TRC approach introduced in Chapter 2.

4.4 Theoretical performance

Here we use the same training settings used in Chapter 3, i.e., we train the algorithm using 60, 80, 100 and 110 patients and for each training size we perform the same 200 permutations of patients to evaluate the performance variability. We compare the theoretical performance by means of the ROC curves between the original (single TRC) algorithm and the generalization presented in this chapter. In Fig. 4.5 we present the ROC curve comparison for the single-TRC approach (blue ROC) and the dual-TRC approach (red ROC) for different number of used eigenfunctions. Please note that we are showing here the average ROC curves over the 200 permutations.

The results for 80, 100 and 110 patients used to train both algorithm are shown in Fig. 4.6, Fig. 4.7 and Fig. 4.8, respectively. The results demonstrate that the dual TRC approach actually improves the theoretical performance upon the single TRC approach when the same combination of patients is utilized to train the skin-cancer detector under both methods. It can be noted from the presented results that the theoretical performance of the dual TRC approach is better that the theoretical performance of the single TRC approach when a few eigenvalue-eigenfunction pair are used. As we increase the number of these pairs, the difference in the performance between both approaches becomes smaller. We compare the two approaches by means of the mean AUC for both hypothesis as we increase the number of eigenfunctions. The mean AUC over 200 permutations when 60, 80, 100 and 110 patients are used in the training is depicted in Fig. 4.9. It can be observed in this figure that the dual-TRC approach requires one eigenpair less than the signal-TRC approach to achieve

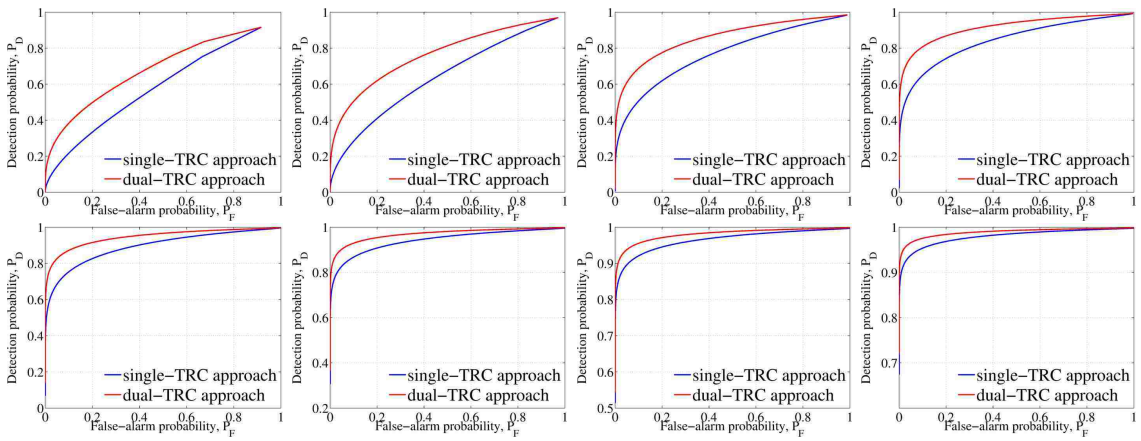


Figure 4.5: Comparison of the mean ROC curves when 60 training are used to train the single-TRC algorithm (blue) and the dual-TRC algorithm (red).

the same performance when 60 patients are used. For example, to achieve an average AUC of 0.9 the single-TRC approach requires 5 eigenpairs. The dual-TRC approach can achieve a theoretical AUC of 0.93 with only 4 eigenpairs.

This observation becomes more relevant as we include more patients in training the algorithms. For example, when we use 110 patients in training the method we

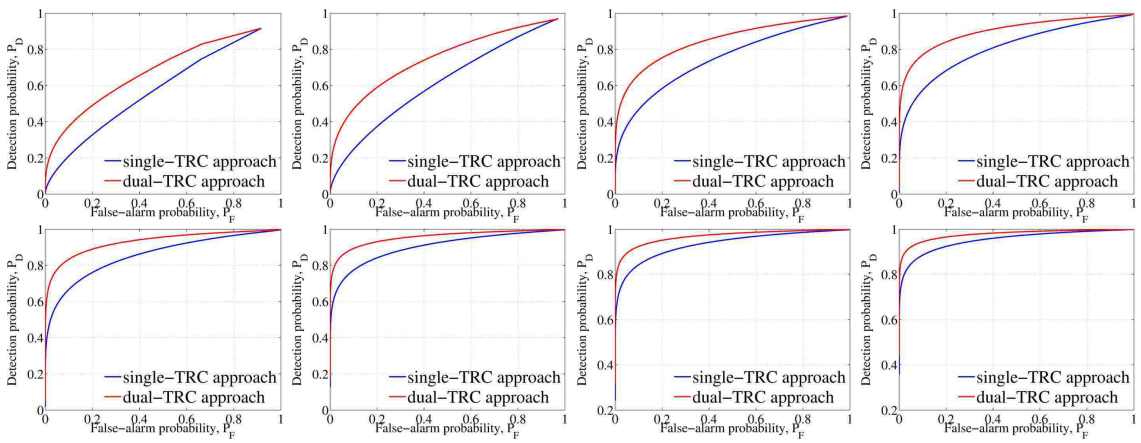


Figure 4.6: Comparison of the mean ROC curves when 80 training are used to train the single-TRC algorithm (blue) and the dual-TRC algorithm (red).

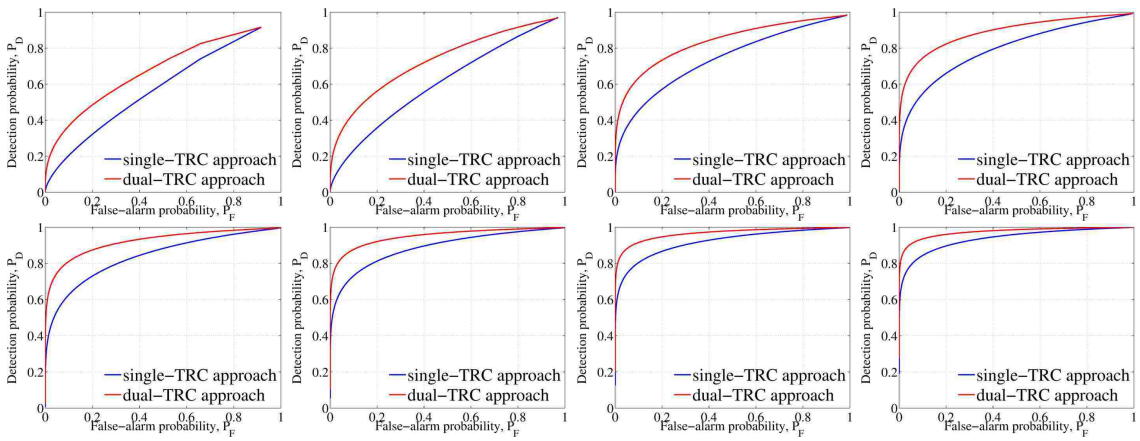


Figure 4.7: Comparison of the mean ROC curves when 100 training are used to train the single-TRC algorithm (blue) and the dual-TRC algorithm (red).

can note in Fig. 4.9(bottom-right) that the dual-TRC approach achieves the same 0.9 AUC with two eigenpair less than the single-TRC counterpart. Please note that the comparison between the single and dual TRC approaches is made only for the first eigenpairs only. The reason to do so is that after certain number of eigenvalues, for a few permutations of patients, the eigenfunctions started to show some imaginary component. Since the imaginary part in the eigenfunctions were in the scale 10^{-13},

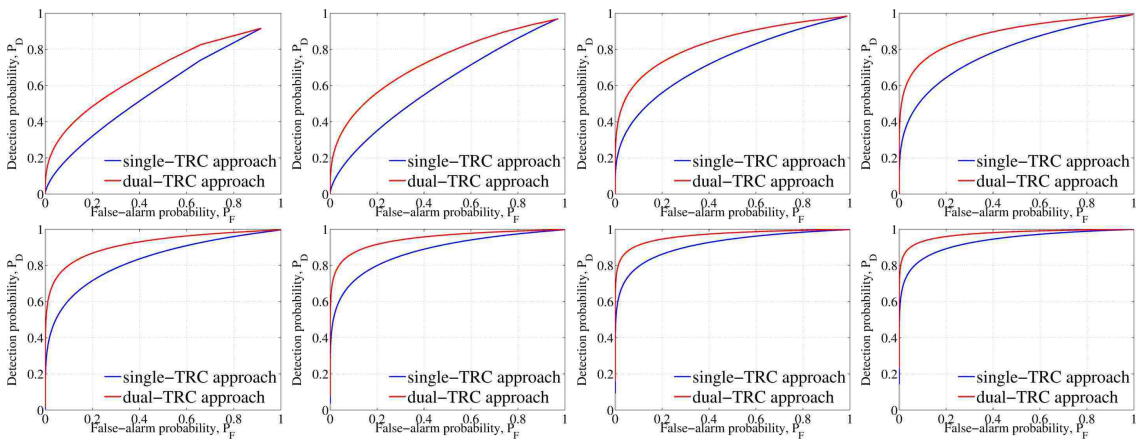


Figure 4.8: Comparison of the mean ROC curves when 110 training are used to train the single-TRC algorithm (blue) and the dual-TRC algorithm (red).

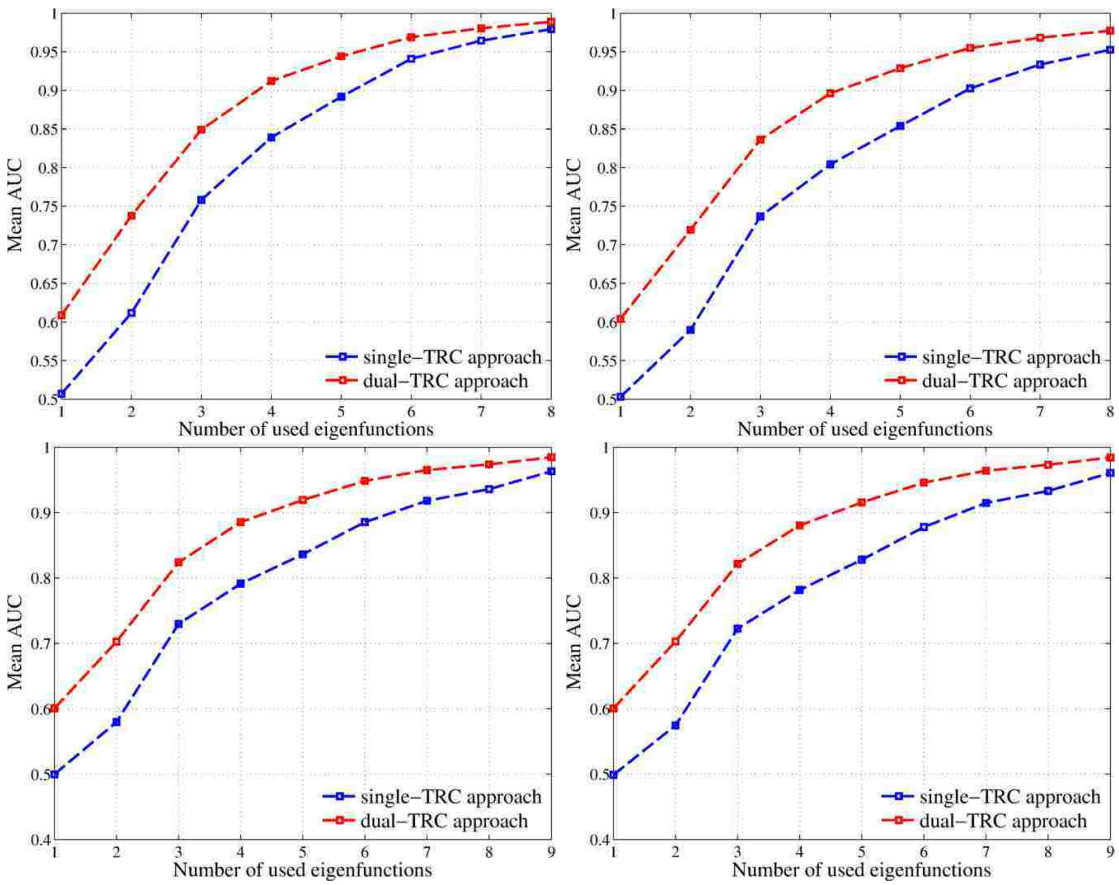


Figure 4.9: Comparison of the single TRC and the double TRC approach by using the mean AUC for different number of used eigenvalue-eigenfunction pairs, using 60 patients in the training (top-left), 80 patients (top-right), 100 patients (bottom-left) and 110 patients (bottom-right).

the author believes that this problem may occur because of the numerical integration performed by the Nystrom method and the trapezoidal quadrature rule. When other rules where explored the complex eigenfunctions were gone but the orthogonality of the components was lost. As a consequence, we utilized the trapezoidal quadrature rule and searched for the maximum number of real eigenfunctions for each case. Moreover, when the training size was increased to 100 or 110 patients, we were able to extract a nine real eigenfunctions, which suggest that the problem may also be related with the *statistical information* contained within the estimated ACFs, which

may not be rich enough to permit the computation of dual-eigenfunctions.

Some conclusions can be drawn with the results presented so far. It is clear that the mean performance over the same 200 permutations is improved with the help of the self-reference signal introduced in this chapter. The most important characteristic of the proposed alternative is that the detection is still performed over an scalar expansion coefficient, which has equivalent statistical features of the dual TRCs in a similar fashion as the KL coefficients characterized the single TRCs. It is clear that this approach present the same limitations that the single TRC approach developed in Chapter 2 regarding the presence of signal under both hypotheses. As before, if the ACFs have their particular sets of eigenvalues and eigenfunctions, the expansion coefficients may not be totally representative of both hypothesis. It may be worth exploring the alternative of whitening one of the ACFs and solve the equivalent problem as it was discussed before. This idea is proposed in the last chapter, under the future work research ideas.

Next, we compute the empirical performance of the dual TRC approach and compare it with the performance of the single TRC approach.

4.5 Empirical performance

In this section we compare the empirical performance, (measured by the empirical detection probability) of the dual TRC and compare the results to those from the single TRC approach. We restrict our attention o the first eigenpairs due to the numerical limitation discussed in the previous section.

We present the mean and mean absolute error of the same 200 permutations when 80, 100 and 110 patients are used to train both algorithm. We use the same permutations of patients to train and test both algorithm in order to highlight only

the difference in their performance without introducing artificial changes due to better training patients.

In Table 4.1 we present the results when 80 patients are used in the training. For comparison purposes we include the empirical detection probability of the single-TRC approach. (Here s-TRC and d-TRC stand for single and dual TRC approach, respectively.) The first difference observed is the clearly improved performance of the dual TRC approach, achieving almost perfect detection with few eigenpairs. For example, for a empirical false-alarm probability of 0.01, the single TRC approach achieves an average detection probability of 0.94 ($\pm$ 0.11) for 7 eigenpairs. The dual TRC approach on the other hand achieves a 0.99 ($\pm$ 0.01) when 5 eigenpairs are utilized. This difference in performance with less eigenpairs is the same observed in the theoretical performance study. Similar observations can be made for the case when 100 patients are utilized in training both algorithms. Interestingly, for the case of 7 eigenpairs, the performance of the single TRC approach is better than the performance of the dual TRC approach. Please recall that for this setting only 40 patients are used in the testing and only 8 of them are malignant. Therefore

Table 4.1: Mean and mean absolute error (given within the parenthesis) of the empirical detection probability for different levels of prescribed false-alarm probability when 80 patients are used in the training and 60 in the testing stage. Here, s-TRC and d-TRC stand for single TRC and double TRC approaches, respectively.

Number of used eigenfunctions	False-alarm probability			
	$P_F = 0.1$		$P_F = 0.01$	
	s-TRC	d-TRC	s-TRC	d-TRC
1	0.00 ($\pm$ 0.00)	0.98 ($\pm$ 0.03)	0.00 ($\pm$ 0.00)	0.95 ($\pm$ 0.05)
2	0.02 ($\pm$ 0.04)	0.97 ($\pm$ 0.04)	0.02 ($\pm$ 0.04)	0.94 ($\pm$ 0.06)
3	0.42 ($\pm$ 0.14)	0.98 ($\pm$ 0.03)	0.42 ($\pm$ 0.14)	0.94 ($\pm$ 0.05)
4	0.46 ($\pm$ 0.12)	0.82 ($\pm$ 0.12)	0.46 ($\pm$ 0.12)	0.78 ($\pm$ 0.12)
5	0.80 ($\pm$ 0.11)	0.99 ($\pm$ 0.01)	0.80 ($\pm$ 0.11)	0.99 ($\pm$ 0.01)
6	0.91 ($\pm$ 0.09)	0.96 ($\pm$ 0.05)	0.91 ($\pm$ 0.09)	0.96 ($\pm$ 0.05)
7	0.94 ($\pm$ 0.11)	0.90 ($\pm$ 0.13)	0.94 ($\pm$ 0.11)	0.90 ($\pm$ 0.13)
8	0.81 ($\pm$ 0.31)	0.90 ($\pm$ 0.11)	0.81 ($\pm$ 0.31)	0.90 ($\pm$ 0.11)

Table 4.2: Mean and mean absolute error (given within the parenthesis) of the empirical detection probability for different levels of prescribed false-alarm probability when 100 patients are used in the training and 40 in the testing stage. Here, s-TRC and d-TRC stand for single TRC and double TRC approaches, respectively.

Number of used eigenfunctions	False-alarm probability			
	$P_F = 0.1$		$P_F = 0.01$	
	s-TRC	d-TRC	s-TRC	d-TRC
1	0.00 (± 0.00)	0.97 (± 0.05)	0.00 (± 0.00)	0.91 (± 0.09)
2	0.00 (± 0.00)	0.96 (± 0.06)	0.00 (± 0.00)	0.90 (± 0.09)
3	0.15 (± 0.09)	0.96 (± 0.05)	0.13 (± 0.07)	0.89 (± 0.09)
4	0.29 (± 0.12)	0.81 (± 0.13)	0.26 (± 0.10)	0.74 (± 0.13)
5	0.69 (± 0.16)	0.99 (± 0.01)	0.67 (± 0.16)	0.99 (± 0.01)
6	0.89 (± 0.09)	0.97 (± 0.05)	0.89 (± 0.09)	0.97 (± 0.05)
7	0.96 (± 0.08)	0.88 (± 0.15)	0.96 (± 0.08)	0.88 (± 0.15)
8	0.85 (± 0.25)	0.84 (± 0.15)	0.85 (± 0.25)	0.84 (± 0.15)
9	0.88 (± 0.22)	0.92 (± 0.12)	0.88 (± 0.22)	0.92 (± 0.12)

an average empirical detection probability of 0.88 indicates that, in average one malignant patient was misclassified as benign.

The real difference is observed when 110 patients are utilized to train the al-

Table 4.3: Mean and mean absolute error (given within the parenthesis) of the empirical detection probability for different levels of prescribed false-alarm probability when 110 patients are used in the training and 30 in the testing stage. Here, s-TRC and d-TRC stand for single TRC and double TRC approaches, respectively.

Number of used eigenfunctions	False-alarm probability			
	$P_F = 0.1$		$P_F = 0.01$	
	s-TRC	d-TRC	s-TRC	d-TRC
1	0.00 (± 0.00)	0.95 (± 0.09)	0.00 (± 0.00)	0.90 (± 0.15)
2	0.00 (± 0.00)	0.94 (± 0.10)	0.00 (± 0.00)	0.89 (± 0.15)
3	0.19 (± 0.19)	0.96 (± 0.07)	0.16 (± 0.18)	0.90 (± 0.15)
4	0.29 (± 0.21)	0.79 (± 0.21)	0.26 (± 0.20)	0.74 (± 0.21)
5	0.76 (± 0.21)	1.00 (± 0.00)	0.73 (± 0.21)	1.00 (± 0.00)
6	0.89 (± 0.15)	0.96 (± 0.06)	0.89 (± 0.15)	0.96 (± 0.06)
7	0.97 (± 0.06)	0.74 (± 0.29)	0.97 (± 0.06)	0.74 (± 0.29)
8	0.84 (± 0.26)	0.86 (± 0.19)	0.84 (± 0.26)	0.86 (± 0.19)
9	0.90 (± 0.18)	0.84 (± 0.23)	0.90 (± 0.18)	0.84 (± 0.23)

gorithm. In this case, a perfect classification of all malignant cases is achieved in all 200 permutations with only 5 eigenpairs per hypothesis. The same performance was achieved by the single-TRC approach with 110 training patients with exactly the double of the eigenpairs. It seems, therefore, that the inclusion of the reference signal can actually compensate some anomalous behavior of the the lesion TRCs.

4.6 Concluding remarks

In this chapter, we have introduced and discussed an approach to represent vector random processes that in the literature is treated as the vector generalization of the KL expansion. Using this representation we have solved the detection problem in which each hypothesis is characterized by a dual TRC. In this work, we have defined the two random signals to be the lesion average TRC and a reference TRC obtained locally from the lesion's surrounding tissue. The approach followed in this chapter have the property of using scalar expansion coefficients, equivalent to the KL coefficients for the scalar TRC approach. Hence, after the expansion is applied to the vector random processes, the dual-TRC detection problem becomes equivalent to the single-TRC detection problem explored and presented in Chapter 2. As a consequence, once the expansion coefficients are characterized, they contain most of the statistical information of both of the signals of each hypothesis, and the solution is trivial because all the properties and limitations discussed for the single TRC approach hold for this new alternative.

We explored both the theoretical and empirical performance of the proposed alternative using the same training settings utilized to test the robustness of the single-TRC approach. By comparing the theoretical performance by means of the AUC for different number of eigenvalue-eigenfunction pairs, the dual-TRC approach achieves the same performances as the single-TRC approach by requiring less eigenpairs. The

empirical performance follows the same trend, allowing, for example, 100% detection of malignant patients using one half of the eigenpairs that the single-TRC approach required for the same training/testing setting.

It is clear that utilizing the same approach, any array of detectors can be exploited to solve different sensing problems. For example, the entire FPA of the IR camera can be considered as a column array of sensors, whose spatial correlations are accounted in the model of the random processes. By using vector eigenfunctions the entire array (or portions of it) can be represented with scalar expansion coefficients allowing, for example, to represent areas with similar conditions with a single scalar (but random) variable. This seems to be an interesting idea to explore in the context of multicolor (spectral) classification.

One major point that must be mentioned here, is that this alternative was proven to be optimal (as stated by Van Tress [72] and Oya *et al.* [77]) but the author of this dissertation did not have access to the original manuscript that does so. As a consequence, some optimality condition may have been overpassed without the required attention. Thus, the results presented in this chapter must be considered as an upper bound for the field performance achievable in a real-life implementation.

Chapter 5

Generalization of the method to include spatial features of the lesion

5.1 Introduction

In this chapter we explore the spatial features of the lesions by analyzing the visible-color image acquired for each patient. This exploration is made with two main purposes in mind: (1) Automatically define a boundary for the every suspicious lesion based on the color contrast between the pigmented and the surrounding tissue and (2) utilize the shape of the boundary to extract spatial features that characterize the lesions; this leads to mimicking what dermatologists perform when diagnosing a suspicious mole. By automatically defining the lesion boundaries, both the single and the dual-TRC approaches (discussed in Chapters 2 and 4, respectively) will be benefiting by minimizing the user interaction and subjectivity, which may affect the expected performance by introducing external factors not addressed in this disserta-

tion.

We divide this chapter in two main parts. In the first part is covered entirely in Section 5.2 and it introduces a novel method to fuse material/tissue classification with an edge-detection algorithm in order to improve the performance of the latter. More precisely, we extend the Spectral-Ratio Contrast (SRC) algorithm (presented by our group) [42, 43] by utilizing spectral classification to further enhance the detection of edges that are solely due to color (not intensity) changes. In the second part of this chapter, we utilize the ASRC algorithm to perform automatic boundary identification of the suspected lesions by exploiting the color contrast between the pigmented lesion and that of the surrounding tissue. After defining the boundary, we define and extract spatial features of the lesions following the general guidelines of the ABCDE rule [3]. We conclude this chapter by developing and testing a method to fuse the spatial features with the temporal features (i.e., the KL coefficients) extracted in the skin-cancer detection method developed in Chapter 2.

5.2 Adaptive spectral-ratio contrast algorithm for boundary detection

5.2.1 Mathematical preliminaries

A multicolor image is normally referred in the literature as a multispectral (MS) or hyperspectral (HS) image, spanning from hundreds to thousands of spectral bands (color slices). A MS or HS image, also termed an image cube, is a 3D array of real numbers that we denote by $\mathbf{u} \in \mathbb{R}^{I \times J \times K}$, where I and J represent the number of horizontal and vertical pixels, respectively, in the spatial domain, and K represents the number of spectral bands. We denote any element of the cube $\mathbf{u}$ as $u_k(i, j)$, where

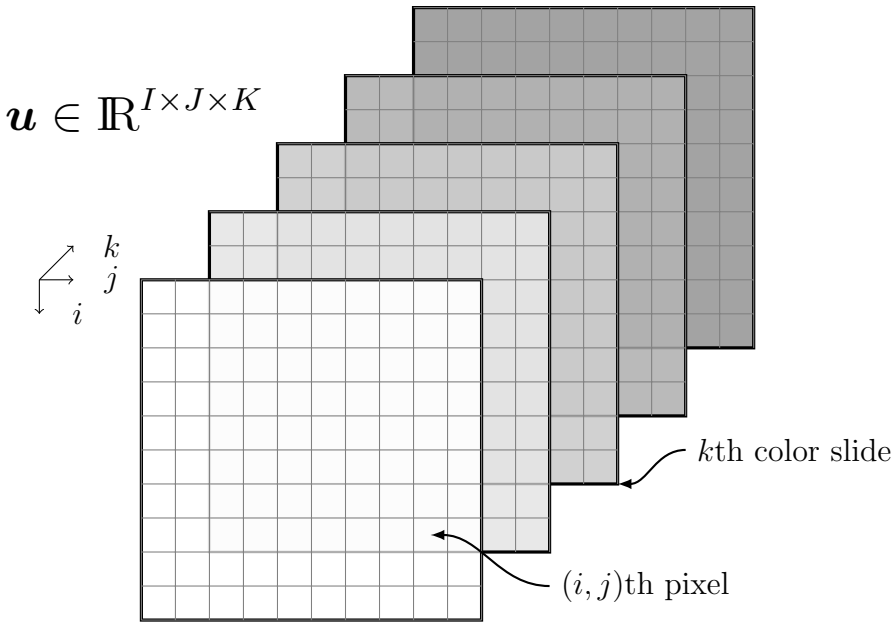


Figure 5.1: Pictorial representation of the three-dimensional array of numbers that describe the datacube in spectral imagery

$1 \leq i \leq I$, $1 \leq j \leq J$ and $1 < k \leq K$. The value of $u_k(i,j)$ is referred to as the intensity of the kth band at spatial location (i,j). For a fixed spatial location (i,j), the K-dimensional vector $\mathbf{u}(i,j) = (u_1(i,j), \ldots, u_K(i,j))$ is termed a hyper-pixel. Meanwhile, for a fixed band index k, the two-dimensional array $u_k(\cdot,\cdot)$ defines the kth image plane (color slice) of the spectral image. Fig. 5.1 depicts the data cube $\boldsymbol{u}$. The goal of this work is to define an edge map,

$$\mathcal{F} : \mathbb{R}^{I \times J \times K} \mapsto \{0,1\}^{I \times J} \ ,$$

that assigns the value 1 to the pixel location (i,j) if $\mathbf{u}(i,j)$ belongs to an edge, while assigning the value 0 otherwise. In this work we define an edge to be either a jump in the broadband intensity (as in the conventional definition of an edge for gray-scale images) or a change in the material that exhibit color contrast but not necessarily luminance contrast (as in isoluminant edges).

5.2.2 The spectral-ratio contrast algorithm

In this section, we review germane aspects of the spectral-ratio contrast (SRC) algorithm in order to later describe the insets into the SRC workflow that the ASRC algorithm introduces. More details about the SRC algorithm can be found in the Appendix B.

For two types of materials A and B, the SRC algorithm builds the edge map $\mathcal{F}_{\mathrm{AB}}$ in three stages: (i) model-based edge signature identification, (ii) sparse spatio-spectral mask development, and (iii) edge discrimination. The SRC and the ASRC algorithms share the first two stages, but they differ in the third stage, where the latter involves spectral classification before discriminate the edges. First, the model-based edge signature identification of the SRC algorithm is a learning step in which a small set of ratios of spectral-band outputs that most profoundly identify edges between each pair of materials is selected. This selection is made judiciously and sparingly recognizing the very few bands, across all bands, that permit a good discrimination between each pair of materials. Through this process, the SRC algorithm achieve substantial levels of data compression at the edge extraction stage. As an example, assume that at this stage the SRC algorithm identifies that the spectral-band ratios a_p/b_q and a_r/b_s are the optimal in terms of separating A and B, then the edge-signature for this problem will be the set of triplets $\{(p, q, \rho_1), (r, s, \rho_2)\}$, where $\rho_1 = a_p/b_q$ and $\rho_2 = a_r/b_s$, respectively.

Second, the sparse spatio-spectral mask is developed by merging the set of bands identified in the previous stage with a regular spatial mask. For example, if a common 3×3 spatial mask is combined with the edge-signature, at the (i, j) spatial location one can form the ratios of pixels (one for each pair of oposite pixels) $u(i-1, j)/u(i+1, j)$ and $u(i+1, j)/u(i-1, j)$ in the vertical direction, $u(i, j-1)/u(i, j+1)$ and $u(i, j+1)/u(i, j-1)$ in the horizontal direction, etc. If the spectral-band ratios are

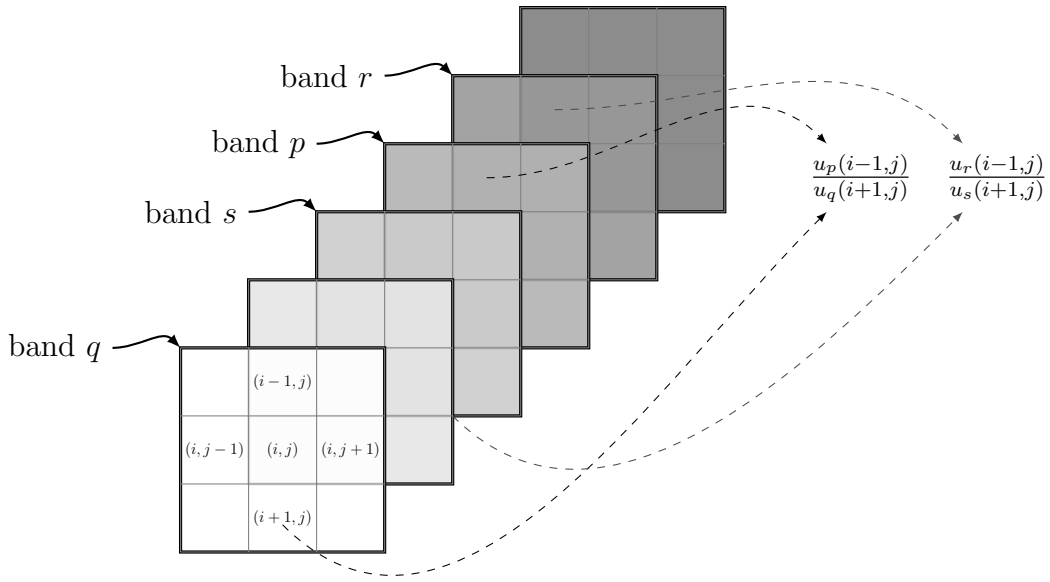


Figure 5.2: Pictorial representation of the fusion between the vertical pixel pair within a 3×3 spatial mask and the edge-signature identified in the first stage of the SRC algorithm to form the sparse spatio-spectral mask

computed for all the band-pairs of the edge signature at all the pixel pairs of oppossite pixels of the spatial mask, then the sparse spatio-spectral mask $\mathcal{K}_{\mathrm{AB}}$ is obtained. A pictorial representation of the development of such a mask for one of the possible vertical pair of pixels is depicted in Fig. 5.2. When this procedure is applied to all the pixel pairs within the spatial mask, the ratios between each of the pixel pairs of the spatial mask, at each of the band pairs given in the edge signature generates a matrix of spectral-ratio signatures. For example, using the horizontal and vertical pairs of hyper-pixels within the 3×3 mask and the example edge-signature explained

above, i.e., $\{(p, q, \rho_1), (r, s, \rho_2)\}$, the resulting matrix will be

$$\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j) = \begin{pmatrix} \frac{u_p(i-1,j)}{u_q(i+1,j)} & \frac{u_r(i-1,j)}{u_s(i+1,j)} \\ \\ \frac{u_p(i,j-1)}{u_q(i,j+1)} & \frac{u_r(i,j-1)}{u_s(i,j+1)} \\ \\ \frac{u_p(i+1,j)}{u_q(i-1,j)} & \frac{u_r(i+1,j)}{u_s(i-1,j)} \\ \\ \frac{u_p(i,j+1)}{u_q(i,j-1)} & \frac{u_r(i,j+1)}{u_s(i,j-1)} \end{pmatrix}, \tag{5.1}$$

where each row is constructed as depicted in Fig. 5.2. The use of both ratios on each direction (e.g., $u(i-1,j,\cdot)/u(i+1,j,\cdot)$ and $u(i+1,j,\cdot)/u(i-1,j,\cdot)$ in the vertical direction) is required to account for the two possible material configurations. In general, the application of the spatio-spectral mask with M pixel pair and R band pairs in the edge signature, to each location (i,j), results in a $2M \times R$ matrix of "features." (Compare this to a gray-scale image when the application of a spatial mask to a pixel results in a scalar.)

Third, the edge discrimination stage utilizes the $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j)$ mask to determine whether the (i,j) location belongs to a edge or not. This is done by matching the outputs of the mask $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j)$ with the ratios of the edge signature. In the ideal case when no noise is present and the image under test is comprised only of hyper-pixels with the exact same value of the characteristic hyper-pixels (i.e., $\mathbf{a}$ and $\mathbf{b}$) as shown in Fig. 5.3, the output of the spatio-spectral mask will perfectly match the values of the ratios obtained from the edge signature. To illustrate this point, assume the same example given earlier, where the spatio-spectral mask is given by (5.1). When we have a horizontal edge, the second and fourth rows of $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j)$ in (5.1) will have a value that is not meaningful, but the first or third rows will match the ratio from the edge signature, $[\rho_1 \quad \rho_2]$. Indeed, if the upper neighbor is from material A and the lower neighbor is from material B, then the first row of (5.1) will be $[a_p/b_q \quad a_r/b_s]$ (which matches the edge signature) and the third row will be $[b_p/a_q \quad b_r/a_s]$. Conversely, when the upper neighbor is from material B and the

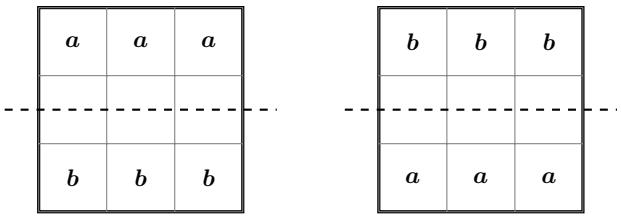


Figure 5.3: Pictorial representation of an horizontal edge (dashed line) within a 3×3 spatial mask for the noiseless case. The drawing on the left represents the case of having material A on top and B on the bottom part. The drawing on the right represent the oppossite situation

lower neighbor is from material A, the computed ratios are switched, which means that the third entry of (5.1) will match the ratios from the edge signature. This example shows that if an edge is present then at least one row of $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j)$ will perfectly match the ratios from the edge signature. When noise is present, we allow a matching tolerance to account for the similarity between the outcome of the mask and the edge signature ratios.

The different ratios of $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j)$ that match the ratios from the edge signature are accounted in a $M \times R$ indicator matrix whose binary entries are defined by

$$\delta_r^m(i,j) = \begin{cases} 1, & \text{if } |\kappa_{\mathrm{AB}}(i,j;m^-,r) - \rho_r| < \epsilon \ , \\ 1, & \text{if } |\kappa_{\mathrm{AB}}(i,j;m^+,r) - \rho_r| < \epsilon \ , \\ 0, & \text{otherwise} \ , \end{cases} \tag{5.2}$$

where $m^{\mp}$ and r represent the mth pixel pair from the spatial mask and the rth band pair of the edge signature, respectively. For the same example worked above, this is translated into $\kappa_{\mathrm{AB}}(i,j;m^{\mp},r) = u_p(i \mp 1, j)/u_q(i \pm 1, j)$ and $\rho_r = \rho_1$, etc. Here, the tolerance parameter, ϵ, accounts for both the natural variability of the data and the presence of noise in the spectral data for materials A and B.

Ideally, if the mth hyper-pixel pair belongs to the same material type, then the test in (5.2) will return the value of zero. Conversely, if the hyper-pixels forming

the pair are from the two different materials, either the entry $\kappa_{\mathrm{AB}}(i,j;m^-,r)$ or $\kappa_{\mathrm{AB}}(i,j;m^+,r)$ will be equal to the corresponding ratio ρ_r from the edge signature. As such, for a given pair of pixels, the number of non-zero elements in the associated mth row of the indicator matrix reveals the number of times the response of the mask $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j)$ has matched (within the specified tolerance ϵ) the spectral ratios from the edge signature. By counting the number of ones of each row (i.e., the number of times this match occurs), we know the strength of the edge on that particular pixel. This is accounted by the matrix infinity norm as explained in detail in the Appendix B.

5.2.3 The adaptive spectral-ratio contrast algorithm

Background

Mainly due to the noise in the spectral image, the computed ratios of the SRC algorithm may match those from the edge signature even when there is no edge present within the 3D mask. If we can somehow adaptively change the SRC threshold depending on the material composition of the scene, these false edges can be avoided. As such, we want to add an new layer of processing within the SRC in order to restrict the algorithm to capture edges that are due solely to material changes (and not intensity changes). By specializing the algorithm to material changes, we increase its tolerance to noise and the corresponding false edges are minimized.

In order to capture solely the changes between materials in the new algorithm, we utilize material classification of neighboring pixels to adaptively pre-qualify the spectral ratios before computing the indicator matrix (5.2). We term the new algorithm the adaptive SRC (ASRC) algorithm, because it adaptively define thresholds for the spectral ratios based on the classification of the very same scene. The use of classification to enhance other tasks such as segmentation is an area that has already been

studied [79, 80]. For example, Loog and Ginneken [79] utilized a k-nearest neighbor classifier to generate an initial segmentation of ribs in chest radiographs, which is iteratively updated using other features such as spatial distribution of the pixels by means of different classifiers. Here, we fuse classification into the SRC algorithm to minimize the effect of the misidentified pixels from the former algorithm and improve the edge identification stage of the latter.

The ASRC algorithm

Let us consider the feature matrix of the (i,j)th location, $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j)$, as given in (5.1). We want to rank the entries $\kappa_{\mathrm{AB}}(i,j;m^{\mp},r)$ of the feature matrix with the following two objectives in mind: (i) promoting the thresholding of ratios (as edge candidates) when the spatio-spectral mask contains hyper-pixels from two distinct materials and (ii) discouraging the thresholding of ratios when the mask contains hyper-pixels from only one type of material.

To do so, we embed the data-dependent multiplicative factors $\gamma_{m,r}^{i,j}$ with the tolerance ϵ in (5.2), which results in its redefinition as

$$\delta_r^m(i,j) = \begin{cases} 1, & \text{if } |\kappa_{\mathrm{AB}}(i,j;m^-,r) - \rho_r| < \gamma_{m,r}^{i,j}\epsilon \ , \\ 1, & \text{if } |\kappa_{\mathrm{AB}}(i,j;m^+,r) - \rho_r| < \gamma_{m,r}^{i,j}\epsilon \ , \\ 0, & \text{otherwise} \ , \end{cases}$$

where we use the same multiplicative factor, $\gamma_{m,r}^{i,j}$, for both tests in order to maintain the independence on the direction of the transition between materials. The mathematical definition of the multiplicative factors $\gamma_{m,r}^{i,j}$ is as follows. Let $N_u(i,j)$, $N_l(i,j)$, $N_r(i,j)$ and $N_\ell(i,j)$ denote four neighborhoods surrounding the (i,j)th hyper-pixel of interest. The subscripts u, l, r and ℓ stand for upper, lower, right and left neighborhoods, respectively. The understanding is that $N_u(i,j)$ contains neighboring pixels above the hyper-pixel (i,j), $N_\ell(i,j)$ contains neighboring pixels to the left of pixel

(i, j), and so on. We do not impose any restrictions on the four neighborhoods at this point. Let the function $\mathcal{C} : \mathbb{R}^S \mapsto \{0, 1\}$ be a classifier that maps each hyper-pixel to a class of materials, where outputs 0 and 1 represent classes A and B, respectively. Now for any collection of indices N, we denote the class of N by $\mathcal{C}(N)$, which is defined according to a certain prescribed classification rule.

Next, define

$$\gamma_{m,r}^{i,j} = \Big(\mathcal{C}(N_u) \oplus \mathcal{C}(N_l)\Big) \bigvee \Big(\mathcal{C}(N_r) \oplus \mathcal{C}(N_\ell)\Big), \tag{5.3}$$

where $\oplus$ denotes the "exclusive OR" operation and the symbol "$\bigvee$" represents the "OR" operation. For simplicity of the notation, we have discarded the (i, j) dependence of each neighborhood set N in (5.3) with the understanding that each neighborhood is defined on a pixel-by-pixel basis.

From (5.3), $\gamma_{m,r}^{i,j}$ will be unity (in which case the (m, r)th pixel-band pair at the (i, j) location qualifies for thresholding as usual) if at least one of the opposite neighborhoods are classified as two different materials. On the other hand, $\gamma_{m,r}^{i,j}$ will be zero (in which case the (m, r)th pixel-band pair at the (i, j) location does not qualify for thresholding) if the declared class of each neighborhood is in agreement with the declared class of its opposite neighborhood. As a consequence, the ASRC will operate as the the SRC algorithm only if the outcome of the classifier indicates the possible presence of an edge, suppressing edges that are due to an intensity change. This will also reduce the detection of false edges.

The fact that the ASRC algorithm is restricted to identifying edges based on color only is similar to that of the HySPADE algorithm [41]; however, the algorithms are conceptually different. A key difference between the ASRC and the HySPADE algorithms is that the former utilizes the sparse, 3D mask of ratios to fuse spectral and spatial information to nonlinearly extract edge information while the latter algorithm utilizes only spectral information to compute spectral angles, which are linear

spatio-spectral features. Even though the HySPADE algorithm performs equally well compared to the SRC algorithm, its performance is worse than that of the ASRC algorithm in the presence of isoluminant edges. Moreover, as presented in the Section 5.2.5, the HySPADE algorithm requires a high number of operations per pixel ($> 10^9$ operations per pixel in our examples), as compared with the operations required by the proposed algorithms (< 90 operations per pixel in our examples).

5.2.4 Application of the algorithm to spectral data

Let us take a temporary departure from the main subject of this dissertation in order to demonstrate the power of the ASRC algorithm in detecting edges from multicolor images. The spectral component is not part of this dissertation, but is utilized here with the purpose of evaluating the ASRC performance and compare it with other similar alternatives.

We compare the outcome of the ASRC algorithm with the edge maps obtained by the Canny algorithm [39] (applied to selected bands), the SRC algorithm, the MCG algorithm [40] and the HySPADE algorithm [41]. We restrict our attention to edge signatures with unity length using two bands (i.e., $S = 2$ and $R = 1$), which is the minimum value required by the SRC and ASRC algorithms. Moreover, we utilize a 3×3 spatial mask to construct the joint spatio-spectral mask, $\mathcal{K}_{\mathrm{AB}}$. Within the spatial mask, we identify four directions (each one associated with a pair of pixels): horizontal, vertical and the two diagonals, i.e., $M = 4$. For the ASRC algorithm, we select the distance-based Euclidean classifier for its simplicity and the good results observed; the neighborhood sets required for the classifier are defined within the same 3×3 spatial mask used in the SRC algorithm. This choice of spatial mask, classifier and neighborhood sets is also considered for the complexity analysis in Section **??**.

The objective of this section is to show that the ASRC algorithm perfoms as

well as the SRC, MCG and HySPADE algorithms for scenes with normal complexity, and in cases when the scene contains isoluminant edges, the ASRC algorithm outperforms all the reference algorithms. In our study, we employ raw imagery from the AHI sensor and the DWELL sensor (more details to follow). In order to create a more challenging scenario for the algorithms, we also normalize the data by their broadband intensity [45]. More precisely, following the notation introduced in Section 5.2.1, the intensity of the kth band at spatial location (i, j), $u_k(i, j)$, is normalized by the broadband intensity $\sum_{k=1}^{K} u_k(i, j)$ such that the normalized intensity is defined as,

$$\underline{u}_k(i, j) = \frac{u_k(i, j)}{\sum_{k=1}^{K} u_k(i, j)} \ . \tag{5.4}$$

The normalized hyperpixels, $\underline{\boldsymbol{u}}_k(i, j) = [\underline{u}_1(i, j) \cdots \underline{u}_K(i, j)]$, will minimize the role of broadband emissivity in the discrimination process and emphasizes the spectral contrast.

For the AHI dataset, we only perform a qualitative comparison of the algorithms since the ground-truth information is not available for this dataset. On the other hand, for the data from the DWELL sensor we perform both qualitative and quantitative assessment of the proposed algorithms and the benchmark algorithms as the ground-truth information is available.

Results using AHI data

The AHI sensor consists of a long-wave IR (7μm-11.5μm) pushbroom HS imager and a visible high-resolution CCD linescan camera. The HS imager has a focal-plane array (FPA) of 256×256 elements with spectral resolution of 0.1μm [81]. For this study, we utilize AHI data that contains three different classes: building (B), ground (G) and road (R). We utilize the 200 low-noise bands out of the 256 available bands. The

calculated edge signatures (band indices and the corresponding responding ratios) for each pair of materials are summarized in Table 5.1.

Figure **??** shows a comparison among the edge maps obtained by the Canny, the HySPADE, the MCG, the SRC and the ASRC algorithms for the raw sensor data (first and second columns) and for the normalized data (third and fourth columns). The Canny algorithm is applied to the same depicted image, which corresponds to the image plane at band 14. Recall that the MCG and the SRC algorithms detect edges characterized by both intensity and spectral changes. The HySPADE and ASRC algorithms, on the other hand, detect edges that exhibit a change in the spectral content only.

From the results presented in Fig. 5.4 we observe that the Canny edge detector performs very well when applied to the AHI raw image for spectral band 14 (row I, column b). However, when the algorithm is applied to the intensity-normalized image, the performance of the Canny algorithm significantly degrades (row I, column d). This degradation is a result of the fact that the Canny algorithm detects intensity changes only, and it is expected to perform optimally for high intensity contrast images such as the image in row I, column a.

The MCG and the SRC algorithms produce virtually the same edge maps when applied to raw sensor data (second row, columns a and b), with a clear computational

Table 5.1: The edge signatures between classes B, G and R obtained for the AHI data

Signatures	Triplets (p_1, q_1, ρ_1)	
	Raw data	Normalized data
$\mathcal{E}_{\mathrm{BG}}$	(17, 16, 0.6941)	(3, 4, 0.8609)
$\mathcal{E}_{\mathrm{RG}}$	(47, 46, 0.7949)	(3, 4, 0.8949)
$\mathcal{E}_{\mathrm{BR}}$	(17, 16, 0.8706)	(16, 17, 0.9588)

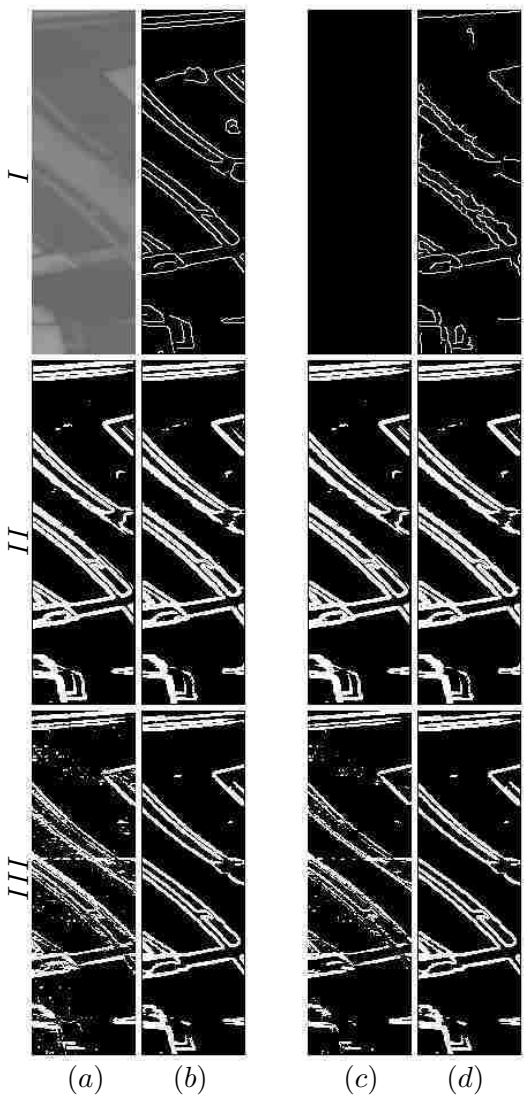


Figure 5.4: Comparison between the ASRC, the SRC, the Canny, the MCG and the HySPADE edge detectors for the raw sensor data (first two columns) and the normalized sensor data (third and forth column) for the AHI imagery. First row, from left to right: raw AHI data (band 14); Canny edge map for raw AHI data at band 14; normalized AHI data at band 14; Canny edge map for AHI normalized data at band 14; Second row from left to right: MCG and SRC edge maps for AHI raw dataset; MCG and SRC edge maps for AHI normalized dataset; Third row from left to right: HySPADE and ASRC edge maps for AHI raw dataset; HySPADE and ASRC edge maps for AHI normalized dataset.

advantage seen in the SRC algorithm by requiring only two spectral bands, whereas the MCG algorithm requires all the 200 available bands. When normalized data is used (second row, c and d columns), few edges in some areas are missed either by the SRC or the MCG algorithms. Nonetheless, the edge maps between the two algorithms are again comparable. Moreover, the results for the normalized case are very similar to those for the raw data case. These results show the advantage of the methods that utilize both intensity and spectral information over purely gray-scale algorithms such as Canny.

The ASRC algorithm (row III, column b) performs significantly better compared to the HySPADE algorithm (row III, column a) when applied to the AHI raw data. The edge map obtained by HySPADE exhibits noise and some of the edges that were detected by SRC, MCG and ASRC are missed by HySPADE. The advantage of ASRC over HySPADE continues to be pronounced when both algorithms are applied to the normalized data (row III, columns d and c). The edge maps obtained by the ASRC algorithm applied to raw and normalized data (row III, columns b and d) are virtually identical. This is due to the fact that the ASRC algorithm detects edges based on changes of the spectral content only. As for HySPADE, the application of the algorithm to the normalized AHI data results in a slight degradation of the edge detection but overall reduction of the noise in the edge map compared to application to the raw AHI data (row III, column a); however, as in the case of ASRC, the edge maps are comparable. It is important to note that the edge maps obtained by the MCG, the SRC and the ASRC algorithms are very similar for both raw and normalized AHI data cases.

The ASRC algorithm offer a performance advantage over the other algorithms for images that contain isoluminant edges as seen next for the DWELL imagery.

Results using DWELL data

The DWELL sensor used in these experiments was designed and fabricated at the Center for High Technology Materials at the University of New Mexico [44, 45]. The DWELL photodetector offers a unique property of spectral tunability that is continuously controllable through the applied bias voltage. This feature of the DWELL is a result of the quantum-confined Stark effect [82]. In essence, a single DWELL photodetector can be thought of as a continuously tunable MS spectral detector, albeit with overlapping spectral bands [45].

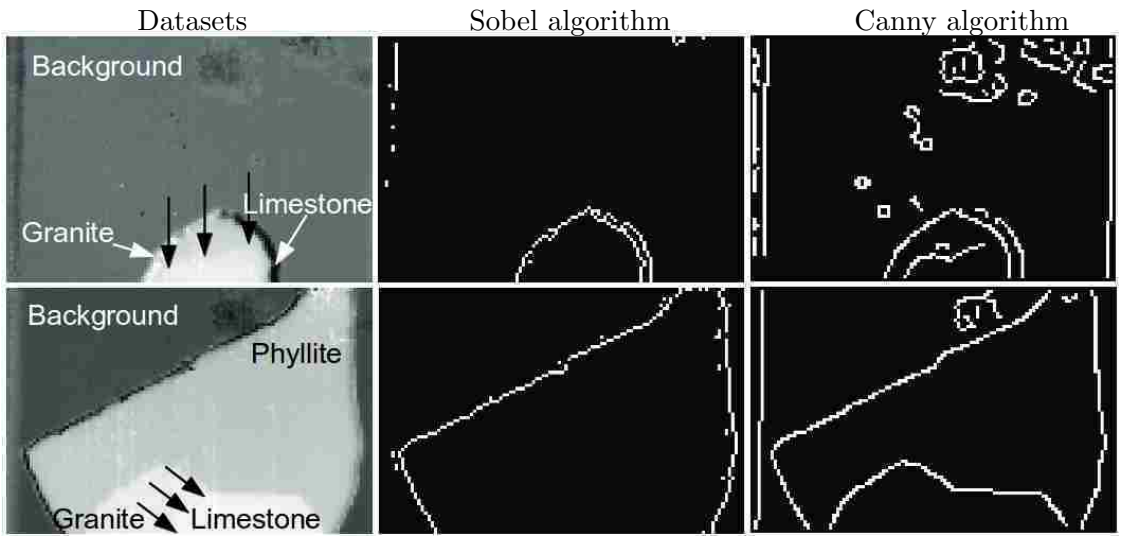


Figure 5.5: Two datasets used in the current study: the first dataset is comprised of B, G and L classes (top row) and the second dataset is comprised of B, G, L and P classes (bottom row). First column: images acquired with the DWELL FPA (with enhanced contrast to show details) operating at an applied bias of $1.0V$. The isoluminant edges (not visible) are marked by the tips of the black arrows. Second column: edge map obtained by the Sobel gray-scale edge detector; third column: edge map obtained by the Canny gray-scale edge detector.

In these experiments we utilize a 320×256 DWELL FPA to image two different arrangements of rocks, as shown in Fig. 5.5 (first column). The first arrangement (top-left) is comprised of granite (G) and limestone (L) rocks (approximately 1–2 inch in diameter). The surrounding background (B) in this image corresponds to the opening of a blackbody source. The second arrangement (bottom-left) is comprised of the rocks phyllite (P), granite (G) and limestone (L), surrounded by the same background (B) as that in the first arrangement. Both examples contain an invisible isoluminant edge between the granite and the limestone rocks that exists on the tip of the black arrows. The edge maps shown in Fig. 5.5 were obtained by using the Sobel (second column) and the Canny (third column) edge detectors applied to raw DWELL-sensor data when the FPA is operated at $1.0V$. The corresponding spectral response of the sensor at the applied bias of $1.0V$ is shown in Fig. 5.6.

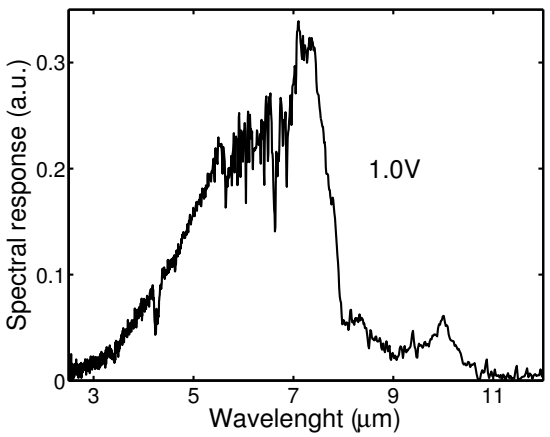


Figure 5.6: Spectral response of the DWELL photodetector at an applied bias of 1.0 V

Note that the Sobel edge detector has entirely missed the edge between granite and limestone rocks in both examples. Moreover, it has also failed to detect strong edges between both the granite-phyllite pair and the limestone-phyllite pair. However, the more sophisticated Canny edge detector picks up these strong edges, and it partially detects the isoluminant edge in the first examples. Nevertheless, it does not detect the isoluminant edge in the second example.

By operating the DWELL sensor at ten different bias voltages, we generated a multispectral cube to test the proposed algorithms. The obtained edge signature triplets for all the possible combinations of material pairs for both datasets are summarized in Table 5.2. In what follows, we will term the DWELL imagery that contains background, granite and limestone classes, as shown in Fig. 5.5 (top-left), the *first DWELL dataset*, and we term the imagery that contains background, phyllite, granite and limestone, as shown in Fig. 5.5 (bottom-left), the *second DWELL dataset*.

The results for the first DWELL dataset for raw sensor data are shown in Fig. 5.7. The first row of edge maps shows the results of the application of the Canny edge detector to four randomly selected bands. It is important to note that some bands present a high number of false edges, whereas for other bands the isoluminant edges are detected. As such, the Canny algorithm can generate good edge maps, depending

Table 5.2: Edge signatures among the B, P, G, and L classes obtained for the DWELL datasets

	Triplets (p_1, q_1, ρ_1)	
Signature	Raw data	Normalized data
$\mathcal{E}_{\mathrm{GB}}$	(6, 7, 0.2747)	(1, 10, 0.1434)
$\mathcal{E}_{\mathrm{LB}}$	(6, 7, 0.2636)	(1, 10, 0.1395)
$\mathcal{E}_{\mathrm{LG}}$	(5, 6, 0.7577)	(9, 10, 0.9109)
$\mathcal{E}_{\mathrm{PL}}$	(4, 5, 0.5703)	(9, 10, 0.8444)
$\mathcal{E}_{\mathrm{PB}}$	(6, 7, 0.3168)	(1, 10, 0.2283)
$\mathcal{E}_{\mathrm{PG}}$	(4, 5, 0.6006)	(9, 10, 0.8590)

on the bands used, but there is no standard way to select those bands that guarantee a good performance.

The second and third rows in Fig. 5.7 show the results for the MCG and HySPADE algorithms, respectively, at different threshold values in order to unveil the isoluminant edge between the granite and limestone rocks. The MCG algorithm (second row) picks up the weak edge only after its tolerance is increased to a degree that results in the detection of a significant number of false edges (second row, fourth column). On the other hand, HySPADE offers a less-noisy edge map compared to the MCG algorithm; nonetheless, the background-granite and granite-limestone edges are not well defined, as shown in the third row, fourth column. Moreover, the high computational cost of the HySPADE algorithm makes it hard for the user to fine tune its tolerances, which is a clear disadvantage of the HySPADE algorithm. (More details regarding computational costs are given in Section **??**.) We also observe that at the cost of a slight increase in the number of false edges, the SRC algorithm can clearly define the background-granite edge with respect to the granite-limestone edge (fourth row, fourth column). Finally, the results of the ASRC algorithm (fifth row, fourth column) are better than all the previous algorithms in terms of clearly defining both the strong and weak edges. The ASRC algorithm also discards all of the false edges in the the background.

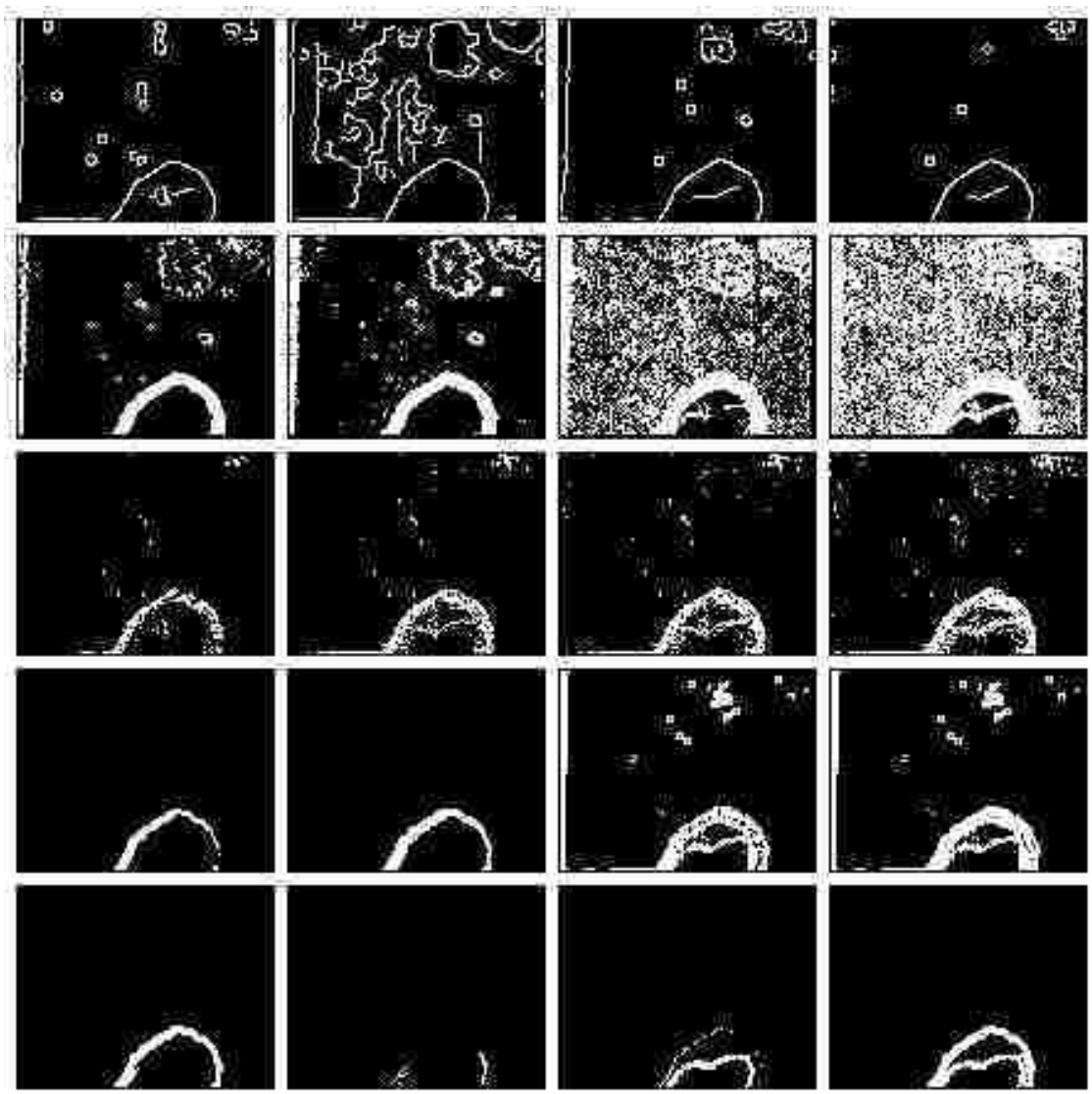


Figure 5.7: Comparison between the Canny algorithm applied to individual bands (first row), MCG algorithm (second row), HySPADE algorithm (third row), SRC algorithm (fourth row) and ASRC algorithm (fifth row) for the dataset containing granite and limestone rocks (first dataset). The Canny algorithm was applied to the images at bands 1, 6, 8 and 9, respectively. The MCG and HySPADE results are presented for a sequence of increasingly permissive tolerances in order to unveil the isoluminant edge. Last two rows show the SRC and ASRC edge maps: first column, the edges E_{GB}; second column, the edges E_{LB}; third column, the edges E_{LG}; fourth column, the combined edge maps.

Table 5.3: Performance comparison table for the results of five algorithms (Canny, MCG, HySPADE, SRC and ASRC) for the dataset containing B, G and L classes (raw data)

Algorithm	Detection probability	False-alarm probability
Canny (band 9)	0.4533	0.0082
MCG	0.9600	0.6112
HySPADE	0.7867	0.0565
SRC	0.9467	0.0862
ASRC	0.9733	0.0244

By utilizing the available ground-truth information for the DWELL datasets, we derived reference edge maps for the scenes under study. These edge maps are utilized to compute the empirical detection and false-alarm probabilities, P_D and P_F, respectively, for the five algorithms (Canny applied on different bands, MCG, HySPADE, SRC and ASRC). The detection probability (also known as the sensitivity of the algorithm) corresponds to the probability that an actual edge (provided by the ground truth) is detected by the algorithm under evaluation. The false-alarm probability (also known as the complement of the specificity of the algorithm) is the probability that the algorithm detects a non-existing edge. For each algorithm, we have tuned the respective parameters in order to unveil the isoluminant edges (the assessment was made by visual inspection). We have conditioned the algorithms' parameters to detect isoluminant edges because they present one of the most challenging problems in multicolor edge detection. The metrics P_D and P_F were computed by comparing the ground-truth edge-map with the algorithm outcome on a pixel-by-pixel basis.

From the results presented in Table 5.3 we see that the best performance achieved by the Canny algorithm is when it is applied to band 9 ($P_D = 0.4533$ and $P_F = 0.0082$). It is important to note that the Canny algorithm, applied to this band, is capable to partially detect the isoluminant edge (see Fig. 5.7, top-right). However, without previous knowledge of the scene and the results of the application of the Canny algorithm to every band, it would be difficult to guess which band gives

the best results. The MCG algorithm, on the other hand, cannot detect the isoluminant edges without producing a high number of false edges. Indeed, when the isoluminant edge is detected (second row, fourth column) the MCG performance is given by a high detection ($P_D = 0.9600$) but also with a high false-alarm probability ($P_F = 0.6112$). At the cost of a tremendous increase of computation complexity (see Section **??**), the HySPADE algorithm outperforms the Canny algorithm in terms of sensitivity ($P_D = 0.7867$) and the MCG algorithm in terms of low false alarm probability ($P_F = 0.0565$). In contrast, the SRC algorithm outperforms the previous algorithms in terms of both simplicity and sensitivity with $P_D = 0.9467$, at the cost of a slight increase in the false-alarm probability ($P_F = 0.0862$) in comparison to HySPADE. The ASRC algorithm outperforms all the other four algorithm in terms of highest detection and lowest false-alarm probability, $P_D = 0.9733$ and $P_F = 0.0244$, respectively.

The edge-detection results for the second DWELL dataset (comprinsing phyllite, granite and limestone) are presented in Fig. 5.8 for intensity-normalized data. Table 5.4 summarizes the detection and false-alarm probabilities achieved by each one of the five algorithms for this dataset. The second dataset is more challenging than the first dataset because the two classes with the isoluminant edge (i.e., granite and limestone rocks) are now positioned against a phyllite backdrop that exhibits less contrast than the blackbody. Moreover, the data is intensity normalized. As before, the Canny edge detector achieves good performance when applied to band 9 ($P_D = 0.7854$ and $P_F = 0.0301$). It is very interesting to note that (for this band) the Canny algorithm is capable detecting the isoluminant edge between the granite and limestone rocks almost fully. This is because the normalization process smooths some intensity peaks and improves the contrast between granite and limestone (for this particular band) as a secondary effect. This result proves that the first category of algorithms (those that do not use spectral information) can achieve good detection as long as the best band is identified through pre-processing of the data, which can

Table 5.4: Performance comparison table for the results of five algorithms (Canny, MCG, HySPADE, SRC and ASRC) for the dataset containing B, P, G and L classes (normalized data)

Algorithm	Detection probability	False-alarm probability
Canny (band 9)	0.7854	0.0301
MCG	0.8802	0.5046
HySPADE	0.7445	0.0833
SRC	0.8593	0.0873
ASRC	0.8919	0.0652

be a very difficult requirement.

As for the MCG-generated edge maps, Fig. 5.8 (second row), the weak edge is detected only when the false-alarm probability reaches unacceptable levels. The HySPADE algorithm performs worst than the MCG algorithm ($P_D = 0.7445$ and $P_F = 0.0833$) and it is not capable of detecting the isoluminant edge. In contrast, the SRC algorithm recovers the strong edges as well as the weak edge between the granite and limestone rocks. Indeed, Fig. 5.8 (fourth row) shows a high-resolution weak edge captured by the SRC algorithm. The achieved detection and false-alarm probabilities of the SRC algorithm ($P_D = 0.8593$ and $P_F = 0.0873$) corroborate this observation. It is important to note that event though the SRC algorithm is able to detect isoluminant edges for challenging scenarios, it still suffers from detecting false edges for each pair of materials, as observed in both examples. However, the ASRC algorithm reduces the detection of false edges substantially ($P_F = 0.0652$ for ASRC compared to $P_F = 0.0873$ for SRC), owing to the fusion of material classification in the edge-detection process. The ASRC is also able to improve the detection of edges, as noted by the improved detection probability ($P_D = 0.8919$ for ASRC compared to $P_D = 0.8593$ for SRC).

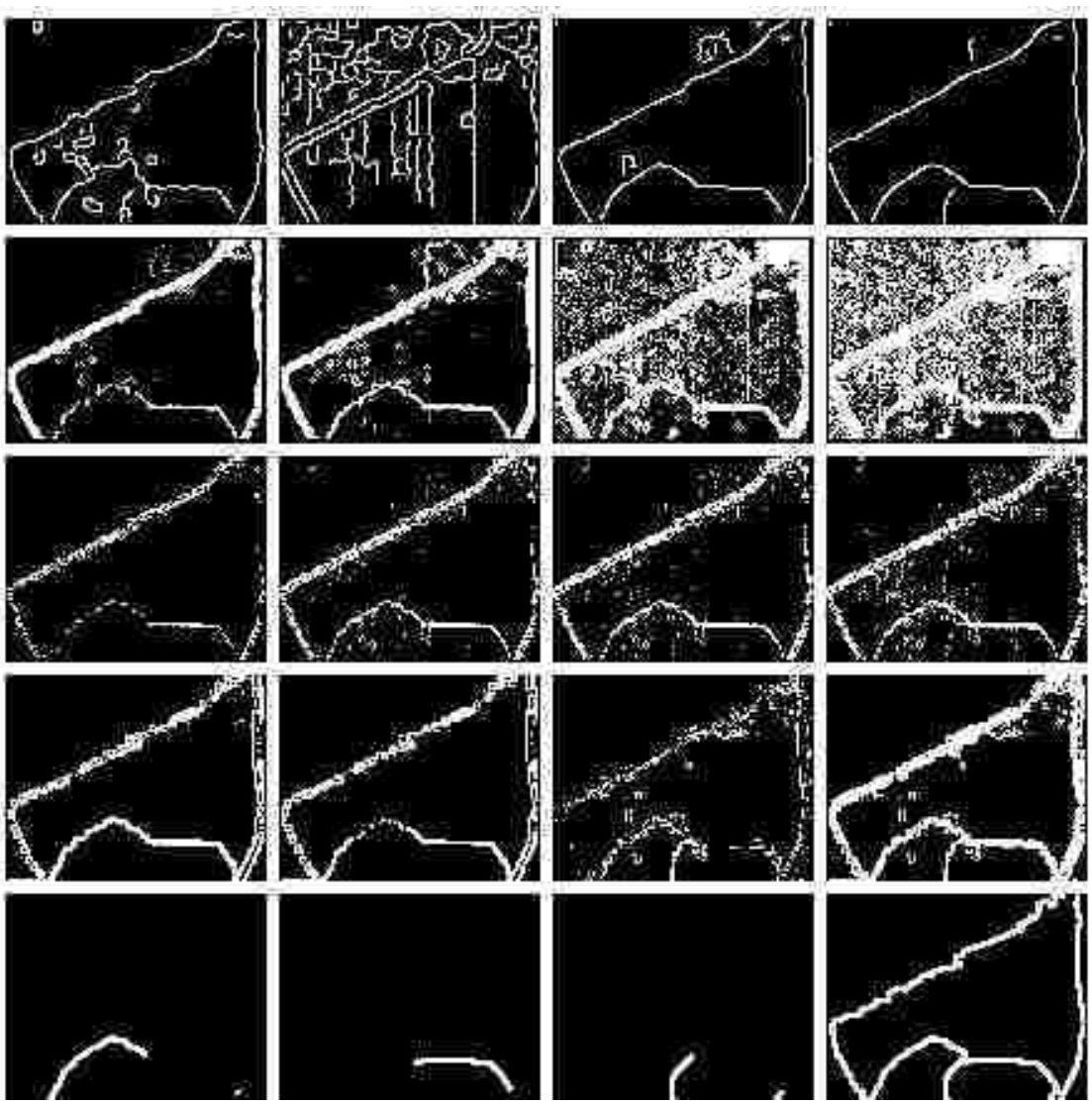


Figure 5.8: Comparison among the Canny algorithm applied to individual bands (first row), MCG algorithm (second row), HySPADE algorithm (third row), SRC (fourth row) and ASRC (fifth row) for the dataset containing Phyllite, Granite and Limestone rocks (second dataset). The MCG and HySPADE results are presented for a sequence of increasingly permissive tolerances in order to unveil the isoluminant edge. Last two rows show the SRC and ASRC edge maps: first column, the edges E_{PG}; second column, the edges E_{PL}; third column, the edges E_{LG}; fourth column, the combined edge maps.

From these results, we can conclude that the SRC algorithm outperforms the

MCG and HySPADE algorithms for the task of detecting edges using spectral data with minimal intensity contrast. Moreover, it performs as good as the Canny edge detector without the difficult requirement for pre-selecting the optimal band. Moreover, at the cost of a slight increase in computational cost, the ASRC algorithm outperforms all other four algorithms presented in this paper.

Next, we compare the multicolor algorithms (SRC, ASRC, MCG and HySPADE) in terms of their computational costs.

5.2.5 Complexity analysis

In this section, we estimate the complexity of the feature extraction stage in the SRC and ASRC algorithms and compare it to those for the MCG and the HySPADE algorithms. Since the edge signature identification is made offline and before the edge-identification stage, we do not include its computational cost. For simplicity, in the cost estimates we will regard the cost of all operations (e.g., multiplication, addition, etc.) as equal.

The SRC operations per pixel include the $2MR$ ratios required to form the matrix $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j)$ in (B.4), plus the $4MR$ computations required to form $\Delta\big(\mathcal{K}_{\mathrm{AB}}(\mathbf{u})\big)(i,j)$ in (B.5), plus the MR operations required to define the edges in (B.7). The total number of operations for the SRC algorithm is $7MR$ operations per pixel.

The ASRC computations include those from the SRC algorithm ($7MR$ operations per pixel) plus those required to compute and utilize the parameters $\gamma_{m,r}^{i,j}$. To calculate these parameters, we first require the classification and label comparison of the pixels within the mask, a task that will cost $2M+6R$ operations per pixel. Next, the computation of the $\gamma_{m,r}^{i,j}$ parameter requires $11MR$ operations per pixel (two XOR operations and one OR operation for each entry in (B.5)). The total number of operations for the ASRC algorithm is therefore $2M+6R+18MR$ operations per

Table 5.5: Comparison table for the total number of operations required for the SRC, ASRC, MCG and HySPADE algorithms

	Total number of operations per pixel		
		Examples ($M = 4$ and $R = 1$)	
		AHI data	DWELL data
Algorithm	General expression	($K = 200$)	($K = 10$)
SRC	$7MR$	28	28
ASRC	$2M + 6R + 18MR$	86	86
MCG	$10K + 9$	2,009	109
HySPADE	$IJ(3IJ + 6K + 3)$	$> 10^9$	$> 10^{10}$

pixel.

Meanwhile, the MCG algorithm requires $10K - 3$ operations to compute the first fundamental form for each hyper-pixel, nine operations to compute the corresponding eigenvalues, and three operations to compute the monitor function and apply the threshold. The total number of operations for the MCG algorithm is therefore $10K+9$ operations per pixel. (Please refer to Appendix **??** for further details.)

Next, for each hyper-pixel, the HySPADE algorithm requires the computation of IJ spectral angles (each spectral angle costs $6K + 1$ operations), plus the $2IJ + 1$ operations per pixel of the SA-cube to compute the one-dimensional derivative approximation, plus the $IJ + 1$ operations required to account for the statistical accumulation of each pixel within the SA-cube. The total number of operations for the HySPADE algorithm is therefore $IJ(3IJ+6K+3)$ operations per pixel. (Please refer to Appendix **??** for further details.)

In Table 5.5 we present a summary of the estimated values for the four algorithms considering the same AHI and DWELL experiments we previously discussed in Section 5.2.4. The proposed algorithms do not change their respective computational costs for the two examples because the edge signature identification removes the dependency of the algorithms on the actual number of bands of the data. From

the presented table, we can observe that the SRC algorithm gives a 71 fold gain in computational efficiency over the MCG algorithm for the two class edge detection problem over the AHI data, whereas the ASRC algorithm gives a 23 fold gain.

5.2.6 Discussion

We have introduced the ASRC algorithm, which is a specialized version of the SRC algorithm, aimed at detecting edges that are due to a change in the material only. The ASRC aims to reduce the detection of false edges due to unwanted changes in the intensity. As an heritage from the SRC algorithm, the ASRC algorithm utilize spectral library information to construct a sparse, non-separable and 3D edge operator while exploiting the concept of spectral ratio contrast.

Whereas the SRC algorithm performs as well as the MCG or HySPADE algorithms for moderately challenging edges the ASRC algorithm, which also involves classification-based step, is capable of minimizing the false-alarm edges, outperforming the SRC, MCG and HySPADE algorithms.

Let's return to the main topic of this dissertation after that brief digression. In the next section, we utilize the ASRC algorithm to automatically select the lesion boundaries by exploiting the RGB color difference between the pigmented and the surrounding tissue.

5.3 Automatic lesion boundary selection

Since the ASRC algorithm exploits the color contrast between two materials, it is an excellent candidate to automatically detect the edge of the lesions under study from the visible-color image. To this end, we only define the edges based on the visible

color image. Let us enumerate the steps and slight modifications that the ASRC algorithm requires to enable the automatic boundary identification of the pigmented lesion.

1. The ASRC involves a learning stage to identify the most relevant pair of bands; nonetheless, for a visible color image there are only 3 bands that are normally labeled as RGB (Red, Green and Blue) so we decided to utilize all three RGB bands.

2. As a consequence, the edge signatures will have nine triplets. Namely, if we denote characteristic pixels of the pigmented lesion and the surrounding skin by $\boldsymbol{a} = [a_R \; a_G \; a_B]$ and $\boldsymbol{b} = [b_R \; b_G \; b_B]$, respectively, then the edge signature becomes $\varepsilon_{AB} = \{(a_R, b_R, \rho_1 = a_R/b_R), \ldots, (a_B, b_B, \rho_9 = a_B/b_B)\}$ as described in (B.2).

3. A preliminary mask is required in order to extract, on a patient-by-patient basis, the hyper-pixels that characterize the pigmented lesion and the surrounding skin, namely $\boldsymbol{a} = [a_R \; a_G \; a_B]$ and $\boldsymbol{b} = [b_R \; b_G \; b_B]$. We refer to the extraction of these characteristic pixels as the training of the ASRC.

4. The characteristic ratios are determined by the nine possible ratios that can be defined by these characteristic pixels, as explained in Step 2.

In preliminary results while utilizing the Euclidean-based classifier (as in the spectral examples presented in the previous section), the boundary of the mole was not correctly selected. Therefore, it was decided to utilize a more sophisticated classifier termed the support-vector machine (SVM) classifier. The SVM classifier is a well-known binary classifier that by using data points that are close to the boundary between the two classes (these data points are termed the support vectors) achieves great performance by searching for the optimal separating plane between these two classes. See, for example, [83] for an introductory coverage of the SVM classifier.

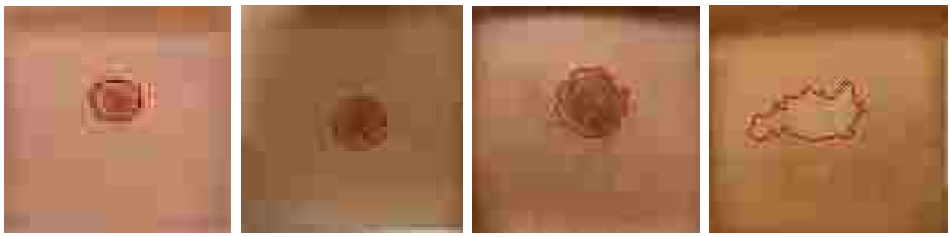


Figure 5.9: Examples of the automatic boundary selection performed by means of the ASRC algorithm with an SVM classifier.

We have trained the ASRC algorithm (including the spectral ratios and the SVM classifier) with the boundaries that were manually defined when computing the mean TRCs in Chapter 2. The ASRC computed the edge by determining those pixels within the color-image that define the greatest contrast in color. Some example images of the automatic boundary selected by the ASRC algorithm are shown in Fig. 5.9.

Next, we utilize the automated selection to define spatial features of the pigmented lesions that mimic those features utilized by the specialists when screening for possible malignant lesions using the ABCDE rule.

5.4 Definition of lesion spatial features

The Skin Cancer Foundation proposes four spatial features that should be considered by the specialists when diagnosing suspicious lesions and by the general public for self-screening [84]. These spatial features are:

Asymmetry of the pigmented lesion is in general accentuated for malignant tissue.

Border of the lesions is, in general, more irregular in malignant lesions than in benign lesions.

Color change can be present in malignant lesions. In general, benign lesions present an uniform color.

Diameter of the lesions is, in general, bigger for malignant lesions that the diameter of benign lesions.

The author believes that the random nature in the cancer cell growth leads to an irregular pigmentation of the malignant cells, which, in turn, leads to lesions that are asymmetric, with irregular border, with different coloration and bigger than their benign counterparts.

We mimic these spatial features for each patient by first defining an ellipse with the minimum area that contains the automated identified boundary. Once this ellipse is constructed, we compute its centroid, minor and major axis length and its area in pixels. With this information from the containing ellipse at hand, we define the spatial features as follows. The asymmetry is accounted by the ellipse eccentricity, e, which is defined as

$$e = \sqrt{1 - \frac{b}{a}}, \tag{5.5}$$

where a and b are the length of the major- and minor-axis lengths, respectively. The eccentricity is zero for perfect circles and close to one if the ellipse is totally asymmetrical. We measure the border irregularity of the lesion by comparing the perimeter of the automatically-detected boundary and the perimeter of the ellipse. (If the automatically-generated boundary of a lesion is irregular then its perimeter length will be high when compared to the perimeter of the containing ellipse.)

The color change is accounted by means of the spectral angle (SA) between pixels. The SA permits to determine the color similarity between two pixels by calculating the angle between them (assuming them as vectors in a space with dimensionality equal to the number of bands) [85]. For visible-color images, the cosine of the spectral

angle between two pixels, say $\boldsymbol{a} = [a_R\ a_G\ a_B]^T$ and $\boldsymbol{b} = [b_R\ b_G\ b_B]^T$, is defined as

$$\cos \alpha_{\boldsymbol{a},\boldsymbol{b}} = \frac{\boldsymbol{a}^T\boldsymbol{b}}{\|\boldsymbol{a}\|\|\boldsymbol{b}\|} \ . \tag{5.6}$$

If the color (spectrum) of these two pixels is similar then their spectral angle will be close to zero. By using the centroid of the containing ellipse as a reference point, we account the color change of the lesion by computing the spectral angle between each pixel and the pixel at the centroid. If the color of the lesion is uniform then the minimum and maximum values of SA will be similar. As the color changes, so does SA and the gap between the minimum and maximum values of SA begins to widen. Hence, we propose as a metric of the variability of the color of a lesion the difference $\max_i \alpha_i - \min_i \alpha_i$, where α_i is the SA between the ith pixel inside the lesion and the centroid pixel.

The diameter of the lesions is simply measured by the length of the major axis of the containing ellipse. Since each pixels sees around 300 microns of the tissue, the diameter in pixels is converted to the approximated physical diameter.

All the qualifications of the spatial features were defined in a way that their value increases as the signs of malignancy appear (based on the recommendations of the Skin Cancer Foundation).

5.5 Fusion of spatial features in the skin-cancer detection method

Here, we develop a method that merges the spatial features described above with the temporal features extracted from the TRCs. To this end, we assume that the spatial and temporal features are separable, i.e., they can be independently extracted and later merges without accounting their possible correlation.

To this end, we assume that the spatial features are also random variables. We denote the spatial features under the jth hypothesis by $\boldsymbol{\Theta}_{s,j} = \left[\theta_{s,j}^{(1)}\ \theta_{s,j}^{(2)}\ \theta_{s,j}^{(3)}\ \theta_{s,j}^{(4)}\right]^T$, where $\theta_{s,j}^{(1)}$ is the measure of lesion asymmetry, $\theta_{s,j}^{(2)}$ is the border irregularity measure, and so on. We use the subscript "s" to emphasize their spatial nature and differentiate them from the model random paramters defined for the stochastic model in Section 2.4.

In both the skin-cancer detection algorithm developed in Chapter 2 and its generalization developed in Chapter 4 all the temporal information was compactly contained in the KL coefficients. As a consequence, regardless of the algorithm selected to extract the temporal components, they will be represented by their KL coefficients, $S_{j,k} \sim \mathcal{N}(0, \lambda_{j,k})$ for $j = 0, 1$ and $k = 1, 2, \ldots$. In this generalization we aim to fuse the spatial and temporal features to define a consolidated detection problem. Since there is no *a priori* knowledge of the spatial random variables at this point, we assume they follow a Gaussian distribution with mean $\boldsymbol{\mu}_{s,j} = \mathsf{E}\left[\boldsymbol{\Theta}_{s,j}\right]$ and covariance matrix $\boldsymbol{\Sigma}_{s,j} = \mathsf{cov}\left(\boldsymbol{\Theta}_{s,j}\right)$. Under the assumption that there is no correlation between the spatial and the temporal components, we propose to state the spatio-temporal detection problem by the discrete binary hypothesis-testing problem:

$$H_0\ :\ \boldsymbol{Y} \sim \mathcal{N}\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\right)\ , \tag{5.7a}$$

$$H_0\ :\ \boldsymbol{Y} \sim \mathcal{N}\left(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1\right)\ , \tag{5.7b}$$

where the mean under the jth hypothesis is

$$\boldsymbol{\mu}_j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mu_{s,j}^{(1)} \\ \vdots \\ \mu_{s,j}^{(4)} \end{bmatrix}\ , \tag{5.8}$$

Table 5.6: Correlation coefficients between the spatial features extracted from patients with benign condition.

	$\theta_{s,j}^{(1)}$	$\theta_{s,j}^{(2)}$	$\theta_{s,j}^{(3)}$	$\theta_{s,j}^{(4)}$
$\theta_{s,j}^{(1)}$	1.00	-0.16	0.03	0.08
$\theta_{s,j}^{(2)}$		1.00	0.02	0.80
$\theta_{s,j}^{(3)}$			1.00	0.06
$\theta_{s,j}^{(4)}$				1.00

and the corresponding covariance matrix is block-diagonal owing the assumption of that the spatial and temporal components are uncorrelated. For the jth hypothesis ($j = 0, 1$) the covariance matrix is

$$\boldsymbol{\Sigma}_j = \left[\begin{array}{ccc:c} \lambda_{j,1} & & & \\ & \lambda_{j,2} & & \\ & & \ddots & \\ & & \lambda_{j,K} & \\ \hdashline & & & \boldsymbol{\Sigma}_{s,j} \end{array}\right], \tag{5.9}$$

where the upper-left block is the covariance matrix of the temporal component and the lower-right block is the covariance matrix of the spatial component. Note that at this stage we are assuming directly that only the K most important eigenvalues of each hypothesis are utilized.

The detection problem presented in (5.7) has a well-known likelihood ratio [86]; nevertheless, in our approach the observation vector, $\boldsymbol{Y}$, is obtained from both eigenfunction sets. Therefore one must be careful before applying off-the-shelf approaches.

By exploring the correlation coefficients between the spatial features, one can note that they have modest correlation except for the border irregularity and the diameter of the lesion (see Table 5.6). The author believes that this is an artificial correlation due to the poor illumination and the natural curvature of the skin, which for large lesions affects the color contrast between the pigmented and surrounding

tissue. Moreover, when the visible images were acquired, there was no control over the illumination, making this correlation more accentuated than what it actually is. As a consequence, we assume that the spatial covariance is also diagonal. If it turns out that the spatial covariance matrix is not diagonal, then we can define the modal matrix

$$\boldsymbol{V} = \left[\begin{array}{c:c} \boldsymbol{I}_K & \\ \hdashline & \boldsymbol{V}_s \end{array}\right] ,$$

where $\boldsymbol{V}_s$ is such that $\boldsymbol{V}_s^T \boldsymbol{\Sigma}_{s,j} \boldsymbol{V}_s = \boldsymbol{\Lambda}_s$, where $\boldsymbol{\Lambda}_s$ is the diagonal spatial eigenvalue matrix for the spatial covariance matrix. After the diagonalization the test-statistics for the spatio-temporal detection problem is

$$Z = \log L(Y) = \sum_{k=1}^{K'} \frac{(Y_{0,k} - \mu_{0,k})^2}{\sigma_{0,k}^2} - \frac{(Y_{1,k} - \mu_{1,k})^2}{\sigma_{1,k}^2}, \tag{5.10}$$

where $K' = K + 4$ (the sum of temporal and spatial features). It is clear that, under the stated assumptions, both expressions in the argument of the summation are χ^2-distributed random variables. In a similar way as in the single-TRC approach, the test-statistics will be a linear combination of χ^2-distributed random variables with coefficients $a_k = \frac{1}{2}\left(1 - \frac{\sigma_{0,k}^2}{\sigma_{1,k}^2}\right)$ under H_0, and with coefficients $b_k = \frac{1}{2}\left(\frac{\sigma_{1,k}^2}{\sigma_{0,k}^2} - 1\right)$ under H_1. Next, we utilize these expressions next to evaluate the theoretical performance of the spatio-temporal approach.

5.6 Theoretical performance

In this section we evaluate the theoretical performance of the spatio-temporal approach. The detection problem is slightly different as compared with the detection problem in continuous-time presented in the previous two chapters but the decision rule is determined in the same manner.

We present a comparison between the theoretical ROC curves of the spatio-

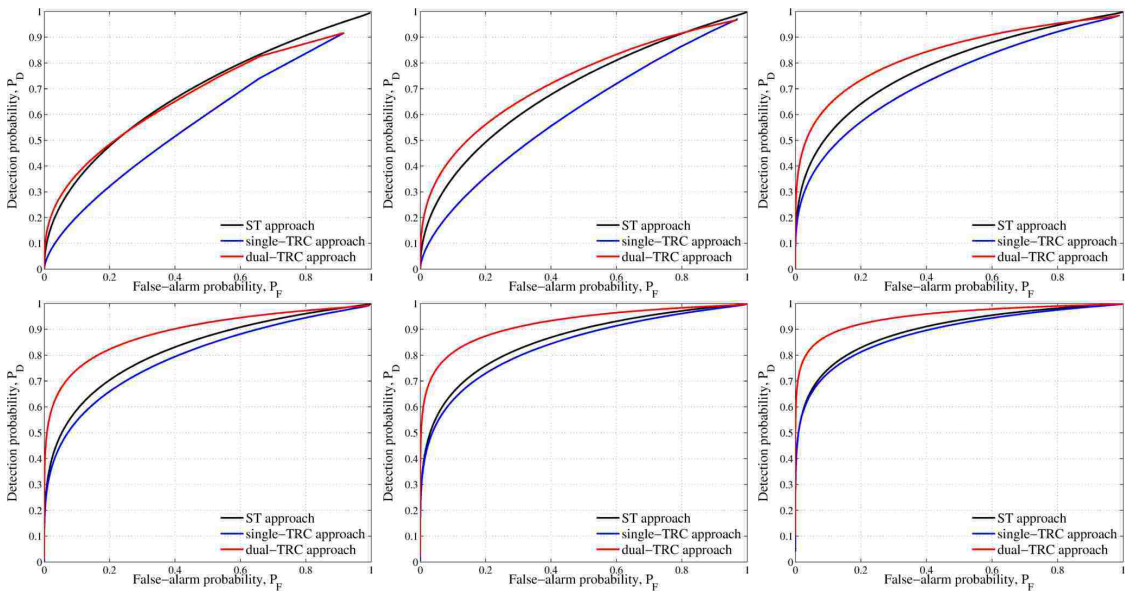


Figure 5.10: Theoretical performance comparison by means of the ROC curves between the single-TRC approach (red), the dual-TRC approach (black) and the spatio-temporal approach (blue) for different number of eigenpairs (1–6) utilized to construct the test-statistics.

temporal (ST) approach and the previous two algorithms in Fig. 5.10 for changing the eigenvalue-eigenfunction pairs from one up to six, and in Fig. 5.11 for seven up to twelve eigenfunctions. For this study the spatial features where always included and we only introduced more temporal features by incrementing the number of eigenpairs. The results depicted in this figures is the average ROC curve over the 200 permutations of 110 patients used in training the algorithm.

It is observed from the presented results that for small number of temporal eigenfunctions, the ST approach performs slightly better than that of the dual-TRC approach (see, Fig. 5.10 (top-left) for the case of one eigenfunction). As the number of temporal components is increased the theoretical performance of the ST and the dual-TRC approaches is similar. After the temporal components are sufficient to undermine the contribution of the spatial features, the ST and the single-TRC approach

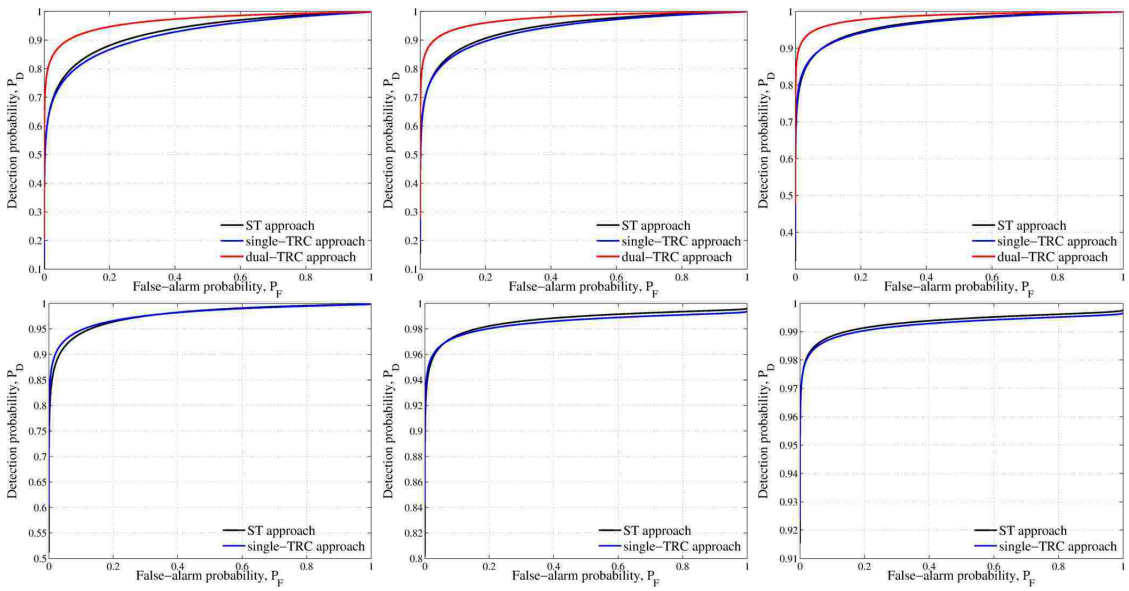


Figure 5.11: Theoretical performance comparison by means of the ROC curves between the single-TRC approach (red), the dual-TRC approach (blue) and the spatio-temporal approach (black) for different number of eigenpairs (7–12) utilized to construct the test-statistics.

achieve virtually the same performance. See, for example, Fig. 5.10 (bottom-right) for the case of six eigenpairs. As expected, for higher number of eigenpairs the ST generalization does not present any improvement over the single-TRC approach.

In Fig. 5.12 we show the mean AUC (over 200 permutations) as a function of the eigenvalue-eigenfunction pairs for the four training settings explained before. We observe that for all the cases, when a single eigenpair is used, the ST approach outperforms the other two alternatives. As we increase the number of eigenpairs, the ST approach presents a performances that appears to be between the other two alternatives. The performance of all three algorithms plateau at an average AUC of 0.99.

Next, we explore the mean empirical detection probability over the same 200 permutations of patients created for all three methods.

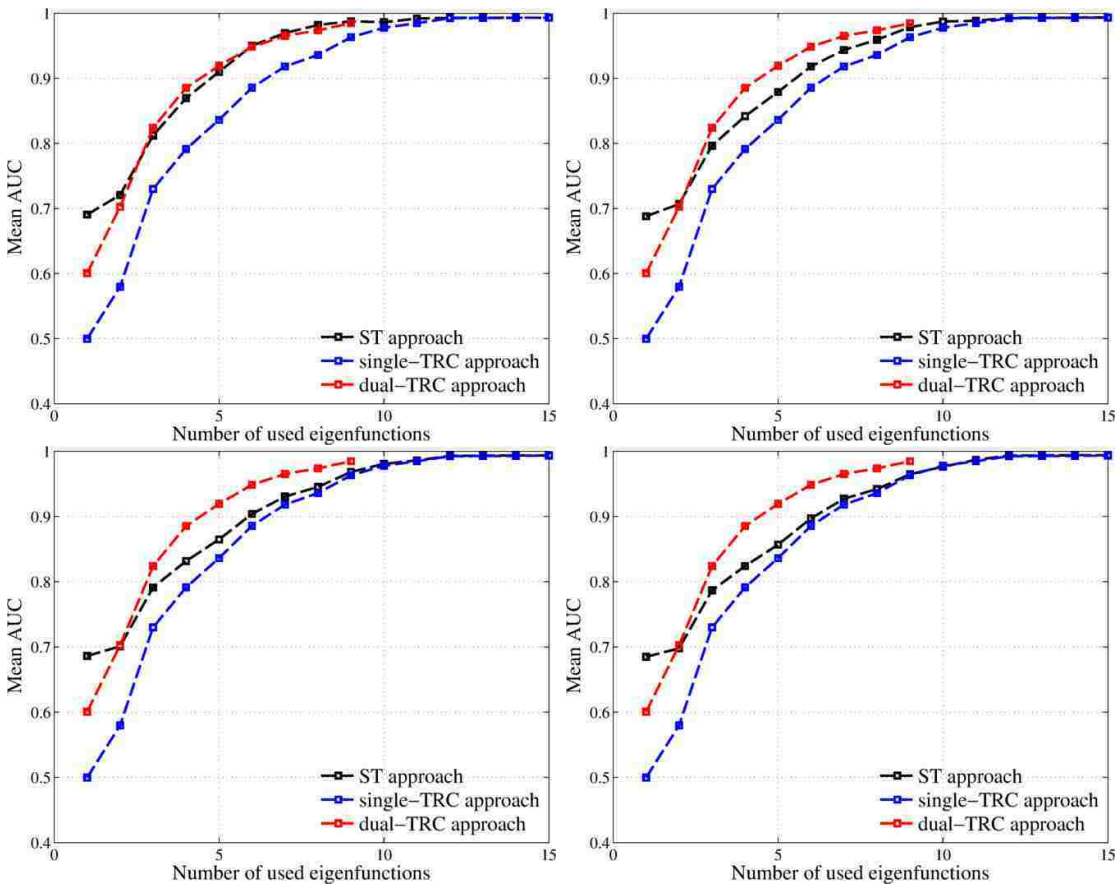


Figure 5.12: Comparison of the spatio-temporal (ST) approach, the single TRC and the double TRC approach by using the mean AUC for different number of used eigenvalue-eigenfunction pairs: using 60 patients in the training (top-left), 80 patients (top-right), 100 patients (bottom-left) and 110 patients (bottom-right).

5.7 Empirical performance

As it was observed in the theoretical performance analysis, the inclusion of spatial features aided the algorithm when a small number of eigenfunctions were used. As the temporal features have sufficient weight to diminish the contribution of the spatial features, the temporal features were dominating the overall performance. Interestingly, we observe the same trend for the empirical results. Let us focus our attention in Table 5.7. For the first three eigenfunctions we note that the ST approach present

Table 5.7: Mean and mean absolute error (given within the parenthesis) of the empirical detection probability for different levels of prescribed false-alarm probability when 80 patients are used in the training and 60 in the testing stage. Here, s-TRC and ST stand for the single-TRC and spatio-temporal approaches, respectively.

Number of used eigenfunctions	False-alarm probability			
	$P_F = 0.1$		$P_F = 0.01$	
	s-TRC	ST	s-TRC	ST
1	0.00 (± 0.00)	0.69 (± 0.10)	0.00 (± 0.00)	0.38 (± 0.12)
2	0.02 (± 0.04)	0.68 (± 0.09)	0.02 (± 0.04)	0.39 (± 0.13)
3	0.42 (± 0.14)	0.56 (± 0.11)	0.42 (± 0.14)	0.26 (± 0.10)
4	0.46 (± 0.12)	0.23 (± 0.09)	0.46 (± 0.12)	0.12 (± 0.07)
5	0.80 (± 0.11)	0.23 (± 0.10)	0.80 (± 0.11)	0.12 (± 0.06)
6	0.91 (± 0.09)	0.25 (± 0.10)	0.91 (± 0.09)	0.14 (± 0.07)
7	0.94 (± 0.11)	0.25 (± 0.09)	0.94 (± 0.11)	0.16 (± 0.07)
8	0.81 (± 0.31)	0.25 (± 0.09)	0.81 (± 0.31)	0.18 (± 0.07)
9	0.86 (± 0.24)	0.63 (± 0.33)	0.86 (± 0.24)	0.56 (± 0.36)
10	0.85 (± 0.25)	0.99 (± 0.01)	0.85 (± 0.25)	0.99 (± 0.01)
11	0.90 (± 0.19)	1.00 (± 0.00)	0.90 (± 0.19)	1.00 (± 0.00)
12	0.86 (± 0.23)	0.99 (± 0.01)	0.86 (± 0.23)	099 (± 0.01)

better empirical performance than the single-TRC approach. Unfortunately, this increased performance is totally lost as more temporal features are introduced. The same can be observed in Table 5.8 and Table 5.9 for the case of training the algorithm with 100 and 110 patients, respectively. (Nevertheless, this performance does not surpass the empirical performance achieved by the dual-TRC approach.) In addition, the spatial features seem to diminish the empirical performance of the presented algorithm, as it can be noted in Table 5.9, when for 10 eigenfunctions, the ST approach does not reach perfection as the single and dual-TRC approaches do.

5.8 Concluding remarks

In this chapter, we have explored the inclusion of spatial features into the skin cancer detection algorithm (termed the single-TRC approach) in order to propose and test

Table 5.8: Mean and mean absolute error (given within the parenthesis) of the empirical detection probability for different levels of prescribed false-alarm probability when 100 patients are used in the training and 40 in the testing stage. Here, s-TRC and ST stand for the single-TRC and spatio-temporal approaches, respectively.

Number of used eigenfunctions	False-alarm probability			
	$P_F = 0.1$		$P_F = 0.01$	
	s-TRC	ST	s-TRC	ST
1	0.00 (± 0.00)	0.69 (± 0.15)	0.00 (± 0.00)	0.38 (± 0.17)
2	0.00 (± 0.00)	0.68 (± 0.15)	0.00 (± 0.00)	0.37 (± 0.16)
3	0.15 (± 0.09)	0.53 (± 0.15)	0.13 (± 0.07)	0.23 (± 0.13)
4	0.29 (± 0.12)	0.19 (± 0.11)	0.26 (± 0.10)	0.10 (± 0.09)
5	0.69 (± 0.16)	0.18 (± 0.12)	0.67 (± 0.16)	0.09 (± 0.08)
6	0.89 (± 0.09)	0.17 (± 0.12)	0.89 (± 0.09)	0.09 (± 0.09)
7	0.96 (± 0.08)	0.16 (± 0.11)	0.96 (± 0.08)	0.09 (± 0.08)
8	0.85 (± 0.25)	0.16 (± 0.10)	0.85 (± 0.25)	0.13 (± 0.09)
9	0.88 (± 0.22)	0.34 (± 0.25)	0.88 (± 0.22)	0.28 (± 0.23)
10	0.93 (± 0.14)	0.98 (± 0.03)	0.93 (± 0.14)	0.98 (± 0.04)
11	0.86 (± 0.24)	1.00 (± 0.00)	1.00 (± 0.00)	1.00 (± 0.00)
12	0.89 (± 0.20)	1.00 (± 0.00)	1.00 (± 0.00)	1.00 (± 0.00)

a spatio-temporal approach to undertake the detection. We introduced an extension to the SRC algorithm that exploits color contrast in multicolor RGB images in order to detect the edges due to color changes instead of intensity changes. This edge-detection algorithm, termed the ASRC algorithm, is virtually insensitive to intensity changes as observed in the presented results. We later exploited this property of the ASRC algorithm to define an automatic boundary selector, which defines the lesions' boundary over the visible image by looking solely on the color difference between the pigmented and the surrounding tissue. This automatic lesion selector requires the user to define a training mask to learn, on a patient-by-patient basis, the color characteristics of each region. When the training mask is defined, the lesion selection is generated as an output.

By using the automatically selected lesion, we proposed and mathematically defined four spatial features that mimic the spatial features utilized by the physicians

Table 5.9: Mean and mean absolute error (given within the parenthesis) of the empirical detection probability for different levels of prescribed false-alarm probability when 110 patients are used in the training and 30 in the testing stage. Here, s-TRC and ST stand for the single-TRC and spatio-temporal approaches, respectively.

Number of used eigenfunctions	False-alarm probability			
	$P_F = 0.1$		$P_F = 0.01$	
	s-TRC	ST	s-TRC	ST
1	0.00 (± 0.00)	0.68 (± 0.20)	0.00 (± 0.00)	0.38 (± 0.24)
2	0.00 (± 0.00)	0.66 (± 0.19)	0.00 (± 0.00)	0.42 (± 0.22)
3	0.19 (± 0.19)	0.57 (± 0.23)	0.16 (± 0.18)	0.27 (± 0.18)
4	0.29 (± 0.21)	0.14 (± 0.18)	0.26 (± 0.20)	0.08 (± 0.12)
5	0.76 (± 0.21)	0.13 (± 0.17)	0.73 (± 0.21)	0.08 (± 0.12)
6	0.89 (± 0.15)	0.12 (± 0.16)	0.89 (± 0.15)	0.08 (± 0.12)
7	0.97 (± 0.06)	0.13 (± 0.16)	0.97 (± 0.06)	0.07 (± 0.11)
8	0.84 (± 0.26)	0.14 (± 0.17)	0.84 (± 0.26)	0.09 (± 0.14)
9	0.90 (± 0.18)	0.26 (± 0.27)	0.90 (± 0.18)	0.21 (± 0.25)
10	1.00 (± 0.00)	0.94 (± 0.11)	1.00 (± 0.00)	0.92 (± 0.14)
11	1.00 (± 0.00)	1.00 (± 0.00)	1.00 (± 0.00)	1.00 (± 0.00)
12	1.00 (± 0.00)	1.00 (± 0.00)	1.00 (± 0.00)	1.00 (± 0.00)

when diagnosing a suspicious lesion. The set of four spatial features were fused with the temporal features extracted by the previous methods to propose and test spatio-temporal approach for skin cancer detection. Unfortunately, the ST-approach only helped with some modest levels of performance improvement when only a few eigenvalue-eigenfunction pairs we used. As more eigenpairs were included, the ST approach performed equally well (theoretical performance) or worse (empirical performance) than the temporal approach proposed originally.

There are several reasons that explain the degraded performance when spatial features are included with eigenpairs. The most important reason is that when the physicians utilize the ABCDE test, they do not only review those four spatial features, but also take into account the clinical history of the patients, their family history and other possible moles that may serve as local reference to see how abnormal a suspicious mole can be.

On one hand, we don't see any advantage of including spatial information into the skin-cancer detection problem. As a consequence, the major contribution of this chapter to the skin cancer detection problem is the automatic lesion boundary identification. On the other hand, both the original skin-cancer detection method, developed in Chapter 2, and its generalization developed in Chapter 4 that includes a self-referenced TRC, had already achieved, to the best of our knowledge, a performance superior to any other alternative proposed in the literature.

Chapter 6

Summary and future work

6.1 Summary

In this dissertation we have proposed and tested a skin-cancer detection algorithm that uses dynamic thermal imaging and achieves promising results. We have developed a physics-based stochastic mathematical model that describe the time evolution of the thermal recovery curves (TRCs) and used the pairwise correlation of those parameters to define an analytical structure of the autocorrelation function (ACF) that describes the TRCs.

With this at hand, we have solved the continuous-time detection problem that aims to classify, in an optimal way, the average TRC of a subject to be either malignant or benign. We have proposed and tested three different alternatives to solve such a problem. The first alternative uses the analytical ACF under each hypothesis and it is solved by means of Grenander's approach. This alternative was studied in detail and it presents one of the major contributions of this dissertation.

The second proposed alternative to solve the detection problem is by introducing

a second TRC under each hypothesis as a reference of the surrounding tissue. This is presented as a generalization of the previous approach. The self-referencing TRC obtained from the tissue that surrounds the mole tissue boosted the performance of the skin-cancer detector when compared to the performance of the first alternative.

The third and last alternative studied here introduces spatial features to the detection problem and it is solved assuming separability between the spatial and temporal components. We have introduced the adaptive version of the SRC algorithm (termed the ASRC algorithm) that fuses classification and edge-detection and used it to automatically detect the lesion boundaries by exploiting the color contrast between the mole tissue and the surrounding skin. After defining the boundary by means of the ASRC algorithm, we define and extract spatial properties of the lesions in a similar fashion as the dermatologist do it when applying the ABCDE test. Moreover, we propose and test a methodology to fuse these spatial features (Asymmetry, Border irregularity, Color change and Diameter) with the TRC (time Evolution) in a single hypothesis-testing problem. While the performance of this alternative was as good as expected the proposed automatic boundary selection constitute another major contribution of this dissertation.

In summary, while providing a rigorous mathematical foundation for the viability of the dynamic thermal recovery approach for skin-cancer detection, the research completed in this dissertation also provides the first reliable, accurate and non-invasive diagnosis method for preliminary skin-cancer detection. This dissertation, therefore, paves the way for future clinical studies to produce new skin-cancer diagnosis practices that minimize the need for unnecessary biopsies without sacrificing reliability.

6.2 Future work

Some possible future research ideas that continue or improve what was presented in this dissertation are briefly described below.

The whitening approach was described as a possible alternative to solve the unequal covariance detection problem. As we discussed it may be hard if not impossible to implement, but it is an approach should be followed with detail before discarding it completely.

Application of machine learning algorithms to the KL coefficients sounds a reasonable idea since they statistically characterize the TRCs under two hypothesis. Binary classifiers such as SVM classifiers may lead to good results in discriminating between benign and malignant conditions.

Estimate the pdf of the KL coefficients numerically should improve the performance of the algorithm once enough patient data is acquired. The direct estimation of their distribution will avoid the introduced errors due to Gaussian assumptions.

The spatio-temporal approach may be explored in more detail considering scenarios like that the spatial and temporal feautures are not separable. By defining joint features by means of a multiple-signal setting with spatial correlation may lead to improved results in the detector.

Appendix A

Probability distribution function of a linear combination of Chi-square–distributed random variables

Here we present the mathematical formulation of the cumulative distribution function (CDF) for a linear combination of Chi-square distributed random variables. We follow the notation presented in the seminal papers presented by Robbins [87] and Pachares [76].

A.1 Mathematical formulation

Let $S_r = \frac{1}{2}\left(a_1\chi^2_{m_1} + a_2\chi^2_{m_2} + \cdots a_r\chi^2_{m_r}\right)$, where the $\chi^2_{m_i}$, for $i = 1, \ldots, r$ are independent random variables having a central chi-square distribution with m_i degrees of freedom. Let $S^*_r = \frac{1}{2}\left(a_1^{-1}\chi^2_{m_1} + a_2^{-1}\chi^2_{m_2} + \cdots a_r^{-1}\chi^2_{m_r}\right)$, $M = \sum_{i=1}^{r} m_i$, and $a_i > 0$.

Then, the CDF of S_r, $G_r(\tau)$, is defined as

$$G_r(\tau) = Pr(S_r \leq \tau) = \sqrt{\frac{\tau^M}{a_1^{m_1} \cdots a_r^{m_r}}} \sum_{k=0}^{\infty} \frac{(-\tau)^k}{k!} \frac{\mathsf{E}\left[S_r^*\right]_k}{\Gamma(M/2+k+1)} \ , \tag{A.1}$$

where $\mathsf{E}\left[S_r^*\right]_k$ represents the kth moment of S_r^* and $\Gamma(\cdot)$ is the Gamma function. The series in (A.1) is proved to converge absolutely by Pachares [76].

A.2 Efficient software implementation

In order to compute $\mathsf{E}\left[S_r^*\right]_k$, we utilize its cumulants. The jth cumulant is given by $\kappa_j = \frac{1}{2}(j-1)! \sum_{i=1}^{r} m_i a_i^{-j}$ [76]. The relationship between the kth moment, $\mathsf{E}\left[S_r^*\right]_k$, and the jth cumulant, κ_j, is given by [88]

$$\begin{aligned}
\mathsf{E}\left[S_r^*\right]_k &= \sum_{j=0}^{k-1} \binom{k-1}{j} \mathsf{E}\left[S_r^*\right]_j \kappa_{k-j} \\
&= \sum_{j=0}^{k-1} \frac{(k-1)!}{j!(k-1-j)!} \mathsf{E}\left[S_r^*\right]_j \left(\frac{1}{2}(k-j-1)! \sum_{i=1}^{r} m_i a_i^{-(k-j)} \right) \\
&= \frac{1}{2} \sum_{j=0}^{k-1} \frac{(k-1)!}{j!} \mathsf{E}\left[S_r^*\right]_j \left(\sum_{i=1}^{r} m_i a_i^{-(k-j)} \right) \ ,
\end{aligned} \tag{A.2}$$

where by convention $\mathsf{E}\left[S_r^*\right]_0 = 1$. Let us define $\boldsymbol{m} = [m_1 \ m_2 \cdots \ m_r]^T$ and $\boldsymbol{a} = [a_1 \ a_2 \cdots \ a_r]^T$, and denote by $\boldsymbol{a}^{-j}$ the vector $\boldsymbol{a}^{-j} = [a_1^{-j} \ a_2^{-j} \cdots \ a_r^{-j}]^T$ then the CDF given in (A.1) can be computed as

$$G_r(\tau) = \sqrt{\frac{\tau^M}{a_1^{m_1} \cdots a_r^{m_r}}} \sum_{k=0}^{\infty} \frac{(-\tau)^k}{2k\Gamma(M/2+k+1)} \left[\sum_{j=0}^{k-1} \frac{\mathsf{E}\left[S_r^*\right]_j}{j!} \boldsymbol{m}^T \boldsymbol{a}^{-(k-j)} \right] \ .$$

The probability density function (PDF) of the distribution is obtained simply by computing the first derivative of (A.1) with respect to the parameter τ.

A.3 Monte Carlo-based validation

Two simple experiments were created to validate the coded CDF and PDF Matlab functions.

In the first experiment, a serie of 10^4 realizations of central chi-squared random variables was generated in Matlab. Each realization was then multiplied by a constant (4 in this example) to generate realization of the random variable $Z = 2\chi^2$. Using all these realizations of the random variable Z, we obtained the estimated PDF by the normalized histogram (blue columns in Fig. A.1(a)). We then utilized the model of the PDF to compute the theoretical distribution. The result is shown by the red line in Fig. A.1(a).

In the second experiment, we wanted to create the random variable $Z = 3\chi_1^2 + \chi_2^2 + 2\chi_3^2 + 4\chi_4^2$, where χ_i^2, for $i = 1, \ldots, 4$ are independent chi-square distributed random variables. The estimated PDF (obtained again by the normalization of the histogram over the 10^4 realizations) and the modeled PDF are shown in Fig. A.1(b) by the blue columns and the red line, respectively.

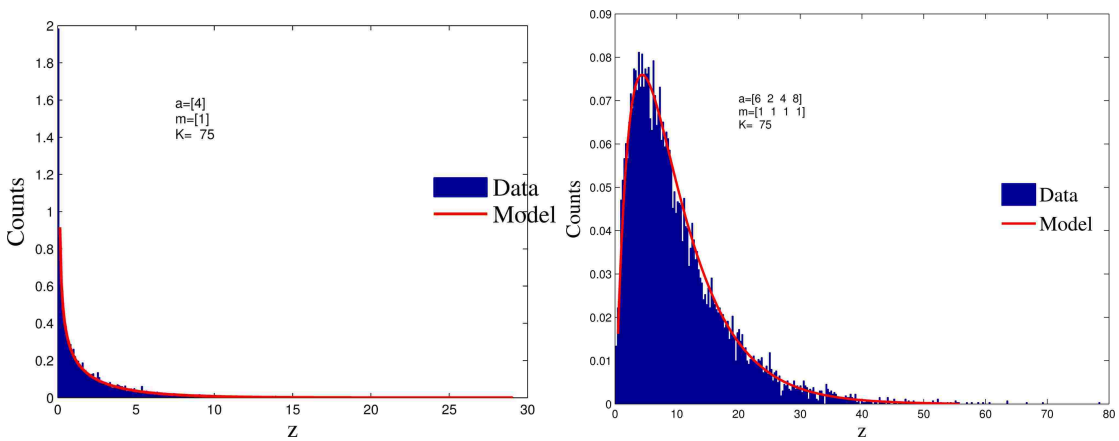


Figure A.1: Experimental validation of the CDF and PDF implementation for a linear combination of Chi-squared distributes random variables. (a) Estimated and modeled PDFs for the random variable $Z = 2\chi^2$, and (b) estimated and modeled PDFs for the random variable $Z = 3\chi_1^2 + \chi_2^2 + 2\chi_3^2 + 4\chi_4^2$ where χ_i^2, for $i = 1, \ldots, 4$ are independent chi-square distributed random variables

Appendix B

The spectral-ratio contrast algorithm

B.1 Introduction

The goal of any edge detector is to define an edge map,

$$\mathcal{F} : \mathbb{R}^{I \times J \times K} \mapsto \{0, 1\}^{I \times J} \ ,$$

that assigns the value 1 to the pixel location (i, j) if $\mathbf{u}(i, j)$ belongs to an edge, while assigning the value 0 otherwise.

For two types of materials A and B, the SRC algorithm builds the edge map $\mathcal{F}_{\text{AB}}$ in three stages: (i) model-based edge signature identification, (ii) sparse spatio-spectral mask development, and (iii) edge discrimination.

For simplicity, we first describe the SRC algorithm assuming only two distinct materials A and B in scenes, and later we describe the extensions to multiple materials.

B.2 Model-based edge signature identification

Given two distinct materials A and B in a scene that is probed by a sensor, we seek those bands for A and B whose ratios can best identify the spatial locations that correspond to the boundary points between the two materials. Let $\mathbf{a} = (a_1, \ldots, a_K)$ and $\mathbf{b} = (b_1, \ldots, b_K)$ denote hyper-pixels corresponding to materials A and B, respectively. For example, the vector $\mathbf{a}$ can be obtained by taking the average of all hyper-pixels as material A is probed by the sensor.

We next define the *spectral ratio index* between materials A and B as the $K \times K$ matrix

$$\mathrm{A/B} \triangleq \begin{pmatrix} \frac{a_1}{b_1} & \cdots & \frac{a_1}{b_K} \\ \vdots & \ddots & \vdots \\ \frac{a_K}{b_1} & \cdots & \frac{a_K}{b_K} \end{pmatrix} . \tag{B.1}$$

We define the *signature* of the edge between materials A and B as a small collection of size R, where $R \ll K$, of elements of (B.1) that can reliably identify, as described below, the spatial indices of the edge between materials A and B. We denote the edge signature by

$$\mathcal{E}_{\mathrm{AB}} = \{(p_1, q_1, \rho_1), \ldots, (p_R, q_R, \rho_R)\}, \tag{B.2}$$

where p_r and q_r are the band indices associated with the ratios $\rho_r = a_{p_r}/b_{q_r}$, $r = 1, \ldots, R$. The integer R is the length of the edge signature.

The selection of the triplets (p_r, q_r, ρ_r), $r = 1, \ldots, R$, is made as follows. First, we select S bands $\{i_1, \ldots, i_S\}$, where the materials A and B exhibit maximum sep-

aration, i.e.,

$$
\begin{aligned}
i_1 &= \arg \max_{1\le i\le K} |a_i - b_i| \ , \\
i_2 &= \arg \max_{\substack{1\le i\le K, \\ i\neq i_1}} |a_i - b_i| \ , \\
&\vdots \\
i_S &= \arg \max_{\substack{1\le i\le K \\ i\neq i_1,\ldots,i_{S-1}}} |a_i - b_i| \ .
\end{aligned}
$$

Clearly, in the case of HS imagery the search for the best S bands will be in a larger space compared to the case of MS imagery, but the same procedure is utilize in both types of imagery. Hence, once the edge signature is obtained the complexity associated with processing MS and HS imagery are identical.

Next, we compute the spectral ratios using all possible band combinations: $\rho_{pq} = a_{i_p}/b_{i_q}$, $1 \le p, q \le S$. Without loss of generality, we may assume that all ratios are less than or equal to unity. (If $\rho_{pq} > 1$ for some p and q, we simply replace it by its reciprocal.) Finally, we define $\mathcal{E}_{\rm AB}$ by selecting the R ratios that exhibit the strongest spectral contrast between the classes. To rank the ratios according to their spectral contrast, we note that owing to the convention that $\rho_{pq} \le 1$, the ratios closest to zero correspond to the strongest spectral contrast between any two bands. Thus, we select the first pair of bands, $\{p_1, q_1\}$, as the pair corresponding to the smallest ratio,

$$\rho_1 = \rho_{p_1 q_1} = \arg \min_{1\le p,q\le S} \rho_{pq} \, ,$$

the second pair of bands $\{p_2, q_2\}$ as the pair corresponding to the next smallest ratio,

$$\rho_2 = \rho_{p_2 q_2} = \arg \min_{\substack{1\le p,q\le S \\ p\neq p_1 \\ q\neq q_1}} \rho_{pq} \, ,$$

and so on. We combine the ordered band indices and the corresponding ranked ratios to define the R triplets in the edge signature (B.2). Since all the ratios are less than or equal to unity, it can be easily shown that the definition of the edge signature is invariant under the change in the order of the materials A and B.

B.3 Sparse spatio-spectral mask development

We denote a spatial mask, $\mathcal{M}$, at a pixel (i, j) as a list of pixel pairs surrounding the pixel of interest. More precisely, $\mathcal{M}(i, j)$ is the union of M neighborhoods of pixels, $\mathcal{M}(i, j) = \cup_{m=1}^{M} \mathcal{N}_m(i, j)$, where each neighborhood consists of two distinct pixels surrounding (i, j), i.e., $\mathcal{N}_m(i, j) = \{u^{m-}(i, j), u^{m+}(i, j)\}$. For example, one can define a 3×3 mask centered at the pixel (i, j) that excludes the center pixel by taking the union of four neighborhoods, $\mathcal{N}_1 = \{u^{1-}(i, j), u^{1+}(i, j)\} = \{u(i - 1, j), u(i + 1, j)\}$, $\mathcal{N}_2 = \{u^{2-}(i, j), u^{2+}(i, j)\} = \{u(i, j - 1), u(i, j + 1)\}$, etc.

Next, we define the operation of the joint spatio-spectral mask at the (i, j)-th pixel by computing the ratios between each of the M pixel pairs of the spatial mask $\mathcal{M}$, at each of the R band pairs given in the edge signature $\mathcal{E}_{AB}$. For example, using the pair of hyper-pixels defined by $\mathcal{N}_1$ and the pair of bands given by first triplet of the edge signature, (p_1, q_1, ρ_1), one can define the ratios $u_{p_1}^{1-}(i, j)/u_{q_1}^{1+}(i, j)$ and $u_{p_1}^{1+}(i, j)/u_{q_1}^{1-}(i, j)$. Namely, the application of the spatio-spectral mask to each location (i, j) results in a $2M \times R$ matrix of "features." (Compare this to a gray-scale image when the application of a spatial mask to a pixel results in a scalar.) Now the application of the spatio-spectral mask to the entire image cube defines the mapping

$$\mathcal{K}_{\mathrm{AB}} : \mathbb{R}^{I \times J \times K} \mapsto \left(\mathbb{R}^{2M \times R}\right)^{I \times J} \ , \tag{B.3}$$

where the (i, j)th entry of $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})$ will be a $2M \times R$ feature matrix of spectral ratios

given by

$$\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j) = \begin{pmatrix} \dfrac{u_{p_1}^{1-}(i,j)}{u_{q_1}^{1+}(i,j)} & \cdots & \dfrac{u_{p_R}^{1-}(i,j)}{u_{q_R}^{1+}(i,j)} \\ \vdots & \ddots & \vdots \\ \dfrac{u_{p_1}^{M-}(i,j)}{u_{q_1}^{M+}(i,j)} & \cdots & \dfrac{u_{p_R}^{M-}(i,j)}{u_{q_R}^{M+}(i,j)} \\ \dfrac{u_{p_1}^{1+}(i,j)}{u_{q_1}^{1-}(i,j)} & \cdots & \dfrac{u_{p_R}^{1+}(i,j)}{u_{q_R}^{1-}(i,j)} \\ \vdots & \ddots & \vdots \\ \dfrac{u_{p_1}^{M+}(i,j)}{u_{q_1}^{M-}(i,j)} & \cdots & \dfrac{u_{p_R}^{M+}(i,j)}{u_{q_R}^{M-}(i,j)} \end{pmatrix} . \tag{B.4}$$

For convenience, we denote the entries of the matrix $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j)$ as

$$\kappa_{\mathrm{AB}}(i,j;m^-,r) = u_{p_r}^{m-}(i,j)/u_{q_r}^{m+}(i,j) \ ,$$

and

$$\kappa_{\mathrm{AB}}(i,j;m^+,r) = u_{p_r}^{m+}(i,j)/u_{q_r}^{m-}(i,j) \ .$$

The use of both $\kappa_{\mathrm{AB}}(i,j;m^-,r)$ and $\kappa_{\mathrm{AB}}(i,j;m^+,r)$ in (B.4) is required to account for the two possible material configurations at the mth hyper-pixel pair. Specifically, the first ratio captures the case when the hyper-pixel $\boldsymbol{u}^{m-}(i,j)$ is, for example, from material A and $\boldsymbol{u}^{m+}(i,j)$ is from material B, whereas the second ratio is needed to account for the possibility that $\boldsymbol{u}^{m-}(i,j)$ is from material B and $\boldsymbol{u}^{m+}(i,j)$ is from material A. Therefore, the use of the two ratios removes dependence on the direction of the transition between A and B, and it is similar to the use of the magnitude in the gradient operator to achieve its rotational invariance.

B.4 Edge identification

The third stage of the SRC algorithm is the utilization of the sparse spatio-spectral mask, $\mathcal{K}_{\mathrm{AB}}$, to identify the edges between materials A and B. The proposed edge-

identification process is based on the following rationale. In the ideal case when no noise is present and the image under test is comprised only of hyper-pixels with the exact same value of the characteristic hyper-pixels **a** and **b**, the output of the spatio-spectral mask will perfectly match the values of the ratios obtained from the edge signature. To illustrate this point, assume the same example given earlier, where the spatio-spectral mask is given by (5.1). When we have a horizontal edge, the second and fourth ratios will have a value that is not meaningful, but the first or third ratio will match the ratio from the edge signature, $\rho_{p_1 q_1} = a_{p_1}/b_{q_1}$. Indeed, if the upper pixel is from material A and the lower pixel is from material B, then the first entry of (5.1) will be $u^{1-}_{p_1}(i,j)/u^{1+}_{q_1}(i,j) = a_{p_1}/b_{q_1}$ (which matches $\rho_{p_1 q_1}$) and the third entry will be $u^{1+}_{p_1}(i,j)/u^{1-}_{q_1}(i,j) = b_{p_1}/a_{q_1}$. Conversely, when the upper pixel is from material B and the lower pixel is from material A, the computed ratios are switched, which means that the third entry of (5.1) will match $\rho_{p_1 q_1}$. This example shows that if an edge is present then at least one row of $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})(i,j)$ will perfectly match the ratios from the edge signature.

When noise is present, we allow a matching tolerance to account for the similarity between the outcome of the mask and the edge signature ratios. This can be accomplished by defining the mapping

$$\Delta : \left(\mathbb{R}^{2M\times R}\right)^{I\times J} \mapsto \left(\{0,1\}^{M\times R}\right)^{I\times J} \ ,$$

where the (i,j)th entry of $\mathcal{K}_{\mathrm{AB}}(\mathbf{u})$ is used to form the $M \times R$ binary indicator matrix

$$\Delta\big(\mathcal{K}_{\mathrm{AB}}(\mathbf{u})\big)(i,j) = \begin{pmatrix} \delta_1^1(i,j) & \cdots & \delta_R^1(i,j) \\ \delta_1^2(i,j) & \cdots & \delta_R^2(i,j) \\ \vdots & \ddots & \vdots \\ \delta_1^M(i,j) & \cdots & \delta_R^M(i,j) \end{pmatrix} \tag{B.5}$$

and the entries $\delta_r^m(i,j)$ are assigned the values of 0 or 1 according to the rule

$$\delta_r^m(i,j) = \begin{cases} 1, & \text{if } |\kappa_{\text{AB}}(i,j;m^-,r) - \rho_r| < \epsilon \ , \\ 1, & \text{if } |\kappa_{\text{AB}}(i,j;m^+,r) - \rho_r| < \epsilon \ , \\ 0, & \text{otherwise} \ . \end{cases} \tag{B.6}$$

Here, the tolerance parameter, ϵ, accounts for both the natural variability and the presence of noise in the spectral data for materials A and B.

Ideally, if the mth hyper-pixel pair belongs to the same material type, then the test in (B.6) will return the value of zero. Conversely, if the hyper-pixels forming the pair are from the two different materials, either the entry $\kappa_{\text{AB}}(i,j;m^-,r)$ or $\kappa_{\text{AB}}(i,j;m^+,r)$ will be equal to the corresponding ratio ρ_r from the edge signature $\mathcal{E}_{\text{AB}}$. As a result, the above test will return the value 1 for the elements δ_r^m in the mth row of (B.5). As such, for a given pair of pixels, the number of non-zero elements in the associated mth row of the indicator matrix reveals the number of times the response of the mask $\mathcal{K}_{\text{AB}}(\mathbf{u})(i,j)$ has matched (within the specified tolerance ϵ) the spectral ratios from the edge signature $\mathcal{E}_{\text{AB}}$.

Because the pixel pairs used to form the rows of the mask correspond to different edge orientations (horizontal, vertical or diagonal), the number of ones in each row of (B.5) indicates the strength of the edge at position (i,j) for that particular direction. One way to account for such strength is by computing the matrix infinity norm of (B.5). Specifically, we define the mapping

$$\Phi : \big(\{0,1\}^{M\times R}\big)^{I\times J} \mapsto \{0,1\}^{I\times J} \ ,$$

which converts the indicator matrix (B.5) into an edge map by

$$\Phi\Big(\Delta\big(\mathcal{K}_{\text{AB}}(\mathbf{u})\big)\Big)(i,j) = \begin{cases} 1, \text{ if } \|\Delta\big(\mathcal{K}_{\text{AB}}(\mathbf{u})\big)(i,j)\|_\infty \geq \widetilde{R} \\ 0, \text{ otherwise} \ , \end{cases} \tag{B.7}$$

where for any matrix $\mathbf{A}$, $\|\mathbf{A}\|_\infty = \max_{1\le i\le M} \sum_{j=1}^{R} |a_{ij}|$, and $\widetilde{R} \le R$ is a specified threshold. With the definition in (B.7), the (i, j) location will belong to the collection E_{AB} of edges if the edge strength in at least one direction, as measured by the number of ones in the rows of (B.5), exceeds the threshold $\widetilde{R}$. If none of the edge strengths exceeds $\widetilde{R}$, then the (i, j) site does not belong to E_{AB}. The value of the threshold $\widetilde{R}$ can be used to adjust the sensitivity of the edge detector to noise. For example, increasing $\widetilde{R}$ makes the algorithm less sensitive to noise but more restrictive.

Finally, we define the edge map as the composition

$$\mathcal{F}_{\mathrm{AB}} = \Phi \circ \Delta \circ \mathcal{K}_{\mathrm{AB}} \ .$$

Note that $\mathcal{K}_{\mathrm{AB}}$ is the only problem-specific component in $\mathcal{F}_{\mathrm{AB}}$; the functions Δ and Φ are not problem specific. Because the edge signatures are determined independently for each pair of materials and the information from different color slices is properly fused, the SRC algorithm is particularly well-suited for scenes that contain both weak edges (e.g., isoluminant edges) and strong edges.

B.5 Extension of the algorithm to multiple materials

We limit the description of the extension of the algorithm to multiple materials to the case of three distinct materials A, B and C. The extension to the general case is straightforward. Due to the invariability of the detector for the order of the materials, for three distinct materials A, B and C there are three possible edges: E_{AB}, E_{AC} and E_{BC}. In this case we obtain three edge signatures, $\mathcal{E}_{\mathrm{AB}}$, $\mathcal{E}_{AC}$ and $\mathcal{E}_{BC}$, from which we define three joint spatio-spectral masks $\mathcal{K}_{\mathrm{AB}}$, $\mathcal{K}_{AC}$, and $\mathcal{K}_{BC}$. We then use these masks to identify the hyper-pixels belonging to the edge E_{AB} between materials A and B, the hyper-pixels from the edge E_{AC} between materials A and C, and the

hyper-pixels from the edge E_{BC} between materials B and C. The final edge map is obtained by the union of the three edges:

$$E_{ABC} = E_{\mathrm{AB}} \cup E_{AC} \cup E_{BC} \, .$$

References

[1] "American Cancer Society cancer facts & figures 2014," http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2014/, Last Accessed: July, 2014.

[2] H. W. Rogers, M. A. Weinstock, A. R. Harris, M. R. Hinckley, S. R. Feldman, A. B. Fleischer, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer in the United States, 2006," *Archives of Dermatology*, vol. 146, no. 3, pp. 283–287, 2010.

[3] N. R. Abbasi, H. M. Shaw, D. S. Rigel, R. J. Friedman, W. McCarthy, I. Osman, A. W. Kopf, and D. Polsky, "Early diagnosis of cutaneous melanoma: Revisiting the ABCDE criteria," *JAMA*, vol. 292, no. 22, pp. 2771–2776, 2004.

[4] L. Thomas, P. Tranchand, F. Berard, T. Secchi, C. Colin, and G. Moulin, "Semiological value of ABCDE criteria in the diagnosis of cutaneous pigmented tumors," *Dermatology*, vol. 197, no. 1, pp. 11–17, 1998.

[5] C. Benellii, E. Roscetti, and V. D. Pozzo, "The dermoscopic (7ffm) versus the clinical (abcde) diagnosis of small diameter melanoma," *European Journal of Dermatology*, vol. 10, no. 4, pp. 282–287, 2000.

[6] H. G. Welch, S. Woloshin, and L. M. Schwartz, "Skin biopsy rates and incidence of melanoma: population based ecological study," *BMJ*, vol. 331, no. 7515, p. 481, 2005.

[7] J. E. Mayer, S. M. Swetter, T. Fu, and A. C. Geller, "Screening, early detection, education, and trends forâmelanoma: Current status (2007-2013) andâfuture directions: Part ii. screening, education, and future directions," *Journal of the American Academy of Dermatology*, vol. 71, no. 4, pp. 611.e1–611.e10, 10 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0190962214015333

[8] V. Ahlgrimm-Siess, M. Laimer, E. Arzberger, and R. Hofmann-Wellenhof, "New diagnostics for melanoma detection: from artificial intelligence to rna microarrays," *Future Oncology*, vol. 8, no. 7, pp. 819–827, 2012.

[9] A. Kardynal and M. Olszewska, "Modern non-invasive diagnostic techniques in the detection of early cutaneous melanoma," *Journal of Dermatological Case Reports*, vol. 8, no. 1, pp. 1–8, 2014.

[10] M. Carrara, A. Bono, C. Bartoli, A. Colombo, M. Lualdi, D. M. N. Santoro, E. Tolomio, S. Tomatis, G. Tragni, M. Santinami, and R. Marchesini, "Multispectral imaging and artificial neural network: mimicking the management decision of the clinician facing pigmented skin lesions," *Physics in Medicine and Biology*, vol. 52, no. 9, pp. 2599–2613, 2007.

[11] D. S. Rigel, M. Roy, J. Yoo, C. J. Cockerell, J. K. Robinson, and R. White, "Impact of guidance from a computer-aided multispectral digital skin lesion analysis device on decision to biopsy lesions clinically suggestive of melanoma," *Archives of Dermatology*, vol. 148, no. 4, pp. 541–543, 2012.

[12] I. Quinzán, J. M. Sotoca, P. Latorre-Carmona, F. Pla, P. García-Sevilla, and E. Boldó, "Band selection in spectral imaging for non-invasive melanoma diagnosis," *Biomedical Optics Express*, vol. 4, no. 4, pp. 514–519, 2013.

[13] L. S. I. Tromme, F. Hammouch, C. Legrand, L. Marot, P. Vereecken, I. Theate, P. van Eeckhout, P. Richez, J. F. Baurain, L. Thomas, and N. Speybroeck, "Availability of digital dermoscopy in daily practice dramatically reduces the number of excised melanocytic lesions: results from an observational study," *British Journal of Dermatology*, vol. 167, no. 4, pp. 778–786, 2012.

[14] A. Sáez, C. Serrano, and B. Acha, "Model-based classification methods of global patterns in dermoscopic images," *IEEE Transactions on Medical Imaging*, vol. 33, no. 5, pp. 1137–1147, May 2014.

[15] G. Pellacani, B. D. Pace, C. Reggiani, A. M. Cesinaro, G. Argenziano, I. Zalaudek, H. P. Soyer, and C. Longo, "Distinct melanoma types based on reflectance confocal microscopy," *Experimental Dermatology*, vol. 23, no. 6, pp. 414–418, 2014.

[16] A. Dancey, B. Mahon, and S. Rayatt, "A review of diagnostic imaging in melanoma," *Journal of Plastic, Reconstructive & Aesthetic Surgery*, vol. 61, no. 11, pp. 1275–1283, 2008.

[17] R. Kleinerman, T. B. Whang, R. L. Bard, and E. S. Marmur, "Ultrasound in dermatology: Principles and applications," *Journal of the American Academy of Dermatology*, vol. 67, no. 3, pp. 478–487, 2012.

[18] R. E. Hunger, R. D. Torre, A. Serov, and T. Hunziker, "Assessment of melanocytic skin lesions with a high-definition laser doppler imaging system," *Skin Research and Technology*, vol. 18, no. 2, pp. 207–211, 2012.

[19] Z. Hamdoon, W. Jerjes, T. Upile, and C. Hopper, "Optical coherence tomography-guided photodynamic therapy for skin cancer: Case study," *Photodiagnosis and Photodynamic Therapy*, vol. 8, no. 1, pp. 49–52, 2011.

[20] L. Themstrup, C. Banzhaf, M. Mogensen, and G. Jemec, "Optical coherence tomography imaging of non-melanoma skin cancer undergoing photodynamic therapy reveals subclinical residual lesions," *Photodiagnosis and Photodynamic Therapy*, vol. 11, no. 1, pp. 7–12, 2014.

[21] K. Stephen and C. Fleming, "An introduction to dermoscopy," *Dermatological Nursing*, vol. 4, no. 4, 2013.

[22] G. Monheit, A. B. Cognetta, L. Ferris, H. Rabinovitz, K. Gross, M. Martini, J. M. Grichnik, M. Mihm, V. G. Prieto, P. Googe, R. King, A. Toledano, N. Kabelev, M. Wojton, and D. Gutkowicz-Krusin, "The performance of melafind: a prospective multicenter study," *Arch Dermatol.*, vol. 147, no. 2, pp. 188–194, 2010.

[23] R. Wells, D. Gutkowicz-Krusin, E. Veledar, A. Toledano, and S. C. Chen, "Comparison of diagnostic and management sensitivity to melanoma between dermatologists and melafind: A pilot study," *Arch Dermatol.*, vol. 148, no. 9, pp. 1083–1084, 2012.

[24] A. R. Cukras, "On the comparison of diagnosis and management of melanoma between dermatologists and melafind," *JAMA Dermatology*, vol. 149, no. 5, pp. 622–623, 2013.

[25] A. Wood, H. Morris, J. Emery, P. N. Hall, S. Cotton, A. T. Prevost, and F. M. Walter, "Evaluation of the molemate training program for assessment of suspicious pigmented lesions in primary care," *Inform Prim Care.*, vol. 16, no. 1, pp. 41–50, 2008.

[26] F. M. Walter, H. C. Morris, E. Humphrys, P. N. Hall, A. Prevost, N. Burrows, L. Bradshaw, E. C. F. Wilson, P. Norris, J. Walls, M. Johnson, A. L. Kinmonth,

and J. D. Emery, "Effect of adding a diagnostic aid to best practice to manage suspicious pigmented lesions in primary care: randomized controlled trial," *BMJ*, vol. 345, 2012.

[27] H. Lui, J. Zhao, D. McLean, and H. Zeng, "Real-time raman spectroscopy for in vivo skin cancer diagnosis," *Cancer Research*, vol. 72, no. 10, pp. 2491–2500, 2012.

[28] D. M. S. N. Nandanan, J. Fisher and O. Markowitz, "Optical coherence tomography imaging of onychomycosis," 2014, presented at EADV, Florence. [Online]. Available: http://www.vivosight.com/clinical-presentations/

[29] B. B. Lahiri, S. Bagavathiappan, T. Jayakuman, and J. Phillip, "Medical applications of infrared thermography: A review," *Infrared Physics & Technology*, vol. 55, no. 4, pp. 221–235, 2012.

[30] C. Herman, "The role of dynamic infrared imaging in melanoma diagnosis," *Expert Rev Dermatol.*, vol. 8, no. 2, pp. 177–184, 2013.

[31] G. A. S. Cruz, J. Bertotti, J. Marin, S. J. González, S. Gossio, D. Alvarez, B. M. Roth, P. Menéndez, M. D. Pereira, M. Alvero, L. Cubau, P. Orellano, and S. J. Liberman, "Dynamic infrared imaging of cutaneous melanoma and normal skin in patients treated with bnct," *Applied Radiat. Isot.*, vol. 67, no. 7-8, pp. S54–S58, July 2009.

[32] M. Cristofolini, B. Perani, F. Piscioli, G. Recchia, and G. Zumiani, "Uselessness of thermography for diagnosis and follow-up of cutaneous malignant melanoma," *Tumori*, vol. 67, no. 2, pp. 141–143, 1981.

[33] M. Ambar, *Quantitative Dynamic Telethermometry in Medical Diagnosis and Management*. CRC Press, Boca Raton, FL, 1994.

[34] T. Buzug, S. Schumann, L. Pfamann, U. Reinhold, and J. Ruhlmann, "Functional infrared imaging for skin-cancer screening," in *Proceedings of the 28th IEEE Engineering in Medicine and Biology Society Conference*, vol. 1, 2006, pp. 2766–2769.

[35] M. P. Cetingul and C. Herman, "Quantification of the thermal signature of a melanoma lesion," *International Journal of Thermal Sciences*, vol. 50, no. 4, pp. 421–431, 2011.

[36] ——, "A heat transfer model of skin tissue for the detection of lesions: sensitivity analysis," *Physics in Medicine and Biology*, vol. 55, no. 1, pp. 5933–5951, 2010.

[37] ——, "The assessment of melanoma risk using the dynamic infrared imaging technique," *Journal of Thermal Science and Engineering Applications*, vol. 3, no. 3, p. 031006, 2011.

[38] U. Grenander, "Stochastic processes and statistical inference," *Arkiv för Matematik*, vol. 1, no. 3, pp. 195–277, 1950.

[39] J. Canny, "A computational approach to edge detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, Nov. 1986.

[40] S. Di Zenzo, "A note on the gradient of a multi-image," *Comput. Vision Graph. Image Process.*, vol. 33, no. 1, pp. 116–125, 1986.

[41] R. G. Resmini, "Hyperspectral/spatial detection of edges (hyspade): An algorithm for spatial and spectral analysis of hyperspectral information," in *SPIE Proceedings. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery X*, vol. 5425, 2004, pp. 433–442.

[42] B. Pasakaleva, W.-Y. Jang, M. Hayat, Y. Sharma, S. Bender, and S. Krishna, "Joint spatio-spectral based edge detection for multispectral infrared imagery," *2010 IEEE International Geoscience and Remote Sensibilitiesng Symposium (IGARSS)*, pp. 2198 – 2201, 2010.

[43] B. S. Paskaleva, S. E. Godoy, W.-Y. Jang, S. C. Bender, S. Krishna, and M. M. Hayat, "Model-based edge detector for spectral imagery using sparse spatiospectral masks," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2315–2327, May 2014.

[44] J. Andrews, W.-Y. Jang, J. E. Pezoa, Y. D. Sharma, S. J. Lee, S. K. Noh, M. M. Hayat, S. Restaino, S. W. Teare, and S. Krishna, "Demonstration of a bias tunable quantum dots-in-a-well focal plane array," *Infrared Phys. Technol.*, vol. in press, 2009.

[45] B. Pasakaleva, W.-Y. Jang, M. Hayat, Y. Sharma, S. Bender, and S. Krishna, "Multispectral classification with bias-tunable quantum dots-in-a-well focal plane arrays," *IEEE Sensors Journal*, vol. 11, pp. 1342 – 1351, 2011.

[46] S. E. Godoy, D. A. Ramirez, S. A. Myers, G. von Winckel, S. S. Krishna, R. S. Padilla, P. Sen, and S. Krishna, "Dynamic infrared imaging for skin cancer screening," *Infrared Physics and Technology*, vol. in press, 2014.

[47] H. H. Pennes, "Analysis of tissue and arterial blood temperature in the resting human forearm," *Journal of Applied Physiology*, vol. 1, no. 2, pp. 93–112, August 1948.

[48] A. D. Carlo, "Thermography and the possibilities for its applications in clinical and experimental dermatology," *Bioengineering of the Skin*, vol. 13, no. 4, pp. 329–338, 1995.

[49] M. A. Fauci, R. Breiter, W. Cabanski, W. Fick, R. Koch, J. Ziegler, and S. D. Gunapala, "Medical infrared imaging – differentiating facts from fiction, and the impact of high precision quantum well infrared photodetector camera systems, and other factors, in its reemergence," *Infrared Physics & Technology*, vol. 42, no. 3-5, pp. 337–344, 2001.

[50] C. Harris and M. Stephens, "A combined corner and edge detector," in *In Proceedings of the Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[51] N. Chumchob and K. Chen, "A robust affine image registration method," *International Journal of Numerical Analysis and Modeling*, vol. 6, no. 2, pp. 311–334, 2009.

[52] A. A. Goshtasby, *2-D and 3-D Image Registration: for Medical, Remote Sensing, and Industrial Applications*. Wiley Press, 2005.

[53] G. C. Holst, *CCD Arrays, Cameras and Displays*. SPIE Optical Eng. Press, 1996.

[54] S. E. Godoy, J. E. Pezoa, and S. N. Torres, "Noise-cancellation-based nonuniformity correction algorithm for infrared focal-plane arrays," *Applied Optics*, vol. 47, no. 29, pp. 5394–5399, October 2008.

[55] N. Nishida, H. Yano, T. Nishida, T. Kamura, and M. Kojiro, "Angiogenesis in cancer," *Vascular Health and Risk Management*, vol. 2, no. 3, pp. 213–219, 2006.

[56] "The Angiogenesis Foundation webpage," http://www.angio.org/, Last Accessed: December, 2014.

[57] "Tumor angiogenesis infographic colorectal cancer webpage," https://www.behance.net/gallery/11816509/Tumor-Angiogenesis-Infographic-Colorectal-Cancer, Last Accessed: December, 2014.

[58] D. J. Watmough and R. Oliver, "Emissivity of human skin in the waveband between 2micra and 6micra," *Nature*, vol. 219, no. 5154, pp. 622–624, 1968.

[59] ——, "Wavelength dependence of skin emissivity," *Phys Med Biol*, vol. 14, no. 2, pp. 201–204, 1969.

[60] ——, "The emission of infrared radiation from human skin–implications for clinical thermography," *Br J Radiol.*, vol. 42, no. 498, pp. 411–415, 1696.

[61] S. B. Wilson and V. A. Spence, "A tissue heat transfer model for relating dynamic skin temperatures changes to physiological parameters," *Phys. Med. Biol.*, vol. 33, no. 8, pp. 895–912, 1988.

[62] T. Kailath, "Rkhs approach to detection and estimation problems - part 1: Deterministic signals in gaussian noise," *IEEE Trans. on Information Theory*, vol. 17, no. 5, pp. 530–549, September 1971.

[63] T. Kailath and H. L. Weinert, "An rkhs approach to detection and estimation problems - part ii: Gaussian signal detection," *IEEE Trans. on Information Theory*, vol. 21, no. 1, pp. 15–23, January 1975.

[64] T. Kailath and H. V. Poor, "Detection of stochastic processes," *IEEE Trans. on Information Theory*, vol. 44, no. 6, pp. 2230–2259, October 1998.

[65] H. L. V. Trees, *Radar–Sonar Signal Processing and Gaussian Signals in Noise*, 2nd ed. John Wiley & Son, Inc., 2001.

[66] A. Ribes and F. Schmitt, "Linear inverse problems in imaging," *IEEE Signal Processing Magazine*, vol. 25, no. 4, pp. 84–99, July 2008.

[67] D. G. Chen, "A bayesian model with a bivariate normal–lognormal prior distribution and a nonlinear mixed-effect model for a regional fish stock-recruitment meta-model," in *Proceedings of the American Statistical Association*, 2002.

[68] M. Yang, "Normal log-normal mixture: Leptokurtosis, skewness and applications," *Applied Economics Letters*, vol. 15, no. 8, pp. 737–742, 2008.

[69] G. Zerovnik, A. Trkov, and I. A. Kodeli, "Correlated random sampling for multivariate normal and log-normal distributions," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 690, no. 0, pp. 75–78, 2012.

[70] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. Springer, 1994, ch. 6, pp. 263–326.

[71] M. Chen, Z. Chen, and G. Chen, *Approximate Solutions of Operator Equations*, C. K. Chui, Ed. World Scientific Publishing Co., 1997, vol. 9.

[72] H. L. V. Trees, *Detection, Estimation and Modulation Theory, Part 1*, 2nd ed. John Wiley & Son, Inc., 2001.

[73] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, second edition ed. Cambridge University Press, 1992, ch. 18, pp. 788–794.

[74] L. M. Delves and J. L. Mohamed, *Computational Methods for Integral Equations*. Cambridge University Press, 1985.

[75] M. M. Hayat, "ECE642 detection and estimation theory," 2014, Lecture Notes.

[76] J. Pachares, "Note on the distribution of a definite quadratic form," *The Annals of Mathematical Statistics*, vol. 26, no. 1, pp. 128–131, 1955.

[77] A. Oya, J. Navarro-Moreno, and J. Ruiz-Molina, "A numerical solution for multichannel detection," *Communications, IEEE Transactions on*, no. 6, pp. 1734 – 1742, 2009.

[78] E. J. Kelly and W. L. Root, "A representation of vector-valued random processes," *J. Math. and Phys.*, vol. 39, pp. 211–216, 1960.

[79] M. Loog and B. Ginneken, "Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification," *IEEE Trans. on Medical Imaging*, vol. 25, no. 5, pp. 602–611, 2006.

[80] T. Rohlfing and C. R. M. Jr., "Multi-classifier framework for atlas-based image segmentation," in *Pattern Recognition Letters*, vol. 26, 2005, pp. 2070–2079.

[81] P. G. Lucey, T. J. Williams, J. L. Hinrichs, M. E. Winter, D. Steutel, and E. M. Winter, "Three years of operation of AHI: the University of Hawaii's Airborne Hyperspectral Imager," *SPIE Infrared Technology and Applications*, vol. XXVII, pp. 112–120, 2001.

[82] D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, "Band-edge electroabsorption in quantum well structures: The quantum-confined stark effect," *Phys. Rev. Lett.*, vol. 53, pp. 2173–2176, Nov 1984.

[83] J. Shawe-Taylor and N. Cristianini, *Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[84] [Online]. Available: http://www.skincancer.org/skin-cancer-information/melanoma/melanoma-warning-signs-and-images/do-you-know-your-abcdes

[85] Y. Sohn and N. S. Rebello, "Supervised and unsupervised spectral angle classifiers," *Photogrammetric Engineering and Remote Sensing*, vol. 68, pp. 1271–1280, 2002.

[86] S. M. Kay, "Fundamentals of statistical signal processing: Detection theory, vol. 2," 1998.

[87] H. Robbins, "The distribution of a definite quadratic form," *The Annals of Mathematical Statistics*, vol. 19, no. 2, pp. 266–270, 1948.

[88] Y. Malevergne and D. Sornette, "High-order moments and cumulants of multivariate weibull asset returns distributions: Analytical theory and empirical tests: Ii," *Finance Letters*, vol. 3, no. 1, pp. 54–63, 2005.