5-2018

Evaluation of Terahertz Imaging for Breast Cancer Detection using Image Morphing

Tanny Andrea Chavez Esparza
University of Arkansas, Fayetteville

Evaluation of Terahertz Imaging for Breast Cancer Detection using Image Morphing

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Electrical Engineering

by

Tanny Chavez Esparza
Escuela Superior Politécnica del Litoral
Bachelor of Science in Electronics and Telecommunications Engineering, 2015

May 2018
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Jingxian Wu, Ph.D.
Thesis Director

Magda El-Shenawee, Ph.D.
Committee Member

Narasimhan Rajaram, Ph.D.
Committee Member

Abstract

This thesis proposes the use of a mesh morphing algorithm for the quantitative evaluation of terahertz (THz) images. This work differs from traditional evaluation methods based on qualitative evaluation because it provides a fair and quantitative measurement of the THz imaging system's performance. The objective of the algorithm is to match the alignment, shape, and resolution of the THz and reference pathology images. Therefore, the proposed morphing method provides a pathology reference for a pixel-by-pixel evaluation of the region classification in the THz image. To achieve this, the morphing algorithm aligns the images using the Pearson's correlation coefficient and reshapes the pathology results using mesh morphing and homography estimation. The results presented in this thesis demonstrate the potential of the algorithm as an evaluation method of THz imaging for breast cancer detection in both fresh and formalin-fixed, paraffin-embedded (FFPE) murine tumors.

Acknowledgments

I would like to thank all the members of the research group for their efforts in obtaining and processing the necessary data for this study. In particular, I would like to express my sincere gratitude to Dr. Wu and Dr. El-Shenawee for guiding me in this process. I would like to thank Tyler Bowman for his advise in the evaluation of the samples and Dr. Rajaram and his students for handling the mice and the tumor samples. In addition, I would like to thank Dr. Chakraborty and Kamrul Khan for developing the statistical model used in this work. Finally, but not least, I would like to thank my family, friends, and labmates for their support along this journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Breast cancer is one of the most commonly diagnosed types of cancer among women in the United States. When it is detected early, it is possible to perform a lumpectomy in the patient, in which a surgeon excises all the cancer tissue while leaving as much healthy breast tissue as possible [2]. After the surgery, a pathologist evaluates the excised tissue to confirm the existence of clear margins and to declare if all the cancer tissue has been excised. Unfortunately, in $20\% - 40\%$ cases, the patient has to go through a second surgery due to the presence of positive margins in the first excised sample, which shows that there are cancer cells remaining in the breast. In this context, the accurate detection of the cancer tissue within the breast is fundamental.

THz imaging has been studied as a suitable option for cancer margins detection for breast conserving surgery in [1, 3, 4, 5, 6, 7]. With THz imaging, it is possible to assess the cancer margins in the operating room right after excising the tumor. If the margins of the excised tissue are positive, the surgeon can excise the rest of the tumor immediately instead of performing a second surgery. Although considerable research has been performed in this area, the quantitatively evaluation of the results remains as a challenge. Given that the main reference in this framework is acquired through the histopathologic analysis of the sample obtained after fixing the excised tissue in paraffin, the evaluation of the results for fresh tissue imaging becomes complicated. Therefore, this thesis proposes the creation of a pathology reference for quantitative evaluation of THz images using mesh morphing and homography estimation.

1.1. Literature Review

To validate the accuracy of breast cancer detection using THz imaging, doctors recommend cancer detection rates of at least 90% with false detection rates no greater than 10% [8]. Therefore, the results should be quantitatively evaluated with respect to the reference image obtained from the histopathology processing of the tumors. The challenges of this evaluation are the resolution

mismatch of the THz and pathology images, their misalignment, and in some cases, their different shapes, which are the results of the shrinking and stretching of the tumors during the histopathology process. The literature presents different evaluation approaches that tackle these issues based on both qualitative and quantitative observations.

The qualitative evaluation has been commonly used to validate the THz imaging as presented in [9, 3, 10]. In this process, the THz and pathology images are manually aligned and rescaled to visually evaluate the tissue classification obtained from the THz image. In cases like [9, 3] the classification of the tissue is not binary (healthy vs. not healthy tissue), hence the visual evaluation is performed as a visual comparison of the images side-by-side or by superimposing the contour of the regions of the pathology images on the THz image. Similarly, [10] performs a visual evaluation of their results obtained from a binary classification of the tumor. In their work, in addition to a side-by-side comparison, the authors analyzed the magnification of certain regions within the pathology image to validate the detection of heterogeneous regions. Although the qualitative evaluation presents a strong correlation between the THz and pathology images, it does not represent a specific measurement of the performance of the imaging system.

To obtain an estimation of the performance of the THz images, a quantitative evaluation is suitable, but challenging due to the differences between the pathology and THz images. [11] presents an evaluation method based on the computation of the Spearman rank correlation coefficient between the number of pixels that represented cancer of the delineated THz and pathology images. In this paper, the THz and pathology images were aligned using inked margins and photographs as a reference. Once aligned, the authors defined the areas of interest in both images and delineated the cancer regions to count the number of pixels that corresponded to cancer in the x and y directions. Using these values, the authors calculated the Spearman rank correlation coefficients between the THz and pathology images as a quantitative measurement of the performance of their results. Similarly, [12] presents a quantitative evaluation method through ROC curves representation for material classification. In this work, the authors manually selected sections of scattered data from each type of material and calculated the probability of detection and the probability of

false alarm sweeping a classification threshold. The implementation of ROC curves in THz imaging evaluation is a more convenient method to achieve the rates for clinical usage, but it becomes more complicated for breast cancer detection given the presence of heterogeneous regions within the tumor.

The application of ROC curves for the quantitative assessment of breast cancer detection was covered in [8, 4, 5, 1]. [8] generates the ROC curves from a pixel-by-pixel comparison between the THz image and a histology map, where this map was manually created from the overlapping of the pathology results on the THz image with software assistance. Although this method overcomes the differences in alignment, size, and shape between the THz and pathology images, its main disadvantage is the manual creation of the map. [4, 5, 1] considered this drawback and proposed the usage of image morphing to automatically reshape the contour of the pathology image into the contour of the THz image, hence making a direct pixel-by-pixel comparison possible. In their work the pathology image was interpolated to stretch its shape horizontally, vertically, or in some cases in both directions. This algorithm performed correctly for most cases, but showed some misalignments in a few cases due to the stretching in a single direction. Given the benefits of the automatic creation of the morphed pathology, this thesis intends to improve the morphing algorithm using a different method more suitable for the pathology reshaping.

Considering that the tumor shrinks and stretches not only vertically or horizontally, but in several directions during the histopathology process, it is necessary to implement a morphing method capable of reshaping this image in any direction. [13] describes different methods for image morphing including mesh warping. Given a few control points, this method creates a source and a target mesh to warp the elements in the source mesh into the shape of the elements of the target mesh. Since mesh warping morphs the image in several directions using a few control points, this thesis explores the benefits of its implementation in the evaluation of the performance of breast cancer THz imaging.

1.2. Problem Formulation

Consider the RGB color model of the pathology image as an array of pixels, where each pixel is represented as $h_r(x_p, y_p)$ with $x_p \in \{1, \cdots, N_x\}$ and $y_p \in \{1, \cdots, N_y\}$ corresponding to the position of the pixel in the horizontal and vertical axis, respectively; and $r \in \{1, 2, 3\}$ representing its red, green, and blue colors intensities, respectively.

During the THz imaging process, the system collects a reflected THz pulse per pixel that is used for the region classification within the tissue. Given that the amount of points per pixel is large, it is necessary to summarize this signal using either the maximum value or the integration of the signal per pixel for block and fresh tissue, respectively. After summarizing the intensities, the THz data becomes an array where the summarized intensity at each pixel is represented as $t(x_t, y_t)$ with $x_t \in \{1, \cdots, M_x\}$ and $y_t \in \{1, \cdots, M_y\}$ corresponding to the position of the pixel in the horizontal and vertical axis, respectively.

As stated before, a direct pixel-by-pixel comparison between the pathology and THz images is not possible due to three main reasons described as follows:

1. Misalignment: Considering that the tumor is moved from one place to another to obtain the pathology and THz images, the images are not aligned with respect to each other. Although the usage of inked margins and photographs to record the position of the tumor can diminish this problem, there is a residual misalignment that should be corrected before performing the comparison between the images.

2. Shape shift: When imaging fresh tissue, there is a considerable shape shifting due to the shrinking and stretching of the tissue during the fixation of the tumor in paraffin in the histopathology process.

3. Resolution mismatch: Since the pathology image is obtained from merging the photos of the histopathology results viewed from a microscope, its resolution $[M_x, M_y]$ is considerable greater than the resolution of the THz image $[N_x, N_y]$.

Therefore, this thesis presents an algorithm that corrects these problems using mesh morphing. Chapter 2 presents the description of the THz imaging system used for this study. Chapter 3 covers the image preparation before morphing, including the correction of the misalignment using a maximum correlation algorithm. Chapter 4 describes the morphing algorithm that corrects the shape shift and the resolution mismatch between the pathology and THz images. Chapter 5 illustrates the experimental results of the algorithm and chapter 6 presents the conclusions of this study.

Chapter 2

THz Imaging System

This chapter describes the TPS Spectra 3000 THz pulsed imaging and spectroscopy system (TeraView, Ltd., UK) located at the University of Arkansas, which represents the THz system used to collect the images presented in this work. For this study, 10 C57BL/6J mice were kept on a high-fat diet until reaching the target weight of 35g. Once the target weight was achieved, the mice were injected with E0771 murine breast adenocarcinoma cells in their flanks to grow tumor xenografts. The tumors were excised under anesthesia once they reached a diameter of 1cm and were imaged (only 9 of them since the last one did not develop any tumor after multiple injections) [1].

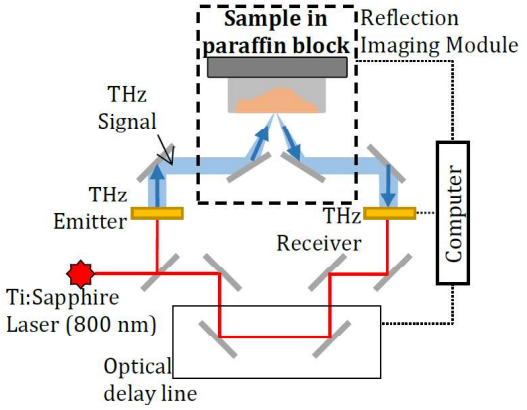
The THz imaging system in reflection mode is shown in Fig. 2. The system setup changes depending on the type of tissue to be imaged. For FFPE tissue, the sample is placed directly on the system, while for fresh tissue, the sample is placed in between two polystyrene plates as shown in Fig. 2.1a and Fig. 2.1b, respectively. The system scans the tissue entirely using $200\mu m$ steps. For each step, the THz signal shown in Fig. 2.1c is directed to the sample and its reflected signal is collected for further processing [1].

The THz image is generated using the reflected THz signal per pixel according to the type of tissue. For fresh tissue, the integration of the frequency domain THz reflected signal is considered using the following equation:

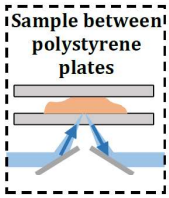
$$S(x_t, y_t) = \int_{f_1}^{f_2} \frac{|E(x_t, y_t, f)|^2}{|E_{ref}(f)|^2} df \tag{2.1}$$

where (x_t, y_t) represents the location of the pixel as explained in section 1.2, $|E(x_t, y_t, f)|$ is the magnitude of the Fourier transform of the reflected signal for the pixel, $|E_{ref}(f)|$ represents the reference THz signal obtained from the reflection of the polystyrene plate, and $[f_1, f_2]$ represents the frequency range for the integration of the signal.

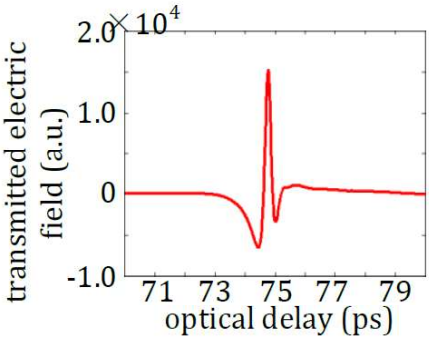
Considering that the FFPE tissue presents multiple reflections using (2.1), the maximum peak of the normalized reflected THz signal in time domain is used to plot the scanned tissue. In this



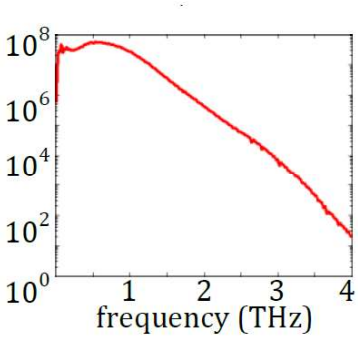
(a) Diagram of the system for FFPE tissue [1].



(b) Sample location for fresh tissue [1].



(c) Transmitted time domain pulse [1].



(d) Transmitted frequency domain pulse [1].

Figure 2.1: THz system diagram [1].

case, the THz reflected signal is normalized with respect to the peak of the reflected signal of a golden mirror [1].

For this study, a total of 11 xenograft breast cancer samples have been analyzed to date. These samples were obtained from 9 murine tumors and the sections collected from the same tumor were named as A or B. [1] and [5] evaluated the results of 8 of these samples using an interpolation-based morphing algorithm, while [6] evaluated the results of 3 samples using the mesh morphing algorithm described in this thesis. This manuscript presents 3 of the samples studied in [1]: tumor 7 section B, tumor 8 section A, and tumor 9 section A.

All the procedures took place in compliance with the Guide for the Care and Use of Laboratory Animals and were approved by the University of Arkansas Institutional Animal Care and Use

Committee (IACUC).

Chapter 3

Image Preparation

This chapter describes the procedures to prepare the pathology image prior the application of the morphing algorithm. The reference pathology image is obtained from the traditional method for cancer detection in which a pathologist analyzes the FFPE tissue through a microscope and formulates the diagnosis as a pathology report. Since this image does not show the classification of the regions, it is necessary to create a new image that includes this information based on the pathologist's evaluation. Hence, instead of using the pathology image directly, this thesis proposes the creation of a classification mask based on the pathology results for the evaluation of the THz imaging system, similar to [8].

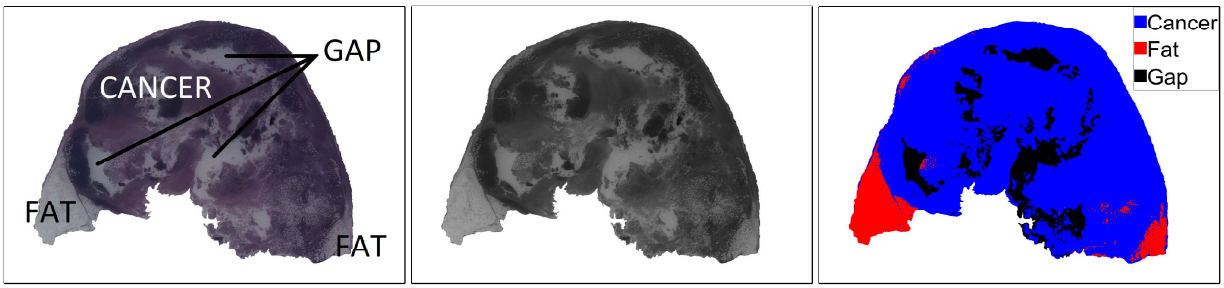
3.1. Mask Creation

To classify the regions within the pathology image, it is necessary to analyze the visual differences between the regions. Considering that the main difference between regions in the pathology images was their color intensities, this thesis presents a classification algorithm based on color discrimination.

Before the mask creation, it is necessary to define a region of interest for the sample in the pathology and THz images, leaving out any region that should not be evaluated in the THz image (e.g. background or skin). It is worth mentioning that the skin is removed from our analysis because they cannot be found in human tumor samples, but just in the murine tumor samples as the ones presented in this work.

3.1.1. Samples with Two Regions

Before the classification process, the background is manually removed from the pathology image, as shown in Fig. 3.1a, which represents the histopathology results obtained from sample 7B. This tumor presents two regions, a clear region that represents fat and a dark purple region that



(a) Pathology image. (b) Greyscale representation of the pathology image. (c) Pathology mask.

Figure 3.1: Creation of the pathology mask for sample 7B with two regions.

represents cancer. In addition, this tumor presents some gaps inside the cancer region that do not constitute a different region, but just an empty space in the sample.

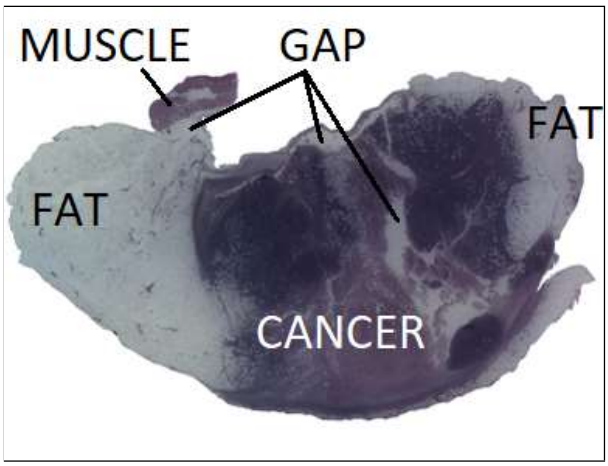
Considering that the difference of these two regions lays on the intensity of their color, the algorithm converts the image into its greyscale representation as shown in Fig. 3.1b. Therefore, a simple threshold is used to classify the intensities for each pixel with respect to the different regions in the sample. The selection of the threshold is subjective to the brightness and contrast of the image and changes from sample to sample.

It is important to highlight that, in this case, the gap was misclassified as fat and some dark regions within the fat (small ducts and vessels) were misclassified as cancer because of their appearance. Hence, some manual corrections were necessary and could only be performed after confirming their true classification in the pathologist's evaluation. The final pathology mask in Fig. 3.1c is generated based on this classification, where each region is assigned an RGB color.

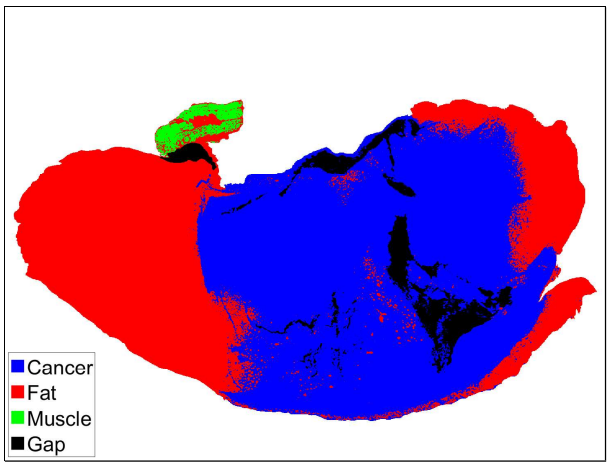
3.1.2. Samples with Three Regions

As done before, the background is manually removed from the pathology image before the tissue classification, as shown in Fig. 3.2a, which represents the histopathology results obtained from sample 9A. This tumor contains three regions: cancer, fat, and muscle.

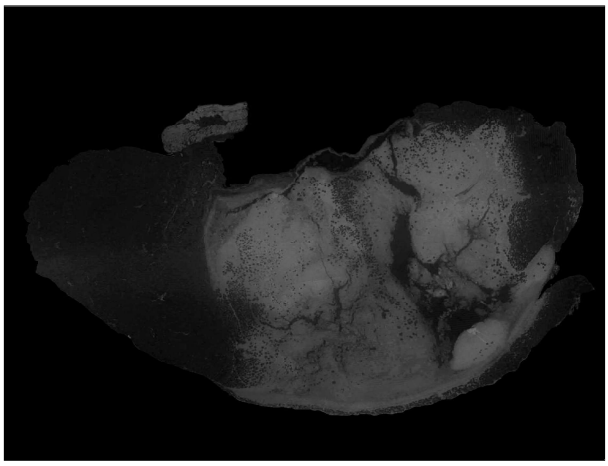
Unlike tumors with two regions, the presence of muscle impedes the use of the grayscale representation for their region classification because the grayscale intensities of cancer and muscle



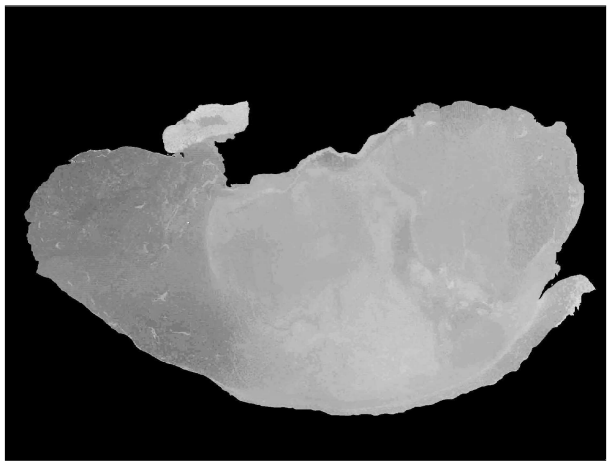
(a) Pathology image.



(b) Pathology mask.



(c) Saturation levels of the pathology image.

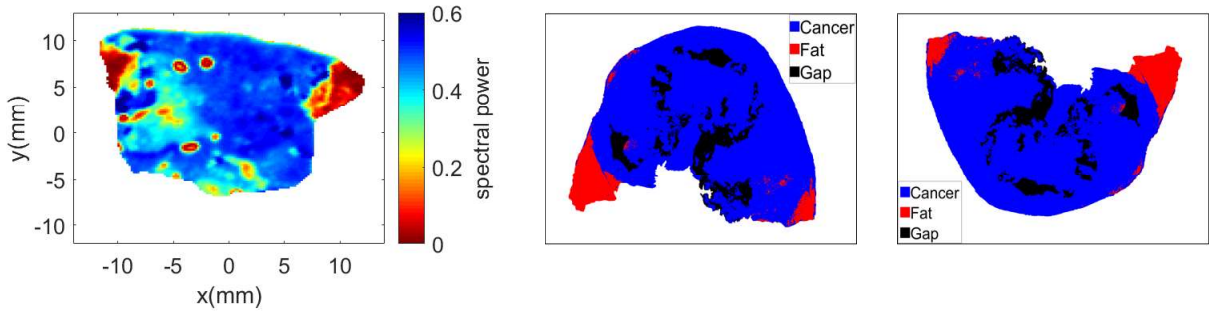


(d) Hue levels of the pathology image.

Figure 3.2: Creation of the pathology mask for sample 9A with three regions.

are similar. Therefore, for tumors with three regions, the pathology image is converted into its HSV color model representation, where the saturation levels are used to identify the fat in the tissue and its hue levels are used to classify muscle and cancer.

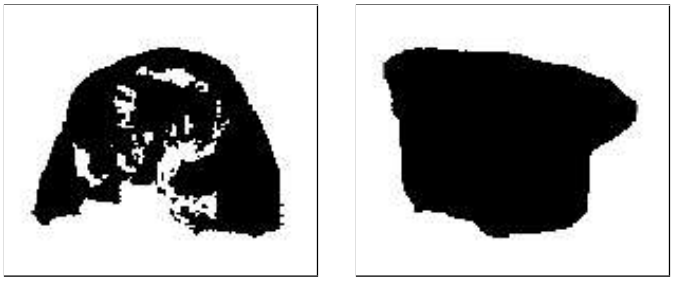
As shown in Fig. 3.2c, the saturation levels of the fat regions are close to zero and are represented as darker regions compared to the rest of the tissue. In addition, the hue levels for muscle are higher compared to cancer, as shown in Fig. 3.2d where the muscle region is brighter compared to the rest of the tissue. Hence, each pixel is classified using a threshold for the saturation levels to identify fat regions and the remaining unidentified pixels (classified as non-fat tissue) are classified using thresholds for the hue levels to identify muscle and cancer regions within the tissue.



(a) THz image.

(b) Pathology mask before alignment.

(c) Pathology mask after alignment.



(d) Pathology monochromatic mask.

(e) THz monochromatic mask.

Figure 3.3: Image misalignment in sample 7B fresh.

Similarly to tumors with two regions, the selected thresholds vary depending on the brightness and contrast of the pathology image and some manual corrections are necessary for the gap and some small regions within the tissue. As shown in Fig. 3.2c, the gaps were manually corrected because they were misclassified as fat. The final pathology mask in Fig. 3.2b is generated based on this classification, where each region is assigned an RGB color.

3.2. Image Alignment

As described in chapter 1, there is an image misalignment between the pathology mask and the THz image due to the different placement of the tissue while performing each process. Fig. 3.3a and 3.3b represent the pathology mask and the THz image obtained from fresh tissue of sample 7B fresh, in which it is shown that the position and orientation of these images are different.

To fix this misalignment, the algorithm uses the Pearson's correlation coefficient [14] to find the angle of maximum correlation between the images. Since this method is used for image processing

Algorithm 1 Image alignment.

Load pathology mask $\mathcal{P}$ and THz image $\mathcal{T}$.
for $\theta_0 = 0 : \Delta\theta_0 : 359$ **do**
  Rotate pathology mask θ_0°.
  Crop borders of THz and pathology mask.
  Monochromatic conversion of cropped images.
  Reduce resolution of the monochromatic pathology mask to match the resolution of the THz monochromatic mask, producing $\mathcal{P}'$ and $\mathcal{T}'$.
  Compute the Pearson's correlation coefficient using (3.1).
end for
Find the angle θ_0 for maximum correlation coefficient.

in monochrome digital images, the algorithm uses monochromatic masks for both images as shown in Fig. 3.3d and 3.3e.

Consider the pathology mask and the THz image as the matrices $\mathcal{P}$ and $\mathcal{T}$, respectively. To create the monochromatic masks, the algorithm crops the border of each image, hides all the internal features in both images using a simple decision rule (colored vs non-colored pixels) and reduces the size of the pathology monochromatic mask to match the THz monochromatic mask's resolution using bicubic interpolation, creating the masks $\mathcal{P}'$ and $\mathcal{T}'$. The importance of generating these monochromatic masks relies on avoiding the difference in the inner features to interfere with the calculation of the correlation coefficient.

As shown in Algorithm 1, the complete alignment process rotates the pathology image, computes the monochromatic masks, and calculates the Pearson's correlation coefficient between them using (3.1). The algorithm looks for the angle that produces the maximum correlation coefficient and establishes the correct alignment of the pathology mask.

$$\rho = \frac{\sum_d \sum_e (\mathcal{P}'_{de} - \overline{\mathcal{P}'})(\mathcal{T}'_{de} - \overline{\mathcal{T}'})}{\sqrt{\left(\sum_d \sum_e (\mathcal{P}'_{de} - \overline{\mathcal{P}'})^2\right)\left(\sum_d \sum_e (\mathcal{T}'_{de} - \overline{\mathcal{T}'})^2\right)}} \tag{3.1}$$

Where ρ represents the Pearson's correlation coefficient, $\mathcal{P}'_{de}$ and $\mathcal{T}'_{de}$ are the values of the pixels in the de position in the pathology and THz monochromatic masks, respectively, and $\overline{\mathcal{P}'}$ and $\overline{\mathcal{T}'}$ are the means of the pixel values in the pathology and THz monochromatic masks, respectively.

Even thought the algorithm is based on an extensive search of the angle, we can reduce its complexity by performing a coarse search first where $\Delta\theta_0 = 1$. After finding the angle θ_0 for this search, the algorithm can proceed to run a fine search changing the value of the iteration to $\theta = \{\theta_0 - 1 : \Delta\theta : \theta_0 + 1\}$, where $\Delta\theta = 0.1$. It is important to consider that any residual misalignment will be corrected during the morphing process. Once the angle of alignment θ is found, the pathology mask rotated θ° is used for the morphing algorithm without reducing its resolution. It is important to keep the large resolution of the pathology mask before performing the homography estimation of the image, as discussed in section 4.2.

Chapter 4

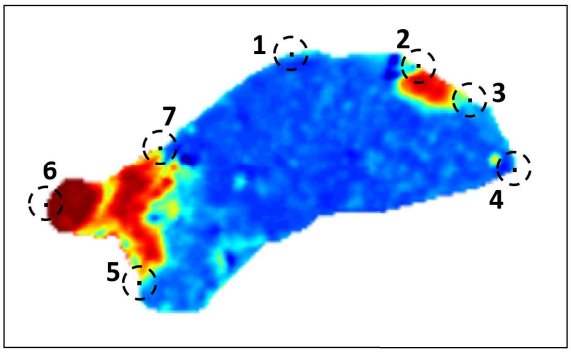
Image Morphing

This chapter describes the procedures to morph the pathology image into the contour of the THz image, which corrects the shape shift and the resolution mismatch between them. After performing the algorithm, a quantitative evaluation through a pixel-by-pixel comparison is possible as shown in chapter 5.

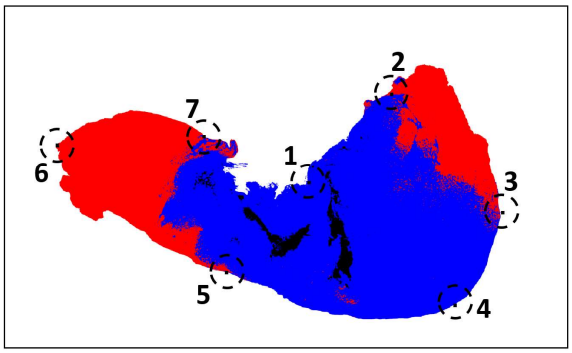
Image morphing is an image processing technique that consists on the reallocation of control points to produce geometric changes within an image. Although this technique is commonly used for the creation of frames for a smooth transition within an animation sequence, it can also be used for the direct metamorphosis of a source image into the shape of a target image. In this thesis, we propose the usage of image mesh warping to fix the shape shifting between the pathology and THz images.

Mesh warping is a morphing technique based on the construction of meshes in the source and target images. This method warps each element in the source mesh into the shape of its corresponding element in the target mesh. To create the meshes, it is necessary to define control points between the two images that could be manually or automatically defined and that are represented as $\mathbf{V} = [\mathbf{v}^{(1)}, ..., \mathbf{v}^{(n)}]$, where $\mathbf{v}^{(i)} \in \mathcal{R}^2$.

The literature presents different approaches for the automatic detection of control points, like the one presented in [15], in which the authors applied neural network for face features recognition. In their work, the authors trained an algorithm using a large amount of data to detect common features in human faces, such as eyes, nose, and, mouth locations, to later use this control points for morphing purposes. Since it is necessary to find common attributes between images for automatic feature identification, it not possible to apply these algorithms for our study. Given that the shape mismatch is a consequence of fixing the fresh tissue in paraffin, the stretching and shrinking varies from sample to sample and most areas change randomly, e.g. some convex sections in the pathology image were concave when they were fresh in the THz image. Therefore, this thesis



(a) Control points in THz image.



(b) Control points in rotated pathology mask.

Figure 4.1: Equivalent selection of control points in sample 8A fresh.

presents an user-defined control points algorithm due to the considerable difference of the features in both images. In addition, to avoid bias in the study, all the selected features were located on the contour of the images. We did not use any features inside the tumor to avoid imposing the shape of each region, instead we used the outline of the images to imitate the effect of the FFPE process in the sample. It is important to consider that the selected control points should represent the position of equivalent features in both images. Fig. 4.1 presents an example for the selection of equivalent features in sample 8A fresh.

For this purpose, consider the aligned pathology mask $\mathcal{S}_r$ and the THz image $\mathcal{T}$ as the source and the target images in this process, respectively. Each pixel within $\mathcal{S}_r$ contains the color intensities of red, green, and blue for $r = 1, 2, 3$, respectively, and take values between $[0, 255]$. Assume that the user has defined the control points $\mathbf{V}_S$ and $\mathbf{V}_T$ on the contour of the source and target images, respectively. Given this information, we can proceed to the creation of the topologically equivalent meshes and to warp of each of their elements.

4.1. Delaunay Triangulation

This section discusses the division of the images into smaller elements prior warping them. Using the control points previously defined, the images are divided into smaller triangular-shaped elements to later warp them element-by-element. For this division, the algorithm creates triangular

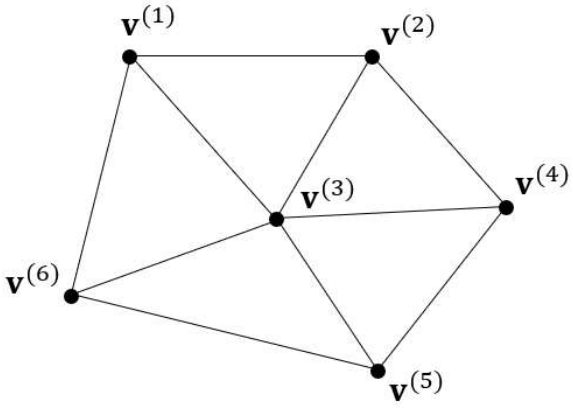


Figure 4.2: Delaunay triangulation example.

meshes in the source and target images using their control points as vertices through a method called Delaunay triangulation. It is important to consider that, to successfully warp the images, the meshes should not present any folding or discontinuities, and therefore should be topologically equivalent [13]. The Delaunay triangulation is suitable for the creation of the meshes because it maximizes the smallest angle among all the triangles in the mesh. Hence this method avoids the creation of thin triangles and provides a fairer division of the image.

Define the Delaunay triangulation of $\mathbf{V}$ control points as $\mathcal{D}(\mathbf{V})$, where $\mathcal{D}$ corresponds to a $M \times 3$ matrix whose rows represent the vertices of each triangle and M represents the total number of triangles in the mesh. Fig. 4.2 shows an example of a Delaunay triangulation with the with $M = 5$ triangles, mathematically represented as follows:

$$\mathcal{D}(\mathbf{V}) = \begin{bmatrix} \mathbf{v}^{(1)} & \mathbf{v}^{(2)} & \mathbf{v}^{(3)} \\ \mathbf{v}^{(2)} & \mathbf{v}^{(3)} & \mathbf{v}^{(4)} \\ \mathbf{v}^{(3)} & \mathbf{v}^{(4)} & \mathbf{v}^{(5)} \\ \mathbf{v}^{(3)} & \mathbf{v}^{(5)} & \mathbf{v}^{(6)} \\ \mathbf{v}^{(1)} & \mathbf{v}^{(3)} & \mathbf{v}^{(6)} \end{bmatrix}$$

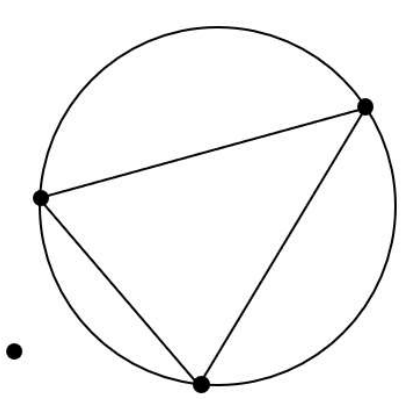
According to [16], the vertices of a Delaunay triangulation satisfy the following two properties:

1. Two points $\mathbf{v}^{(i)}, \mathbf{v}^{(j)} \in \mathbf{V}$ represent an edge of a triangle within the mesh if there is an

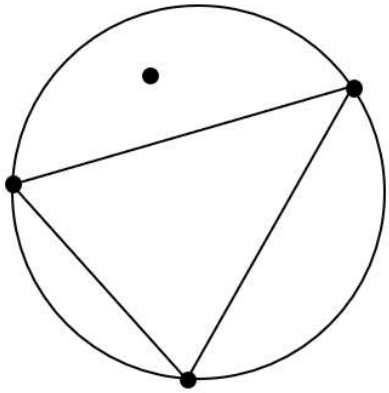
antipodal circle that includes both points on its boundary without containing any other point within $\mathbf{V}$.

2. Three points $\mathbf{v}^{(i)}, \mathbf{v}^{(j)}, \mathbf{v}^{(k)} \in \mathbf{V}$ represent the vertices of a triangle in the mesh if their circumcircle does not contain any other point within $\mathbf{V}$.

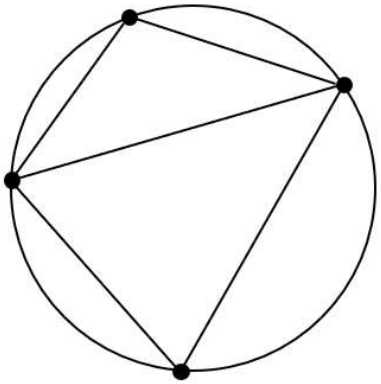
Fig. 4.3 shows some examples of Delaunay triangulations, for non-cocircular and cocircular points. Fig. 4.3a represents an example of a valid Delaunay triangulation that follows the properties described above, while Fig. 4.3b does not follows the second property and therefore, it does not represent a Delaunay triangulation. As shown in the examples, if the control points are not cocircular with respect to each other, there is an unique Delaunay triangulation for this set of points.



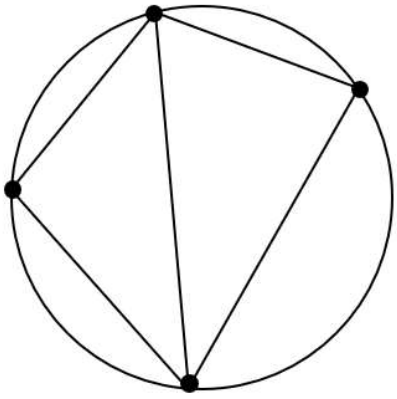
(a) Delaunay triangulation.



(b) Non-Delaunay triangulation.

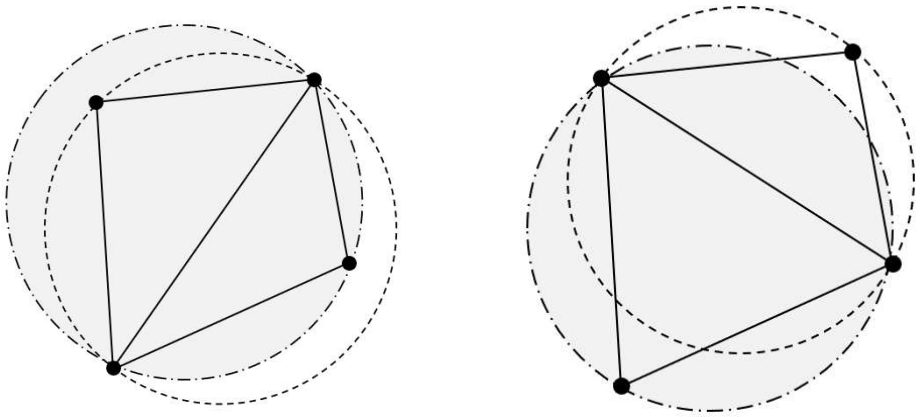


(c) Cocircular triangulation - version 1.



(d) Cocircular triangulation - version 2.

Figure 4.3: Delaunay triangulation.



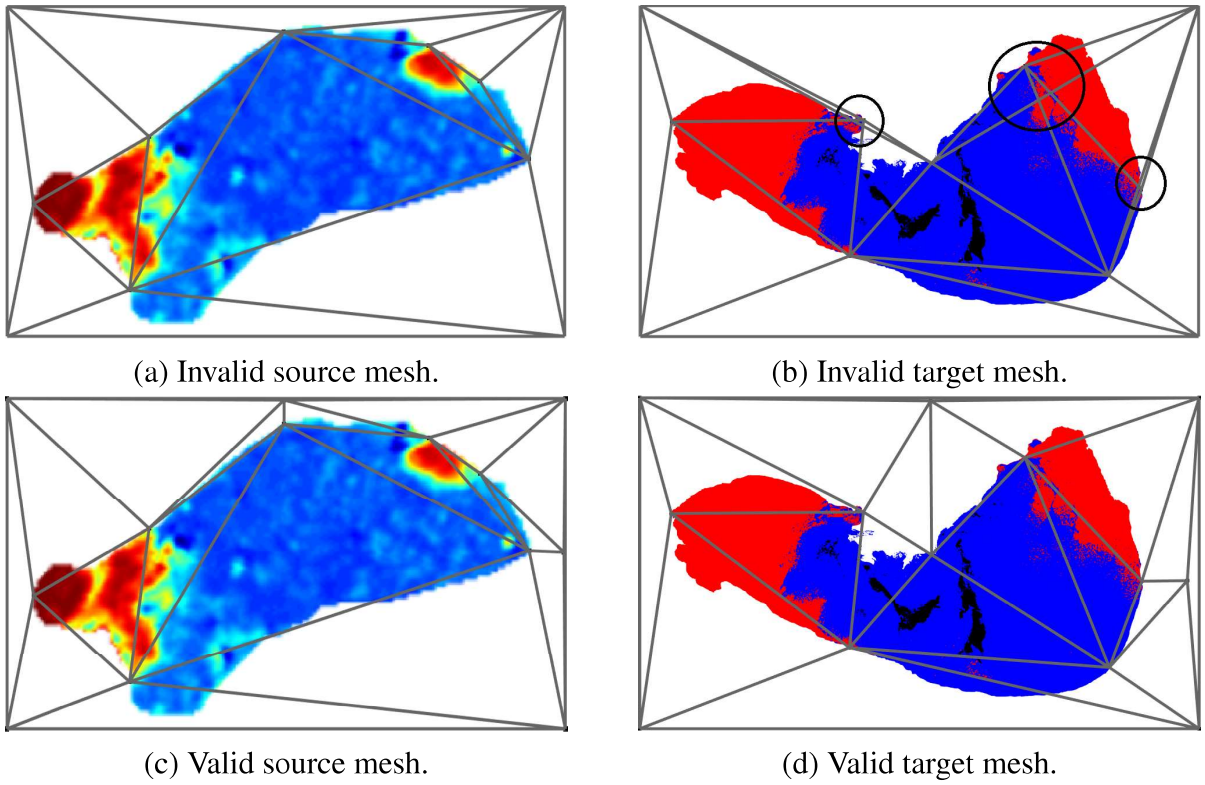
(a) Edge before flipping. (b) Edge after flipping.

Figure 4.4: Flip algorithm.

Otherwise, as shown in Figs. 4.3c and 4.3d, there are different possible Delaunay triangulations for that set of points.

Given the control points, the flip algorithm is a simple, but effective technique to create an optimum Delaunay triangulation. This algorithm starts by establishing an arbitrary initial triangulation. A commonly used initial triangulation method is the lexicographically technique, in which the control points are sorted either by their x-coordinates or their y-coordinates (in case some points share the same x-coordinates) and the algorithm follows this sequence to establish the triangulation. Once this initial triangulation is established, the algorithm checks every edge in the triangulation to make sure it is locally Delaunay, if not, the algorithm flips the edge as shown in Fig. 4.4 and keeps performing this procedure until all the edges are locally Delaunay. To do this, the algorithm considers the edges located at the diagonals of the convex quadrilaterals in the mesh, as shown in Fig. 4.4a. If the algorithm detects that the edge is not locally Delaunay, it flips it to the location of the other diagonal of the quadrilateral, as shown in Fig. 4.4b. It is important to highlight that although the edge flip is legal only when both diagonals of the quadrilateral cross each other, the edges that cannot be flipped are always locally Delaunay [17].

Even though it is possible that the surrounding edges of the mesh are not locally Delaunay after flipping an edge, it has been proved that this algorithm achieves a fully Delaunay triangulation after $\mathcal{O}(n^2)$ flips for n control points [17].



(a) Invalid source mesh.

(b) Invalid target mesh.

(c) Valid source mesh.

(d) Valid target mesh.

Figure 4.5: Mesh validation for sample 8A fresh.

Considering the application of Delaunay triangulation in image mesh morphing, the source mesh is computed using this approach, while the target mesh is defined as topologically equivalent to the source mesh. Therefore, the selection of the control points plays an important role in the validation of the meshes prior their warping, as shown in Fig. 4.5. Figs. 4.5a and 4.5b show invalid source and target meshes for sample 8A fresh due to the folding of some triangles in the target mesh. It is important to consider that although the source mesh in Fig. 4.5a is a correct Delaunay triangulation, it does not represent a valid mesh because of the folding of the topologically equivalent mesh in the target image. Therefore, the control points should be selected so that the meshes do not present any folding as shown in Figs. 4.5c and 4.5d.

4.2. Homography Estimation

This section describes the procedure to reallocate the control points from the source to their target locations and how each element within the mesh is warped according to this reallocation using

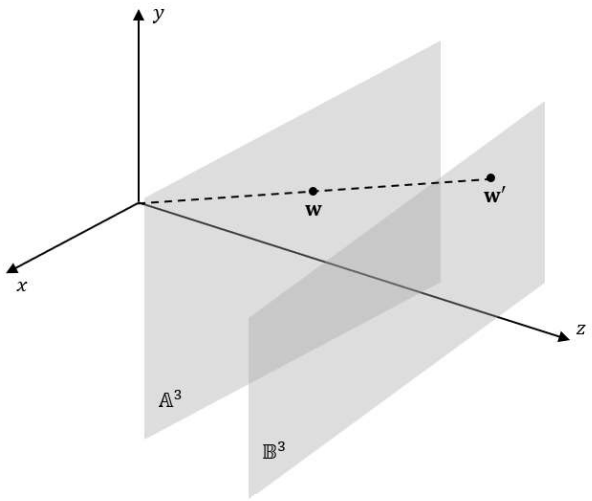


Figure 4.6: Projection example.

homography estimation.

Consider the homogeneous representation of point $\mathbf{v} = (x, y)$ as $\mathbf{w} = (x_w, y_w, z_w)^T$, where $x = x_w/z_w$ and $y = y_w/z_w$. According to [18], an homography is an invertible mapping $h : \mathbb{P}^2 \rightarrow \mathbb{P}^2$ if and only if there is a non-singular matrix $\mathbf{H}_{3\times3}$ such that $h(\mathbf{w}) = \mathbf{H}\mathbf{w}$, where $\mathbb{P}^2$ corresponds to a projective plane, $\mathbf{w}$ represents the homogeneous representation of a point $\mathbf{v} \in \mathbb{P}^2$, and $\mathbf{H}$ is an homogeneous matrix with 8 degrees of freedom [19].

To estimate the homogeneous matrix $\mathbf{H}$, this algorithm uses the linear direct transformation of the homogeneous representation of the points [20]. As shown in Fig. 4.6, this approach can be represented as a linear projection of $\mathcal{R}^3$, where the homogeneous representation $\mathbf{w} \in \mathbb{A}^3$ is projected as $\mathbf{w}' = (x'_w, y'_w, z'_w)^T \in \mathbb{B}^3$ using the following expression:

$$\mathbf{w}' = \mathbf{H}\mathbf{w} \tag{4.1}$$

Since the multiplication by an arbitrary non-zero factor does not affect the expression, e.g. $\mathbf{w}' = a\mathbf{H}'\mathbf{w}$ for $a \neq 0, a \in \mathcal{R}$, the homography matrix $\mathbf{H}$ is considered homogeneous with 8 degrees of freedom [20]. Once the projection of the homogeneous coordinates of the point $\mathbf{w}'$ has been established, it is possible to define the projection $\mathbb{P}^2 \rightarrow \mathbb{P}^2$ of the point $\mathbf{v} \in \mathbb{P}^2$ as

$\mathbf{v}' = \left(\frac{x'_w}{z'_w}, \frac{y'_w}{z'_w}\right) \in \mathbb{P}^2$.

The application of homography estimation in image mesh morphing focuses on the procedures to warp each element within the mesh according to the reallocation of the control points. For this purpose, consider that the Delaunay triangulation $\mathcal{D}_S(\mathbf{V_S})$ has been computed for the control points $\mathbf{V_S}$ in the source image $\mathcal{S}$, as described in section 4.1. Given that the source and target images should be topologically equivalent, the number of triangles in both meshes is the same, M, and the indexes of the mesh $\mathcal{D}_T(\mathbf{V_T})$ are equivalent to $\mathcal{D}_S(\mathbf{V_S})$.

To start warping each triangle within the source mesh into the location of the triangles in the target mesh, it is necessary to compute their homography matrices. Since there are M triangles in each mesh, one homography matrix $\mathbf{H}_m$ should be computed per triangle $m = 1, \cdots, M$.

To illustrate the calculation of $\mathbf{H}_m$, consider the 3 vertices of the m-th triangle in the source mesh as $\mathcal{D}_S(m) = (\mathbf{v_S}^{(i)}, \mathbf{v_S}^{(j)}, \mathbf{v_S}^{(k)})$ and its centroid $\mathbf{C}_S(m) = (a_S, b_S)$, calculated as follows:

$$(a_S, b_S) = \left(\frac{x_S^{(i)} + x_S^{(j)} + x_S^{(k)}}{3}, \frac{y_S^{(i)} + y_S^{(j)} + y_S^{(k)}}{3}\right)$$

where $i \in \{1, \cdots, n\}$, $j \in \{1, \cdots, n\}$ and $k \in \{1, \cdots, n\}$. Similarly, consider the vertices of the m-th triangle in the target mesh as $\mathcal{D}_T(m) = (\mathbf{v_T}^{(i)}, \mathbf{v_T}^{(j)}, \mathbf{v_T}^{(k)})$ and its centroid $\mathbf{C}_T(m) = (a_T, b_T)$. It is possible to represent the projection of each pair of control points $\mathbf{v_S}^{(i)} = (x_S^{(i)}, y_S^{(i)})$ and $\mathbf{v_T}^{(i)} = (x_T^{(i)}, y_T^{(i)})$ as:

$$\begin{bmatrix} x_T^{(i)} \\ y_T^{(i)} \\ 1 \end{bmatrix} = \begin{bmatrix} h_{m1} & h_{m2} & h_{m3} \\ h_{m4} & h_{m5} & h_{m6} \\ h_{m7} & h_{m8} & h_{m9} \end{bmatrix} \begin{bmatrix} x_S^{(i)} \\ y_S^{(i)} \\ 1 \end{bmatrix} \tag{4.2}$$

Rewriting eq. (4.2), the following two equations are obtained based on $\mathbf{v_S}^{(i)}$ and $\mathbf{v_T}^{(i)}$:

$$\begin{cases} -x_S^{(i)} h_{m1} - y_S^{(i)} h_{m2} - h_{m3} + x_S^{(i)} x_T^{(i)} h_{m7} + y_S^{(i)} x_T^{(i)} h_{m8} + x_T^{(i)} h_{m9} = 0 \\ -x_S^{(i)} h_{m4} - y_S^{(i)} h_{m5} - h_{m6} + x_S^{(i)} y_T^{(i)} h_{m7} + y_S^{(i)} y_T^{(i)} h_{m8} + y_T^{(i)} h_{m9} = 0 \end{cases}$$

Considering the 4 reference points in the m-th triangle in the source and target meshes, it is possible to formulate the following equation system.

$$\begin{bmatrix} -x_S^{(i)} & -y_S^{(i)} & -1 & 0 & 0 & 0 & x_S^{(i)}x_T^{(i)} & y_S^{(i)}x_T^{(i)} & x_T^{(i)} \\ 0 & 0 & 0 & -x_S^{(i)} & -y_S^{(i)} & -1 & x_S^{(i)}y_T^{(i)} & y_S^{(i)}y_T^{(i)} & y_T^{(i)} \\ -x_S^{(j)} & -y_S^{(j)} & -1 & 0 & 0 & 0 & x_S^{(j)}x_T^{(j)} & y_S^{(j)}x_T^{(j)} & x_T^{(j)} \\ 0 & 0 & 0 & -x_S^{(j)} & -y_S^{(j)} & -1 & x_S^{(j)}y_T^{(j)} & y_S^{(j)}y_T^{(j)} & y_T^{(j)} \\ -x_S^{(k)} & -y_S^{(k)} & -1 & 0 & 0 & 0 & x_S^{(k)}x_T^{(k)} & y_S^{(k)}x_T^{(k)} & x_T^{(k)} \\ 0 & 0 & 0 & -x_S^{(k)} & -y_S^{(k)} & -1 & x_S^{(k)}y_T^{(k)} & y_S^{(k)}y_T^{(k)} & y_T^{(k)} \\ -a_S & -b_S & -1 & 0 & 0 & 0 & a_S a_T & b_S a_T & a_T \\ 0 & 0 & 0 & -a_S & -b_S & -1 & a_S b_T & b_S b_T & b_T \end{bmatrix} \begin{bmatrix} h_{m1} \\ h_{m2} \\ h_{m3} \\ h_{m4} \\ h_{m5} \\ h_{m6} \\ h_{m7} \\ h_{m8} \\ h_{m9} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.3}$$

Since the homography matrix $\mathbf{H}_m$ has 8 degrees of freedom, the equation system in eq. (4.3) provides enough equations to find a solution. Additionally, this equation can be represented as:

$$\mathbf{A}\mathbf{h}_m = \mathbf{0} \tag{4.4}$$

where $\mathbf{h}_m$ represents the elements of matrix $\mathbf{H}_m$ arranged in a column vector. To solve this equation and find the homography matrix for the m-th triangle in the mesh, consider the singular value decomposition (SVD) of the 8×9 matrix $\mathbf{A}$ as:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^T \tag{4.5}$$

where $\mathbf{U}$ is an 8×8 matrix formed by the eigenvectors of $\mathbf{A}\mathbf{A}^T$, $\mathbf{W}$ is an 9×9 matrix formed by the eigenvectors of $\mathbf{A}^T\mathbf{A}$, and $\mathbf{\Sigma}$ is an 8×9 diagonal matrix that contains the singular values $\{\sigma_1, \cdots, \sigma_8\}$ of $\mathbf{A}$, as shown in eq. (4.6). These singular values represent the non-zero square roots of the eigenvalues of $\mathbf{A}^T\mathbf{A}$, and therefore $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_8$ [21].

$$\mathbf{\Sigma} = \text{diag}(\sigma_1, \cdots, \sigma_8) \tag{4.6}$$

According to the SVD approach, it is possible to solve eq. (4.4) considering that $\mathbf{h}_m$ is equal to the last column of $\mathbf{W}$ [22]. Once $\mathbf{h}_m$ is rearranged into $\mathbf{H}_m$, the algorithm proceeds to estimate the location of each point $(x_S^{(l)}, y_S^{(l)})$ for $l = 1, \cdots, L_m$ that lies inside the triangle $\mathcal{D}_S(m)$ using eq. (4.7), where L_m is the total number of pixels that lie inside $\mathcal{D}_S(m)$.

$$\begin{bmatrix} x_w^{(l)} \\ y_w^{(l)} \\ z_w^{(l)} \end{bmatrix} = \begin{bmatrix} h_{m1} & h_{m2} & h_{m3} \\ h_{m4} & h_{m5} & h_{m6} \\ h_{m7} & h_{m8} & h_{m9} \end{bmatrix} \quad \begin{bmatrix} x_S^{(l)} \\ y_S^{(l)} \\ 1 \end{bmatrix} \tag{4.7}$$

Therefore, it is possible to estimate the target location $(x_T^{(l)}, y_T^{(l)})$ of $(x_S^{(l)}, y_S^{(l)})$ considering that $x_T^{(l)} = \frac{x_w^{(l)}}{z_w^{(l)}}$ and $y_T^{(l)} = \frac{y_w^{(l)}}{z_w^{(l)}}$. Once the target location is known, the code proceeds to project every point from the source triangle into the positions of the target triangle in a new image $\mathcal{F}_r$, as follows:

$$\mathcal{F}_r(x_T, y_T) = \mathcal{S}_r(x_S, y_S); \quad r = \{1, 2, 3\} \tag{4.8}$$

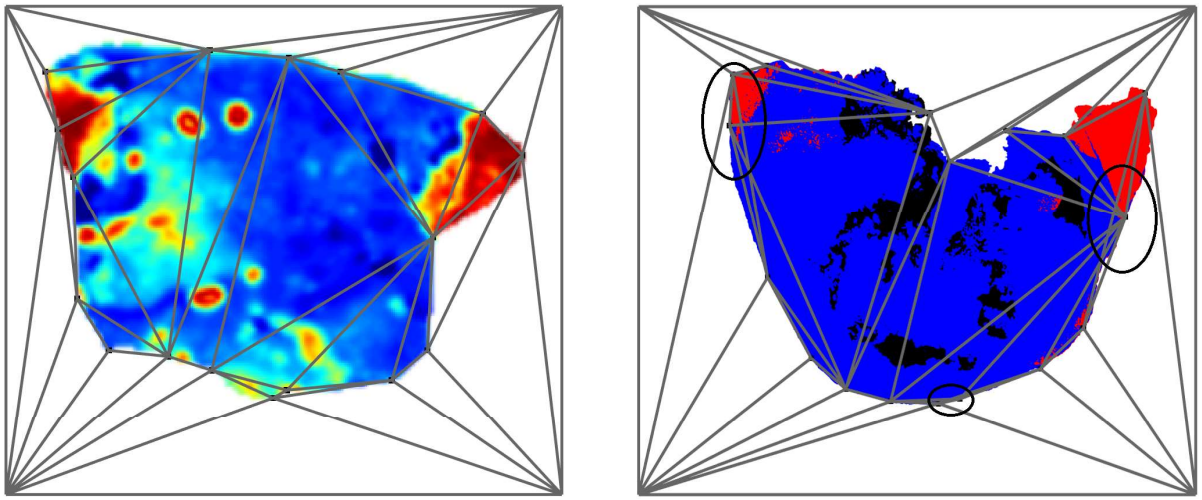
$\mathcal{F}_r$ represents the morphed pathology mask that matches the shape of the contour of the tissue in the THz image and its resolution, where $r = \{1, 2, 3\}$ represent the red, green, and blue color intensities per pixel. The complete process of the homography estimation is described in Algorithm 2. For the correct performance of the algorithm, the resolution of the source image must be larger than the target image.

4.3. Pathology Mask Morphing

This section describes the application of homography estimation to completely reshape the pathology mask into the contour of the THz image. As introduced in section 4.2, it is feasible to use homography estimation to reallocate the pixels at the source image into a new target image according to the prior knowledge of the location of a finite number of control points between the

Algorithm 2 Image morphing.

Define the control points $\mathbf{V}_S$ and $\mathbf{V}_T$.
Define the source and target triangulations $\mathcal{D}_S$ and $\mathcal{D}_T$.
for $m = 1, \cdots, M$ **do**
  Define the homography matrix $\mathbf{H}_m$ using eq. (4.5).
  for $l = 1, \cdots, L_m$ **do**
    Estimate $(x_T^{(l)}, y_T^{(l)})$ using eq. (4.7).
    Place the pixel in its estimated target position using eq. (4.8).
  end for
end for



(a) THz image mesh. (b) Pathology mask mesh.

Figure 4.7: Non-iteratively pathology morphing for sample 7B fresh.

two images. Although this approach is sufficient to completely morph the image, the requirement of having topologically equivalent meshes in both images represents a challenge. As shown in Fig. 4.7, when trying to morph all the mismatched areas in the source image of sample 7B fresh at once, the triangulations of the source and target images are compromised due to the folding of some triangles in the mesh (circled areas in Fig. 4.7b). Hence, an iterative morphing approach is recommended, in which the user starts morphing a few regions in the source image and keeps morphing the obtained results until reaching the shape of the contour of the target image.

As shown in Fig. 4.8, the iteratively morphing process for sample 7B fresh ensures the creation of topologically equivalent meshes and the correct reallocation of all the regions within the tumor in the final result. The idea of this procedure is to morph specific challenging regions one at the

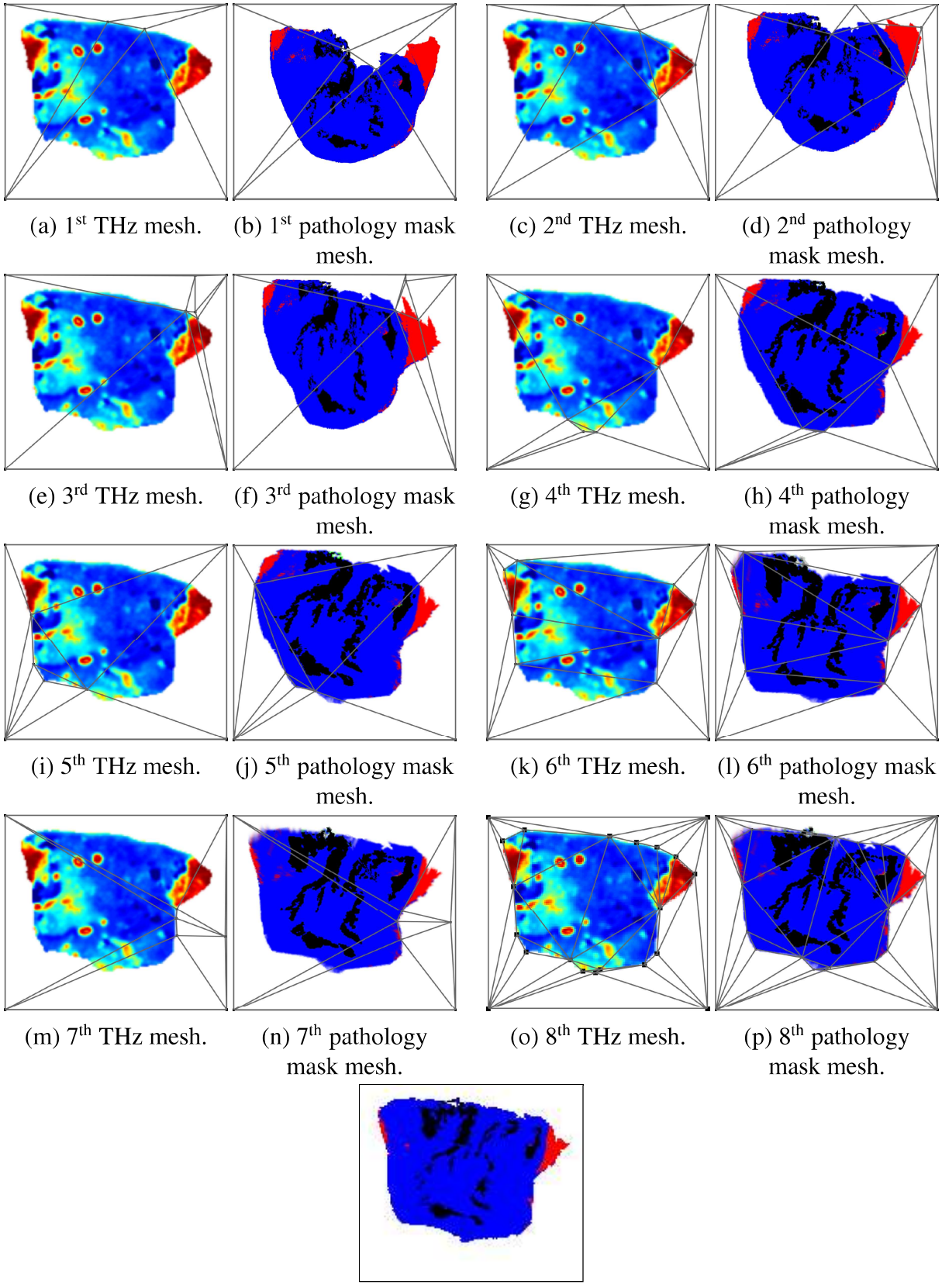
time, e.g. the gap next to the green region at the top of Fig. 4.8b. As shown in 4.8d, the gap is almost gone and the user moved on to correct the alignment of the fat area at the right of the image. It is worth mentioning that the number of iterations and number of regions to morph at the time varies depending on the severity of the mismatch between the images and the experience of the user.

Once the pathology mask has been morphed, it is possible to create a reference matrix such that it summarizes the tissue classification of each pixel as presented in the histopathology report, e.g. pixels labeled as 1 represent cancer, 2 represent muscle, etc. This reference matrix $\mathbf{D}$ is created based on the RGB color assigned during the creation of the pathology mask. Considering that the colors become darker or lighter after their morphing, the code classifies each pixel according to the minimum difference between its color in the morphed pathology mask and all the possible assigned colors within the pathology mask prior morphing, as follows:

$$\mathbf{D}(x_q, y_q) = \underset{p:1,\ldots,P}{\operatorname{argmin}} \sqrt{\sum_{r=1}^{3} |f_r(x_q, y_q) - c_r(p)|^2} \tag{4.9}$$

where (x_q, y_q) for $x_q \in \{1, \ldots, Q_x\}$ and $y_q \in \{1, \ldots, Q_y\}$ represent the location of the pixel in the morphed pathology image, P is the total number of regions in the tissue, $f_r(x_q, y_q)$ for $r = 1, 2, 3$ represent the red, green, and blue color intensities of the pixel in the morphed pathology mask, and $c_r(p)$ represent the color intensities that represent each region (e.g. blue, $c(1) = (0, 0, 255)$, corresponds to cancer).

Once the reference matrix $\mathbf{D}$ has been created and considering that it has the same size as the THz image, it is possible to perform a pixel-by-pixel comparison between them to quantitatively evaluate the results.



(a) 1st THz mesh. (b) 1st pathology mask mesh. (c) 2nd THz mesh. (d) 2nd pathology mask mesh.

(e) 3rd THz mesh. (f) 3rd pathology mask mesh. (g) 4th THz mesh. (h) 4th pathology mask mesh.

(i) 5th THz mesh. (j) 5th pathology mask mesh. (k) 6th THz mesh. (l) 6th pathology mask mesh.

(m) 7th THz mesh. (n) 7th pathology mask mesh. (o) 8th THz mesh. (p) 8th pathology mask mesh.

(q) Final result.

Figure 4.8: Iteratively pathology morphing for sample 7B fresh.

Chapter 5

Experimental Results

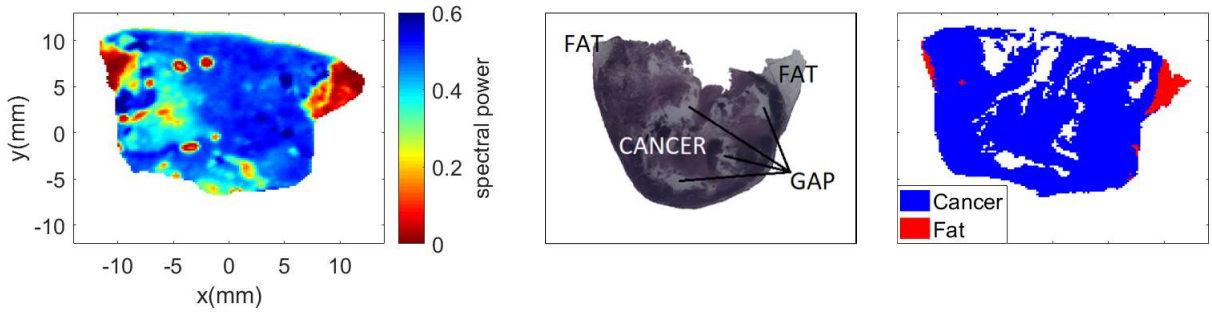
This chapter presents the experimental results obtained from the morphing algorithm described in chapters 3 and 4. As mentioned in chapter 2, this thesis evaluates 3 samples: tumor 7 section B, tumor 8 section A, and tumor 9 section A, to demonstrate the effectiveness of the mesh morphing algorithm for the quantitative evaluation of the THz imaging system.

For this purpose, the THz images were classified using the unsupervised Bayesian learning algorithm in [6]. In this work, a Markov chain Monte Carlo (MCMC) approach was employed to classify the regions within the tumor according to the summarized intensity of each pixel in the THz image (e.g. maximum value or integration of the signal per pixel for block and fresh tissue, respectively). This method, known as mixture components, considers that the pixels that belong to certain region follow a specific probability distribution different from the rest of the regions and uses this characteristic to estimate the probability of each pixel belonging to each region. The classification of the regions in this thesis was performed considering the Gaussian mixture and the t-mixture models described in [6]. The reference matrix $\mathbf{D}$ described in section 4.3 was used as the histopathologic reference to quantitatively evaluate the results obtained from the unsupervised Bayesian model.

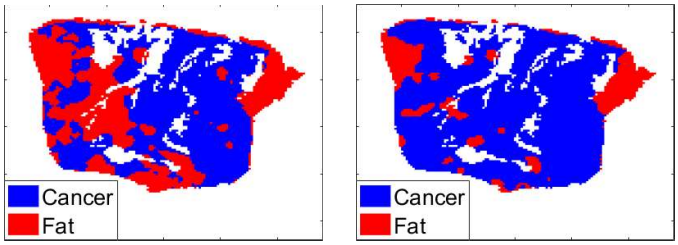
5.1. Samples with Two Regions

Two freshly excised murine tumors corresponding to samples 7B and 8A, were selected to evaluate the performance of the mesh morphing algorithm in tumors with two regions.

Fig. 5.1a presents the THz image collected from sample 7B, which was obtained using the spectral power of the reflected THz signal per pixel. Fig. 5.1b shows the histopathologic results obtained after fixing the tumor in paraffin, where it is confirmed that the tissue presents only two regions, cancer and fat. Fig. 5.1c represents the morphed pathology image that was obtained using the mesh morphing algorithm, in which the blue region represents cancer and the red represents



(a) THz image. (b) Pathology image. (c) Morphed Pathology.



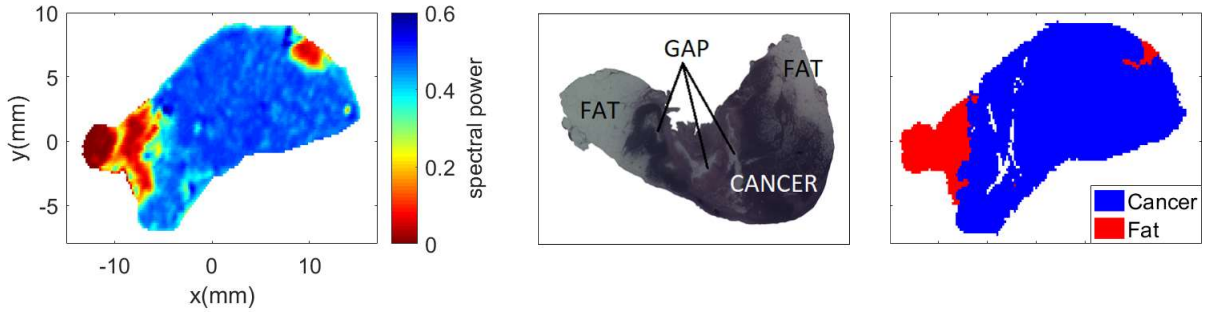
(d) Gaussian mixture model. (e) t-mixture model.

Figure 5.1: Sample 7B fresh.

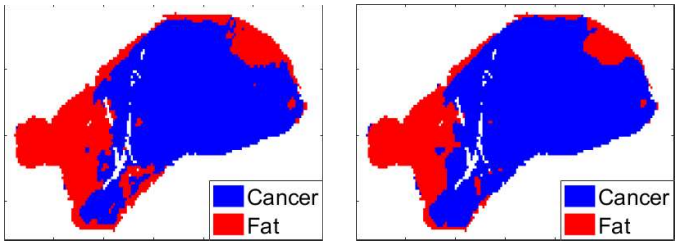
fat. As shown in Figs. 5.1a and 5.1c, there is a significant correlation between the THz image and the morphed pathology results. Figs. 5.1d and 5.1e represent the classification results obtained from the Gaussian mixture and the t-mixture models, respectively.

Similarly, Fig. 5.2a presents the THz image collected from sample 8A. Fig. 5.2b shows the histopathologic results obtained after fixing the tumor in paraffin. Fig. 5.2c represents the morphed pathology image that was produced using the mesh morphing algorithm, in which the blue region represents cancer and red represents fat. As shown in Figs. 5.2a and 5.2c, there is a significant correlation between the THz image and the morphed pathology results. Figs. 5.2d and 5.2e represent the classification results obtained from the Gaussian mixture and the t-mixture models.

It is worth mentioning that the white areas inside the tissue in Figs. 5.1c, 5.1d, 5.1e, 5.2c, 5.2d, and 5.2e represent the gaps that were formed during the histopathologic process in samples 7B and 8A (see Figs. 5.1b and 5.2b). Since the gaps were produced due to fixing the tissue in paraffin, they were nonexistent while the tissue was fresh. To avoid compromising the evaluation of the results, the pixels that corresponded to gaps were not considered during the evaluation process.



(a) THz image. (b) Pathology image. (c) Morphed Pathology.

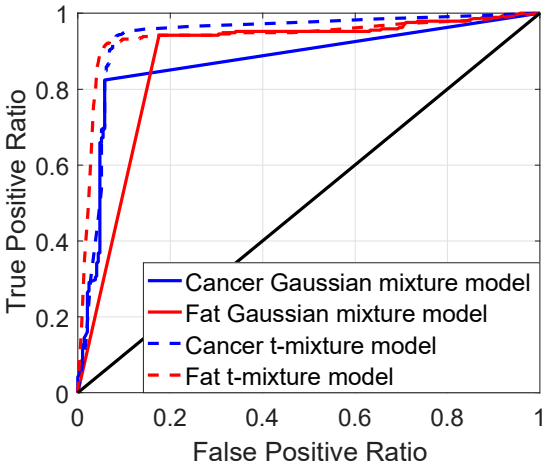


(d) Gaussian mixture model. (e) t-mixture model.

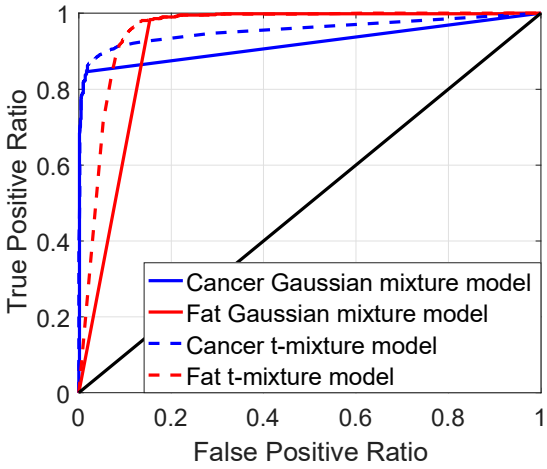
Figure 5.2: Sample 8A fresh.

Considering that the morphed pathology has the same resolution as the results obtained from the Gaussian mixture and the t-mixture models, a pixel-by-pixel comparison was performed and ROC curves were generated to represent the probability of detection of each region vs. its probability of false alarm. Figs. 5.3a and 5.3b present the ROC curves for samples 7B and 8A, respectively. For sample 7B, the ROC curves show that it is possible to achieve a 90% probability of detection with a 6.8% and 5.1% probability of false alarm for cancer and fat, respectively, using the t-mixture model. Likewise, the t-mixture model for sample 8A achieves a 90% probability of detection with a 7.2% and 9.1% probability of false alarm for cancer and fat, respectively. Additionally, Figs. 5.3c and 5.3d represent the probability distributions of the regions for samples 7B and 8A, respectively. For both images, the distributions of cancer and fat are considerable different from each other and hence, its differentiation is possible using the mixture model.

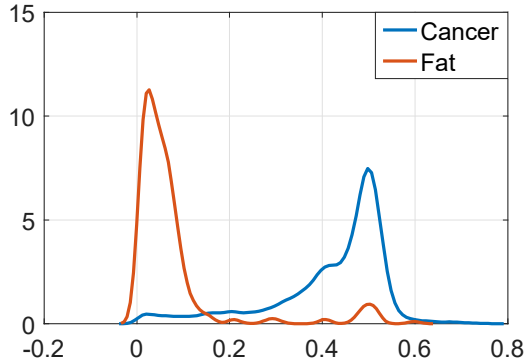
To better present the performance of the Bayesian algorithm, the areas under the ROC curves are presented in Table 5.1. For both samples, the t-mixture model performed better than the Gaussian mixture model. All areas for both samples are greater than 87%, providing quantitative evi-



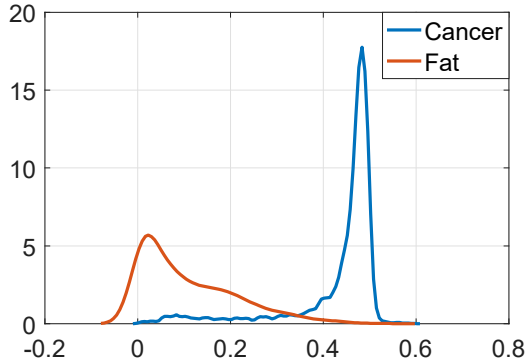
(a) ROC curves for sample 7B fresh.



(b) ROC curves for sample 8A fresh.



(c) Probability distribution for sample 7B fresh.



(d) Probability distribution for sample 8A fresh.

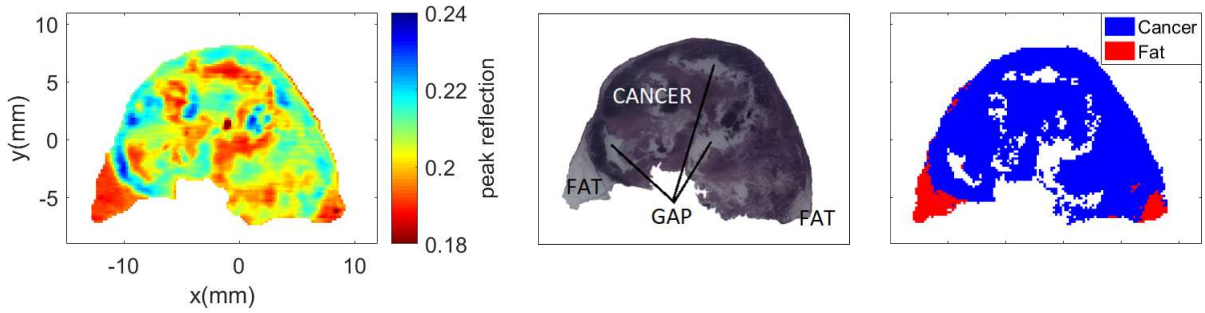
Figure 5.3: ROC curves and probability distributions for samples 7B and 8A fresh.

dence of the good performance of the Bayesian algorithm for fresh samples with two regions.

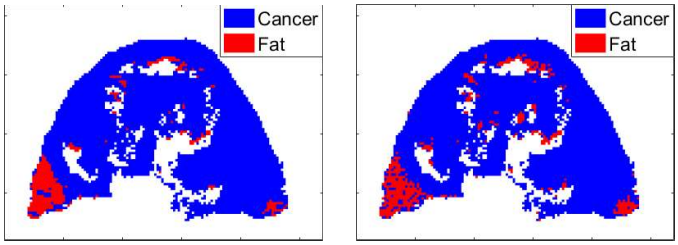
Although fixed tissue does not present severe shape shifting, there is still a significant misalignment and resolution mismatch with respect to the pathology image. Therefore, the FFPE sample corresponding to tumor 7B was selected to evaluate the performance of the morphing algorithm

Table 5.1: Area under the ROC curve for samples 7B and 8A fresh.

Type of tissue: Model	Sample 7B	Sample 8A
Cancer: Gaussian mixture model	0.8767	0.9199
Fat: Gaussian mixture model	0.8767	0.9199
Cancer: t-mixture model	0.9377	0.9558
Fat: t-mixture model	0.9377	0.9558



(a) THz image. (b) Pathology image. (c) Morphed Pathology.

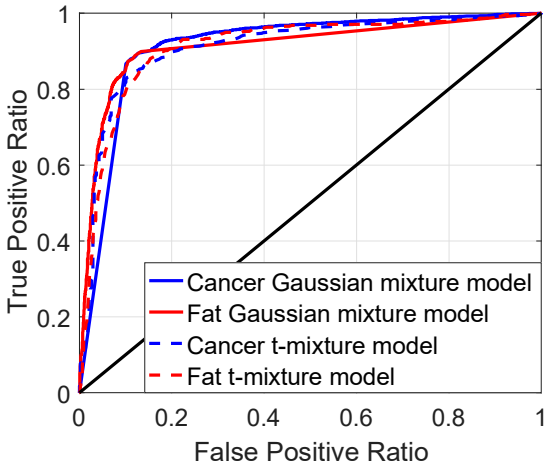


(d) Gaussian mixture model. (e) t-mixture model.

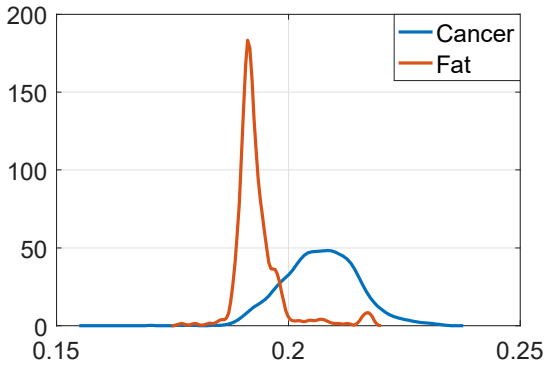
Figure 5.4: Sample 7B FFPE.

in fixed tissue. Fig. 5.4a presents the THz image collected from FFPE sample 7B, which was obtained using the peak of the normalized THz signal in time domain per pixel. Fig. 5.4b shows the histopathologic results obtained after fixing the tumor in paraffin. Fig. 5.4c represents the morphed pathology image that was produced using the mesh morphing algorithm, in which the blue region represents cancer and the red represents fat. As shown in Figs. 5.4a and 5.4c, there is a significant correlation between the THz image and the morphed pathology results. Figs. 5.4d and 5.4e represent the classification results obtained from the Gaussian mixture and the t-mixture models, respectively.

The performance of the unsupervised Bayesian model was evaluated for sample 7B FFPE using ROC curves, as shown in Fig. 5.5a. These curves show that it is possible to achieve a 90% probability of detection with a 14.6% and 14.0% probability of false alarm for cancer and fat, respectively, using the Gaussian mixture model. Fig. 5.5b represents the probability distributions of the regions for sample 7B FFPE. Similarly to the previous samples, given that the distributions are different from each other, it is possible to differentiate them. The areas under the ROC curves



(a) ROC curves for sample 7B FFPE.



(b) Probability distribution for sample 7B FFPE.

Figure 5.5: ROC curves and probability distributions for sample 7B FFPE.

are presented in Table 5.2. For this case, all areas are greater than 91% and the performance of both models is similar.

5.2. Samples with Three Regions

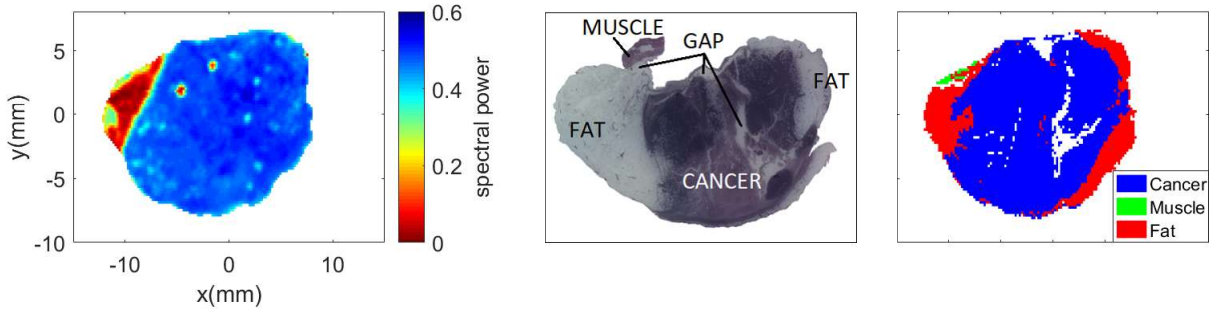
This section evaluates the performance of the mesh morphing algorithm in tumors with 3 regions. For this purpose, the freshly excised sample 9A was selected due to the presence of cancer, fat, and muscle in it.

Fig. 5.6a presents the THz image collected from sample 9a, which was obtained using the spectral power of the reflected THz signal per pixel. Fig. 5.6b shows the histopathologic results

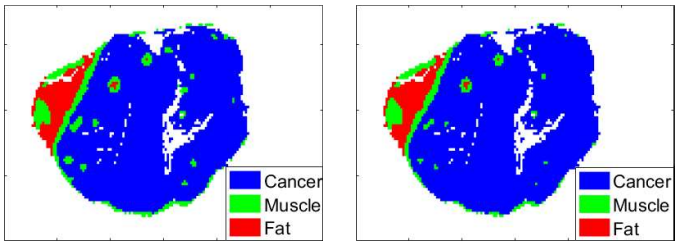
Table 5.2: Area under the ROC curve for sample 7B FFPE.

Type of tissue: Model	Sample 7B
Cancer: Gaussian mixture model	0.9136
Fat: Gaussian mixture model	0.9136
Cancer: t-mixture model	0.9121
Fat: t-mixture model	0.9121

obtained after fixing the tumor in paraffin, where it is confirmed that the tissue presents three regions, cancer, muscle, and fat. Fig. 5.6c represents the morphed pathology image that was produced using the mesh morphing algorithm, in which the blue, green, and red regions represent cancer, muscle, and fat, respectively. As shown in Figs. 5.6a and 5.6c, there is small correlation between the THz image and the morphed pathology results. Although the morphed pathology image clearly represents the histopathologic results of the sample, the fluid reflection in the right part of the tissue during the THz imaging process covered most of the fat region in this case. Figs. 5.6d and 5.6e represent the classification results obtained from the Gaussian mixture and the t-mixture models, respectively.

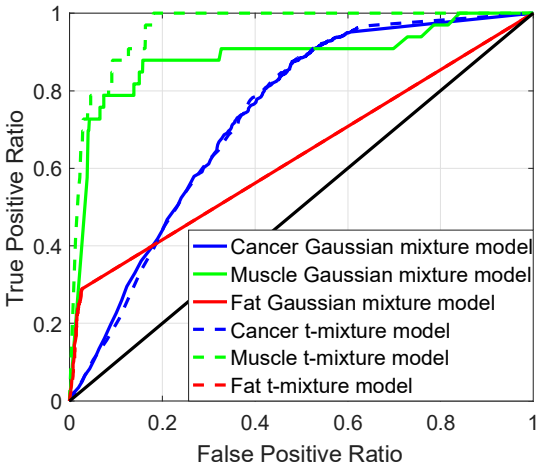


(a) THz image. (b) Pathology image. (c) Morphed Pathology.

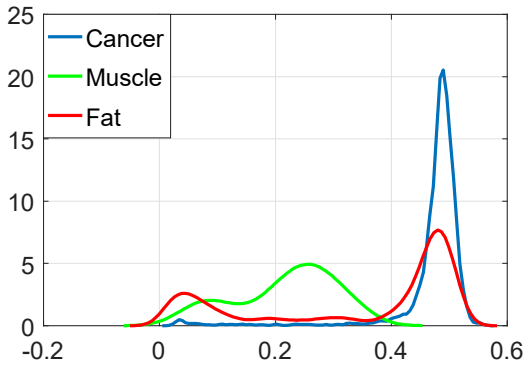


(d) Gaussian mixture model. (e) t-mixture model.

Figure 5.6: Sample 9A fresh.



(a) ROC curves for sample 9A fresh.



(b) Probability distribution for sample 9A fresh.

Figure 5.7: ROC curves and probability distributions for sample 9A fresh.

Similarly to samples with two regions, the performance of the unsupervised Bayesian model was evaluated for sample 9A using ROC curves, as shown in Fig. 5.7a. This curves show that the performance of the model seems outstanding for muscle (90% probability of detection with 12.7% probability of false alarm), while the detection of cancer and fat is compromised due to the presence of fluid in the sample. For this sample, the areas under the ROC curve for cancer and fat are lower than 80%, while the area for muscle is greater than 88%.

Considering that the muscle area was small compared to the rest of the regions and the fact that its detection on other samples was not good in [1] and [6], the results obtained from this sample were particular and nonconclusive. In addition, the probability distributions of each of the regions

Table 5.3: Area under the ROC curve for sample 9A fresh.

Type of tissue: Model	Sample 9A
Cancer: Gaussian mixture model	0.7342
Muscle: Gaussian mixture model	0.8854
Fat: Gaussian mixture model	0.6314
Cancer: t-mixture model	0.7341
Muscle: t-mixture model	0.9608
Fat: t-mixture model	0.6313

are shown in Fig. 5.7b, where the overlap of the fat and cancer distributions is given by the fluid reflection that covered the right fat region in the sample. Hence, as mentioned in [6], the Bayesian modeling should be further improved for the detection of muscle within a tumor.

Chapter 6

Conclusions

This thesis presented a pathology mesh morphing algorithm based on homography estimation for THz imaging evaluation. Since the reference pathology image of a sample shows different alignment, shape and resolution compared to the collected THz image, a direct pixel-by-pixel evaluation is not possible. Therefore, this algorithm provides an effective alternative to create a reference matrix that corrects all these problems from the pathology image and provides a reference for the evaluation of the THz image.

The algorithm starts by using the Pearson's correlation coefficient to correct the misalignment between the pathology and THz images. Once the images are aligned, it proceeds to create topologically equivalent meshes for the THz image and its corresponding aligned pathology using Delaunay triangulation. Finally, the algorithm morphs the contour of the aligned pathology image into the shape of the THz image using homography estimation. In addition, this thesis proposed the creation of a pathology reference matrix based on the results obtained from the mesh morphing algorithm that can be used for a pixel-by-pixel comparison between the THz and the pathology images.

The performance of the mesh morphing algorithm was evaluated in chapter 5, in which the reference matrix was used to quantitatively evaluate the classification results of the unsupervised Bayesian learning algorithm described in [6]. For all the samples, the implementation of the mesh morphing algorithm provided a summary of the histopathologic results which presented a good correlation with respect to the THz image. Hence, the results demonstrated that the algorithm provides a good evaluation reference for the classification of the regions within the THz image.

Although all the morphed pathology images represented a summary of the histopathologic results in the shape of the THz image, there were some cases like sample 9A fresh where the pathology results after fixing the tissue were considerable different from the THz image collected from the same sample while it was fresh. For future studies, instead of using the pathology image

of the FFPE tissue, the use of a fairer reference image for fresh tissue will be considered.

Chapter 7

Bibliography

[1] T. Bowman, T. Chavez, K. Khan, J. Wu, A. Chakraborty, N. Rajaram, K. Bailey, and M. El-Shenawee, "Pulsed terahertz imaging of breast cancer in freshly excised murine tumors," *Journal of Biomedical Optics*, vol. 23, no. 2, p. 026004, 2018.

[2] Breastcancer.org. (2018) U.S. Breast Cancer Statistics. [Online]. Available: https://www.cancer.org/cancer/breast-cancer/treatment/surgery-for-breast-cancer/breast-conserving-surgery-lumpectomy.html

[3] T. Bowman, M. El-Shenawee, and L. K. Campbell, "Terahertz transmission vs reflection imaging and model-based characterization for excised breast carcinomas," *Biomed. Opt. Express*, vol. 7, no. 9, pp. 3756–3783, Sep. 2016. [Online]. Available: http://www.osapublishing.org/boe/abstract.cfm?URI=boe-7-9-3756

[4] T. Bowman, T. Chavez, K. Khan, A. Chakraborty, J. Wu, K. Bailey, and M. El-Shenawee, "Statistical signal processing for quantitative assessment of pulsed terahertz imaging of human breast tumors," in *2017 42nd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, Aug. 2017.

[5] T. Bowman, K. Alhallak, T. Chavez, K. Khan, D. Lee, N. Rajaram, J. Wu, A. Chakraborty, K. Bailey, and M. El-Shenawee, "Terahertz imaging of freshly excised breast cancer using mouse model," in *2017 42nd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, Aug. 2017.

[6] T. Chavez, T. Bowman, K. Khan, J. Wu, A. Chakraborty, K. Bailey, and M. El-Shenawee, "Unsupervised bayesian learning for cancer detection with terahertz imaging," 2018, unpublished manuscript.

[7] T. Bowman, M. K. Khan, T. Chavez, K. Alhallak, D. Lee, N. Rajaram, A. Chakraborty, J. Wu, K. Bailey, and M. El-Shenawee, "Terahertz imaging and segmentation of freshly excised xenograft mouse tumors," in *32nd International Union Radio Science General Assembly & Science Symposium*, Aug. 2017.

[8] B. S. Peter, S. Yngvesson, P. Siqueira, P. Kelly, A. Khan, S. Glick, and A. Karellas, "Development and testing of a single frequency terahertz imaging system for breast cancer detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 4, pp. 785–797, July 2013.

[9] T. C. Bowman, M. El-Shenawee, and L. K. Campbell, "Terahertz imaging of excised breast tumor tissue on paraffin sections," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 5, pp. 2088–2097, May 2015.

[10] M.-A. Brun, F. Formanek, A. Yasuda, M. Sekine, N. Ando, and Y. Eishii, "Terahertz imaging applied to cancer diagnosis," *Physics in Medicine & Biology*, vol. 55, no. 16, p. 4615, 2010. [Online]. Available: http://stacks.iop.org/0031-9155/55/i=16/a=001

[11] A. J. Fitzgerald, V. P. Wallace, M. Jimenez-Linan, L. Bobrow, R. J. Pye, A. D. Purushotham, and D. D. Arnone, "Terahertz pulsed imaging of human breast tumors," *Radiology*, vol. 239, no. 2, pp. 533–540, 2006, pMID: 16543586. [Online]. Available: https://doi.org/10.1148/radiol.2392041315

[12] S. C. Henry, L. M. Zurk, and S. Schecklman, "Terahertz spectral imaging using correlation processing," *IEEE Transactions on Terahertz Science and Technology*, vol. 3, no. 4, pp. 486–493, July 2013.

[13] G. Wolberg, "Image morphing: a survey," *The Visual Computer*, vol. 14, no. 8, pp. 360–372, Dec. 1998. [Online]. Available: https://doi.org/10.1007/s003710050148

[14] A. Kaur, L. Kaur, and S. Gupta, "Image recognition using coefficient of correlation and structural similarity index in uncontrolled environment," *International Journal of Computer Applications*, vol. 59, pp. 32–39, Dec. 2012.

[15] S. Karungaru, M. Fukumi, and N. Akamatsu, "Morphing human faces: Automatic control points selection and color transition," in *International Conference on Computational Intelligence*, 2004.

[16] J. Fisher, "Visualizing the connection among convex hull, voronoi diagram and delaunay triangulation," 03 2018.

[17] S.-W. Cheng, T. K. Dey, and J. Shewchuk, *Delaunay Mesh Generation*, 1st ed. Chapman & Hall/CRC, 2012.

[18] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[19] E. Dubrofsky, “Homography estimation,” Master’s thesis, Carleton University, Canada, 2007.

[20] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[21] E. J. Ientilucci, “Using the singular value decomposition,” 2003.

[22] C. Tomasi, “Lecture notes in computational modeling for the sciences,” 2007.

[23] *”System Requirements - Release 2016a”*, ”The MathWorks, Inc.”, ”2016”.

Appendix A

User's Manual

Pathology morphing is a software that provides a quantitative evaluation reference for cancer detection in THz imaging. In particular, the program summarizes the histopathology results of certain tumor sample into a matrix that matches the resolution of its corresponding THz image. Hence, making it possible to quantitatively evaluate the THz results through a pixel-by-pixel comparison.

Considering that the histopathology results of a tumor sample are produced after fixing the sample in paraffin, there is a significant difference in shape between these results and the THz image that was collected while the sample was still fresh. Therefore, this software implements mesh morphing and homography estimation to:

1. Match the resolution of the histopathologic results and the THz image.
2. Reshape the outline of the histopathologic results into the contour of the THz image.
3. Create a reference matrix that contains the location of each region within the tumor.

A.1. System Requirements

The system requirements for this software are described in Table A.1 [23].

Table A.1: System requirements.

MATLAB	64-Bit MATLAB R2016a or higher.
Operating system	Windows 7 Service Pack 1, Windows 8, Windows 8.1, Windows 10, Windows Server 2008 Service Pack 2, Windows Server 2008 R2 Service Pack 1, Windows Server 2012.
Processors	Any Intel or AMD x86-64 processor.
RAM	2GB or higher.

A.2. Getting started

Before starting, you should have the pathology mask and the THz image of the sample you need to evaluate as .png files (to create the pathology mask refer to chapter 3.1). Make sure that the THz image has the exact same resolution as its data.

To represent the THz image as a .png file, use the peak of the time domain reflected signal for FFPE tissue or the integration of the frequency domain reflected signal for fresh tissue (see chapter 2). In addition, you need to manually define the region of interest within the THz image. In this case, it is suitable to represent the area of interest as a matrix **TissueMask**, such that:

$$\textbf{TissueMask}(x_t, y_t) = \begin{cases} 1 & (x_t, y_t) \text{ belongs to the region of interest} \\ 0 & \text{otherwise} \end{cases} \tag{A.1}$$

A.2.1. Pathology Alignment

Use the MATLAB script *Rotation.m* to align the pathology and THz images. This function implements the maximum correlation alignment explained in section 3.2. You can find the description of its inputs and outputs in table A.2.

Table A.2: Arguments description of *Rotation.m*.

	Name	Format	Description
Input	Pathology_image	.png	Pathology mask.
	TissueMask	.mat	Region of interest **TissueMask**.
Output	Pathology_Mask_rotated	.png	Pathology mask aligned with the region of interest.

Note: Make sure the name of the matrix containing the region of interest is TissueMask.

Once the image is aligned and as a result of the rotation, there will be some black areas in the border of the image that can be manually deleted.

A.3. Morphing a Pathology Image

Use the MATLAB script *main.m* to morph the pathology mask into the shape of the THz image. This function implements the mesh morphing algorithm described in chapter 4. You can find the description of its inputs and outputs in table A.3.

Table A.3: Arguments description of *main.m*.

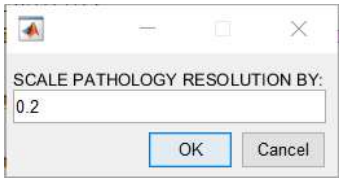
	Name	Format	Description
Input	Pathology_image	.png	Pathology mask.
	THz_image	.png	THz image with region of interest.
Output	Mesh	.mat	MATLAB struct that contains the control points location in the THz and pathology images and their triangulation.
	Morphed _image	.png	Final result of the morphing process.
	Pathology_Reference	.mat	Group of data that contains: 1. matrix.mat: Reference matrix **D**. 2. matrix_key_number.mat: Numbers corresponding to each region. 3. matrix_key_tissue.mat: Regions within the tumor.

A.3.1. Creating a Mesh

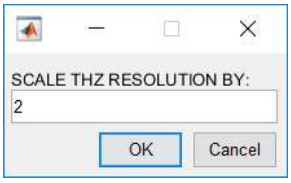
To create a mesh, run the MATLAB script *main.m* and select CREATE MESH in the main menu, as shown in Fig. A.1.

Figure A.1: Main menu in *main.m*.

Load the aligned pathology mask and the THz image in the script and select the scaling factor for each one, as shown in Fig. A.2. The scaling factors are subjective to the resolution of your pathology mask and THz image. As a reference, when the resolution of the pathology mask is approximately 30 times greater than the THz image, the recommended scaling values are 0.2 for pathology and 2 for THz in the first iteration, and 8 for pathology for the rest of iterations. For the last iteration, the scale factor for the THz image should be 1 to make sure that the final result has the same resolution of the THz image.



(a) Scaling factor for pathology mask.



(b) Scaling factor for THz image.

Figure A.2: Scaling factors.

Once the scaling factors have been defined, you can proceed to select the control points one at the time. The command description for this function is described in table A.4.

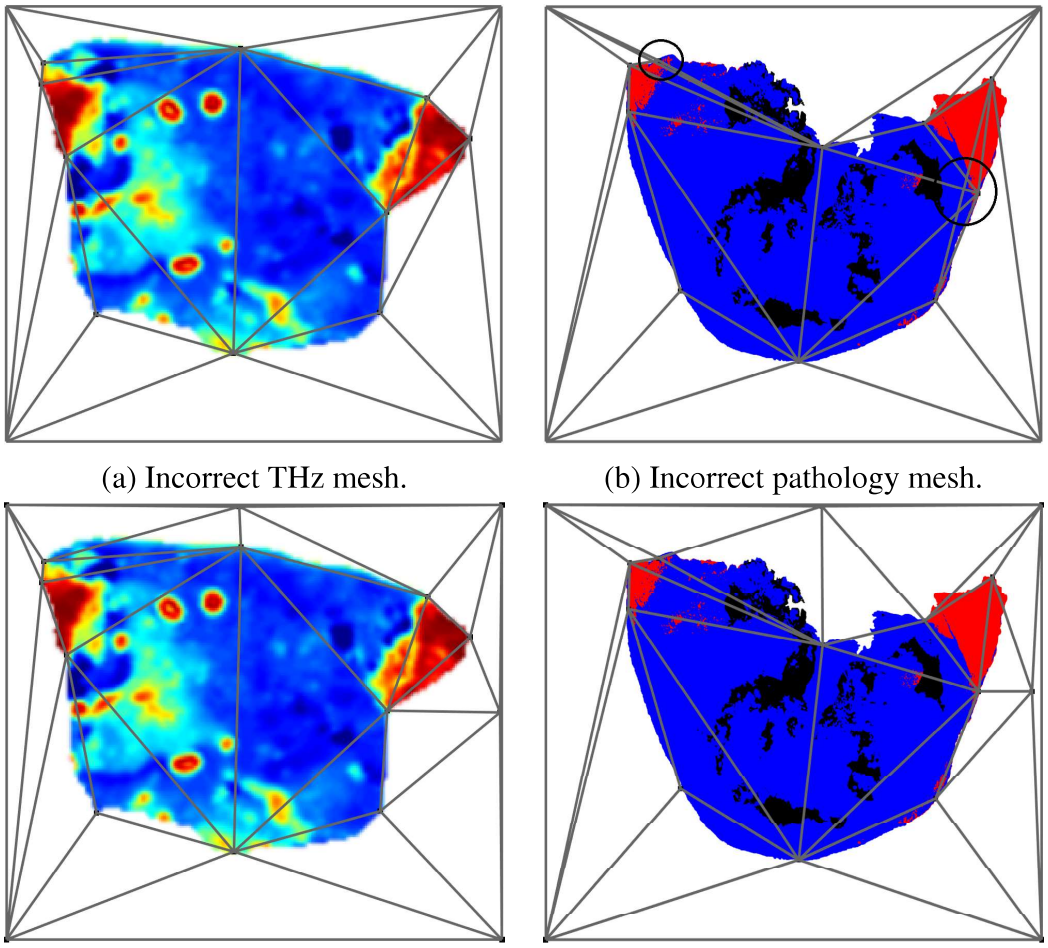
Table A.4: Control point selection.

Command	Action
Left click	Selects a control point.
Right click	Deletes the last control point.
Enter	Finishes the control point selection.

After selecting the control points, you have the option of creating automatic boundaries. These boundaries refer to 4 additional control points at the edges of both images to make sure that the

mesh includes the entire tissue. The creation of the boundaries is optional and its usage should not compromise the topological equivalence between the source and target meshes.

In case the boundaries affect the mesh, you can delete the boundary points or you can create more control points. For example, Figs. A.3a and A.3b show invalid meshes due to the folding between some triangles in the pathology mask. To correct these meshes, you can add two control points to modify the mesh into a valid triangulation, as shown in Figs. A.3c and A.3d.



(a) Incorrect THz mesh.

(b) Incorrect pathology mesh.

(c) Corrected THz mesh.

(d) Corrected pathology mesh.

Figure A.3: Control points allocation.

Once the mesh has been created, you can proceed to save it and morph the pathology image using it. In addition, you can keep the current data as long as the script is running and keep working on the mesh until reaching the desired results.

A.3.2. Modify Mesh

If the mesh presents any problems, e.g. it is not topologically equivalent or there is folding, you can modify it by adding or deleting some control points of the mesh until reaching a valid triangulation. To modify a mesh, select the option MODIFY MESH in the main menu (see Fig. A.1). You can either modify the current mesh in the program or load a previous mesh.

Adding a new control point

The command description for this function is described in table A.5. The script allows the addition of one control point at the time. After adding the point in both images, the script shows the new mesh and you can add new points, save the current mesh or return to the main menu.

Table A.5: Adding a control point.

Command	Action
Left click	Selects a control point.

Deleting an existing control point

The command description for this function is described in table A.6. The script allows the removal of one control point at the time. When deleting a point in one image, the pair point in the other image is deleted as well. After the deletion, the script shows the new mesh and you can delete more points, save the current mesh or return to the main menu.

Table A.6: Deleting a control point.

Command	Action
Left click	Deletes a control point.

A.3.3. Morph Pathology

Once the mesh is valid, you can morph the pathology image based on the control points you selected earlier. To morph the image, select the option MORPH PATHOLOGY in the main menu

(see Fig. A.1). You can either use the current mesh in the script or you can load a previous mesh. Once the morphing process starts, the script presents an indicator of the progress of the code until completion.

As described in section 4.3, it is recommended to follow an iterative process to complete the morphing of the pathology image. For this purpose, after the morphing process of one iteration is completed, you can save the partial results and reuse them for the next iteration until reaching the expected results. At the last iteration, you can create a pathology reference matrix $\mathbf{D}$, as shown in Fig. A.4.

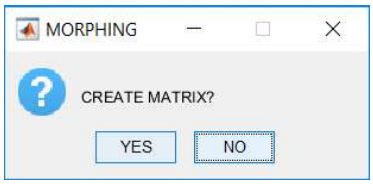


Figure A.4: Reference pathology matrix creation.

To create the reference matrix, follow the instructions of the script and select all the regions within the pathology image including the background. Enter the region name and its number representation, e.g. in Fig. A.5 cancer will be represented as 1 in the matrix.

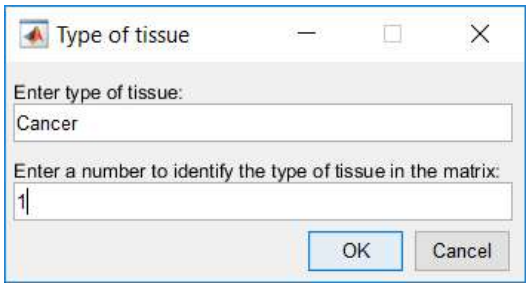


Figure A.5: Matrix key: region representation.

Once the region name and number representation have been established, you can select the color representation of the region in the pathology image, as shown in Fig. A.6, in which blue was selected as the color representation of cancer. The script allows the entry of a finite number of regions until the user indicates there are no more regions to add.

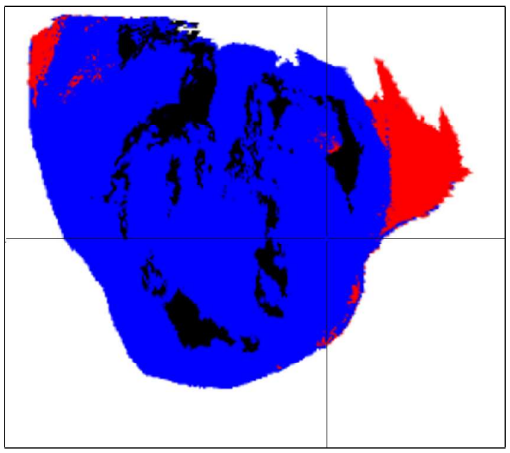


Figure A.6: Region color representation.

Finally, the matrix and the final morphed pathology image are created. The description of the output data is shown in table A.3.