Multimodal registration of optical microscopic and infrared spectroscopic images from different tissue sections: an application to colon cancer

Francisco Peñaranda[a], Valery Naranjo[a,*], Rafael Verdú-Monedero[b],
Gavin R. Lloyd[c], Jayakrupakar Nallala[d], Nicholas Stone[d]

[a]*Instituto de Investigación e Innovación en Bioingeniería (I3B),
Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain*
[b]*Dpto. Tecnologías de la Información y Comunicaciones,
Universidad Politécnica de Cartagena, Cartagena, Spain*
[c]*Biophotonics Research Unit,
Gloucestershire Hospitals NHS Foundation Trust, Gloucester, United Kingdom*
[d]*Biomedical Physics, School of Physics,
University of Exeter, Exeter, United Kingdom*

Abstract

Fourier transform infrared (FTIR) spectroscopic images provide rich information of the biochemical tissue composition that can be analyzed together with other microscopy modalities to perform objective pathological diagnoses. Hematoxylin and Eosin (H&E) stained images are the reference images that pathologists use to make a final diagnosis of most diseases, such as cancer. Therefore, H&E images may be the most interesting imaging modality to be fused with FTIR images. Unfortunately, H&E stain introduces severe confounding artifacts in the FTIR spectra. Thus, in repeatable studies different slices of tissue must be used to acquire images for each imaging modality, which must be aligned so that the different regions of tissue spatially match. The main objective of this manuscript is to establish a complete pipeline where the two types of images (H&E and FTIR) from different tissue sections are aligned or registered. The proposed automatic framework starts by obtaining grayscale images from both FTIR raw data and H&E images where analogous anatomical structures are easily distinguishable. In the first alignment step, a feature-based registra-

---

*Corresponding author
Email addresses:* `frapeago@upv.es` (Francisco Peñaranda), `vnaranjo@upv.es` (Valery Naranjo)

tion produces a fast coarse rigid alignment by using the Scale Invariant Feature Transform (SIFT) algorithm to automatically find and match relevant keypoints in both grayscale images. Due to the spatial variability between samples, different combinations of SIFT parameters are explored and the best combination is selected through the maximization of a similarity measure between the aligned images. In the second alignment step, an intensity-based registration refines the initial alignment and compensates for the local spatial differences between the tissue sections by iteratively estimating a non-rigid transformation. This methodology was used to register 47 colon samples from three different pathological groups (16 normal, 16 intermediate and 15 tumoral) with good overall results, which were quantitatively evaluated for both registration steps. In the first rigid alignment step, the global median of difference in positioning compared to a manual registration was under 1 pixel. In the second registration step, the global median gain in mutual information between the registered images was 0.125 bits. In contrast to existing approaches, the proposed method does not need a prior segmentation step that may introduce errors and reduce the spatial information content, which is crucial when different sections of tissue are used. It can improve the accuracy to combine the spatial information extracted from both the traditional H&E stained images and the emerging FTIR spectroscopy.

1. Introduction

Colorectal cancer is the third most diagnosed cancer worldwide with 1.36 million of new cases detected in 2012 [1]. In this type of cancer, the survival rate is highly dependent on the stage at which the disease is diagnosed. For instance, in the former Anglia Cancer Network the five-year relative survival decreased from 95% at initial Stage I to 7% at metastatic Stage IV for patients diagnosed during 2002-2006 [2].

The final diagnosis of colorectal cancer is performed by expert pathologists, who examine histological sections extracted from suspicious biopsies acquired during colonoscopy. These slices of tissue are commonly stained with Hematoxylin and Eosin (H&E) to color the tissue structures (nuclei in blue or purple, cytoplasm and connective tissue in pink) and observed with an optical microscope. However, this assessment is still done in a qualitative manner, relying heavily on the judgment and experience of the pathologist. Indeed, several studies [3, 4] have revealed a suboptimal inter-observer variability in the differentiation and reporting of colorectal polyps, suggesting that more objective criteria should be applied for risk stratification in screening and surveillance guidelines.

Fourier transform infrared (FTIR) spectroscopy is a technology widely used in industry that has recently shown an increasing capability to improve the diagnosis of different types of cancer [5]. This technique takes advantage of the ability of near- and mid-IR light to excite the vibrational modes of the chemical bonds that form the tissue. The main generated signal consists of an absorption spectrum that informs of the chemical composition of the illuminated tissue, fitting all the characteristics to become a clinically useful biomarker [6]. When combined with array detectors and appropriate optics, a hyperspectral image with spatially resolved chemical information can be acquired [7].

FTIR images are information-rich data structures that can be analyzed, alone or together with other imaging modalities, to provide objective pathological diagnoses. In order to develop new diagnostic algorithms, the different regions of tissue should be correctly labeled and must spatially match between images of distinct modalities. As commented, the H&E stained samples are the gold standard where the pathologists can distinguish and label different anatomical and pathological structures within the tissue. Therefore, H&E images may be the most interesting imaging modality to be fused with FTIR images. However, the H&E stain irreversibly changes the chemical composition of the tissue and introduces confounding artifacts in the FTIR spectra [8]. In some cases, the tissue is measured by FTIR spectroscopy before applying the H&E stain.

3

Nevertheless, this option is technically challenging and it impedes performing posterior FTIR measurements and retrospective studies. Therefore, a common solution is to measure different slices of tissue with each imaging modality.

1.1. Objective

The main objective of this manuscript is to establish a methodology where the two types of images (H&E and FTIR) from different tissue sections are aligned or registered. The aim of this alignment is to compute the spatial transformation that communicates the coordinate systems of the two images. This transformation makes possible to translate or fuse the spatial information contained within each imaging modality. In particular, the position and auxiliary information (e.g. pathological labels) of different regions of interest, identified by an expert pathologist in the H&E images, can be transferred to the FTIR coordinates with the calculated spatial transformation. In doing so, objective diagnostic algorithms may be created in future studies by employing the fused pixels' information from each imaging modality.

The developed registration method must be robust against the different sources of variability intrinsic to the problem (Sec. 1.2). Therefore, another important objective is to assess the robustness and effectiveness of the proposed methodology in a real-world dataset.

1.2. Problem overview

Fig. 1 outlines the problem faced in this work. Different histological sections are extracted with a microtome from a colon biopsy that has been chemically treated and fixed in a paraffin block for preservation purposes. Some sections follow the H&E staining process and one of them is carried to an optical microscope to take RGB images in the visible spectrum range. This color image is composed of three channels (red, green and blue) so that a three-valued intensity vector is associated with each pixel.

Another section is directly analyzed by means of FTIR spectroscopy without any further chemical processing. In the ideal case, this last section would be

4

adjacent to the one used in the optical microscope. However, this ideal case cannot be guaranteed in clinical routine due to problems in the handling and cutting of the biopsy sample. FTIR sensors can measure a large quantity of wavelengths in the near and mid-IR range, providing hyperspectral images that may have hundreds or even thousands of channels. As can be seen in Fig. 1, FTIR images are three-dimensional data structures with two spatial dimensions (x, y) that define the position of each pixel and a third spectral dimension (ν), which consists of the recorded wavenumbers (inverse of wavelengths). Thus, each pixel has an associated absorption spectrum that represents the overall chemical composition of a tissue portion of several squared microns.

The variability between the images of both modalities can be divided into four main sources:

- *Intensity values:* due to the distinct range of wavelengths covered by each image detector. Finding a correct relationship between intensities of corresponding pixels is one of the challenges of this *multimodal* problem.

- *Pixel size:* owing to the characteristics of the optics and the detectors. This can be solved by applying a proper scaling if the exact resolutions in both images are known or by introducing a scaling factor in the spatial transformations.

- *Coarse spatial differences:* because each image device has its own spatial reference. Therefore, large differences in spatial positions and orientation may exist between corresponding anatomical structures.

- *Local spatial differences:* as a consequence of not using the same tissue sections. These dissimilarities may be divided into anatomical differences naturally present in the tissue and local deformations introduced by the physical manipulation, such as cutting with the microtome. If the sections are not adjacent, these local differences are further exaggerated.

The process of finding the alignment between images acquired with different types of sensors and matching their spatial references is called *multimodal re-*
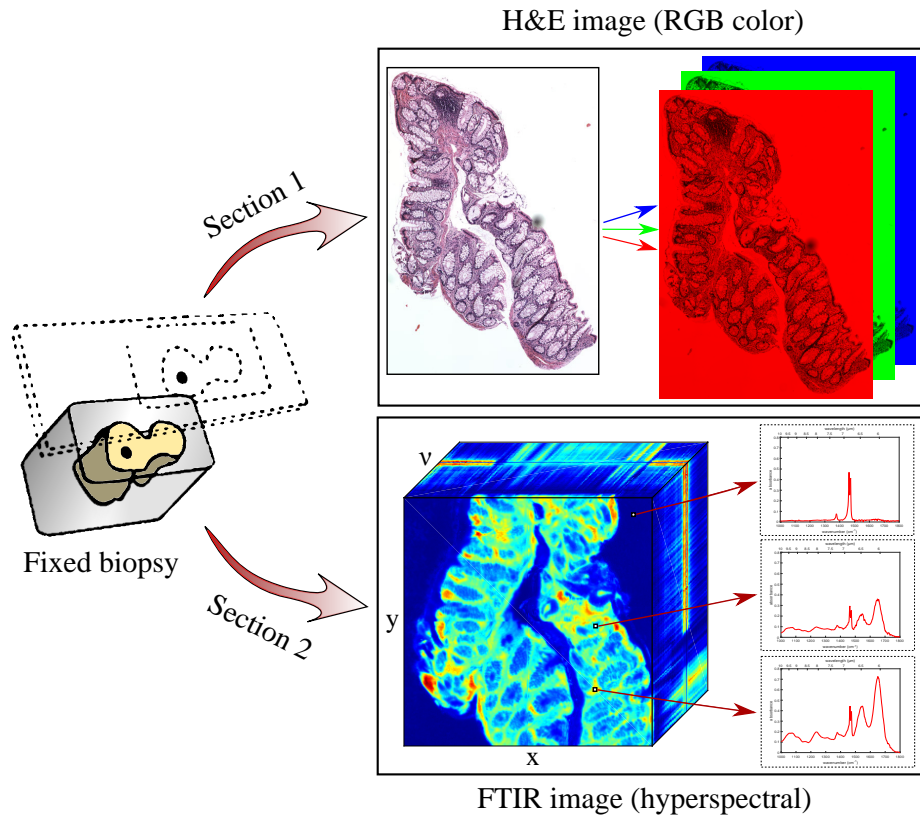
Figure 1: Problem overview. Several histological sections are extracted from a paraffin-fixed colon biopsy. One of the sections is stained with H&E and analyzed with an optical microscope to take a three-channel RGB image. A different section is directly measured with a FTIR spectrometer to obtain a hyperspectral image, where each pixel has an associated spectrum of hundreds or thousands of channels.

gistration. Classical reviews of image registration can be found in [9–12] and a recent comprehensive overview in [13, 14], where the most recent advances in this field are described as well as the techniques applied to medical images. In the problem addressed in this paper, the use of images from different modalities along with the fact that they come from different sections of tissue complicate this task.

1.3. Related work

Many approaches exist to deal with the problem of multimodal registration, mainly because the decisions made in each step of the process depend on the application and on the characteristics of the involved images. To our knowledge, only two recent studies [15, 16] tackle the problem of registering H&E and FTIR images. Nevertheless, in both of them the same histological sections were firstly measured by FTIR spectroscopy and later stained with H&E to be analyzed by optical microscopy. In [15], the images were taken from tissue microarrays (TMAs) of prostate cancer and converted into binary form by trying to separate the pixels that contained tissue from the background. As stated in [15], this binarization bypasses the multimodal problem because the only reliable features to establish a spatial correspondence between these kinds of images are macroscopic sample shapes and empty spaces (glandular lumens and breaks) inside the tissue. A more general problem is treated in [16], where images of small regions of tissue obtained using FTIR or other spectroscopic modalities are registered with whole-slide H&E images by a *template matching* procedure. In that study a pre-segmentation step is performed on each image through *k-means* clustering, taking the FTIR spectrum and the RGB vector as input features for each pixel. The spatial arrangement of the segmented pixels is then matched through an exhaustive sparse search with geometrical restrictions (only translations or little rotations up to ± 30 degrees) due to its computational complexity. However, the pre-segmentation step diminishes the spatial local information and can also introduce severe mistakes because the natural clusters may not reflect the same spatial structures in both images due to the different information contained in the RGB vectors and the FTIR spectra.

A different multimodal registration problem involving microscopic images from adjacent sections of prostatic tissue is addressed in [17]. In that study, several adjacent sections with different grades of cancer are stained with either H&E, immunohistochemical or fluorescence dyes. The spatial correspondence between adjacent sections is obtained by means of a rigid registration based on the features automatically provided by the Scale Invariant Feature Transform

7

(SIFT) [18, 19] and filtered using the Random Sample Consensus (RANSAC) algorithm [20]. This approach is similar to the one used in the first step of the registration framework proposed here (Sec. 2.5).

A previous study was performed to resolve the alignment problem faced here [21]. That work presented the preliminary application of a two-step registration framework similar to the one addressed in this paper to a cohort of 7 samples. However, in that study a less robust algorithm based on Speeded Up Robust Features (SURF) descriptor [22] was used in the first alignment step and thorough evaluation of the results was not performed.

1.4. Proposed framework

As stated in Sec. 1.1, the objective of this study is to establish a complete pipeline where the two types of images of the different sections of tissue are aligned. After setting some parameters, this alignment is performed in an automatic way through two registration steps which analyze different characteristics of the grayscale images obtained from the initial images. The key step of the suggested method is to obtain these grayscale images from the FTIR hyperspectral data cube.

The proposed framework starts by obtaining grayscale images from the two imaging modalities, which are the inputs of the registration method. The main aim of this preliminary step is to get images with similar local contrast where analogous anatomical structures are easily distinguishable.

The registration process consists of two steps. The first step produces a fast coarse alignment that offers good initial conditions for the finer registration performed in the second step. The first registration step is a *feature-based registration* that uses the SIFT algorithm [18, 19] to automatically find and match relevant keypoints in both images. These matches are filtered with the RANSAC algorithm citefischler1981 to estimate a coarse rigid transformation. Several combinations of SIFT parameters were considered due to the wide spatial variability between the studied sections of tissue, which may not be immediately adjacent. The best combination of parameters in each sample was

8

chosen through the maximization of a multimodal similarity measure between the aligned images.

The second registration step consists of an *intensity-based registration* that seeks to refine the alignment and to compensate for the local spatial differences between the tissue sections of the two imaging modalities. In this last stage, a non-rigid transformation is iteratively estimated to maximize again the same multimodal similarity measure. Several parameters must be adjusted in the intensity-based algorithm, but this time a global combination of parameters was empirically fixed to reach a reasonable fine alignment.

The obtained results for the available dataset were qualitatively and quantitatively assessed. The rigid transformation obtained with the first registration step was evaluated by comparison with a *gold standard* rigid transformation that was estimated based on manually selected landmarks. The evaluation of the second non-rigid registration step is more difficult because no *gold standard* can be practically established for comparison. Therefore, the results were assessed by studying the modification of a different multimodal similarity measure before and after this last registration step.

2. Materials and methods

2.1. Dataset

The available dataset consists of 47 colon samples from different pathological groups: 16 normal (non-cancerous), 16 intermediate (comprising adenoma and hyperplastic tissue) and 15 tumoral (cancerous). All samples were fixed and embedded in paraffin blocks. One slice of 7 μm thickness and several contiguous slices of 3 μm were extracted from each block with a microtome. The thicker slice was not further treated and was measured with an FTIR imaging system. This system consisted of an Agilent 620 FTIR microscope coupled to an Agilent 670 FTIR spectrometer with a Globar® light source and a liquid nitrogen cooled Focal Plane Array (FPA) detector of 128×128 pixels. IR light was transmitted through a $15\times$ Cassegrain reflective condenser and objective (NA = 0.62) giving

an effective pixel size of 5.5×5.5 μm^2 and a corresponding field of view (FOV) of 704×704 μm^2. The samples were measured in transmission mode and the absorption spectra were acquired between 1000-3900 cm^{-1} with a wavenumber interval of 4 cm^{-1}. Multiple frames were acquired and combined to allow a larger overall FOV, which varied depending on the size of the sample and the region of interest.

The rest of sections were chemically deparaffinated and stained with H&E. One of these sections, ideally contiguous to the one measured by FTIR spectroscopy, was chosen to acquire RGB images with an optical microscope. A Philip Harris DMSK211 microscope with a coupled digital camera of 1280×1024 pixels was used to record sequential images that covered the full H&E sample. An objective lens of 20× was used, giving a pixel size of 0.4×0.4 μm^2. The movement of the microscope stage was automatized with a custom hardware based on Raspberry Pi. The tiled images were stitched with Microsoft Image Composite Editor (ICE) in order to obtain the final RGB image of the whole sample.

2.2. Registration pipeline

A block diagram with the main steps of the proposed registration pipeline is shown in Fig. 2. In order to illustrate the outputs of each step, a representative image from a specific sample is shown next to each block.

The inputs of the process are the two raw data structures of each imaging modality (RGB and FTIR), which were described in Sec. 1.2. The first stage of the framework consists of obtaining *representative grayscale images* from the two raw inputs. This is the key step in the pipeline and enables us to apply registration methods that have proven to be effective in other medical areas that operate with multimodal grayscale images [13, 14, 23, 24]. The aim is to obtain two images with similar local contrast where analogous anatomical structures can be easily observed. These two grayscale images, named T_1 (RGB) and R (FTIR), are the inputs of the next step.

A *feature-based registration* with a rigid spatial transformation is performed
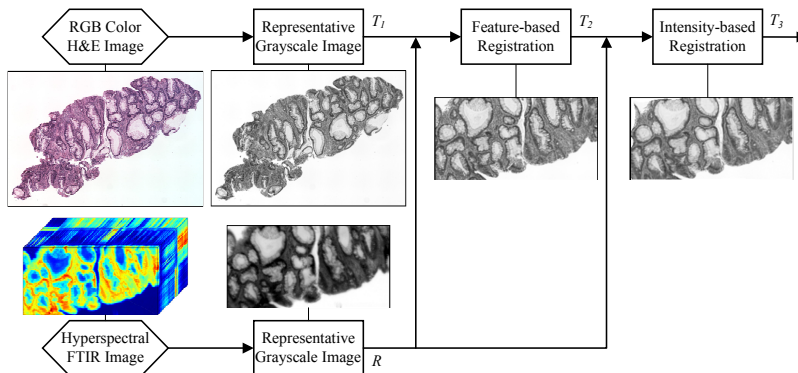
10

Figure 2: Block diagram of the proposed registration framework. The outputs of each step are illustrated with a representative image from a single sample.

in the second level of the pipeline. Its main objectives are to compensate for the coarse spatial differences between images and to provide an initial alignment based on the correspondence of keypoints, which are automatically detected and matched in both images using SIFT and RANSAC as described in Sec. 2.5. In this initial alignment, the representative FTIR grayscale image R acts as the *reference* image and the H&E grayscale image T_1 is the *target* image that is shifted through a rigid transformation to produce image T_2.

In the third and final step of the framework, an *intensity-based registration* is applied to perform a non-rigid spatial transformation. It takes as inputs the unaltered *reference* image R and the *target* image T_2 to produce the final output image T_3. The aim of this step is twofold: on the one hand, to refine any remaining misalignment and on the other hand, to diminish local spatial differences due to the use of different slices of tissue.

2.3. Representative images

The critical step in this work is converting the data structures of each modality into grayscale images such that registration can be performed in the grayscale domain. The goal is to create two images with similar local contrast where corresponding anatomical structures can be easily observed. The process to obtain

11

these grayscale images is different for each imaging modality.

2.3.1. H&E grayscale image

Two steps were applied to construct the H&E grayscale image:

1. *Resizing:* A bi-cubic interpolation with an antialiasing filter was performed to downscale the original H&E color images to the same pixel size as FTIR images.

2. *Transformation to grayscale:* The RGB values of the resized images were converted to grayscale by computing the luma $\mathcal{Y}$, which is the achromatic component representing luminance in the YIQ color space [25]. In the studied dataset, luma component demonstrated to have a high rate of convergence to an optimal solution, especially in the first registration step, as described in Sec. 3.2. In addition, it demonstrated to be more robust against different illumination conditions (some samples had uneven spatial illumination) and staining variability (e.g. cancerous regions stain darker than normal ones) than other chromatic components. In a theoretical way, human visual system is more sensitive to luminance differences rather than chromatic differences. The luminance is closely related to the perceptual attribute called *brightness*, which is the visual sensation according to which a source appears to emit more light or less than another does [25]. On the other hand, absorbance values of the FTIR images may be interpreted as the capability of an object to absorb more light or less than another does. Therefore, theoretically luma component may be a good candidate to be inversely related to a transformation that condenses the absorbance values of the FTIR images (Sec. 2.3.2). As commented before, this theoretical intuition was empirically supported by the results obtained and compared with a manual gold standard registration. The specific conversion to obtain $\mathcal{Y}$ is a weighted combination of the three-valued vector $(\mathcal{R}, \mathcal{G}, \mathcal{B})$:

$$\mathcal{Y} = 0.299 \cdot \mathcal{R} + 0.587 \cdot \mathcal{G} + 0.114 \cdot \mathcal{B} \tag{1}$$

2.3.2. FTIR grayscale image

Before transforming the FTIR hyperspectral data into a single grayscale image, the spectra must be pre-processed. Pre-processing is the most relevant stage in FTIR data analysis and interpretation [26, 27]. It is essential to remove the unwanted effects involved in the spectral data acquisition and to highlight specific information within the spectra. No universal pre-processing method exists [28]; the specific combination of pre-processing steps depends on the application.

In order to illustrate the applied pre-processing, the final representative FTIR grayscale image of one sample is shown in Fig. 3a. In this image five pixels have been coarsely marked with different colors and numbers. Pixels 1 and 2 exemplify two regions without tissue (substrate), pixel 3 corresponds to a region where there is a low amount of sample (interior of lumen), pixel 4 represents a region where the presence of organic material is higher than in pixel 3 (outer region of lumen) and, finally, pixel 5 illustrates a region with the highest content of tissue of the selected pixels. The same colors and numbers are used to represent the corresponding raw spectra in Fig. 3b. Strong peaks due to contaminants such as CO_2 ($\sim$2300-2400 cm^{-1}) and paraffin ($\sim$2800-3000 cm^{-1}) can be easily observed in the raw spectra because they are even present in the substrate regions (pixels 1 and 2).

The pre-processing steps applied in these images, with the reasoning behind their use, were:

1. *Spectral window selection:* The raw absorption spectra were cropped to the so-called *fingerprint region* (1000-1800 cm^{-1}), which contains the vibrational frequencies of the chemical bonds from the most relevant biomolecules [6]. Thus, attention is focused on the most informative wavenumbers and the stronger contaminant bands are rejected (Fig. 3c).

2. *Removal of remaining spectral contaminants:* Relatively strong peaks due to paraffin still appear in the *fingerprint region* (1360-1390 and 1430-1490 cm^{-1}). The absorbance values corresponding to those ranges of
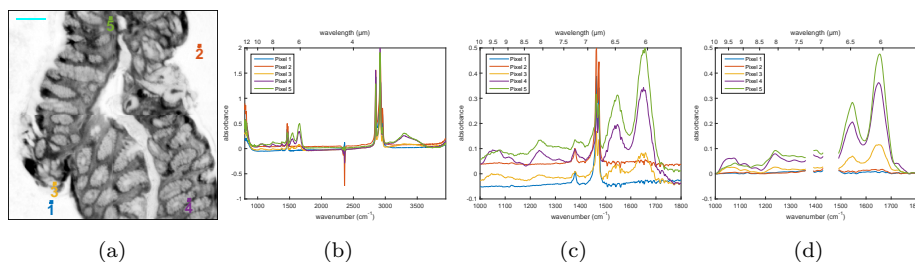
13

Figure 3: (a) Representative FTIR grayscale image of a colon sample where five pixels have been marked and numbered with different colors. Cyan scale bar represents 200 μm. (b) Full-range raw spectra drawn in the same color of their corresponding pixels. (c) Raw spectra cropped to *fingerprint region*. (d) Pre-processed spectra after cropping to the fingerprint region (1000-1800 cm^{-1}), smoothing by Savitzky-Golay filtering, rejecting the paraffin ranges (1360-1390 and 1430-1490 cm^{-1}) and applying rubberband baseline correction.

wavenumbers are removed from the spectra to reduce the noise produced in the final images by the interference of these contaminants and the rest of meaningful biochemical peaks.

3. *Denoising:* A Savitzky-Golay filter was applied to smooth the spectra and reduce the random noise. Its parameters were fixed to a window size of 15 points and a 2nd order fitting polynomial, which were a good tradeoff between noise attenuation and signal distortion in the studied dataset.

4. *Baseline correction:* It is crucial to diminish the spectral baseline effects due to scattering, heterogeneity in external illumination, supporting substrate or sensor's sensitivity and other changing conditions during data collection [27]. There is a large variety of baseline correction algorithms. Here the *rubberband baseline correction* method was employed [26], which subtracts a convex polygonal line whose edges are minima within the spectrum. In particular, the parameterless implementation of the *rubberband baseline correction* included in the IRootLab toolbox [29] was used.

The final pre-processed spectra of the selected pixels are shown in Fig. 3d. The selected pixels have been numbered in increasing order of maximum absorbance values. Absorbance is related to the concentration and thickness of the

14

biological content as stated by the famous Beer-Lambert law. There are some metrics commonly used in vibrational spectroscopy to condense the absorbance measurements, such as peak height or peak integral [30]. In this study, different combinations of these metrics were explored, such as finding the maximum or computing the integral of the complete *fingerprint region* or only the strongest peak (Amide I peak, between 1630-1670 cm^{-1}). However, it was found that the standard deviation (*std*) of each spectrum was the most robust metric (in terms of noise, sharpness and independence of previous pre-processing steps) to produce an image with a local contrast similar to the corresponding H&E grayscale image. In particular, the intrinsic removal of the mean absorbance value performed by *std* reduces the influence of spectral baseline artifacts (e.g. produced by scattering) inefficiently removed in the baseline correction pre-processing step. More specifically, if $\mathbf{A} = \{A_1, \ldots, A_{N_w}\}$ is the associated absorption spectrum of a pixel with N_w wavenumbers, then the assigned value to that pixel was computed as:

$$std = \sqrt{\frac{1}{N_w - 1} \sum_{i=1}^{N_w} \left(A_i - \overline{A}\right)^2}, \tag{2}$$

where $\overline{A} = \frac{1}{N_w} \sum_{i=1}^{N_w} A_i$ is the mean absorbance value of $\mathbf{A}$. The obtained *std* values were linearly mapped to get a grayscale image with intensities between 0 and 255. However, this gives the substrate low gray values, as opposed to the H&E grayscale image where the substrate has high values and appears white. For this reason the gray values were inverted to obtain the reference image R in the proposed registration method.

2.4. Similarity measures

Images from different imaging modalities have different intensity characteristics due to the specific properties of each sensor. Thus, multimodal registration problems require similarity measures based on statistical relationships between the pixels of the images rather than direct intensity relationships as in

monomodal scenarios [31]. Two similarity measures widely used in multimodal problems were employed: the correlation ratio and the mutual information.

The correlation ratio (CR) assumes a functional relationship between the intensities of the registered images [32, 33]. CR does not consider any specific relationship, whereas the correlation coefficient does. The correlation coefficient assumes a linear relationship between intensities, making it more appropriate for monomodal registration problems. CR can take values from 0 (no functional dependence) to 1 (purely deterministic dependence) [33]. The closer CR is to 1, the more similar R and T are and, in the extreme case ($CR = 1$), it would inform of a hypothetical perfect alignment. To compute CR, the reference image R and the target image T are viewed as random variables. Let $\mathbf{x}$ denote a pixel that has an intensity $T(\mathbf{x})$ in image T; Ω the set of pixels in the overlapping region between R and T; N the total number of pixels in Ω; Ω_i the subset of Ω whose pixels have the same intensity level i in R, that is, $\Omega_i = \{\mathbf{x} \in \Omega, R(\mathbf{x}) = i\}$; N_i the number of pixels in Ω_i; then, CR is defined as:

$$\mathrm{CR} = 1 - \frac{1}{N\sigma^2} \sum_i N_i \sigma_i^2, \tag{3}$$

where

$$\sigma^2 = \frac{1}{N} \sum_{\mathbf{x} \in \Omega} T(\mathbf{x})^2 - m^2, \quad m = \frac{1}{N} \sum_{\mathbf{x} \in \Omega} T(\mathbf{x}),$$
$$\sigma_i^2 = \frac{1}{N_i} \sum_{\mathbf{x} \in \Omega_i} T(\mathbf{x})^2 - m_i^2, \quad m_i = \frac{1}{N_i} \sum_{\mathbf{x} \in \Omega_i} T(\mathbf{x}).$$

Mutual information (MI) is a measure from information theory that characterizes the amount of shared information between the registered images [34]. It assumes a probabilistic relationship between the intensities of the registered images R and T. MI can range from 0 (statistical independence) to any positive quantity and, as in CR, the higher MI, the more similar R and T are. However, interpretation of MI is difficult because there is not a maximal value that can be used as a reference for a hypothetical *perfect* alignment. MI is computed in

16

terms of image entropies:

$$\text{MI} = H_R + H_T - H_{RT}. \tag{4}$$

The three involved entropies are defined as:

$$H_R = -\sum_i P_i \log(P_i),$$

$$H_T = -\sum_j P_j \log(P_j),$$

$$H_{RT} = -\sum_{i,j} P_{ij} \log(P_{ij}),$$

where P_i is the probability of an intensity level i occurring in image R; P_j is the probability of intensity level j occurring in image T; P_{ij} is the joint probability of both intensity levels i in image R and j in image T occurring at the same position.

2.5. Feature-based registration

Scale Invariant Feature Transform (SIFT) is an algorithm to detect and describe local features in images [18, 19]. The main characteristic of SIFT is its ability to find distinctive keypoints that are invariant to location, scale and rotation, and robust to affine transformations (changes in scale, rotation, shear, and position) and changes in illumination, which makes it usable for object/pattern recognition. SIFT is the core algorithm employed for keypoint selection and feature extraction. In the SIFT algorithm, a series of keypoints that are invariant to scale and orientation are firstly detected by seeking extrema in a difference-of-Gaussian (DoG) transformation. At each candidate keypoint, a local descriptor relative to scale-invariant coordinates is computed based on local image gradients. The feature descriptors of the candidate keypoints in both images are matched by a nearest-neighbor strategy through minimum Euclidean distance.

The values of four relevant SIFT parameters have been explored to get different rigid transformations for each sample:

- *Number of orientations (r) and width of the descriptor (n):* these two parameters change the size (rn^2) of the local descriptor vector, which is computed in a $n \times n$ array of histograms of r orientations. When these values increase, the descriptor complexity and discriminative properties grow. The highest rates of convergence were achieved with values of $r = \{4, 8\}$ and $n = \{4, 6\}$.

- *Number of scale samples per octave (s):* this parameter modifies the number of extrema detected in the DoG transformation. The higher its value, the more candidate keypoints are detected. Although Lowe [19] experimentally found an optimal value of 3, more robust results (higher rates of convergence) were obtained with values of $s = \{6, 10\}$ in this dataset.

- *Ratio of distances closest/next closest (rod):* each candidate keypoint is only retained if the ratio of distances between the first and the second nearest matched neighbor is below rod. When this threshold decreases, a higher number of false matches is discarded although good matches can also be rejected. The highest rates of convergence were obtained with $rod = \{0.8, 0.9\}$.

Afterwards, the Random Sample Consensus (RANSAC) algorithm [20] is used to filter the candidate keypoints matched by means of the SIFT descriptors. In RANSAC, a spatial transformation model between both images must be imposed in order to estimate its parameters. In this case, a rigid transformation $\mathcal{T}$ is considered, which follows the equation:

$$\mathcal{T}(\mathbf{x}) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}, \tag{5}$$

where the original spatial coordinates $\mathbf{x} = (x_1, x_2)$ are converted into the new ones by applying a global translation (t_1, t_2) and a rotation of angle θ.

RANSAC computes the parameters t_1, t_2 and θ by considering a minimum percentage of inlier matched keypoints (5% in this case) and a maximal alignment error between the transformed keypoints and their corresponding matches.

For this work we limited the alignment error to a maximum of 15 pixels.

Due to the spatial variability between samples, a rigid transformation may not be constructed for some combinations of parameters because no inlier correspondences are detected. The selection of the best combination of parameters for each sample was therefore based on the maximization of a similarity measure (Sec. 2.4) between the output target image T_2 and the reference image R. In particular, the correlation ratio (CR) was the variable to maximize.

2.6. Intensity-based registration

The second registration step is based on a variational approach which has been formulated in the frequency domain [35, 36], and also implemented in the frequency domain [37] providing a fast and efficient registration method. This method produces a non-rigid displacement field $\mathbf{u}$ that compensates for the remaining differences between the reference R and the target T_2 images caused by an imperfect alignment in the first registration step and the local spatial differences between tissues. The obtained non-rigid displacement field $\mathbf{u} : \mathbb{R}^2 \to \mathbb{R}^2$ will make the transformed target image similar to the reference image, $T_2(\mathbf{x} - \mathbf{u}(\mathbf{x})) \approx R(\mathbf{x})$, where $\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}))^\top$ and $\mathbf{x}$ is the spatial position $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$.

The non-parametric registration can be approached in terms of calculus of variations by defining the following energy functional to be minimized:

$$\mathcal{J}[\mathbf{u}] = \mathcal{D}[R, T_2; \mathbf{u}] + \alpha \mathcal{S}[\mathbf{u}]. \tag{6}$$

The energy term $\mathcal{D}$ measures the distance between the deformed target and the reference images; $\mathcal{S}$ is a penalty term which acts as a regularizer and determines the smoothness of the displacement field; and $\alpha > 0$ weighs the influence of the regularization.

The distance measure D is chosen depending on the datasets to be registered. In this application, since the H&E and FTIR datasets do not share the same intensity range (*multimodal* problem), statistical-based measures are more appropriate. Among the similarity measures described in Sec. 2.4, the correlation

ratio (CR) was used in this case because it provided more accurate alignments and had a better behavior against wrong global minima during its optimization. The regularization term $\mathcal{S}$ gives the smoothness characteristics to the displacement field [38]. In this problem, we used the diffusion term, which is given by the energy of the first derivatives of $\mathbf{u}$ [35].

As described in [35], the energy functional (Eq. (6)) can be translated into the frequency domain by means of Parseval's theorem. Then $\mathcal{J}[\mathbf{u}] = \tilde{\mathcal{J}}[\tilde{\mathbf{u}}]$, with $\tilde{\mathbf{u}}(\boldsymbol{\omega}) = (\tilde{u}_1(\boldsymbol{\omega}), \tilde{u}_2(\boldsymbol{\omega}))^\top$ being the frequency counterpart of the displacement field, $\boldsymbol{\omega} = (\omega_1, \omega_2)$ is the two-dimensional variable in the frequency domain. The minimization of the energy functional $\tilde{\mathcal{J}}[\tilde{\mathbf{u}}]$ leads to the translation of the Euler-Lagrange equation into the frequency domain and provides the following iteration for the Fourier transform of the l-th component of the displacement field:

$$\tilde{u}_l^{(\xi)}(\boldsymbol{\omega}) = H(\boldsymbol{\omega})\big(\tilde{u}_l^{(\xi-1)}(\boldsymbol{\omega}) - \alpha\, \tilde{f}_l^{(\xi-1)}(\boldsymbol{\omega})\big), \tag{7}$$

where $\xi \in \mathbb{N}$ is the iteration index, $l = \{1, 2\}$ in this 2D problem, $H(\boldsymbol{\omega})$ is a low pass filter in the frequency domain and $\tilde{\mathbf{f}}(\boldsymbol{\omega}) = (\tilde{f}_1(\boldsymbol{\omega}), \tilde{f}_2(\boldsymbol{\omega}))^\top$ is the Fourier transform of the external forces field. For further details, please refer to [35, 36].

From Eq. (7), the target image is iteratively modified by $\mathbf{u}$ and goes through different intermediate states T_i until the final image $T_3(\mathbf{x}) = T_2(\mathbf{x} - \mathbf{u}(\mathbf{x}))$ is obtained. Eq. (7) provides a stable implementation for the computation of a numerical solution for the displacement field in a more efficient way than existing approaches if the two-dimensional fast Fourier transform is used [36].

The values of the most relevant parameters in this step were: a maximum number of iterations (ξ_{max}) equal to 400 and $\alpha = 500$. The images R and T_2 were pre-processed with a 3×3 and a 5×5 median filter, respectively, before applying this last registration step to avoid excessive local distortions that may be induced by noise artifacts.
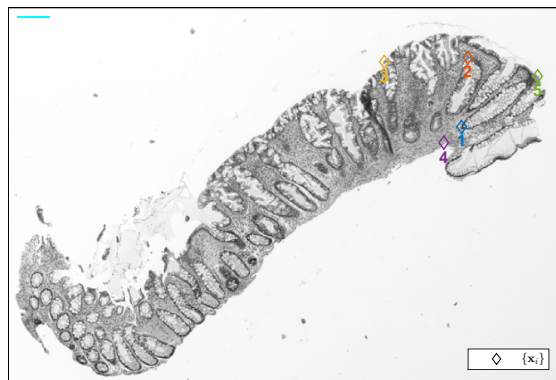
2.7. Evaluation

No *ground truth* for alignment evaluation exists in this problem mainly for two reasons: placing external fiducial marks in the microscopic tissue is not trivial; the deformations produced during the processing of the different slices of tissue can be highly variable. Even so, special care was taken to evaluate the obtained results in both registration stages.

The rigid registration step (Sec. 2.5) was assessed by a procedure similar to the one used in [17]: an artificial rigid transformation was computed in each sample as a *gold standard* for comparison. This *gold standard* transformation was estimated by applying Procrustes analysis [39–41] to a set of manually selected landmarks or control points. These sets of landmarks were manually chosen by trying to select anatomical structures shared by both images to register (R and T_1), such as distinctive points in the glands or in the tissue border. Formally, a set of N_p points $\{\mathbf{x}_i\}$ were selected in the original target image T_1 and their corresponding set of points $\{\mathbf{y}_i\}$ were marked in the reference image R, where $i = 1, \ldots, N_p$. For each sample, five corresponding points ($N_p = 5$) were selected in both images. As an illustration, Fig. 4 shows the images T_1 and R of one tissue sample where the two corresponding sets of points $\{\mathbf{x}_i\}$ and $\{\mathbf{y}_i\}$ are marked with different colored diamonds.

The set $\{\mathbf{x}_i\}$ can be mapped into the reference space of image R with a generic rigid transformation (Eq. (5)) to give a set of transformed points $\{\mathcal{T}(\mathbf{x}_i)\}$. The root-mean-square of the distances between the points of a generic transformed set $\{\mathcal{T}(\mathbf{x}_i)\}$ and their corresponding $\{\mathbf{y}_i\}$ points is the Fiducial Registration Error (FRE) [41]:

$$\text{FRE} = \sqrt{\frac{1}{N_p} \sum_{i=1}^{N_p} \left(\mathcal{T}(\mathbf{x}_i) - \mathbf{y}_i\right)^2}, \tag{8}$$

Two FRE values can be computed for each sample by applying each rigid transformation: one for the *gold standard* transformation computed with Procrustes analysis (FRE_{PA}) and one for the transformation obtained with the proposed feature-based registration (FRE_{FB}). These values respectively inform

21

(a) T_1
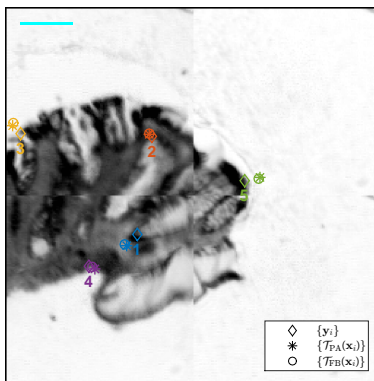
(b) R

Figure 4: Illustration of the sets of manual landmarks used in the evaluation of the feature-based registration step. Cyan scale bars represent 200 μm. (a) Target image T_1 with five manually selected points $\{\mathbf{x}_i\}$ marked with diamonds. (b) Reference image R with the corresponding set of manually selected points $\{\mathbf{y}_i\}$ (diamonds), the transformed set $\{\mathcal{T}_{\mathrm{PA}}(\mathbf{x}_i)\}$ (asterisks) obtained by Procrustes analysis and the transformed set $\{\mathcal{T}_{\mathrm{FB}}(\mathbf{x}_i)\}$ (circles) obtained by the feature-based registration step.

of the mean alignment error in pixel units of the transformed sets $\{\mathcal{T}_{\mathrm{PA}}(\mathbf{x}_i)\}$ and $\{\mathcal{T}_{\mathrm{FB}}(\mathbf{x}_i)\}$. Procrustes analysis [39] computes the optimal linear transformation (rigid in this case) by least squares minimization of the distances between the cloud of landmarks selected in the reference image R ($\{\mathbf{y}_i\}$) and their corresponding transformed landmarks from the target image T_1 ($\{\mathcal{T}(\mathbf{x}_i)\}$), i.e., by minimizing FRE [40, 41]. Thus, $\mathrm{FRE}_{\mathrm{PA}}$ represents the lowest possible error (in

22

terms of minimum least squares) for the manually selected sets of landmarks, what justifies the use of Procrustes analysis as the *gold standard* rigid transformation. The difference error ($\Delta\mathrm{FRE} = \mathrm{FRE_{FB}} - \mathrm{FRE_{PA}}$) was also computed for each sample to show explicitly the disagreement between both rigid registration methods. The transformed sets $\{\mathcal{T}_\mathrm{PA}(\mathbf{x}_i)\}$ and $\{\mathcal{T}_\mathrm{FB}(\mathbf{x}_i)\}$ for one tissue sample have been represented in Fig. 4b with asterisks and circles, respectively. As can be seen, the agreement between the two rigid registrations in the chosen sample is very high.

The evaluation of the non-rigid transformation performed in the intensity-based registration step (Sec. 2.6) is not an easy task and no well-established validation methods exist [11, 31]. The comparison with a manual *gold standard* transformation based on selected landmarks is not appropriate due to the local deformations produced by the non-rigid transformation. These spatially uneven deformations may lead to a deceptive estimation of the errors in the positioning of the manually selected landmarks, which are actually very difficult to pick due to the use of different slices of tissue. A usual validation method for non-rigid intensity-based approaches is the computation of a similarity measure (Sec. 2.4) between the target and the reference images before and after applying the non-rigid registration [31]. In order to avoid false conclusions, the similarity measure used for this evaluation must be different from the one (CR) maximized during the registration. In this case, MI was computed for each sample and its improvement was statistically analyzed by pathological group and over the whole dataset. In each group, the right-tailed Wilcoxon signed-rank test [42] was applied, whose null hypothesis states that the differences before and after applying the intensity-based registration step have zero median, meanwhile the alternate hypothesis states that the median of these differences is positive. A p-value was obtained for each group to assess the statistical significance of rejecting the null hypothesis in favor of the alternate hypothesis of this non-parametric test.

Finally, careful visual assessment remains the first and most important validation check [11, 31] especially in this multimodal registration problem which involves non-rigid deformations. The best visual results were obtained when

23

CR was considered as the similarity measure to maximize through the complete registration pipeline. It must be remarked that the subjectivity in this last assessment is unavoidable due to the high variability introduced by the non-rigid distortions.

2.8. Implementation

All the steps that compose the suggested registration framework, including its evaluation, were implemented in MATLAB. As external programs, the IRootLab toolbox [29] was used to perform the pre-processing of FTIR images (Sec. 2.3) and the implementation of SIFT and RANSAC algorithms included in the software Fiji/ImageJ [43, 44] were linked with our in-house MATLAB algorithms through the MIJ Java package [45].

The implemented code was applied to the available dataset (Sec. 2.1) in a computer with a processor Intel® Core™ i7-4790K @ 4.00GHz and 16GB of RAM memory, running under Windows 8.1 and with the version R2016b of MATLAB installed. Fig. 5 shows the running times in this computer platform versus the size of the reference image R, in terms of total number of pixels. As described in Sec. 2.1, the FTIR reference image R is composed of different frames of 128×128 pixels which cover the corresponding region of interest in each sample. As can be seen in Fig. 5a, the computation time for the complete proposed framework varies from around 40 to 360 seconds, with an increasing but not clear relationship with the size of the reference image R. However, as presented in Fig. 5b the running time in the intensity-based registration step does follow a linear relationship with the size of R with an estimated slope by least-squares (coefficient of determination $R^2 = 0.99$) of 1.6 miliseconds/pixel. The variability in the total running time comes from the feature-based registration step, whose computation not only depends on the size of the images to be registered but also on their content. Images with higher number of relevant structures increases the detection of candidate keypoints and the time for matching and filtering them with SIFT and RANSAC algorithms (Sec. 2.5).

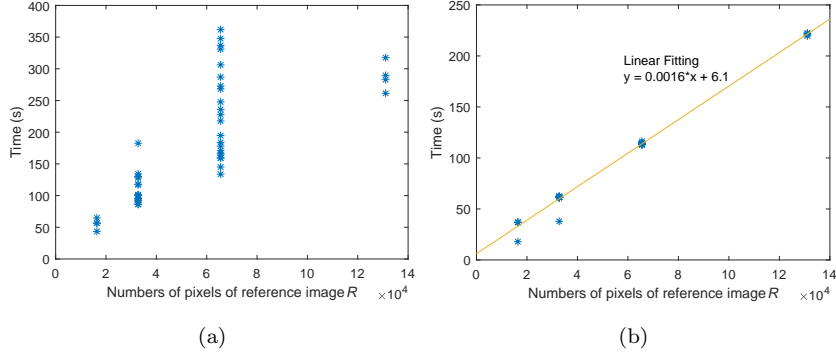Tab. 1 summarizes the most relevant parameters of both registration steps,

24

Figure 5: Running times (in the computer platform described in the main text) vs. number of pixels of reference image R using the samples of the available dataset for (a) the complete proposed framework and (b) the intensity-based registration step.

including the applied values in the available dataset and the risks of using too low or too high values. The relevance of SIFT parameters in the feature-based registration step were described in Sec. 2.5. Choosing incorrect values of these parameters would create incorrect or insufficient matches between keypoints, leading to the computation of wrong rigid transformations or, even worse, the failure of finding any transformation. In the intensity-based registration step, the main parameter that can be tuned is α, which controls the smoothness of the applied deformation. The higher α, the smoother the deformation, at the risk of not compensating sufficiently for the local spatial differences. On the contrary, the lower α, the more unrealistic excessive deformations may be applied. Finally, ξ_{max} is mainly related to the chosen α. In this case, a relatively high value of ξ_{max} was selected to be conservative.

As last remarks, the suggested parameter values should be a good starting point for other datasets with different sources of variability between images. In the feature-based registration step, the indicated methodology of exploring different combinations of parameters and choosing the one that maximizes a similarity measure between images (CR in this case) demonstrated to be able to handle satisfactorily different scenarios. In the intensity-based registration step, the results did not change significantly up to a double or a half of the applied

25

Table 1: Summary of the most relevant parameters of each registration step.

Registration step	Symbol	Description	Applied values	Risk of low values	Risk of high values
Feature-based	r	Number of orientations	$\{4, 8\}$	Descriptor with lacking information	Descriptor with excessive noise
	n	Width of descriptor	$\{4, 6\}$		
	s	Number of scale samples per octave	$\{6, 10\}$	Not enough keypoints	Many confounding keypoints
	rod	Ratio of distances closest/next closest	$\{0.8, 0.9\}$	Right matches discarded	Wrong matches retained
Intensity-based	α	Weight of regularization	500	Excessive local deformation	Insufficient local deformation
	ξ_{max}	Maximum number of iterations	400	Optimal solution not reached	Excessive computation time

values.

3. Results

3.1. Qualitative results

One sample per pathological group has been chosen to present the most relevant images obtained in the different steps of the proposed registration framework. Fig. 6 shows the images of the chosen intermediate sample, Fig. 7 the normal sample and Fig. 8 the tumoral sample. In these figures, the automatically-detected inlier keypoints of the feature-based registration step (Sec. 2.5) have been marked in the original target T_1 and reference R images. The output target images of the feature-based T_2 and the intensity-based T_3 registrations are also shown. These output images have been independently overlaid with the reference image R to create composite RGB images whose gray levels denote regions with similar intensities in the two overlaid images, whereas magenta and green regions correspond to different intensities. Although the presence of magenta or green in some cases does not mean a bad alignment (due to the multimodality of the overlaid images), these representations allow a better visualization of the alignment produced by each registration step. Finally, and also for better visualization of the applied deformation, the figures show synthetic grids with the same dimensions than T_2 and T_3, which were deformed with the non-rigid displacement fields **u** computed in the intensity-based registration step. Each

(a) T_1

(b) R

(c) T_2

(d) T_3

(e) R and T_2 composite

(f) R and T_3 composite

(g) Deformation
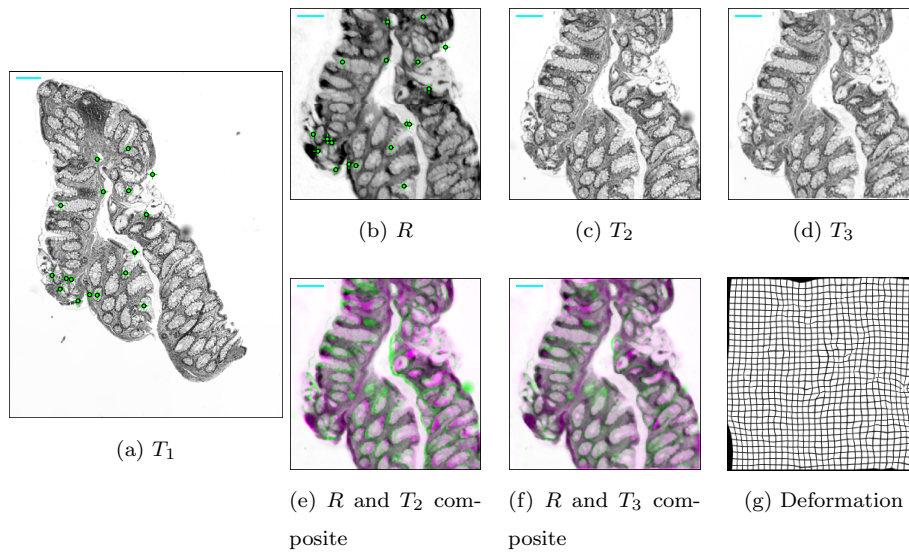
Figure 6: Registration results for an intermediate sample. Cyan scale bars represent 200 μm. (a) Target image T_1 with automatically-detected inlier keypoints. (b) Reference image R with automatically-detected inlier keypoints. (c) Feature-based registration output image T_2. (d) Intensity-based registration output image T_3. (e) Composite RGB image of overlaid images R and T_2. (f) Composite RGB image of overlaid images R and T_3. (g) Artificial grid deformed with the non-rigid displacement field computed in the intensity-based registration step.
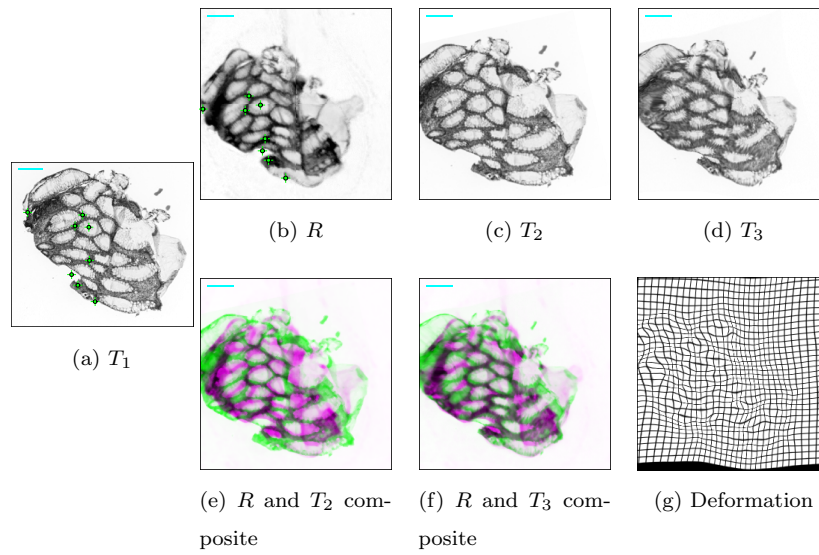
(a) T_1

(b) R

(c) T_2

(d) T_3

(e) R and T_2 composite

(f) R and T_3 composite

(g) Deformation

Figure 7: Registration results for a normal sample. Cyan scale bars represent 200 μm. (a) Target image T_1 with automatically-detected inlier keypoints. (b) Reference image R with automatically-detected inlier keypoints. (c) Feature-based registration output image T_2. (d) Intensity-based registration output image T_3. (e) Composite RGB image of overlaid images R and T_2. (f) Composite RGB image of overlaid images R and T_3. (g) Artificial grid deformed with the non-rigid displacement field computed in the intensity-based registration step.
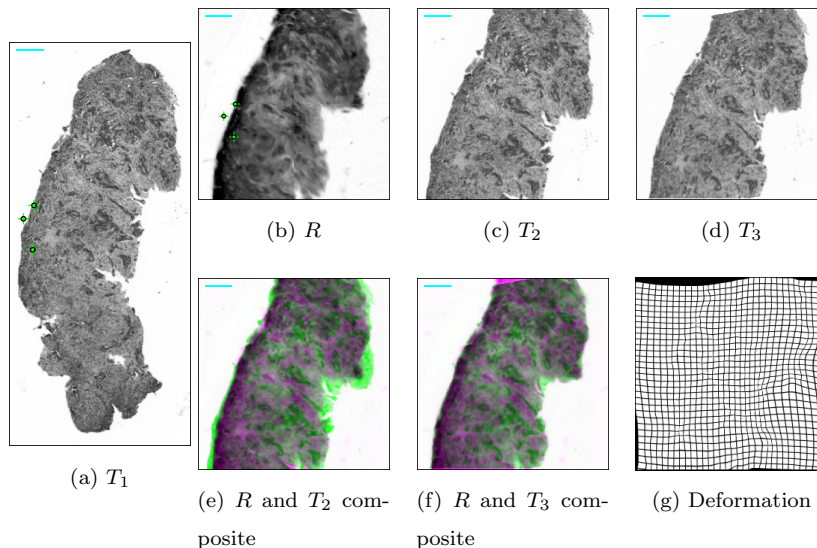
Figure 8: Registration results for a tumoral sample. Cyan scale bars represent 200 µm. (a) Target image T_1 with automatically-detected inlier keypoints. (b) Reference image R with automatically-detected inlier keypoints. (c) Feature-based registration output image T_2. (d) Intensity-based registration output image T_3. (e) Composite RGB image of overlaid images R and T_2. (f) Composite RGB image of overlaid images R and T_3. (g) Artificial grid deformed with the non-rigid displacement field computed in the intensity-based registration step.

squared interval of the undeformed synthetic grids was chosen to occupy 16 pixels both in horizontal and vertical directions.

As can be seen in figures 6 to 8, the feature-based registration step is capable of compensating for coarse misalignment (very relevant in the chosen intermediate sample) and cropping the initial target image to the region of interest of the reference image. It can be observed that the detected inlier keypoints of the tumoral sample are placed near the edge of the tissue. In the intermediate and the normal samples, the keypoints are also identified in the inner region of the tissue, mainly close to the glandular regions. As the SIFT algorithm detects keypoints at different resolutions, the ones which appear outside the tissue at this resolution belong to the border of the tissue at a lower resolution.

It can also be seen that the intensity-based registration step refines the general misalignment (better appreciated in the chosen normal sample) and also produces local deformations that improve the correspondences inside the tissue. These local modifications can be better viewed in the glands of the intermediate and normal tissues.

3.2. Quantitative evaluation

3.2.1. Feature-based registration

Due to the variability between tissue sections in the studied samples, 16 combinations of the SIFT parameters (Sec. 2.5) were considered in order to increase the probabilities of convergence and optimization of the first registration step. Tab. 2 shows the 16 combinations (IDcomb) of the four SIFT parameters (r, n, s and rod) explored for each sample. The table also presents the number (#) and the percentage (%) of samples which have chosen each combination as optimal (in terms of maximal CR) in the feature-based registration step. As can be seen, although three combinations of parameters have not been selected as optimal, there is not a clearly prevailing combination. This fact justifies the exploration of different SIFT parameters in this dataset. The exact combination (IDcomb) of SIFT parameters that each sample has chosen as optimal is shown in the supplementary-file.

By using the manually selected landmarks, the fiducial registration error for the Procrustes analysis ($\mathrm{FRE_{PA}}$) considered as the *gold standard* registration, the proposed feature-based registration ($\mathrm{FRE_{FB}}$) and the differences between them ($\Delta\mathrm{FRE} = \mathrm{FRE_{FB}} - \mathrm{FRE_{PA}}$) were computed for each sample (see the supplementary-file). Fig. 9 presents the box plots [46] which summarize the distributions of these error metrics if the samples are divided into the three pathological groups or if all samples are considered. In these boxplots, whiskers extend to the most extreme data point that is no more than 1.5 times the interquartile range (IQR) from the edge of the box. All the values outside the whiskers range are considered as outliers (red plus signs).

Table 2: Explored combinations, identified by a number (IDcomb), of the four SIFT parameters (r, n, s and rod) with the number (#) and the percentage (%) of samples which has chosen each combination as optimal in the feature-based registration step.

IDcomb	SIFT parameters				#	%
	r	n	s	rod		
1	4	4	6	0.8	6	13.0
2	4	4	6	0.9	0	0.0
3	4	4	10	0.8	4	8.7
4	4	4	10	0.9	2	4.3
5	4	6	6	0.8	2	4.3
6	4	6	6	0.9	5	10.9
7	4	6	10	0.8	3	6.5
8	4	6	10	0.9	9	19.6
9	8	4	6	0.8	1	2.2
10	8	4	6	0.9	0	0.0
11	8	4	10	0.8	2	4.3
12	8	4	10	0.9	0	0.0
13	8	6	6	0.8	3	6.5
14	8	6	6	0.9	2	4.3
15	8	6	10	0.8	4	8.7
16	8	6	10	0.9	3	6.5

As can be seen in Fig. 9a, the FRE_{PA} values (blue box plots) are very similar between pathological groups, being their medians around 6 pixels. Only one outlier exists for the intermediate group, although three samples (with FRE_{PA} values above 15 pixels) are considered as outliers in the group of all samples. A higher dispersion exists in FRE_{FB} values (green box plots) although their distributions are also similar between pathologies and their median values remain under 10 pixels for all the groups. Three outliers are present in the intermediate group with values between 20 and 50 pixels and one extreme outlier with a FRE_{FB} value over 90 pixels exists in the normal group. In Fig. 9b, the distributions of the paired differences of errors (ΔFRE) show a bit higher dispersion in the normal group with a median of 2.2 pixels meanwhile the rest of pathological groups have a median error difference under 1 pixel. If all the samples are considered, the median of (ΔFRE) is also under 1 pixel with a low dispersion, excluding the 5 outlier samples with (ΔFRE) values above 10 pixels. It must be remarked that there was one tumor sample (TUM13; see supplementary-file) where no evaluation was performed because no corresponding points could be manually identified.
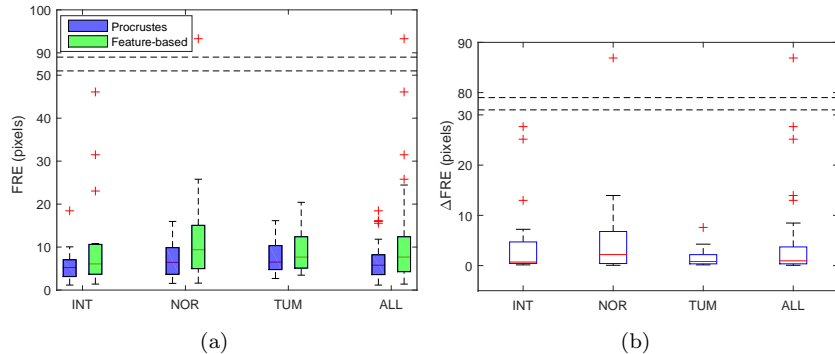
Figure 9: Evaluation of the feature-based registration step. Box plots for the intermediate (INT), normal (NOR), tumoral (TUM) and ALL samples groups representing (a) the fiducial registration error for the Procrustes analysis (FRE$_\text{PA}$) *gold standard* registration (blue), the proposed feature-based (FRE$_\text{FB}$) registration (green) and (b) their differences (ΔFRE = FRE$_\text{FB}$ − FRE$_\text{PA}$). In both subfigures, the empty regions of the vertical axes have been cropped to improve the visualization.

3.2.2. Intensity-based registration

In order to assess the improvement due to the non-rigid transformation provided by the intensity-based registration, the reference image R was compared with the initial target image T_2 and with the final target image T_3. To that end, the mutual information before (MI$_\text{before}$; computed between R and T_2) and after (MI$_\text{after}$; computed between R and T_3) the last registration step as well as their difference (ΔMI = MI$_\text{after}$ − MI$_\text{before}$) were computed for each sample (see the supplementary-file). Fig. 10 shows the box plots of the distributions of these MI values divided by pathological groups and considering all the samples. The rules regarding the notches, whiskers and outliers of these box plots are the same as in Fig. 9.

As observed in Fig. 10a, the distributions of MI$_\text{after}$ are above their respective distributions of MI$_\text{before}$ for all the groups. Although these tendencies suggest an improvement in the alignment of the images, it is not possible to state confidently such an improvement due to the partial overlap caused by the spread of the corresponding distributions. The analysis of the paired differences ΔMI removes the within-sample variability and reduces the source of uncertainty to
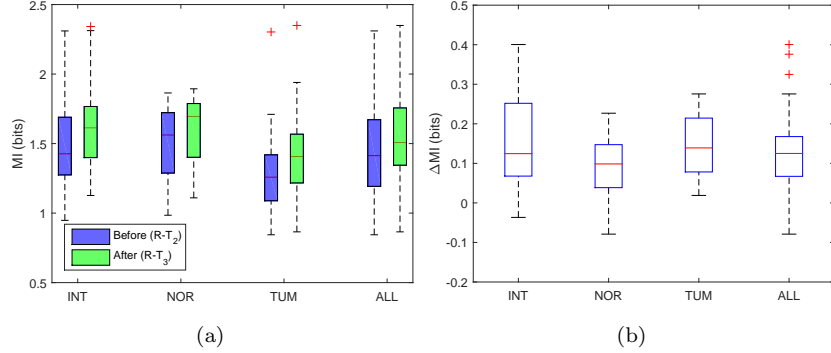
Figure 10: Evaluation of the intensity-based registration step. Box plots for the intermediate (INT), normal (NOR), tumoral (TUM) and ALL samples groups representing (a) the mutual information before ($\text{MI}_{\text{before}}$; computed between R and T_2) the intensity-based registration (blue), after (MI_{after}; computed between R and T_3) the intensity-based registration (green) and (b) their differences ($\Delta\text{MI} = \text{MI}_{\text{after}} - \text{MI}_{\text{before}}$).

the spread in these differences [47]. The distributions of ΔMI (Fig. 10b) present predominant positive values (at exception of one intermediate and two normal samples; see the supplementary-file) with medians around 0.1 bits in the three pathological groups and a median of 0.125 bits in the group of all samples. The right-tailed Wilcoxon signed-rank test [42] returned p-values under 0.001 for all the groups, supporting the alternate hypothesis that the medians of all ΔMI distributions are greater than 0. This result suggests that there is a statistically significant improvement in the alignment of the images.

4. Discussion

The main novelty of this work is the use of a representative grayscale image extracted from the FTIR hyperspectral data cube which condenses the most relevant information of the biological structures of the tissue. The pre-processing to extract this grayscale image is crucial to get a spatial contrast similar to the grayscale image obtained from the H&E image. This grayscale image provides higher spatial information than the binary or the clustered images which were used as inputs in other previous approaches that treated to solve similar

multimodal registration problems [15, 16]. These increase in local information is essential in this work in order to tackle the additional problem of aligning different sections of tissue.

In the feature-based registration step a medium level of information at different levels of resolution is explored by the SIFT algorithm to automatically detect relevant landmarks. No optimal combination of SIFT parameters clearly prevailed in the studied samples (Tab. 2), which justifies the use of different SIFT values and an optimization procedure in terms of CR. This fact also confirms the wide morphological variability between samples, whose sections of tissue may not even be adjacent. Regarding the differences in pathology, normal and intermediate samples have more distinctive references inside the tissue due to the presence of glandular structures, meanwhile tumoral tissue contains more heterogeneous structures as the severity of cancer increases. As a consequence, the automatically detected inlier keypoints in the tumoral samples are mainly located at the tissue borders, which may be a problem if not enough border regions are captured in the image. Apart from that, similar results were obtained for all the pathological groups in the comparison of the feature-based registration step with a *gold standard* manual registration. Most of samples obtained a difference in positioning error (ΔFRE) under 10 pixels and the median of the distribution of all samples was under 1 pixel. These differences may be considered quite satisfactory taking into account that the main purpose of the first registration step is to provide a coarse initial alignment for the second registration step.

The final intensity-based registration step is essential to refine the initial coarse alignment provided by the detected keypoints and to compensate for the spatial unevenness between the different sections of tissue. This step is a complement of the first registration step because it considers the images at the lowest level of information given by their intensities. The values of the parameters used in this step were empirically tuned in the studied dataset as a good global tradeoff between achieving a finer alignment and avoiding excessive local deformations. The statistical significant increase of the mutual information

34

shared between the registered images confirmed the alignment improvement that had been already visually observed, independently of the pathological state.

4.1. Advantages and limitations

Taking into account the previous considerations, the main advantages of the proposed method may be divided in the following points:

- *Simplification of the multimodal problem*: the multimodal problem of finding the correspondence between the three-valued H&E pixels and the hyper-valued FTIR pixels is simplified by condensing their information in grayscale images. It allows the use of registration methods in the grayscale domain, which for example have been successfully and efficiently applied in other medical areas.

- *Gain of spatial information*: using grayscale images augments the spatial information, compared to other related approaches recently published. In particular, a simple binarization is applied in [15], meanwhile previous independent segmentations by clustering within each image are proposed in [16]. It must be remarked that in those studies the same tissue section was measured by FTIR microspectroscopy and then stained to get the H&E image. Therefore, relatively simple spatial transformations had to be calculated; specially restricted in [16], where only translations and rotations up to ±30 degrees were considered due to its computational complexity and probably to the uncertainty introduced by the pre-segmentation step. In this problem, the gain in spatial information provided by the grayscale image is crucial to compensate for the local spatial differences caused by employing distinct tissue sections for H&E and FTIR images. In addition, the suggested approach avoids pre-segmentation steps, which would increase the complexity of the problem (segmentation is maybe the most challenging task in medical image analysis) and may accumulate errors coming from the incorrect partition of corresponding anatomical structures.

35

- *Estimation of local deformations*: the computation of non-rigid spatial transformations by the intensity-based registration step is another improvement compared to [17], where distinct tissue sections treated with different stains are aligned only with rigid transformations. The computed non-rigid displacements are deciding to improve the alignment of more distant sections and compensate for the deformations due to the handling and cutting of the biopsies.

- *Robustness against pathological variability*: the methodology is versatile and robust against distinct sources of variability, including different morphological configurations depending on the pathological state. This robustness is increased in the first registration step by considering different combinations of parameters and maximizing a similarity measure (CR).

On the other hand, the main limitations of the framework are:

- *Presence of distinctive anatomical structures*: the feature-based registration step tries to find relevant keypoints normally present inside the tissue in normal or intermediate pathological states, such as glandular structures. However, the tissue heterogeneity increases with the pathological state and the distance between tissue sections. Therefore, in those cases the presence of tissue borders in the recorded images may be crucial to find an optimal spatial transformation.

- *Global regularization in non-rigid registration step*: the smoothness of the displacement field is equally defined in the whole image by the parameter α. This characteristic may be problematic in samples with uneven local deformations, for example, due to the presence of different pathological regions. Therefore, there may be zones where under- or over-deformations may be computed.

- *Possible unrealistic deformations*: because the intensity-based registration method only considers the low-level information of pixels' intensities. A possible solution would be to create a model of tissue deformation with

higher levels of information, which may require challenging tasks such as pre-segmenting inner tissue structures (e.g. nuclei or glands) or estimating mechanical properties of the biological material.

- *Automatic parameter optimization in non-rigid registration step*: single values for the parameter α and the related ξ_{max} were fixed for all the samples as a good global tradeoff, but better results may be obtained by individually tuning these parameters. However, the main challenge is to pick the optimal values automatically without applying over-deformations. Opposing to the feature-based registration step, where the best combination of parameters is chosen by the maximization of CR, there is no auxiliary reliable metric that may indicate the optimal configuration for each sample.

5. Conclusions

A multimodal registration framework for the automatic alignment of FTIR spectroscopic and H&E stained images from different histological sections has been presented. This methodology was used to register 47 colon samples from three different pathological groups (16 normal, 16 intermediate and 15 tumoral) with good overall qualitative and quantitative results. The proposed method exploits the information of concentration and thickness contained in the absorbance FTIR spectra to generate a grayscale image with a contrast similar to the one obtained from the H&E image. Thus, the morphological structures are highlighted and can be used as a reference for the alignment of the two multimodal images. This approach does not need a prior segmentation step that may introduce errors and reduce the spatial information content.

This automatic method can be easily extrapolated to other kinds of pathologies, such as prostate or breast cancer, where the inner part of the tissue contains relevant morphological structures. Moreover, it can also be applied to more heterogeneous tissues if their borders are also captured in the image. The values of the parameters used in the two steps of the registration framework may be tuned

37

depending on the intrinsic morphological variability of the tissue. In particular, the required deformations computed in the second registration step should be higher as the distance of the sections of tissue to register increases.

The proposed method can improve the accuracy to combine the spatial information extracted from both the traditional H&E stained images and the emerging FTIR spectroscopy, even if different sections of tissue are used. These combinations can result in richer diagnostic algorithms which may consider complementary aspects of the pathological tissue, following the same philosophy of other approaches that fuse different medical imaging modalities.

Acknowledgements

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, F. Bray, Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012, International Journal of Cancer 136 (5) (2015) E359–E386. doi:10.1002/ijc.29210.

[2] Cancer Research UK, Bowel cancer survival statistics - Bowel cancer survival by stage at diagnosis, http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/survival/, (accessed 08.03.17) (2016).

[3] P. G. van Putten, L. Hol, H. van Dekken, J. Han van Krieken, M. van Ballegooijen, E. J. Kuipers, M. E. van Leerdam, Inter-observer variation in the histological diagnosis of polyps in colorectal cancer screening, Histopathology 58 (6) (2011) 974–981. doi:10.1111/j.1365-2559.2011.03822.x.

[4] J. K. Turner, G. T. Williams, M. Morgan, M. Wright, S. Dolwani, Inter-observer agreement in the reporting of colorectal polyp pathology among bowel cancer screening pathologists in wales, Histopathology 62 (6) (2013) 916–924. `doi:10.1111/his.12110`.

[5] C. Kendall, M. Isabelle, F. Bazant-Hegemark, J. Hutchings, L. Orr, J. Babrah, R. Baker, N. Stone, Vibrational spectroscopy: a clinical tool for cancer diagnostics, Analyst 134 (2009) 1029–1045. `doi:10.1039/B822130H`.

[6] G. Bellisola, C. Sorio, Infrared spectroscopy and microscopy in cancer research and diagnosis, American Journal of Cancer Research 2 (1) (2012) 1–21.

[7] R. Bhargava, Infrared Spectroscopic Imaging: The Next Generation, Applied Spectroscopy 66 (10) (2012) 1091–1120. `doi:10.1366/12-06801`.

[8] J. Pijanka, G. D. Sockalingum, A. Kohler, Y. Yang, F. Draux, G. Parkes, K.-P. Lam, D. Collins, P. Dumas, C. Sandt, D. G. van Pittius, G. Douce, M. Manfait, V. Untereiner, J. Sule-Suso, Synchrotron-based FTIR spectra of stained single cells. Towards a clinical application in pathology, Laboratory Investigation 90 (5) (2010) 797–807. `doi:10.1038/labinvest.2010.8`.

[9] L. G. Brown, A survey of image registration techniques, ACM computing surveys 24 (4) (1992) 325–376. `doi:10.1145/146370.146374`.

[10] J. Maintz, M. A. Viergever, A survey of medical image registration, Medical Image Analysis 2 (1) (1998) 1–36. `doi:http://dx.doi.org/10.1016/S1361-8415(01)80026-8`.

[11] B. Zitová, J. Flusser, Image registration methods: a survey, Image and Vision Computing 21 (11) (2003) 977 – 1000. `doi:10.1016/S0262-8856(03)00137-9`.

[12] J. Modersitzki, Numerical methods for image registration, Oxford university press, 2004.

[13] A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration: A survey, IEEE Transactions on Medical Imaging 32 (7) (2013) 1153–1190. `doi:10.1109/TMI.2013.2265603`.

[14] A. Sotiras, Y. Ou, N. Paragios, C. Davatzikos, Handbook of Biomedical Imaging: Methodologies and Clinical Research, Springer US, Boston, MA, 2015, Ch. Graph-based Deformable Image Registration, pp. 331–359. `doi: 10.1007/978-0-387-09749-7_18`.

[15] J. T. Kwak, S. M. Hewitt, S. Sinha, R. Bhargava, Multimodal microscopy for automated histologic analysis of prostate cancer, BMC Cancer 11 (1) (2011) 1–16. `doi:10.1186/1471-2407-11-62`.

[16] C. Yang, D. Niedieker, F. Großerüschkamp, M. Horn, A. Tannapfel, A. Kallenbach-Thieltges, K. Gerwert, A. Mosig, Fully automated registration of vibrational microspectroscopic images in histologically stained tissue sections, BMC Bioinformatics 16 (1) (2015) 1–14. `doi:10.1186/ s12859-015-0804-9`.

[17] G. Lippolis, A. Edsjö, L. Helczynski, A. Bjartell, N. C. Overgaard, Automatic registration of multi-modal microscopy images for integrative analysis of prostate tissue sections, BMC Cancer 13 (1) (2013) 408. `doi: 10.1186/1471-2407-13-408`.

[18] D. G. Lowe, Object recognition from local scale-invariant features, in: Proceedings of the Seventh IEEE International Conference on Computer Vision, Vol. 2 of ICCV '99, IEEE Computer Society, Washington, DC, USA, 1999, pp. 1150–1157. `doi:10.1109/ICCV.1999.790410`.

[19] D. G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110. `doi:10. 1023/B:VISI.0000029664.99615.94`.

[20] M. A. Fischler, R. C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Communications of the ACM 24 (6) (1981) 381–395. `doi:10.1145/358669.358692`.

[21] F. Peñaranda, V. Naranjo, R. Verdú, G. R. Lloyd, J. Nallala, N. Stone, A two-step framework for the registration of HE stained and FTIR images, Vol. 9703, SPIE-Intl Soc Optical Eng, 2016, p. 970305. `doi:10.1117/12.2208869`.

[22] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, Speeded-up robust features (surf), Computer Vision and Image Understanding 110 (3) (2008) 346 – 359, similarity Matching in Computer Vision and Multimedia. `doi:10.1016/j.cviu.2007.09.014`.

[23] J. M. R. S. Tavares, Analysis of Biomedical Images Based on Automated Methods of Image Registration, Springer International Publishing, Cham, 2014, pp. 21–30. `doi:10.1007/978-3-319-14249-4_3`.

[24] A. P. Keszei, B. Berkels, T. M. Deserno, Survey of non-rigid registration tools in medicine, Journal of Digital Imaging 30 (1) (2017) 102–116. `doi:10.1007/s10278-016-9915-8`.

[25] R. M. Rangayyan, B. Acha, C. Serrano, Color Image Processing With Biomedical Applications (SPIE Press Monograph Vol. PM206), SPIE Press, 2011. `doi:10.1117/3.887920`.

[26] J. Trevisan, P. P. Angelov, P. L. Carmichael, A. D. Scott, F. L. Martin, Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives, Analyst 137 (2012) 3202–3215. `doi:10.1039/C2AN16300D`.

[27] P. Lasch, Spectral pre-processing for biomedical vibrational spectroscopy

and microspectroscopic imaging, Chemometrics and Intelligent Laboratory Systems 117 (2012) 100 – 114. `doi:10.1016/j.chemolab.2012.03.011`.

[28] M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner, F. L. Martin, Using Fourier transform IR spectroscopy to analyze biological materials, Nature Protocols 9 (8) (2014) 1771–1791. `doi:10.1038/nprot.2014.110`.

[29] J. Trevisan, P. P. Angelov, A. D. Scott, P. L. Carmichael, F. L. Martin, IRootLab: a free and open-source MATLAB toolbox for vibrational biospectroscopy data analysis, Bioinformatics 29 (8) (2013) 1095–1097. `doi:10.1093/bioinformatics/btt084`.

[30] D. Mayerich, M. Walsh, M. Schulmerich, R. Bhargava, Real-time interactive data mining for chemical imaging information: application to automated histopathology, BMC Bioinformatics 14 (1) (2013) 1–11. `doi:10.1186/1471-2105-14-156`.

[31] W. R. Crum, T. Hartkens, D. L. G. Hill, Non-rigid image registration: theory and practice, The British Journal of Radiology 77 (suppl_2) (2004) S140–S153. `doi:10.1259/bjr/25329214`.

[32] A. Roche, G. Malandain, X. Pennec, N. Ayache, The correlation ratio as a new similarity measure for multimodal image registration, in: W. M. Wells, A. Colchester, S. Delp (Eds.), Medical Image Computing and Computer-Assisted Intervention — MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, 1998, pp. 1115–1124. `doi:10.1007/BFb0056301`.

[33] A. Roche, G. Malandain, N. Ayache, X. Pennec, Multimodal image registration by maximization of the correlation ratio, Tech. Rep. 3378, INRIA (1998).

[34] F. Maes, D. Vandermeulen, P. Suetens, Medical image registration using mutual information, Proceedings of the IEEE 91 (10) (2003) 1699–1722. doi:10.1109/JPROC.2003.817864.

[35] J. Larrey-Ruiz, R. Verdú-Monedero, J. Morales-Sánchez, A Fourier domain framework for variational image registration, Journal of Mathematical Imaging and Vision 32 (1) (2008) 57–72. doi:10.1007/s10851-008-0075-4.

[36] R. Verdu-Monedero, J. Larrey-Ruiz, J. Morales-Sanchez, Frequency implementation of the Euler-Lagrange equations for variational image registration, IEEE Signal Processing Letters 15 (2008) 321–324. doi:10.1109/LSP.2007.913588.

[37] A.-G. Legaz-Aparicio, R. Verdú-Monedero, J. Larrey-Ruiz, J. Morales-Sánchez, F. López-Mir, V. Naranjo, Á. Bernabéu, Efficient variational approach to multimodal registration of anatomical and functional intra-patient tumorous brain data, International Journal of Neural Systems 0 (0) (2017) 1750014. doi:10.1142/S0129065717500149.

[38] B. Fischer, J. Modersitzki, A unified approach to fast image registration and a new curvature based registration technique, Linear Algebra and its Applications 380 (2004) 107 – 124. doi:10.1016/j.laa.2003.10.021.

[39] P. H. Schönemann, A generalized solution of the orthogonal Procrustes problem, Psychometrika 31 (1) (1966) 1–10. doi:10.1007/BF02289451.

[40] D. L. G. Hill, P. G. Batchelor, M. Holden, D. J. Hawkes, Medical image registration, Physics in Medicine and Biology 46 (3) (2001) R1.

[41] J. M. Fitzpatrick, J. B. West, C. R. Maurer, Predicting error in rigid-body point-based registration, IEEE Transactions on Medical Imaging 17 (5) (1998) 694–702. doi:10.1109/42.736021.

[42] J. D. Gibbons, S. Chakraborti, Nonparametric Statistical Inference, Fifth Edition (Statistics: Textbooks and Monographs), Chapman and Hall/CRC, 2010.

[43] C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH image to ImageJ: 25 years of image analysis, Nature Methods 9 (7) (2012) 671–675. `doi:10.1038/nmeth.2089`.

[44] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis, Nature Methods 9 (7) (2012) 676–682. `doi:10.1038/nmeth.2019`.

[45] D. Sage, D. Prodanov, J.-Y. Tinevez, J. Schindelin, MIJ: making interoperability between ImageJ and Matlab possible, in: ImageJ User & Developer Conference, 2012, pp. 24–26.

[46] M. Krzywinski, N. Altman, Points of significance: Visualizing samples with box plots, Nature Methods 11 (2) (2014) 119–120. `doi:10.1038/nmeth.2813`.

[47] M. Krzywinski, N. Altman, Points of significance: Comparing samples—part i, Nature Methods 11 (3) (2014) 215–216. `doi:10.1038/nmeth.2858`.