UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

LIZETH ANDREA CASTELLANOS BELTRAN

Characterization of Structures in Confocal Images Datasets Obtained from Bile Ducts

Thesis presented in partial fulfillment
of the requirements for the degree of
Doctor of Computer Science

Advisor: Prof. Dr. Carla Maria Dal Sasso Freitas

Porto Alegre
August 2020

"So Long,
and Thanks for All the Fish."
— Dolphins

ACKNOWLEDGMENTS

ABSTRACT

Confocal microscopy is a useful tool for acquiring 3D datasets of fluorescent specimens. In hepatology, researchers have been using confocal microscopy for investigating the microanatomy of bile ducts. Since confocal images are difficult to segment because of the noise introduced during the specimen preparation, traditional quantitative analyses used in other medical datasets are difficult to perform on confocal microscopy data and require extensive user intervention. Thus, the analysis of bile ducts pose challenges for hepatology research, requiring different methods.

This thesis provides methods for characterizing structures in confocal image datasets obtained from bile ducts. In our motivating case study, the characterization of such structures is likely to help hepatologists to distinguish specimens affected by biliary atresia, a disease that requires a liver transplant to avoid premature death. Our data consists of 3D image datasets containing several slices of mouse bile ducts organized as two fluorescence channels. The red channel contains a network of small vessels named Peribiliary Vascular Plexus (PVP), and the green channel contains the internal bile duct with Peribiliary Glands (PBGs). Our approach for characterizing the bile ducts structures includes a three-stage process: a stage to enhance the 3D visualization of bile ducts, a stage for extracting important structures and a stage to quantify specific structures of interest.

In the first stage, we proposed an approach to enhance noisy confocal images of bile ducts by applying anisotropic diffusion. The significant result in this stage was the enhanced volumetric visualization of the bile duct microanatomy, which allowed the visualization of details that are hardly seen in the original data.

In the second stage, we have explored the density-based spatial clustering for applications with noise (DBSCAN) algorithm, using gradient information for guiding the clustering. As a result, we discovered a representative cluster for each dataset containing the most prominent vessels (for the red channel) and internal structures (for the green channel).

Finally, we have explored the concepts of fractal dimension and multiscale fractal dimension applied to the structures obtained from clustering, which we found useful for extracting quantitative information aiming at characterizing relevant structures. Our analyses give us some evidence that the fractal dimension is a measure that can be used for quantification and characterization of bile ducts.

Keywords: Confocal microscopy. clustering. DBSCAN. fractal dimension.

Caracterização de Estruturas em Volumes de Imagens Confocais de Dutos Biliares

RESUMO

A microscopia confocal é uma ferramenta útil para adquirir dados 3D de amostras fluorescentes. Na hepatologia, pesquisadores vêm usando microscopia confocal para investigar a microanatomia dos ductos biliares. Como as imagens confocais são difíceis de segmentar devido ao ruído introduzido durante a preparação das amostras, as análises quantitativas tradicionais, em geral, são difíceis de serem executadas e requerem extensa intervenção do usuario. Assim, a análise dos ductos biliares representam um desafio na pesquisa em hepatologia, exigindo diferentes métodos.

Nesta tese, são propostos métodos para caracterizar estruturas em imagens confocais de ductos biliares. No estudo de caso motivador, supõe-se que a caracterização dessas estruturas ajudará os hepatologistas a distinguir amostras afetadas por atresia biliar, uma doença que requer transplante de fígado para evitar a morte prematura. Nossos dados consistem em volumes de imagens de ductos biliares de camundongos organizados em dois subconjuntos, um para cada canal de fluorescência. O canal vermelho contém uma rede de pequenos vasos denominados Plexo Vascular Peribiliar (PVP), e o canal verde representa o ducto biliar interno com as Glândulas Peribiliárias (PBGs). Nossa abordagem para caracterizar as estruturas dos ductos biliares inclui um processo de três estágios: um estágio para melhorar a visualização 3D dos ductos biliares, um estágio para extrair estruturas importantes e um estágio para quantificar estruturas específicas de interesse.

Na primeira etapa, propusemos uma abordagem para realçar as imagens confocais dos ductos biliares, aplicando difusão anisotrópica. O resultado significativo nesta etapa foi a visualização volumétrica aprimorada da microanatomia do ducto biliar, que permitiu a visualização de detalhes que dificilmente são vistos nos dados originais.

No segundo estágio, exploramos o agrupamento espacial baseado no método conhecido como *Density-based spatial clustering of applications with noise* (DBSCAN), usando, porém, informações de gradiente para orientar o agrupamento. Como resultado, descobrimos um cluster representativo para cada conjunto de dados que contém os vasos mais representativos (para o canal vermelho) e estruturas internas (para o canal verde).

Por fim, exploramos os conceitos de dimensão fractal e dimensão fractal multiescala aplicados às estruturas obtidas do agrupamento, que consideramos úteis para extrair informações quantitativas com o objetivo de caracterizar estruturas relevantes. Nossas análises

nos dão algumas evidências de que a dimensão fractal é uma medida que pode ser usada para quantificação e caracterização dos ductos biliares.

Palavras-chave: Microscopia Confocal, clustering, dbscan, dimensão fractal.

LIST OF ABBREVIATIONS AND ACRONYMS

ANOVA   Analysis of variance.

CLSM   Confocal Laser Scanning Microscopy.

DBSCAN   Density-based spatial clustering of applications with noise.

FD   Fractal Dimension.

LIF   Leica Image File Format.

LSM   Laser Scanning Microscope.

MFD   Multiscale Fractal Dimension.

MSE   Mean Square Error.

PBGs   Peribiliary Glands.

PVP   Peribiliary Vascular Plexus.

PSNR   Peak Signal-to Noise Ratio.

SD   Standard Deviation.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

The advent of confocal microscope systems in the 1980s opened the door to the analysis of cells using three-dimensional (3D) images. However, it was not until the 1990s, when computers became powerful enough to handle 3D data and complex 2D data, that the image processing and computer vision communities really began to take up the challenge of confocal images processing (MEIJERING, 2012).

Confocal microscopy offers a powerful means to address biological problems and gives a new understanding of cellular structure and function (MATSUMOTO, 2003). Qualitative and quantitative characterization of cell images is useful for clinical applications (e.g., vaccine development and diagnosis and treatment of disease) and biological research (e.g., to understand the input mechanisms of the virus in cells). Several algorithms have been developed for cell identification using different image processing techniques (GAMARRA et al., 2017).

Some applications of confocal microscopy include:

- **Diagnosis and treatment of diseases**: determination of corneal thickness, diagnostic virology as a survey tool to identify areas of necrosis, studying cell organelles, and tissue changes during cancer (MATSUMOTO, 2003).

- **Biological research**: analysis of gene expression, analysis of changes in membranes, cell division (MATSUMOTO, 2003).

- **Automated classification of cells**: automated identification and tracking of cellular features (TSYGANKOV et al., 2014).

- **Morphometry and Reconstruction** : measure structures in 2D images and 3D reconstructions, three-dimensional reconstruction, four-dimensional imaging (MATSUMOTO, 2003).

Confocal microscopy has several advantages over conventional optical microscopy, due to the ability to control the depth of field and the capability of allowing the collection of serial optical slices from fluorescent specimens. Because of these advantages, confocal microscopy has been used in research to obtain 3D datasets, mostly from cells (CLAXTON; FELLERS; DAVIDSON, 2006), but also for semiconductor inspection (RIDEOUT, 2007)(HONG, 2019) and in materials science(HOHEISEL et al., 2001). Despite these applications, confocal microscopy is a relatively young field (PRICE; JEROME, 2011).

It is well known that it is difficult to analyze complex 3D tissues by viewing indi-

vidual 2D slices (DREBIN; CARPENTER; HANRAHAN, 1988). Capturing the spatial organization of complex tissues, while crucial to understanding their biological function, represents a challenge for current microscopy imaging and analysis systems (KELCH et al., 2015).

Microscopy experiments are producing large volumes of data that subsequently need to be systematically analyzed. Manual image analysis is laborious and not always objective. Analyzing images with algorithms has become popular for it offers automation and reproducibility, and often has higher sensitivity than manual observers (WU; MERCHANT; CASTLEMAN, 2010).

The analysis of fluorescent images involves several stages, each of which may influence the results of the subsequent ones. Figure 1.1 shows a typical workflow in biomedical images analysis. Many steps are needed to extract new knowledge from the images (MEIJERING et al., 2016):

- image acquisition

- preprocessing (denoising, deconvolution)

- detection, i.e., determining the presence of objects based on image-filtering techniques

- segmentation, i.e., grouping of pixels relating to the same object or class

- tracking, i.e., linking detected objects from frame to frame in a movie

- quantification (of shape, dynamics, texture, and other object properties)

- recognition (clustering or classifying objects and patterns)

- visualization (rendering high-dimensional images)

- analytics (statistical processing of the extracted information), and

- modeling (constructing high-level descriptions of the results).

Quantitative analysis of large 3D datasets is not trivial, and the paucity of powerful post-processing and analysis tools for large 3D image datasets is considered as a bottleneck in the field (KELCH et al., 2015).

In this context, our work aimed at creating methods for characterizing structures in confocal microscopy volumetric data. We focus our study on datasets obtained from bile ducts. Bile ducts are thin tubular structures that carry the bile, and studying their microanatomy is a hot topic in hepatology research (DIPAOLA et al., 2013; HAMMAD et al., 2014; LAMETSCHWANDTNER et al., 2015; MORALES-NAVARRETE et al., 2015; VARTAK et al., 2016). The analysis of microscopic morphological changes in the

Figure 1.1: Common steps in bioimage analysis.
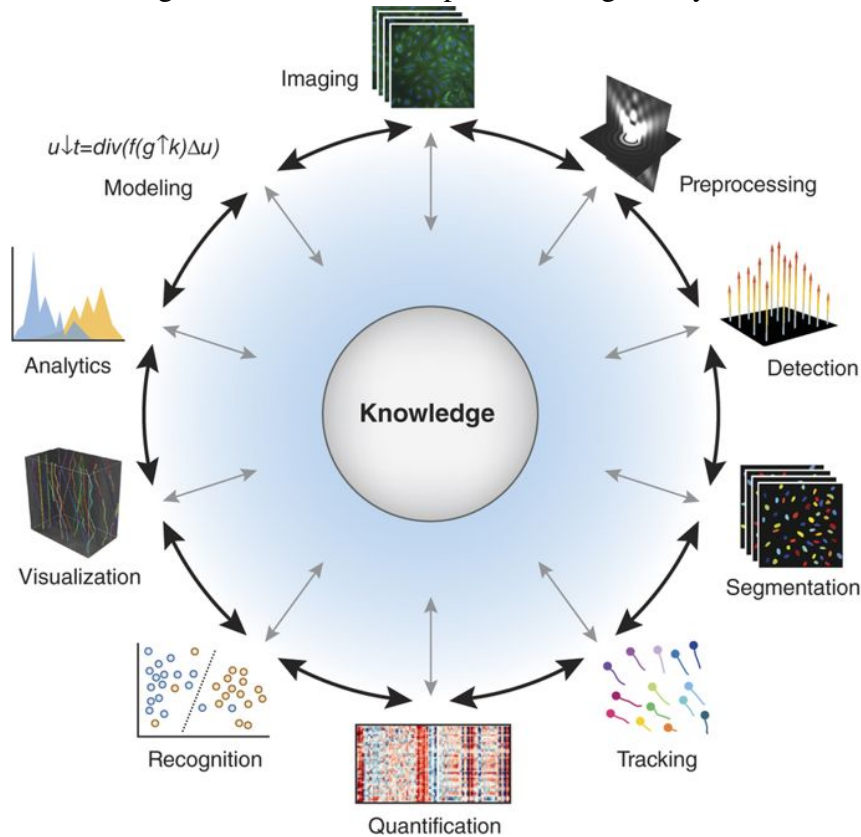
Source: (MEIJERING et al., 2016).

bile ducts is important in the study of biliary diseases.

Our focus on bile ducts is because the motivating case study was the research conducted by collaborators at the Hospital de Clínicas de Porto Alegre (the University hospital) and the Cincinatti Children's Hospital in search of the causes of the biliary atresia, a disease that requires liver transplant to avoid premature death.

We worked more on methods for detection, segmentation, quantification, and recognition, although there are connections with the other stages presented in Figure 1.1:

- **Detection:** We propose a method based on anisotropic diffusion filtering to enhance details of the bile duct structures.

- **Segmentation:** We group pixels based on thresholding and clustering methods to improve the segmentation of the bile ducts structures

- **Quantification:** We explore the fractal dimension as a technique for quantification of relevant structures.

- **Recognition:** We implement a method for extracting clusters of relevant structures, using the density-based spatial clustering for applications with noise (DBSCAN).

- **Visualization:** We use direct volume rendering techniques to visualize all the

resulting volumetric data.

Although a few existing platforms provide standard tools for 3D segmentation and methods to process 2D surface layers of cells, the challenges posed by dense and thick tissue specimens require the development of new algorithms (MORALES-NAVARRETE et al., 2016). We give further details about such issue in Chapter2.

1.1 Goal and Research Questions

The goal of this thesis is to develop methods for characterizing structures in confocal microscopy volumetric data. Our work focuses on datasets obtained from bile ducts.

Aiming at this goal, we state our research questions as follows:

1. The details from bile ducts are hardly visualized in the original images. Thus, the first challenge in our research is associated with the noise in confocal microscopy datasets. In this context, our first research question is about how can we improve the visualization of multichannel confocal datasets to provide a better distinction of the structures of interest?

2. Our second question is related to which measurements can we use for quantifying and characterizing 3D structures in confocal images datasets?

To guide the answers of our research questions we hypothesized the following:

- **H1:** Regarding the 3D visualization of confocal images datasets, it is possible to achieve quality by enhancing structures using a pre-processing step with appropriate techniques to deal with the noise.
  Our first hypothesis assumed that volume rendering of confocal microscopy images could be improved by appropriate image enhancement techniques. We believed that treating the noise by considering the properties of confocal data, we could reveal the 3D organization of bile duct structures using direct volume rendering.

- **H2:** Regarding the 3D visualization of confocal images datasets, it is possible to improve the distinction of the relevant structures using an unsupervised image segmentation method before rendering.
  We intended to prove with our second hypothesis that we could separate the relevant structures in confocal images. In some cases, it is common that medical image seg-

mentation can be performed by manually labeling. However, confocal microscopy images have complex tissue and cellular structures, and manual segmentation is only possible in a limited way. This fact motivated our interest in unsupervised image segmentation algorithms.

- **H3:** Fractal dimension analysis can be used for the quantification and characterization of structures in confocal images datasets.

  Our third and last hypothesis was the most challenging one and was concerned with finding the measures that could be used to characterize the bile ducts. We wanted to verify that fractal dimension analysis makes it easy to quantify the relevant structures found in bile ducts confocal images and allow the comparison of specimens.

1.2 Overview

In this thesis, we describe the significant steps performed to address our research questions, focusing on confocal images taken from bile ducts samples:

- The first step was related to specimen preparation and image acquisition. That process was performed by the medical team using confocal microscopy, and provided the input datasets for our work;

- The acquired datasets required the investigation of pre-processing approaches for revealing the structures of interest;

- Subsequent three-dimensional image reconstructions evidenced the spatial organization and morphology of the structures within the bile duct;

- A clustering technique was used for segmenting the structures of interest;

- Fractal dimension analyses were explored for quantification and characterization of bile ducts.

The remainder of this document is organized as follows: Chapter 2 introduces the necessary background to understand the context of the thesis, while Chapter 3 presents the related work on image filtering, clustering techniques, and fractal dimension analysis. Chapter 4 describes our approach, and Chapters 5 and 6 present and discuss our results. Finally, in Chapter 7 we revisit our research questions and hypotheses to summarize our contributions, and draw comments on future work.

2 BACKGROUND

The thesis is focused on the processing and analysis of 3D confocal microscopy data (Figure 2.1). In this chapter, first, we describe the most relevant concepts of confocal microscopy imaging. This knowledge provides an adequate comprehension of the advantages and limitations associated with such an acquisition process, which directly affects the next stages, which correspond to image processing and image analysis techniques. Then, we describe the necessary techniques for image processing and analysis and, finally, volume rendering.

2.1 Confocal Microscopy

Along the last years, there has been an increase in the use of confocal microscopy (CLAXTON; FELLERS; DAVIDSON, 2006; MASTERS, 2008; PRICE; JEROME, 2011), more specifically in cell biology applications using fixed and living cells and tissues. Confocal microscopy data have their characteristics, which differ from other biomedical data (like optical microscopy, tomography, ultrasound among others) (WAN et al., 2012).

In confocal microscopy, the images are acquired point-by-point, using lasers and based on the principle of fluorescence (Figure 2.2). Fluorescence is the property of some atoms and molecules to absorb light at a particular wavelength and to subsequently emit light of a longer wavelength after a brief interval (Michael W. Davidson., 2014). Our data consists of several slices of extrahepatic bile ducts (Figure 2.3). The biological samples can be labeled with several appropriate fluorescent antibodies during the staining process (Figure 2.4), which allows marking different tissues or cells (Figure 2.5) .

Figure 2.1: Representation of a 3D Confocal Image.
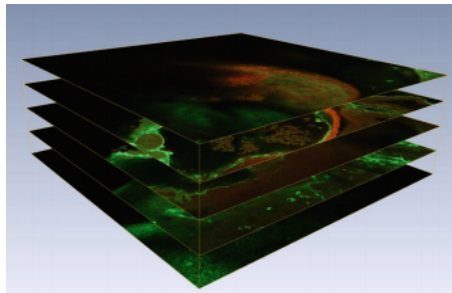
Source: (PRICE; JEROME, 2011).

The characteristics of a confocal microscope offer several advantages over conventional optical microscopy such as:

Figure 2.2: General diagram of the mechanism for collecting signals from a sample labeled with three fluorescent antibodies using three laser beams.
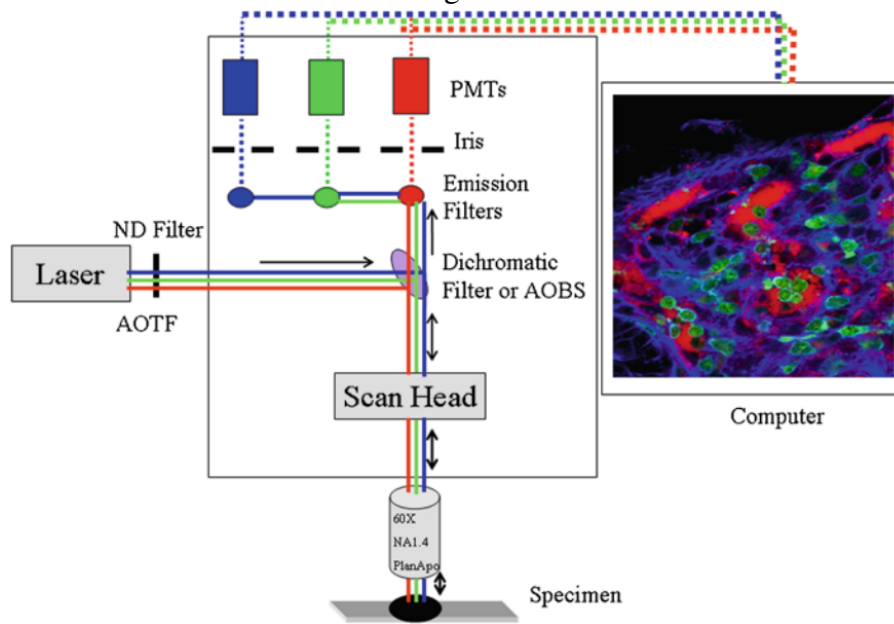
Source: (PRICE; JEROME, 2011).

- Confocal Microscopy removes out-of-focus light from the image. When used correctly, this results in all planes of the final image being in focus(PRICE; JEROME, 2011).

- Capability of controlling the depth of field.

- Capability of collecting several aligned images of the same sample.

However, confocal images are normally affected by several artifacts and noise sources:

- Low signal-to-noise ratio: confocal images have a strong decrease in the signal-to-noise ratio over the slices depth (RAMESH; OTSUNA; TASDIZEN, 2013).

- Diversity of density values: the physical meaning of density values is not limited to image subjects (TORIWAKI; YOSHIDA, 2009). Confocal images have an inhomogeneous density inherent to the fluorescent staining process (CHEN; CHEN; CHIANG, 2008).

- Visual occluders: structures irrelevant to the analysis may also be labeled through the fluorescent staining process, resulting in visual occluders that obscure the structures to be visualized (WAN et al., 2009).

- Subtle boundaries: meaningful boundaries may be only subtly presented in the confocal data (WAN et al., 2009).

Figure 2.3: Extrahepatic Bile Ducts Anatomy.
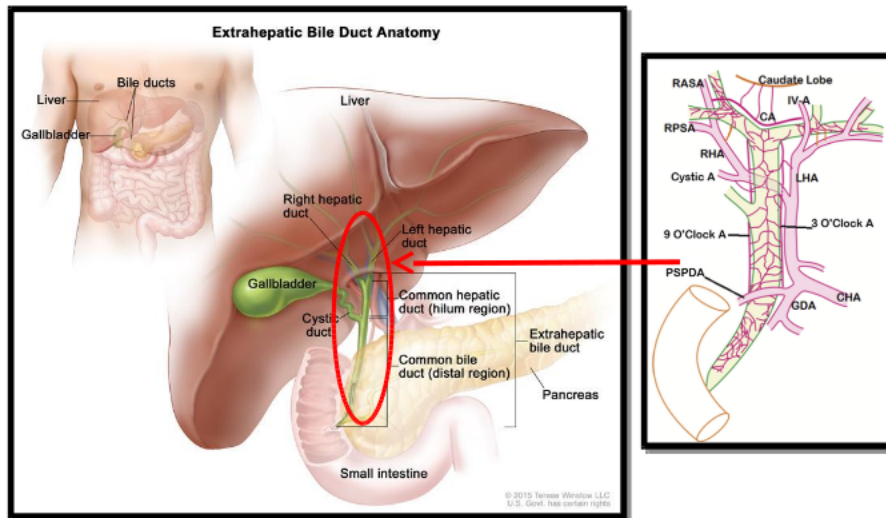
Source: Adapted from (BOARD, 2019) and (BABU; SHARMA, 2014).

Confocal microscopes use proprietary formats, which encodes the information about the images and the microscope configuration used during acquisition. The information regarding the process of image acquisition is important for further interpretation of the obtained data by the hepatologist. However, the downside of proprietary formats is that they need proprietary software to decode the stored images (PRICE; JEROME, 2011). We use datasets acquired by two kinds of confocal microscope: a Leica and a Zeiss confocal microscope. These microscopes produce datasets in the formats named *LIF* (Leica Image File Format) and *LSM* (Laser Scanning Microscope), respectively.

Figure 2.4: A mouse bile duct after the specimen preparation.
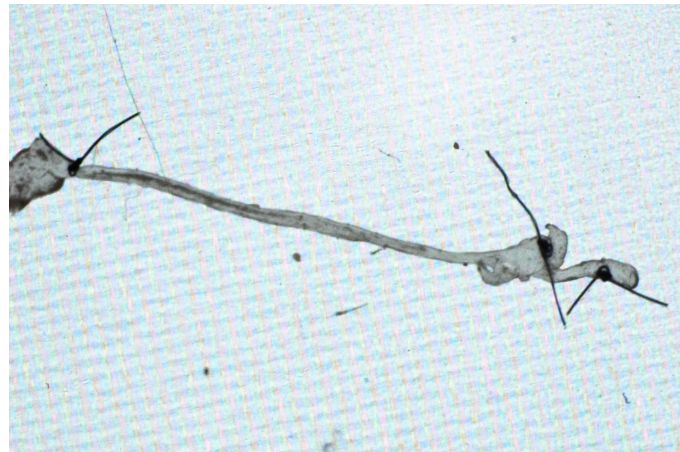
Source: The author.

Figure 2.5: Example of confocal microscopy slices of a mouse bile duct.
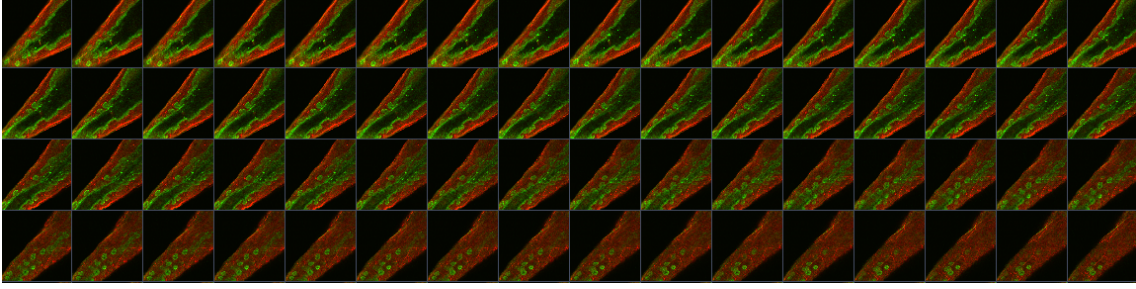
Source: The author.

2.2 Anisotropic diffusion

Since anisotropic diffusion was used to reduce the noise and enhance details of the bile duct images, we also present the basic principles of anisotropic diffusion filtering. Anisotropic diffusion was introduced by Perona and Malik (PERONA; MALIK, 1990), and since then, has been used as an effective approach in image processing and computer vision for noise removal, edge detection, and image restoration (WEICKERT, 1998) (FRANGAKIS; HEGERL, 2001) The main idea behind this approach is that smoothing should be low on relevant edges and stronger in regions dominated by noise (FORMAGGIA; QUARTERONI; VENEZIANI, 2010). In the classical formulation (PERONA; MALIK, 1990), the anisotropic diffusion equation is given by the following Partial Differential Equation (PDE):

$$\frac{\partial I}{\partial t} = \nabla \cdot (c(\nabla I)\nabla I) \qquad (2.1)$$

where t is the time parameter, ∇I is the gradient of the image at time t and c is the diffusivity. The diffusivity can be expressed as a decreasing function of the image gradient magnitude, such as:

$$c(x,y,z,t) = e^{-\frac{\|\nabla I\|^2}{k}}, c(x,y,z,t) = \frac{k^2}{k^2 + \|\nabla I\|^2} \qquad (2.2)$$

where k is the gradient magnitude threshold parameter that controls the rate of the diffusion and serves as a soft threshold between the image gradients that are attributed to noise and those attributed to edges (TSIOTSIOS; PETROU, 2013). The great success of the Perona and Malik's model can be mainly attributed to its excellent performance in edge preservation and noise removal (YUAN; WANG, 2016).

Anisotropic diffusion process can be improved with the proper choice of param-

eters of the anisotropic diffusion equation (VOCI et al., 2004) (TSIOTSIOS; PETROU, 2013). According to Formaggia et al. (FORMAGGIA; QUARTERONI; VENEZIANI, 2010), the anisotropic filter must be tuned for specific applications in terms of *k* and *t*. This implicates that empirical evaluation of the effects of the filter is necessary. Several papers discuss aspects aimed at optimizing the anisotropic diffusion. However, none of those proposals are applied to confocal microscopy images.

There are two models for the estimation of the ***k* parameter**:

- **Estimation of the k parameter using Perona and Malik's model**: The model for the *k* estimation proposed by Perona and Malik (PERONA; MALIK, 1990) is based on a noise estimator using the histogram of the gradient. This noise estimator consists of calculating the histogram of the absolute values of the gradient for every image, and the *k* parameter value is equal to the 90% value of its integral (PERONA; MALIK, 1990).

- **Estimation of the k parameter using Voci et al. model**: The model for the *k* estimation proposed by Voci et al. (VOCI et al., 2004) is based on mathematical morphology. The idea of using a morphological approach derives from the fact that morphology can be used for an estimation of noise intensity in the image. Their model is based on opening and closing operations from mathematical morphology. The *k* is given by the following equation:

$$k = \sum_{i,j \in I} \frac{(I(i,j) \circ st)}{(r.c)} - \sum_{i,j \in I} \frac{(I(i,j) \bullet st)}{(r.c)} \qquad (2.3)$$

where $I(i,j)$ refers to the image consisting of r rows and c columns, a structuring element st (we use a st with size 5x5), and the symbols $\circ$ and $\bullet$ represent the opening and closing operations, respectively.

2.3 Density-Based Spatial Clustering of Noisy Images

The process of clustering plays an important role in the fields of knowledge discovery and data mining (MEHMOOD et al., 2016). There are several approaches to clustering described in the literature of unsupervised learning, including partitioning algorithms, hierarchical methods, density-based algorithms, and models-based algorithms. Density-based Spatial Clustering has been successfully applied in images datasets ob-

Figure 2.6: Anisotropic Diffusion using Perona Model and Voci Model.

(a) Original Image   (b) Noisy versionl   (c) Perona and Malik's Model   (d) Voci et al. Model
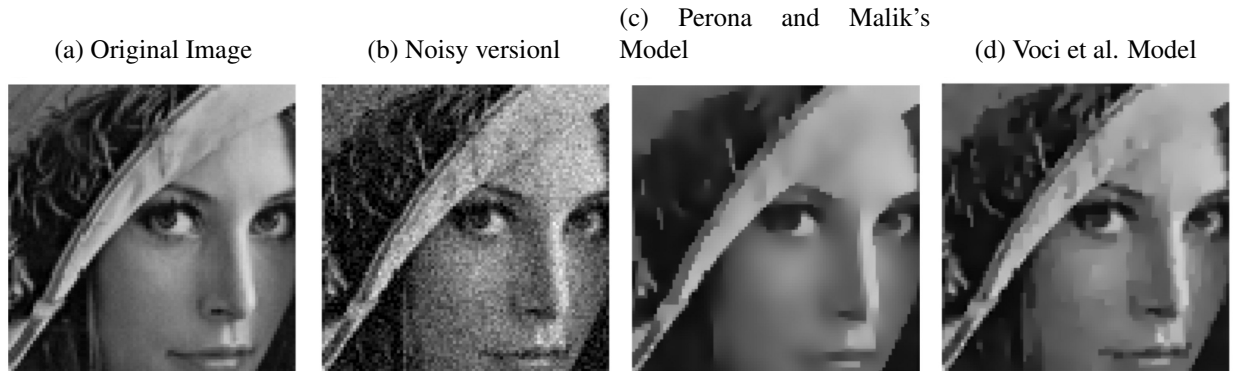
Source: (TSIOTSIOS; PETROU, 2013)

tained from different sources for application in distinct domains like spectrometer, tomography, dermascopy, angiography (HUI; LIU, 2017; HUI; LIU; PARK, 2018; LI et al., 2017; CELEBI; ASLANDOGAN; BERGSTRESSER, 2005; METE; KOCKARA; AYDIN, 2011; TRAN et al., 2012), including also confocal images (MU et al., 2009; CHAN; CHENG; POON, 2007).

Figure 2.7: Concepts used the DBSCAN Clustering: a)Shows examples for the three point classes, core, border, and noise points. b)Shows the concept of density-reachability and density-connectivity

(a) Core, border, and noise points   (b) Density-reachability and Density-connectivity
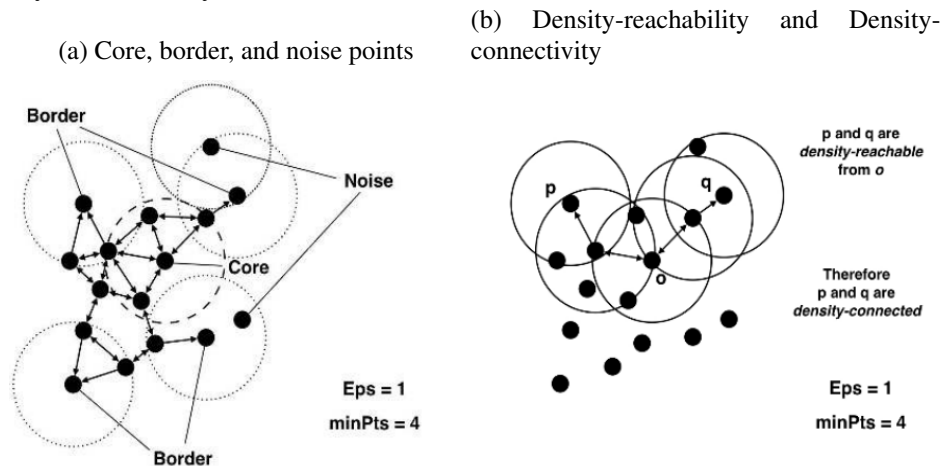
Source: (HAHSLER; PIEKENBROCK; DORAN, 2017)

Since our approach is exploratory, and we do not know a priori the number of clusters to partitioning the data, herein, we focus on density-based algorithms. Since confocal data is often noisy, we choose a density-based clustering algorithm known as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (ESTER et al., 1996), which we briefly introduce in this section.

DBSCAN was proposed for knowledge discovery in databases (KDD) and has the following characteristics:

- Minimal requirements of domain knowledge: *a priori* number of clusters is not necessary.

- It discovers clusters of arbitrary shapes.

- It is robust against noise: less sensitive to outliers and can identify the noise points.

- Deal with large datasets: was proposed to be used in large databases with thousands of objects.

- Based on only two global parameters: *eps*, that is the radius around a pixel for the density calculation, i.e., the size of the *eps* neighborhood, and *minPts*, which corresponds to the minimum number of points required to form a cluster.

The key idea of DBSCAN is that for each point of a cluster the neighborhood of a given radius has to contain at least a minimum number of points, i.e. the density in the neighborhood has to exceed some threshold (ESTER et al., 1996).

2.4 Direct Volume Rendering

Direct volume rendering techniques allow visualization of volume data without extracting intermediate geometry. These methods process the volume data based on fuzzy segmentation through transfer functions. This means that one group of points can belong to more than one structure/tissue with different degrees of membership.

Volume rendering is based on the emission-absorption optical model defined by the Equation 2.4:

$$I(D) = I_0 e^{-\int_{s_0}^{D} k(t)dt} + \int_{s_0}^{D} q(s) e^{\int_{s}^{D} k(t)dt} ds \tag{2.4}$$

With optical properties k (absorption coefficient) and q (source term describing emission) and integration from entry point into the volume, $s = s_0$, to the exit point toward the camera, $s = D$.

The basic idea of the ray casting algorithm (LEVOY, 1988)(HADWIGER et al., 2006) is to directly evaluate the volume-rendering integral along with the rays that are traced from the camera into the object space. For each point in a projection plane (i.e., for each pixel in the image), a single ray is cast into the volume. Then, the volume data is resampled at discrete positions along the ray. Viewing rays are fired through the data to sample the volume. After that, the data is evaluated by using a function in order to compute the final pixel value. The final pixel color is obtained from the accumulation of

values obtained using a *transfer function* (TF) that maps data values to optical properties.

The most common type of transfer function is the one-dimensional TF, which assigns optical properties (usually color and opacity) to voxels based only on their scalar value. Notwithstanding, one-dimensional TFs have a limited classification power because they can not distinguish between volume regions defined by scalar values within the same range. On the other hand, multi-dimensional transfer functions can perform better classification because they can take into account not only the scalar value of a voxel (KNISS; KINDLMANN; HANSEN, 2002), but also other attributes like gradient magnitude, directional second derivative, curvature (HLADUVKA; KÖNIG; GRÖLLER, 2000; KINDLMANN et al., 2003) and statistical measures (TENGINAKAI; LEE; MACHIRAJU, 2001).

2.5 Fractal Analysis

Fractal analysis have been applied in many areas since the seminal work by Mandelbrot(MANDELBROT, 1983). Literature on its application in biology and medicine is large, and a thorough review is beyond the scope of this thesis. When it comes to medical images, we also found the extensive use of fractal and multifractal geometries in the analysis of medical signals (1D, 2D or 3D) for supporting pattern recognition, texture analysis and segmentation (LOPES; BETROUNI, 2009).

The main attraction of fractal geometry stems from its ability to describe the irregular or fragmented shape of natural features as well as other complex objects that traditional Euclidean geometry fails to analyze. Fractal objects have three main features (MANDELBROT, 1967)(KUIKKA, 2002):

- large degree of heterogeneity,

- self-similarity over many scales of observation, and

- the lack of a well-defined scale

As mentioned by Kuikka (KUIKKA, 2002), the three features have important implications when one wants to interpret experimental observations. First, certain measurements do not have a single 'true' value, but the value depends on the resolution used for the measurement. So, there is a scaling relationship between the resolution and the value measured. Second, self-similarity (or scale invariance), the most known characteristic of fractals, is generated by a recursive expression, determining a non-linear scaling relationship. The last implication mentioned by Kuikka (KUIKKA, 2002) is the fact that natural

objects are seldom regular and geometric, but they present an inherent variation in structure and function, which is related to their scaling properties. The scaling property $L(r)$ of a fractal object depends non-linearly on the scale r at which some measurement (of the object) is taken.

Fractals have fractional dimensions, known as *fractal dimension*, which are a measure of the complexity of the fractal objects (MANDELBROT, 1983)(MANDELBROT, 1967)(BACKES et al., 2010). A fractal dimension (referred to as FD or D) is said to be fractional because it is a ratio, i.e., a statistical index of complexity that measures how the details in a pattern changes with the scale at which it is measured. As so, fractal dimensions are usually non-integer values.

2.5.1 Fractal Dimension

The fractal dimension (FD) of images can measure the distribution of pixels intensities. Thus, it can be used as a tool to describe the image content (Zhou; Liang, 2014). There are many methods to compute the fractal dimension. Although the applied algorithms differ, they obey to the same basis summarized by three steps (LOPES; BETROUNI, 2009):

- Measure the quantities of the object using various step sizes
- Plot log (measured quantities) versus log (step sizes) and fit a least-squares regression line through the data points
- Estimate FD as the slope of the regression line.

We restrain ourselves to briefly review two methods of calculating the fractal dimension of an image, the Box-counting (BCM) and the Bouligand-Minkowski methods.

- **Box-counting method (BCM):** This is the most popular method (LOPES; BETROUNI, 2009), following principles that are simple and easy to develop (TRICOT, 1994). (LI; SUN; DU, 2006). It consists of covering the image A with a mesh of squares of side r, and count how many squares $N_r(A)$ intercept the image. FD is estimated according to Equation 2.5.

$$FD = -\lim_{r \to 0} \frac{\log\left(N_r(A)\right)}{\log\left(r\right)} \tag{2.5}$$

The relationship between r and $N_r(A)$ generates a log–log curve, and the slope of

Figure 2.8: Illustration of the box-counting method for fractal dimension in a binary image.
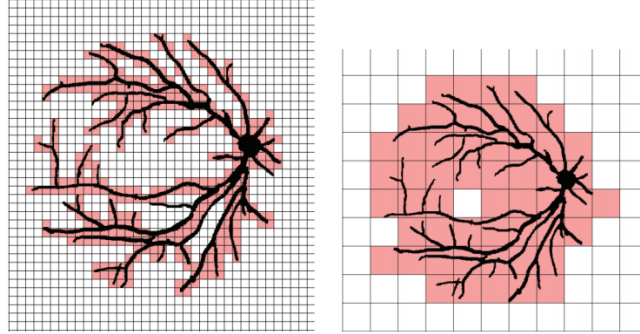
Source: (COSTA; CESAR, 2000).

this curve is the Box-Counting fractal dimension FD of the image A.

The original formulation has limitations because it requires signal binarization; for this reason, some adaptations have been proposed (LOPES; BETROUNI, 2009).

- **Bouligand-Minkowski method:** This method is based on the study of an influence area *A(r)* in image A, which is created by the dilation of the image by a disc of radius *r*. The Bouligand-Minkowski fractal dimension FD is estimated according the Equation 2.6:

$$FD = N - \lim_{r \to 0} \frac{\log\left(A(r)\right)}{\log\left(r\right)} \qquad (2.6)$$

with $A(r) = \{p' \in R^N \mid \exists p \in A : \mid p - p' \mid \leq r\}$ where p is a point from image A, p' is a point in R^N whose distance from p is smaller or equal to *r* and *N* is the number of dimensions of the space where image *A* is inserted. For binary images, $N = 2$. As for gray-scale imagens, to compute the Bouligand-Minkowski fractal dimension, one takes a 3D approach. A surface *S* is generated by transforming each image pixel into a point $p = (x, y, z), p \in S$, where x and y correspond to the coordinates of the pixel in image A and $z = A(x, y)$ is its intensity. By performing the dilation of surface S by a radius r, the Bouligand-Minkowski fractal dimension FD of S can be estimated as Equation 2.7:

$$FD = 3 - \lim_{r \to 0} \frac{\log\left(V(r)\right)}{\log\left(r\right)} \qquad (2.7)$$

with $V(r) = \{p' \in R^3 \mid \exists p \in S : \mid p - p' \mid \leq r\}$ where $p' = (x', y', z')$ is a point in R^3 whose distance from $p = (x, y, z)$ is smaller or equal to r and $V(r)$ is the influence volume calculated by dilating each point of S by a sphere of radius r

(BACKES; CASANOVA; BRUNO, 2009).

Equation 2.7 can also be used to estimate the FD for volumetric objects ((BACKES et al., 2010)). By using the Bouligand-Minkowski method for 3D obects, the object is dilated using spheres of every possible radius. As we show in Figure 2.9, as the radius increases, more interaction among the spheres occurs, thus producing an influence volume for the set of points. The influence volume $V(r)$ is then used to calculate the fractal dimension FD, and includes information from various spatial scales r.

To compute the Minkowski–Bouligand method at the 3D level, for a given radius $r = R$, the algorithm visits each point of the 3D object as follows: For each point (x, y, z), the method calculates $V(r)$ for r ranging from 0 to R. $V(r)$ is equal to the number of points of the object within the sphere of radius r centered on (x, y, z). In other words, it works like the box-counting method but considering the spheres.

Figure 2.9: Example of the influence volume $V(r)$ for different radius values ($r = 1, 3, 5$).
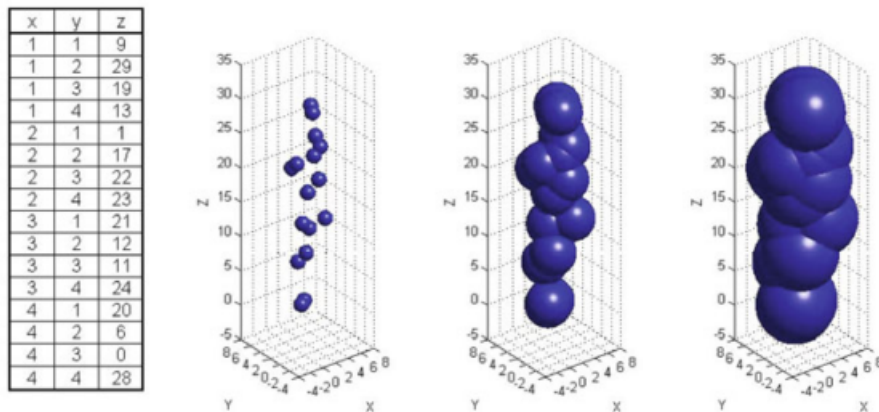
Source: (BACKES et al., 2010).

2.5.2 MultiScale Fractal Dimension

The log–log curve calculated by Bouligand–Minkowski method provides information about the complexity of the shape for different values of r, which can be the size of the box or the radius of a sphere. These values corresponds to various scales. However, for computing the FD, usually only part of this information is used (the slope of the curve defined by a specific r). The rest of the information intrinsic to the log-log curve is discarded. The shapes of objects can have many details, and a single number may not be enough to discriminate it (BACKES; BRUNO, 2010).

The Multiscale Fractal Dimension (MFD) approach (EMERSON, 1998; PLOTZE et al., 2005; BRUNO et al., 2008) can be used to describe the complex shape of an object by means of several values of FDs. The multiscale fractal dimension is calculated by applying a multiscale transformation over the log-log curve $u(t)$ obtained from the Bouligand–Minkowski method.

Recall that the $u(t)$ curve depicts the logarithm of the area of influence of a given shape A, $A(r)$, originated from its dilation by a disk of radius r. From the first derivative of the curve $u(t)$, a function capable of representing the variation of shape complexity from an observation scale is obtained. This function is called multiscale fractal dimension and can be defined, as:

$$MFD = N - \frac{dlog(V(r))}{dr} \tag{2.8}$$

where $du(t)/dt$ is the derivative of log–log curve $u(t)$ (Equation 2.6). To compute the MFD, it is necessary to calculate the derivate of $u(t)$. One way to calculate this derivative is using the finite differences method, where the derivative of a point is obtained by analyzing its neighborhood.

So, given a curve $C = \{c_i = (x_i, y_i) \,|\, i = 1, 2, ...n\}$ and a sampling window h, for each point i of the curve, the slope of the line formed by the points $(i - h)$ and $(i + h)$ is calculated as the derivative at that point on the curve as described in Equation 2.9, with k being the dimension of the data space, in this case, is 3.

$$c(i) = k - \frac{y_{i+h} - y_{i-h}}{x_{i+h} - x_{i-h}} \tag{2.9}$$

An important aspect to consider with the derivative is its tendency to emphasize high-frequency signals, such as noise. Therefore, it is necessary to apply a smoothing filter to the curve before calculating the derivative (Equation 2.10, where w is the size of the window used to average the neighborhood of a point (SMITH; SMITH; SMITH, 1985).

$$c(i) = \frac{\sum_{j=i-w}^{i+w} c(j))}{w} \tag{2.10}$$

As a final observation, we should notice that for calculating the MFD for a 3D object, the log-log curve is given by Equation 2.7.

3 RELATED WORK

In this chapter, we present related research on confocal microscopy data. It is worthy to emphasize that most of the research devoted to biological/medical studies based on confocal microscopy data focuses on small structures such as cells. In our work, differently from these studies, we analyze a macro-structure, the bile duct. Since cells are thinner than bile ducts samples, this poses a challenge in data processing because the noisy sources affect the image quality in thick specimens to a greater extent than in thinner specimens such as cells.

In the following sections, we summarize related works and point out the differences between them and ours.

3.1 Confocal Microscopy Data in Hepatology Research

We found a few works using confocal microscopy to study the micro-anatomy of bile ducts, and they are briefly reviewed herein.

DiPaola et al. (DIPAOLA et al., 2013) identified peribiliary glands (PBGs) residing within the bile duct walls (Figure 3.1). However, the images were visualized using the confocal microscopy proprietary software, which provided limited features for image post-processing, like we need for our purposes.

Hammad et al. (HAMMAD et al., 2014) and Vartak et al. (VARTAK et al., 2016) used confocal microscopy images to visualize intrahepatic bile ducts that are much smaller than the extrahepatic bile ducts we work with. Figures 3.2 and 3.3 illustrates their works.

3.2 Image Processing: Filtering Confocal Microscopy Images

The median filter is a traditional filter used for noise reduction in confocal microscopy images. Parazza et al. (PARAZZA; HUMBERT; USSON, 1993) used a 3D median filter for noise reduction in confocal microscopy images from cell nuclei. A median filter was also used for noise reduction in images from rat brain (MADDAH; SOLTANIAN-ZADEH; AFZALI-KUSHA, 2003). In another work, Paul et al. (PAUL et al., 2010) used median filtering to estimate the global noise variance in images from cells.

Araujo et al. (ARAUJO et al., 2000) propose the use of blurring filters, histogram

Figure 3.1: Manual identification of the peribiliary glands (PBGs) in a mouse bile duct: a) A cystic bile duct densely populated by PBGs that are juxtaposed to the mucosa epithelium (white arrowheads). b) An intersection of the cystic duct and hepatic ducts to form the common bile duct CBD. c) The CBD has PBGs that are unilobulated or multilobulated (white arrowhead and yellow arrow, respectively) and connect with the main epithelium or form the peribiliary network containing tubular structures within the wall (white arrows) and with narrow lumen (yellow arrowheads) connecting different segments of the duct. d) The junction of the CBD and the pancreatic duct (PD).
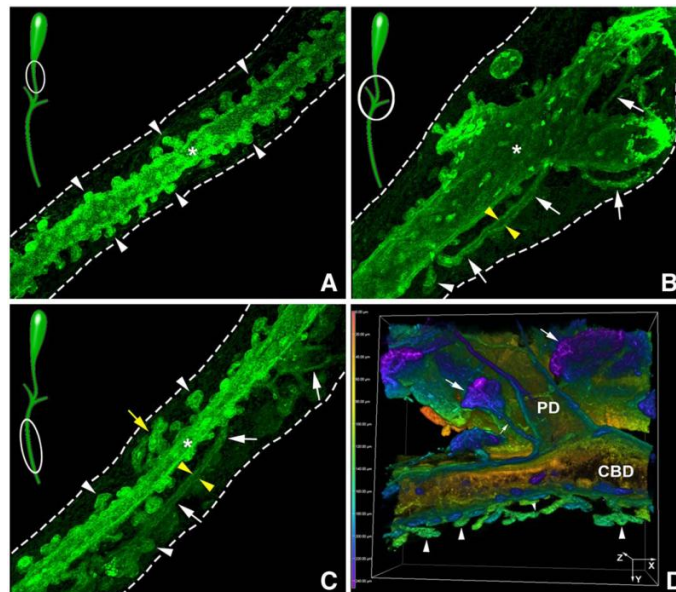
Source: (DIPAOLA et al., 2013).

Figure 3.2: Imaging and analysis pipeline for reconstruction and quantification of liver microarchitecture.
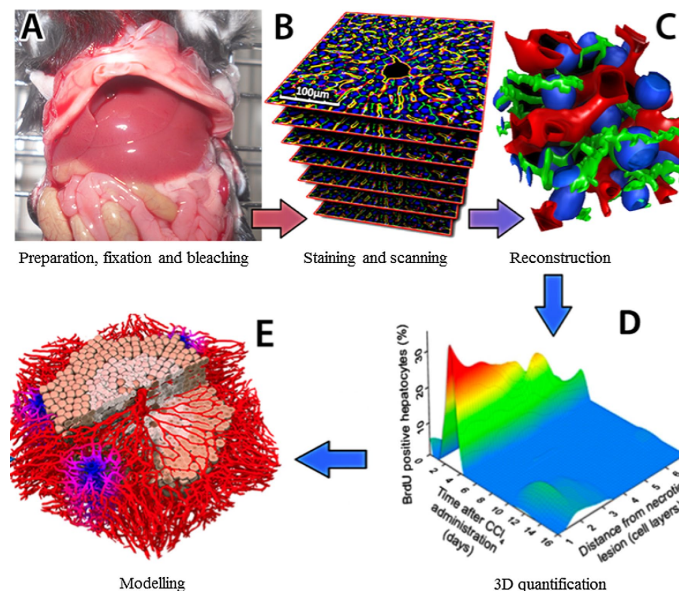
Source: (HAMMAD et al., 2014).

equalization, and arithmetic operations to enhance images of cells from the nervous system.

In our work, we tried traditional filters, such as the median filter, but they did not

Figure 3.3: Imaging and analysis pipeline for reconstruction and quantification of liver microarchitecture: a,b) Location of tissue sectioning. c) Bile duct cross sections. d) Cells quantification. e) confocal micro-graph tissue section with architectural staining
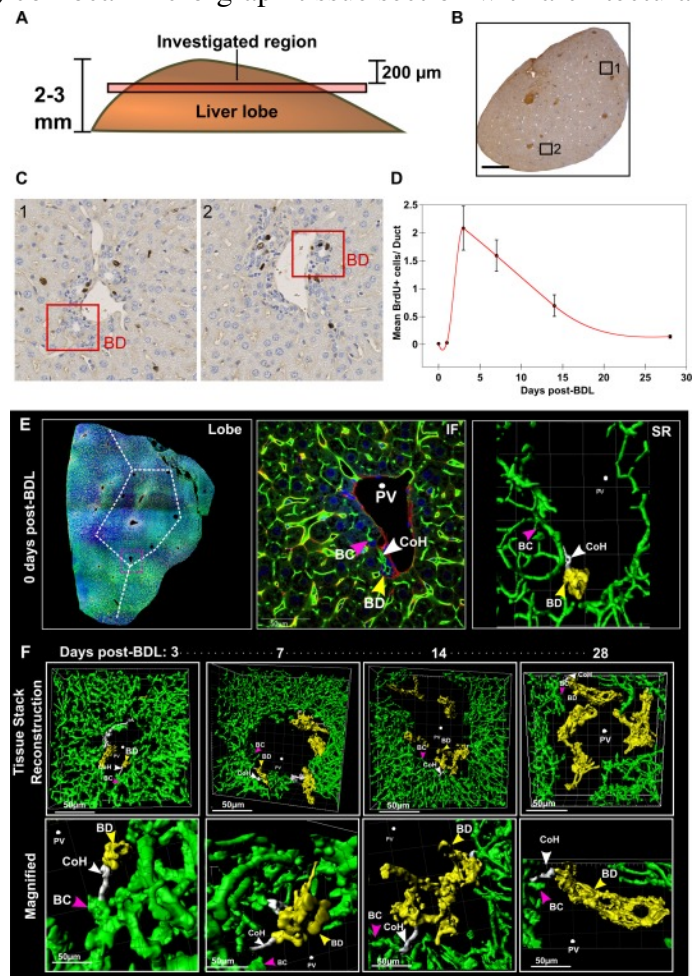
Source: (VARTAK et al., 2016).

provide significant results in enhancing details like we need for our purposes. That is because they are filters performing in all the data and do not discriminate the important structures we need to improve.

In our work, we propose a pipeline to enhance the structures present in the confocal images and visualize the microvasculature of bile ducts. The pipeline included a non-linear filtering step based on anisotropic diffusion that we reviewed in Section 2.2. We present details of this pipeline in Section 5.1.

3.3 Image Segmentation: Using Density-based Spatial Clustering

Density-based Spatial Clustering (Section 2.3) has been successfully applied in images datasets obtained from different sources for application in distinct domains (HUI;

LIU, 2017; HUI; LIU; PARK, 2018; LI et al., 2017; CELEBI; ASLANDOGAN; BERGSTRESSER, 2005; METE; KOCKARA; AYDIN, 2011; TRAN et al., 2012), including confocal images (MU et al., 2009; CHAN; CHENG; POON, 2007). In Table 3.1, we summarize the main characteristics of these works.

Mu et al. report that the density-based spatial clustering approach is useful for image segmentation of blood thrombus (MU et al., 2009). They did not use DBSCAN (ESTER et al., 1996), but a generalized version of the density-based clustering proposed by Chen et al. (CHEN; SMID; XU, 2005). Figure 3.4 illustrates their work.

Figure 3.4: DBSCAN clustering of blood thrombus: a)One slice of the confocal dataset. b)A reconstructed 3D clot. c)a 2D example of comparison: expert-produced result (solid curve) and Mu et al. algorithm (dashed curve)
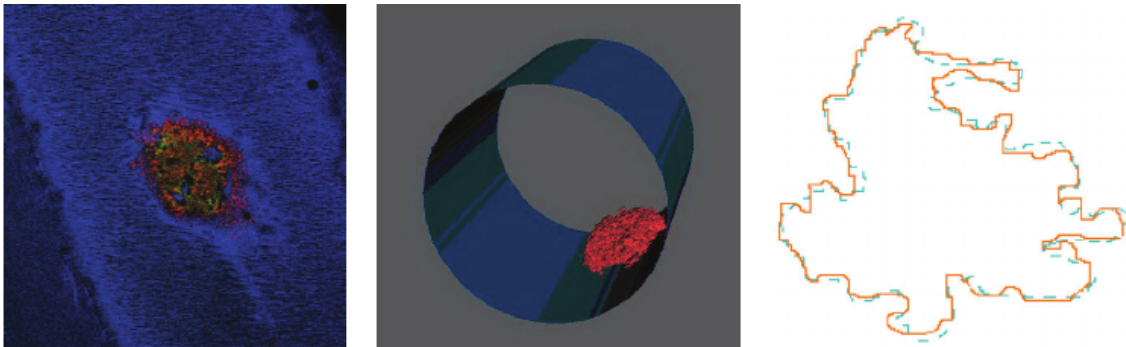
Source: (MU et al., 2009).

Chan et al. also modified a different density-based clustering method, known as DENCLUE (HINNEBURG; KEIM, 1998), to perform segmentation in confocal images to study gene expression in zebrafish (CHAN; CHENG; POON, 2007). The original method is based on a set of density distribution functions, which are, in fact, influence functions that model the influence of a given data point in its neighborhood. In the work by Chan et al., the Density-Based Segmentation (DBS) method, the density function of each pixel is calculated using the differences of pixel intensity between the neighboring pixels, which is an approximation of the gradient of each pixel like we did in our approach.

In the three surveyed papers based on the original version of DBSCAN (HUI; LIU, 2017; HUI; LIU; PARK, 2018; LI et al., 2017), two features are used to guide the clustering: pixel location (LI et al., 2017) and pixel intensity (with pixel location)(HUI; LIU, 2017; HUI; LIU; PARK, 2018). Figures 3.5 and 3.6 illustrate the works by Hui and Liu (HUI; LIU, 2017) and Li et al. (LI et al., 2017), respectively.

In our work, in addition to the spatial position and the size of the neighborhood of

the voxel, we also use its gradient magnitude to guide the clustering.

Figure 3.5: DBSCAN clustering of 3D Neutron Data: a) Clusters identified by DBSCAN. b) The two most prominent clusters. c) The next group of eight prominent clusters
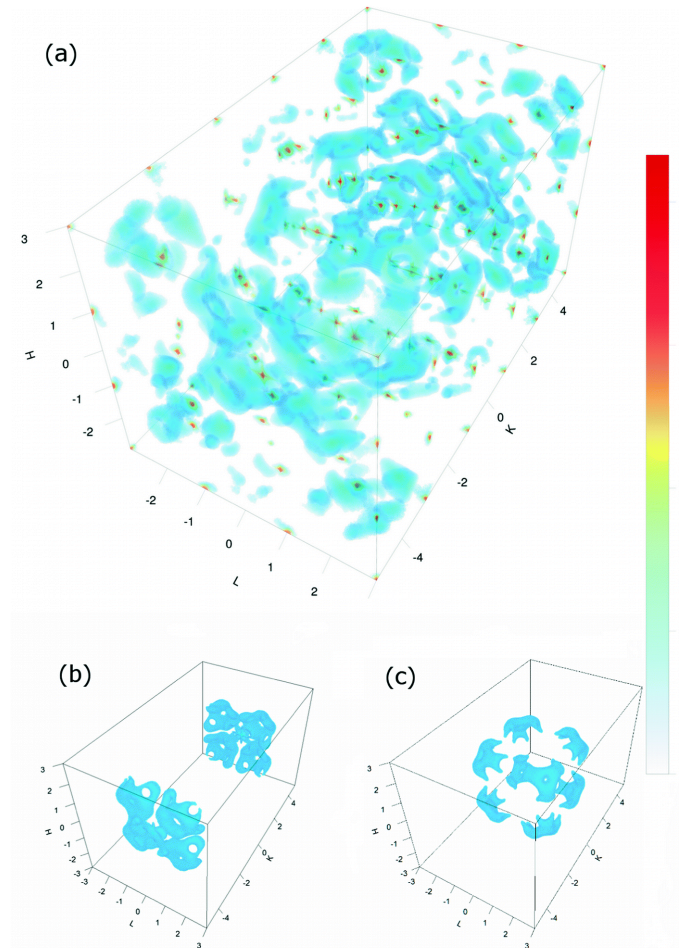
Source: (HUI; LIU, 2017).

Figure 3.6: DBSCAN clustering of CT angiography Data: a) Original volume rendering. b) Clustering result based on DBSCAN. c) Volume rendering after heart isolation. d)Heart cross section.
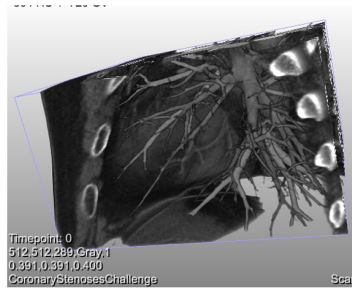
(a)
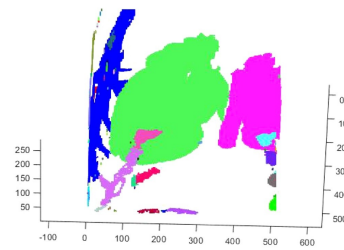
(b)
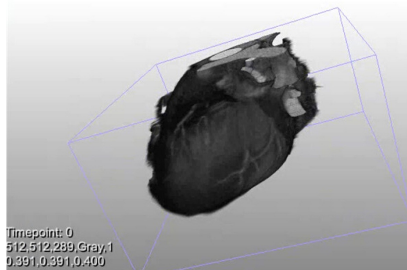
(c)
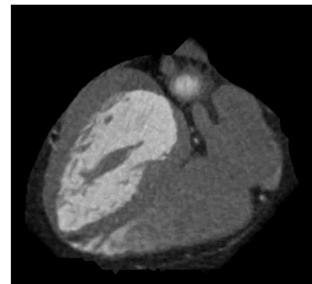
(d)

Source: (LI et al., 2017).

Table 3.1: Summarization of papers reporting density-based clustering in image datasets.

Ref.	Image Acquisition	Database	Image Size	3D Stack?	Clustering Algorithm	Clustering use	Features for clustering
(CELEBI; ASLANDOGAN; BERGSTRESSER, 2005)	Dermascopy.	135 color skin lesions images.	256 x 256 pixels.	No	GDBSCAN (Generalized DBSCAN) (SANDER et al., 1998).	Segmentation.	Color and pixel location.
(CHAN; CHENG; POON, 2007)	Confocal Microscopy.	4 images of zebrafish embryos.	Not mentioned.	No	Density-Based Segmentation (DBS).	Segmentation.	Intensity, pixel location.
(MU et al., 2009)	Confocal Microscopy.	15 Z-stacks of thrombi (clots). 80 2D slices per stack.	512 x 512 pixels.	Yes	Density-Based Clustering (DBC) algorithm (CHEN; SMID; XU, 2005).	Segmentation.	Pixel location.
(METE; KOCKARA; AYDIN, 2011)	Dermascopy.	100 RGB color images.	From 577 x 397 to 1921 x 1285 pixels.	No	Boundary driven density-based clustering (BD-DBSCAN).	Edge detection.	Pixel location.
(TRAN et al., 2012)	X-ray micro- tomography.	A synthetic 3D binary image containing 144 rod-like particles.	Not mentioned.	Yes	DBSCAN for binary 3D images XMT-DBSCAN.	Edge detection.	Pixel location.
(LI et al., 2017)	CT angiography.	12 datasets of coronary arteries.	Not mentioned.	Yes	Original formulation of DBSCAN (ESTER et al., 1996).	Segmentation.	Pixel location.
(HUI; LIU, 2017)	Diffuse Scattering Spectrometer and Tomography.	1 diffuse scattering dataset and 1 neutron tomography dataset.	701 x 701 x 701 and 1997 x 1997 x 1997 pixels.	Yes	Original formulation of DBSCAN (ESTER et al., 1996).	Remove noise and Segmentation.	Intensity and pixel location.
(HUI; LIU; PARK, 2018)	Neutron Single Crystal Diffraction.	1 dataset.	501 x 501 x 501 pixels.	Yes	Original formulation of DBSCAN(ESTER et al., 1996).	Data Reduction and Segmentation.	Intensity and pixel location.

Source: (BELTRAN et al., 2019).

3.4 Image Analysis: Fractal Dimension

In the field of microscopy, the effects of magnification and scaling on morphology and the methodological approaches necessary to measure self-similar structures are related to fractal geometry. Some degrees of self-similarity are present in numerous biological structures like plants, the spatial distribution of vessels of the circulatory system, the boundaries of some tumor types, the bronchial airways, the duct system of many secretory glands and the dendritic branching of neurons to name just a few (LANDINI, 2011).

Fractal dimension is used in image analysis studies to characterize a wide range of medical and biological signals. Some examples are electrocardiogram (ECG) and electroencephalogram (EEG) signals, brain imaging, mammography, retinal images and bone imaging (LOPES; BETROUNI, 2009). Fractal analysis is used in histopathology to interpret histological images (ANNADHASON, 2012).

Figure 3.7: Generalized fractal dimensions obtained from different images using the Box Counting Method: a,d) straight line, checkerboard, Sierpinski carpet and retinal vasculature. e,h) Fractal dimensions obtained from images a-d, respectively.
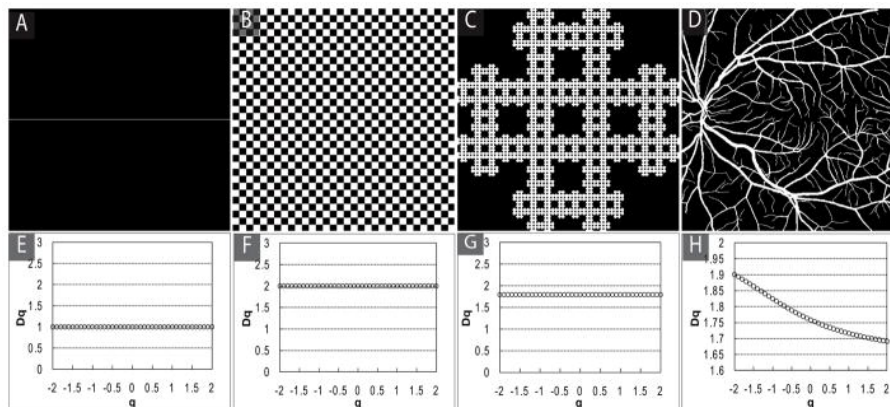
Source: (GOULD et al., 2011).

The use of fractal geometry in vivo confocal microscopy[1] has permitted the study of the microscopic morphology of the cornea with a resolution comparable to that of histological examination (IEVA et al., 2016).

Some authors apply fractal dimensions in the analyses of patterns (GOULD et al., 2011), such as the branching patterns in vascular structures of the human retina (ȚĂLU; GIOVANZANA, 2012). Texture and shape analysis are other applications of fractal analysis (BACKES, 2017) used in different contexts such as characterization of 3d shapes(BACKES

---

[1]In vivo confocal microscopy is a noninvasive imaging and diagnostic tool.

Figure 3.8: Fractal dimension plot of an epithelial cell nuclei. The plot scale is in log(pixels) and linear regression line (blue).
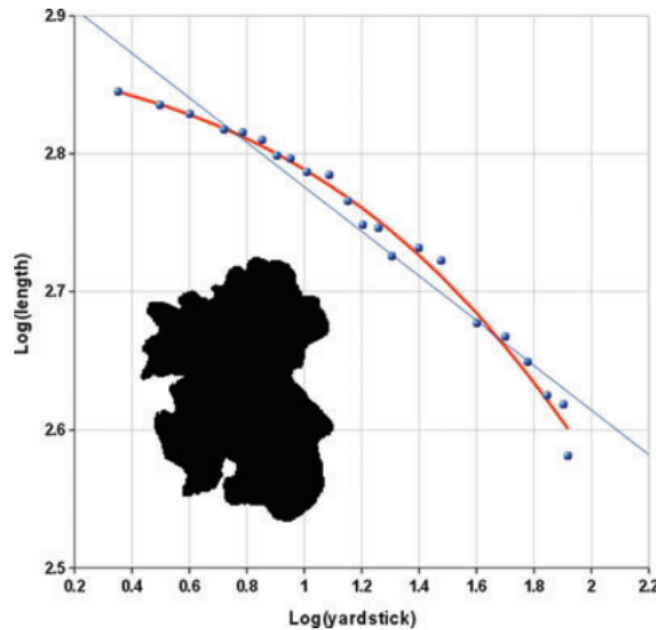
Source: (LANDINI, 2011).

et al., 2010), plant leaves identification (BACKES; CASANOVA; BRUNO, 2009), classification of coral specimens analyses (REICHERT et al., 2017), and medical images retrieval(BACKES; BRUNO, 2010). In some of these applications, multiscale fractal dimensions were also employed (REICHERT et al., 2017; BACKES; BRUNO, 2010).

3.5 Volume Visualization: Direct Volume Rendering of Confocal Microscopy Data

Volume rendering of microscopic data has been a research topic along the years due to its inherent visual complexity (BEYER et al., 2013). In addition to this complexity, confocal microscopy produces multi-channel data, and this creates a large amount of information. Transfer functions are the key to volume rendering of medical datasets (HADWIGER et al., 2006). As mentioned before, some approaches can use one-dimensional transfer functions and, in contrast, other approaches can use multidimensional transfer functions.

Regarding confocal data, Wan et al. developed an application for enhancement and rendering of confocal microscopy data, but their approach aimed at specific research on neuron cells (WAN et al., 2009; WAN et al., 2012). Figure 3.9 illustrates their work.

Kim et al. used gradient information combined with other properties such as curvature and texture to guide the transfer functions specification in the visualization of brain

cells (KIM et al., 2010). In Figure 3.10, we can observe different transfers functions being used to enhance different parts of the volumetric dataset.

Figure 3.9: Volume Rendering of Confocal Datasets: A) a zebrafish head dataset, rendered without any enhancement. B) the same dataset rendered from the same view directions, with enhancements applied. C) groups and different rendering modes can create clear visualizations when derived channels are presented.
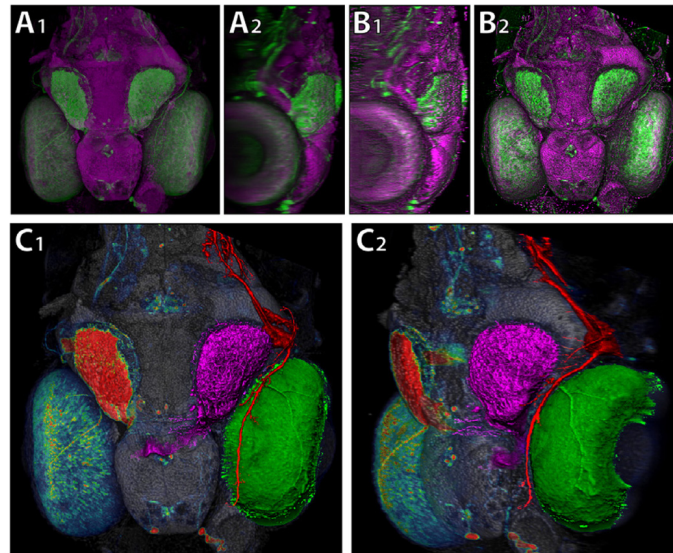
Source: (WAN et al., 2012).

Figure 3.10: Volume Rendering of mouse brain tissue (Confocal Datasets): All the images depict the same volume but with different transfer functions.

(a) Traditional Volume Rendering with only Three Channel Intensity

(b) Volume Rendering with PCA Domain. Connexin43 Expression within Astrocytes

(c) Volume Rendering with Isomap Domain. Pannexin1 Expression

(d) Volume Rendering with Isomap Domain. Astrocytes Enhancement
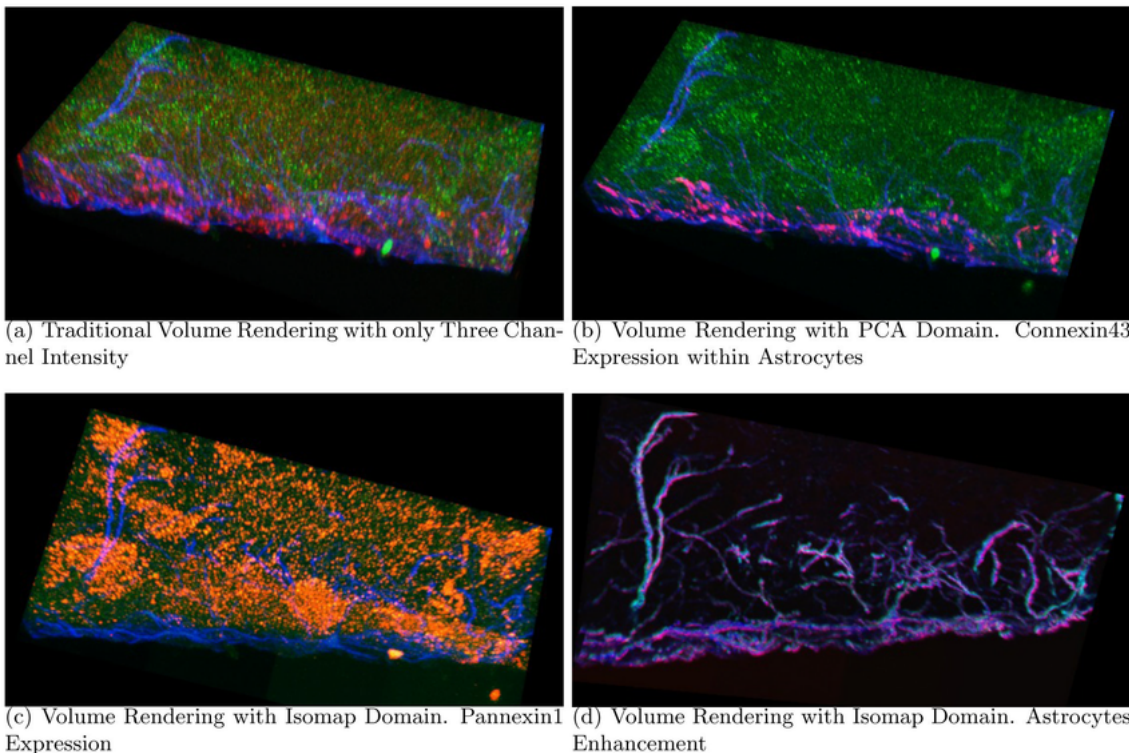
Source: (KIM et al., 2010).

As we observed, visualization strategies in these works have been applied to cells. Our work aims at the analysis and visualization of a macro-structure, the bile duct, and there are some factors associated with the staining process and image acquisition that affect the image quality in such thick specimens. Confocal images datasets are noisy and although they are multichannel data, each channel does not have much variation in intensity. This characteristic affects image segmentation as well as transfer functions design, and ultimately, the visualization of the structures of interest.

4 METHODOLOGY

Our research questions and hypotheses were stated in Chapter 1. In this chapter, we introduce the methodology we adopted for developing the methods that aimed at the characterization of confocal images datasets, with emphasis on bile ducts. Herein, we present the choices we made for investigating alternatives for visualization and analyses of bile ducts datasets. Specific details about the methods are presented in Chapters 5 and 6.

4.1 Improving the Visualization of Bile Ducts Confocal Images

The acquired confocal images are affected by several sources of noise. For improving the visualization of bile ducts micro-anatomy and validating our first hypothesis, we investigated methods for improving the quality of the dataset for later characterization.

We proposed an approach to enhance noisy confocal images of bile ducts by applying anisotropic diffusion (PERONA; MALIK, 1990) in our datasets. We compared the obtained image quality using quantitative measures such as the Peak Signal-to Noise Ratio (PSNR) and the Mean Square Error (MSE). We also have the results evaluated qualitatively by an expert. These approach and results are described in section 5.1.

4.2 Improving the Image Segmentation of Bile Ducts Confocal Images

The visual exploration of bile ducts in the liver is of relevant clinical interest, as it provides information related to the biliary diseases. However, the study of biliary disease has been limited by the inability to observe the bile ducts' structures. Moreover, very little is known about the internal structure of the bile duct, and there is a lack of image datasets providing a gold standard of the bile duct microanatomy. Figure 4.1 shows that in confocal microscopy, cellular structures visualization can be improved due to the good resolution in the z-axis. However, in the case of tissues like bile ducts, only a very thin layer through the tissue is visualized (ZIEGLER; BITTERMANN; HOECHLI, 2013).

Considering this scenario, segmenting images from bile ducts' tissue represent a challenge in hepatology research. Since cells are thinner than bile ducts, this poses a challenge associated with the data because the noisy sources affect the image quality in thick

Figure 4.1: Characteristics of confocal laser scanning microscopy of cellular structures and tissue.

(a) Cell culture: Cellular structures can be resolved due to the good resolution of depth dimension.
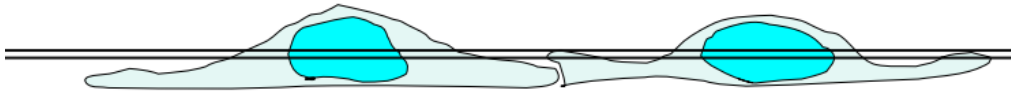
(b) Tissue: Only a very thin layer of the tissue is visualized.
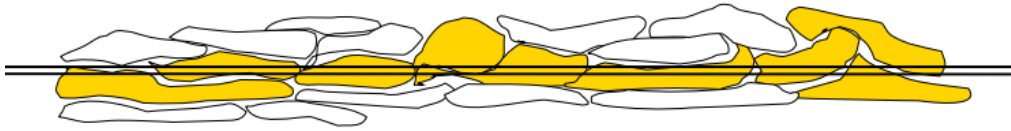
Source: (ZIEGLER; BITTERMANN; HOECHLI, 2013).

specimens to a greater extent than thinner specimens such as cells. Existing software provides standard methods for 3D segmentation of cells, and the challenges associated with tissue specimens require the development of new approaches and new algorithms. Although in some cases image segmentation can be performed by manually labeling the images, on confocal laser scanning microscopy (CLSM) images showing complex cellular structures such as microtubules, cytoplasm, or high multiplicity of features, manual segmentation is only possible in a limited way. Also, in many applications, CLSM image segmentation is a highly repetitive task motivating a growing interest in new CLSM images segmentation procedures (CALAPEZ; ROSA, 2010).

We explore the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm (ESTER et al., 1996), for extracting structures from confocal datasets. We propose an approach to adapt the DBSCAN algorithm using gradient information for guiding the clustering in the bile ducts. We describe our approach for improving the segmentation of bile duct structures and the results we obtained in section 5.2.

4.3 Improving Quantification of Bile Ducts Images

As mentioned before, digital microscopy, such as CLMS, allows biologists and biomedical end users to obtain high-resolution 3D data sets of biological objects, such as

cells and tissues (LEEUW; VERSCHURE; LIERE, 2006). However, there is still a lack of post-processing techniques to quantify specific structures of interest.

In the case of CLSM images, the intensity range of the desired structures is not known in advance (like the Hounsfield scale for CT scans), which leads to the situation that pre-segmentation can not be obtained automatically. For this reason, we can not use the intensity as the first feature to guide the quantification process. Then, quantification also represents a challenge in hepatology research.

According to the literature, a potential approach to assess shape quantification and spatial complexity characterization of 3D organisms are 3D fractal dimension analyses. This approach combines information from various spatial scales, thus enabling a holistic shape quantification (REICHERT et al., 2017). Fractal dimension analyses include the Fractal Dimension FD and Multiscale Fractal Dimension MFD as measures for characterization.

We explore the two measures, FD and MFD, for (volumetric) image quantification and obtain a characterization of bile ducts structures. We describe our approach for bile ducts characterization in Chapter 6.

5 IMPROVING THE VISUALIZATION AND THE IMAGE SEGMENTATION OF BILE DUCTS CONFOCAL IMAGES

In this chapter, we present the results we obtained in our investigation that aimed at image visualization and segmentation of bile ducts based on sets of confocal microscopy images. We have investigated the application of anisotropic diffusion in confocal microscopy images to deal with the noise found in confocal images. We explored different models for estimating the parameters of this method, which we describe in section 5.1. In section 5.2, we present the unsupervised machine learning method we explored for finding the relevant structures in the bile ducts images datasets.

5.1 Enhancing Confocal Microscopy Images

In our first study, we explored the challenge associated to the noise in confocal microscopy images. We summarize our results in a paper already published (BELTRAN et al., 2016). The full paper is available in Appendix A. We aimed at enhancing the quality of images of bile ducts and associated vessels, both obtained from confocal microscopy. In this context, image processing techniques are essential to remove noise and enhance the acquired images. We proposed the use of anisotropic diffusion to enhance the confocal images, and volumetric visualization techniques to create projections of the bile duct samples in a 3D space so the experts can examine the microvasculature distribution and interact with the bile ducts.

The main contribution of this study is the enhancement of the volumetric visualization of the Peribiliary Vascular Plexus (PVP) obtained from confocal microscopy images. We explore two models to estimate appropriate parameters for the anisotropic diffusion equation used to improve the original volume. As result, we enhance details that are hardly visualized in the original data. Moreover, using interactive manipulation like rotation and zooming operations on the resulting volumetric visualization, the hepatologists can have different views of the microvasculature.

5.1.1 Image Acquisition

The specimen preparation of bile ducts included a technique, developed at the Cincinnati Children's Hospital (DIPAOLA et al., 2013), which is innovative in relation to the mounting of the histologic sample for in situ staining in order to preserve the anatomic structures. Image acquisition was conducted at Hospital de Clinicas de Porto Alegre (HCPA-Brazil) using the confocal microscopy Leica TCS SP5. An example of a mouse bile duct after the specimen preparation is shown in Figure 5.1. Table 5.1 illustrates the information about the datasets obtained for each bile duct. The number of slices varies depending on the bile duct size and the microscope settings.

Figure 5.1: An extrahepatic mouse bile duct after the clearing and staining process.
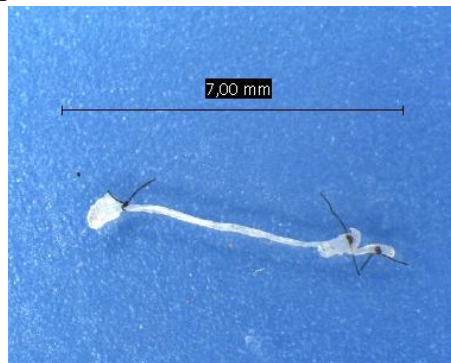
Source: The author.

Table 5.1: Size of Acquired Datasets from Extrahepatic Mice Bile Ducts.

Dataset Id	Image size (pixels)	# Slices
mouse1-day5	512x512	85
mouse2-day5	512x512	102
mouse3-day5	512x512	116
mouse2-day7	512x512	140
mouse3-day7	512x512	100
mouse4-day7	512x512	117

5.1.2 Anisotropic Diffusion in Confocal Microscopy Images

Anisotropic diffusion only affects parts where the gradient value is below a certain threshold. Therefore, the estimation of this parameter plays an important role in the anisotropic diffusion process. Methods for estimating a suitable value for the *k* parameter are useful in cases in which we have no idea about an appropriate value of the diffusion coefficient, and we would like to perform noise reduction with low loss of details (VOCI

et al., 2004). We compared the two approaches described in Section 2.2 to explore the *k* estimation in our images: the first model proposed by Perona and Malik (PERONA; MALIK, 1990) and a second model, proposed by Voci et al. (VOCI et al., 2004). Then, we use this *k* values as the input parameter to calculate the anisotropic diffusion. We calculated the *k* parameter in several slices from the same dataset. The parameter *t* that represents the time in the anisotropic diffusion was experimentally established as ten iterations. Figure 5.2 illustrates some results obtained with our approach, showing the volume rendering of the original dataset and the resulting volume after the anisotropic diffusion. The network of vessels surrounding the bile duct can be clearly observed.

For quantitative analyses, we calculated the Peak Signal-to-Noise Ratio (PSNR) and the Mean Square Error (MSE) for comparing the enhanced images with the original images. As for quality, the measured values of MSE should be small and PSNR should be large. The PSNR and the MSE are defined by:

$$MSE = \frac{\sum_{i=1}^{r} \sum_{j=1}^{c} \mid I(i,j) - \hat{I}(i,j) \mid^2}{r.c} \tag{5.1}$$

where $I(i,j)$ is the original image, $\hat{I}(i,j)$ is the enhanced image, and $r.c$ the size of the image.

$$PSNR = 10 \log_{10}(\frac{MAX_I^2}{MSE}) \tag{5.2}$$

where $MAX_I = 2^n - 1$ and n is the number of bits. Since the confocal images are 8-bits depth, n is set to 255.

Table 5.2 and Table 5.3 show the *k* values and the respective measures of MSE and PSNR of the enhanced images.

After applying the anisotropic diffusion with both models, we verified that PSNR values are very similar in the resulting images. In terms of image quality, this means that the two models are adequate for enhancing our images.

For enhancing the whole confocal dataset using the two models presented in Section 2.2, we calculated the average value of k considering all slices. The results for the enhanced volume are shown in 5.2. As for qualitative analyses, we invited a senior hepatologist to describe how he found the enhanced volume in comparison to the original volume. According to the hepatologist, the resulting volumetric visualization solves some problems associated with the original data such as noise and superposition of vessels. He also commented that the microvasculature was clearly discernible, which gives a better

Figure 5.2: Comparison of the anisotropic diffusion using the two models for estimating the *k* parameter: (a) Original reconstructed volume, (b) Anisotropic filtering results using Perona and Malik's model and (c) Anisotropic filtering results using Voci et al.'s model.

(a) Original reconstructed volume.
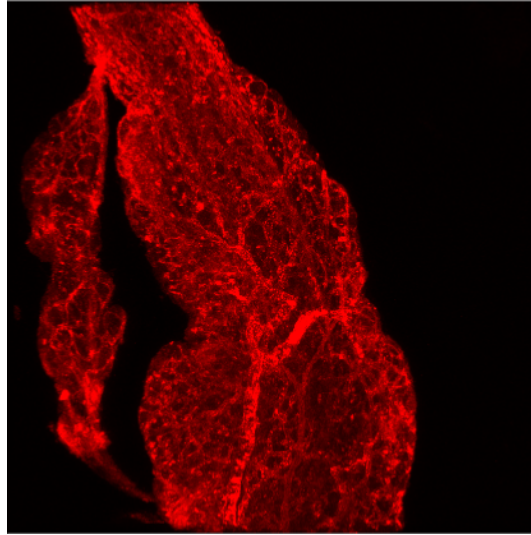
(b) Anisotropic filtering results using Perona and Malik's model.
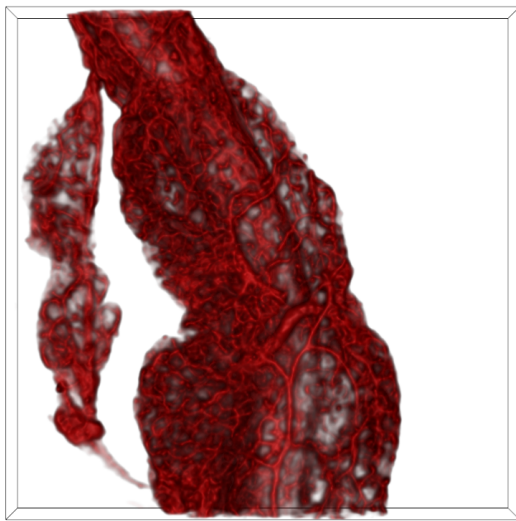
(c) Anisotropic filtering results using Voci et al.'s model.
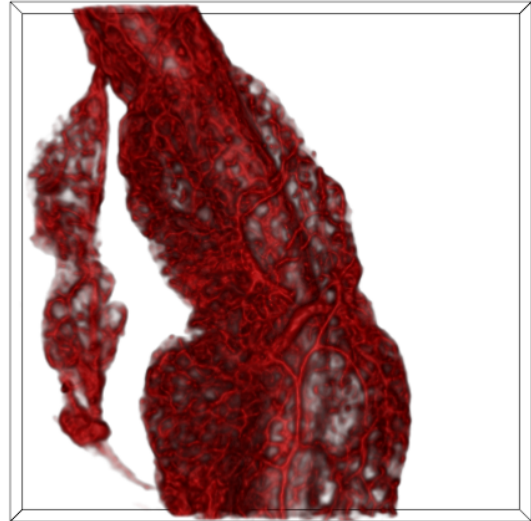
Source: The author.

idea of the 3D distribution of the vessels. This observation is really important because it represents that our method allows hepatologists to evaluate morphological alterations in the bile ducts.

Table 5.2: *k* parameter according to Perona and Malik's model and MSE, PSNR of the enhanced images

mouse2-day7	*k* parameter	MSE	PSNR
Slice #1	131	25.0112	34.1494
Slice #10	139	25.7968	34.0151
Slice #20	139	25.9635	33.9871
Slice #30	138	26.9295	33.8285
Slice #43	137	29.4593	33.4385
Slice #50	155	30.9956	33.2177
Slice #60	156	32.7181	32.9829
Slice #70	145	34.3088	32.7767
Slice #80	138	34.8454	32.7093
Slice # 90	144	35.4471	32.6349
Slice #100	155	34.242	32.7852
Slice #110	137	31.7267	33.1165
Slice #120	147	27.1433	33.7941
Slice #130	141	22.3949	34.6293
Slice #140	144	18.1274	35.5474

Table 5.3: *k* parameter according to Voci et al.'s model and MSE, PSNR of the enhanced images

mouse2-day7	*k* parameter	MSE	PSNR
Slice #1	243	25.0208	34.1477
Slice #10	240	25.8104	34.0128
Slice #20	242	25.9736	33.9854
Slice #30	242	26.9542	33.8245
Slice #43	236	29.4543	33.4393
Slice #50	238	30.9819	33.2197
Slice #60	223	32.7058	32.9845
Slice #70	124	34.2854	32.7797
Slice #80	123	34.8371	32.7103
Slice # 90	123	35.4506	32.6345
Slice #100	126	34.2397	32.7854
Slice #110	139	31.7235	33.1169
Slice #120	140	27.1388	33.7948
Slice # 130	159	22.4014	34.628
Slice #140	182	18.141	35.5441

5.2 Extracting Relevant Structures from Confocal Microscopy Images

In our second study, we investigated the application of unsupervised machine learning to extract relevant structures from confocal microscopy datasets representing bile ducts. Our approach consists of pre-processing, clustering, and 3D visualization. For clustering, we explore the density-based spatial clustering for applications with noise (DBSCAN) algorithm, using gradient information for guiding the clustering. We obtained a better visualization of the most prominent vessels and internal structures. We summarize these results in our paper (BELTRAN et al., 2019), which is available in Appendix B.

To the best of our knowledge, there is no reported application of this technique in the study of data from hepatological samples. Our approach is based on adapting the DBSCAN method for extracting structures from confocal images of bile ducts. The main challenge is to find the appropriate similarity features between voxels that allow for differentiating such structures.

The main contributions of this study are the use of gradient information as a feature to guide the clustering process and the proposal of a specific pre-processing step that can also be used in other applications involving confocal microscopy images.

5.2.1 Image Acquisition

The input datasets that we use in this study were acquired at the Cincinnati Children's Hospital (DIPAOLA et al., 2013). The mice bile duct was stained with two different fluorescent antibodies, α-tubulin and Cytokeratin CK, to mark different tissues. The resulting dataset consists of two channels: the first one (red channel) represents the microvasculature or blood vessels around the bile duct with (α-tubulin staining) (Figure 5.3a); the second one (green channel) represents the bile duct wall containing the peribiliary glands with CK staining (Figure 5.3b).

5.2.2 Discovering structures in bile ducts confocal images datasets

In hepatology research, the *a priori* labels (ground truth) on the pixels are not available. Creating labels by hand is a hard task due to the complexity of the structures and

Figure 5.3: View of a single slice from a bile duct dataset: (a) the red channel encodes the microvasculature, while (b) the green one encodes the peribiliary glands. The dataset has 192 slices (512 x 512 image each) ≈ 50 millions of points.

(a) Red channel   (b) Green channel   (c) Superimposed channels
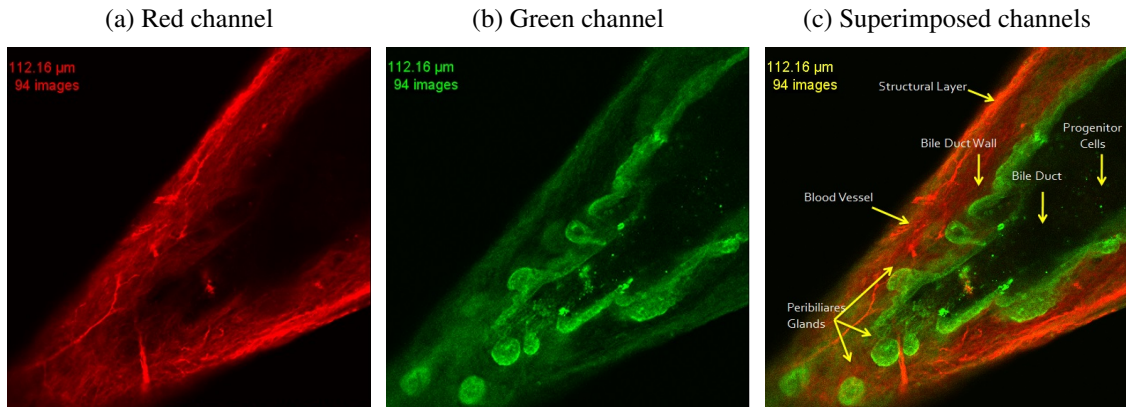
Source: the author

the high dimensionality of data. Thus, we formulate our problem of extracting structures from these data sets as a clustering problem.

5.2.2.1 Pre-processing

We use two operations to normalize the image stacks and prepare the data for the clustering process.

- *Normalization*: We apply contrast stretching to increase the visibility of the structures.

- *Data Reduction*: We remove all points with intensity 0 (background), for eliminating unnecessary points and reducing the amount of data that will undergo the clustering phase.

5.2.2.2 Density-Based Spatial Clustering

The spatial information, i.e., the coordinates (x,y,z) are a typical candidate **clustering feature**. As for images, any kind of pixel (or voxel) attribute can be used as a clustering feature. Confocal images have an inhomogeneous intensity inherent to the fluorescent staining process (TORIWAKI; YOSHIDA, 2009), and the gradient was investigated as a more robust candidate feature. After experimenting with the intensity and gradient values, we found out that the gradient was a richer source of information for distinguishing the regions of interest. Then, we adopted the gradient magnitude to guide the clustering process (BELTRAN et al., 2019).

- **Determining the parameters for 3D clustering**: In the original DBSCAN algorithm (ESTER et al., 1996), the key idea is that, for each point of a cluster, the neighborhood defined by a given radius *(eps)* around it has to contain at least a minimum number of points *(minPts)*, i.e., the local density in the neighborhood has to exceed some threshold. Based on some heuristics we determined the appropriate *eps* parameter, and set *minPts* empirically. In the following, we give details about the configuration of DBSCAN for clusterizing our dataset.

  The ***eps*-neighborhood of a point** dictates the maximum distance (radius) between two points for them to reside in the same neighborhood. A general heuristic to establish the value for *eps* is by computing the *k*-nearest neighbor distances. However, in a recent application of DBSCAN (HUI; LIU, 2017), a simplified calculation for *eps* was proposed. The author's idea is that the coordinates of the data points in the case of 3D image datasets are uniformly distributed voxels. Then, it is possible to use the Cartesian coordinate system and Euclidean distance to obtain the neighborhood. Values of *eps* in the interval $[1, \sqrt{2}]$ includes the six first nearest neighbors, values in $[\sqrt{2}, \sqrt{3}]$ to include the twelve second nearest neighbors, and so on. Based on this last approach, we fixed the *eps* to $1.7 \approx (\sqrt{3})$. This value means that the local density function uses 18 nearest neighbors of a given point data in the clustering.

  minPts denotes the minimum number of points located in an *eps*-neighborhood, and is data dependent. If we select a low *minPts* value, we get more clusters from noise. We have experimented *minPts* values from 50 to 300, and finally set it to 200 points for the green channel and 50 points for the red channel.

  The density in a neighborhood is just the sum of the **weights** of the points inside the neighborhood. By default, each data point has weights 1, so the density estimate for the neighborhood is just the number of data points inside the neighborhood. We can use the parameter *weight* to change the importance of points (HAHSLER; PIEKENBROCK; DORAN, 2017). The *weight* is an optional parameter to perform clustering based on a specific feature.

  As described before, we have chosen the gradient magnitude as a feature to guide the clustering. We follow the model for the *weight* parameter proposed by (HUI; LIU, 2017). However, we use the gradient magnitude instead of intensity. We calculate the *gradient magnitude* for every point of the dataset considering the x, y, and z dimensions. Then, we take a specific value of gradient magnitude as a *threshold*. We fixed the threshold empirically as 20 for the red and the green channel. Any

data point with gradient magnitude less than the threshold will take the weight of 1, while the data points with gradient magnitude greater than the threshold will have their weights assigned to the difference

- **DBSCAN applied to 3D data points using gradient information**: We used the DBSCAN R package (HAHSLER; PIEKENBROCK; DORAN, 2017) to perform the clustering on the 3D data points. As mentioned before, we configured the *eps* and *minPts* parameters and feed the algorithm with a list of data points containing their x, y, and z coordinates and the weight obtained from the gradient information. It is important to recall that background voxels are not considered in the clustering phase.

  In this way, the clustering method uses both information (gradient and spatial location) to obtain at least the *minPts* data points for each cluster. The output is the list of points labeled with the cluster identification of each point as well as basic numbers about the clusters detected. Then, we use the original volumetric dataset again, and voxels belonging to the cluster of interest form a new volume that is passed to the visualization module.

5.2.3 Results and Discussion

5.2.3.1 Microvasculature: Red Channel

Figure 5.4 shows 3D visualizations of selected regions in the dataset that contain the microvasculature (red channel) of the bile duct. We obtained a total of 2478 clusters from the clustering process in the red channel. Due to the large number of clusters detected by DBSCAN, we summarize the results in the plot shown in Figure 5.4a, and use it to select the clusters for 3D visualization. Figure 5.4b shows the original dataset rendered with direct volume rendering.

We identified most of the points as belonging to clusters 0 and 1. The cluster 0 is composed by 250,081 noise points, which can be discarded for visualization and analyses purposes. In other words, cluster 0 contains all the points that do not satisfy the conditions to belong to a cluster. Since clusters 1 to 2478 represent the detected objects, and cluster 1 is the largest one among them representing a connected region, it is the one that best represents the microvasculature (Figure 5.4c). Figure 5.4a shows that the clusters 2 to 2478 contain a lower quantity of points, and so we can also consider these points as noise

(Figure 5.4d).

Figure 5.4: 3D visualization of the microvasculature of a bile duct:(a) summarization of clusters detected by DBSCAN (b) original volume (c) represent the most prominent vessels extracted as cluster 1, and those shown in (d) are considered noise and were detected as clusters 2 to 2478.

(a) Number of points detected per cluster.

(b) Original Volume.

(c) Cluster 1.

(d) Clusters 2 to 2478.
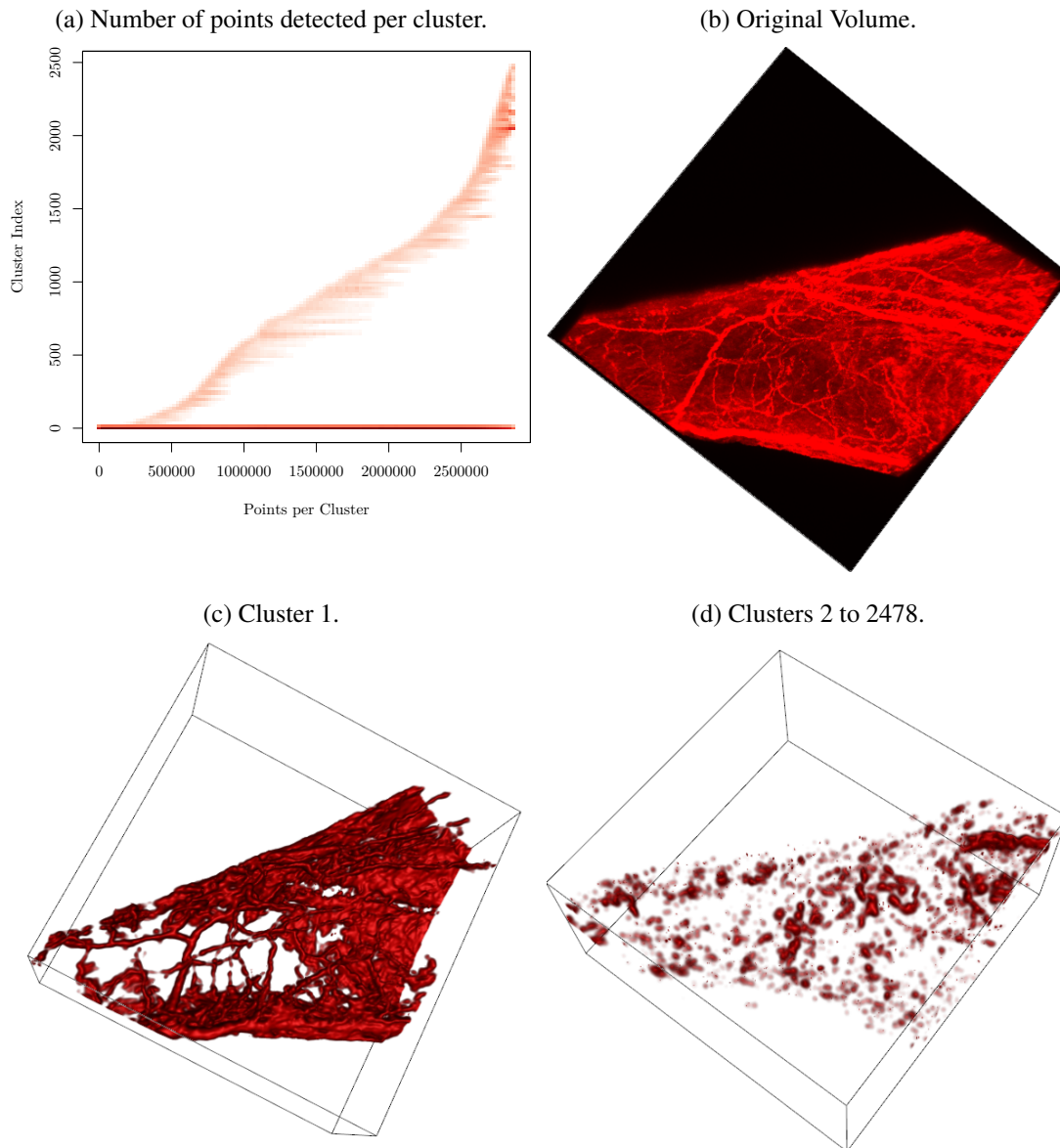
Source: the author

5.2.3.2 Peribiliary Glands (PBGs): Green Channel

In the case of the green channel, we obtained a total of 3,603 clusters (more clusters than in the red channel). In this case DBSCAN detected 1,998,026 noise points. Figure 5.5 shows the 3D visualization of data points belonging to clusters chosen among the ones that were detected in the green channel dataset. As we did in the processing of

the red channel, we summarized the DBSCAN result in a plot (Figure 5.5a) that allowed us to analyze and select for visualization only the relevant clusters. Cluster 0 corresponds to the noise points. For the other clusters, we find a behavior similar to the red channel: cluster 1 is the most prominent one, representing a connected region containing the internal bile duct wall and the peribiliary glands. The other clusters, i.e., clusters 2 to 3,603, contain a lower quantity of points, and we can also consider them as noise. While Figure 5.5b shows the original volume, cluster 1 representing the internal bile duct wall and the PBGs are presented in Figure 5.5c. Figure 5.5d present the clusters $[2 - 3603]$, considered as noise.

5.2.3.3 Discussion

When comparing our work to others that adapt DBSCAN for their application domain, we found different approaches. For example, Celebi et al. used the original DBSCAN method for segmenting 2D digital images of skin lesions (CELEBI; ASLAN-DOGAN; BERGSTRESSER, 2005), while Tran et al. presented a version of DBSCAN to process 3D binary images, using the coordinates of the original image data and solving a known instability issue of the original DBSCAN in classifying border points of adjacent objects (TRAN et al., 2012). Our method is not limited to binary images and also uses the original data points' coordinates. Regarding the use of additional features to guide clustering with DBSCAN, only two works adopt this approach. Hui and collaborators (HUI; LIU, 2017; HUI; LIU; PARK, 2018) use the intensity value as a feature for selecting the points during the clustering. In our work, besides the spatial position and the size of the neighborhood, we use the gradient information to select the points during the clustering.

Figure 5.5: 3D visualization of the bile duct wall and PBGs: (a) summarization of clusters detected by DBSCAN (b) original volume (c) represent mostly the PBGs identified as cluster 1, and those shown in (d) are also considered noise and were detected as clusters 2 to 3,603.

(a) Number of points detected per cluster.

(b) Original Volume.

(c) Cluster 1.

(d) Clusters 2 to 3603.
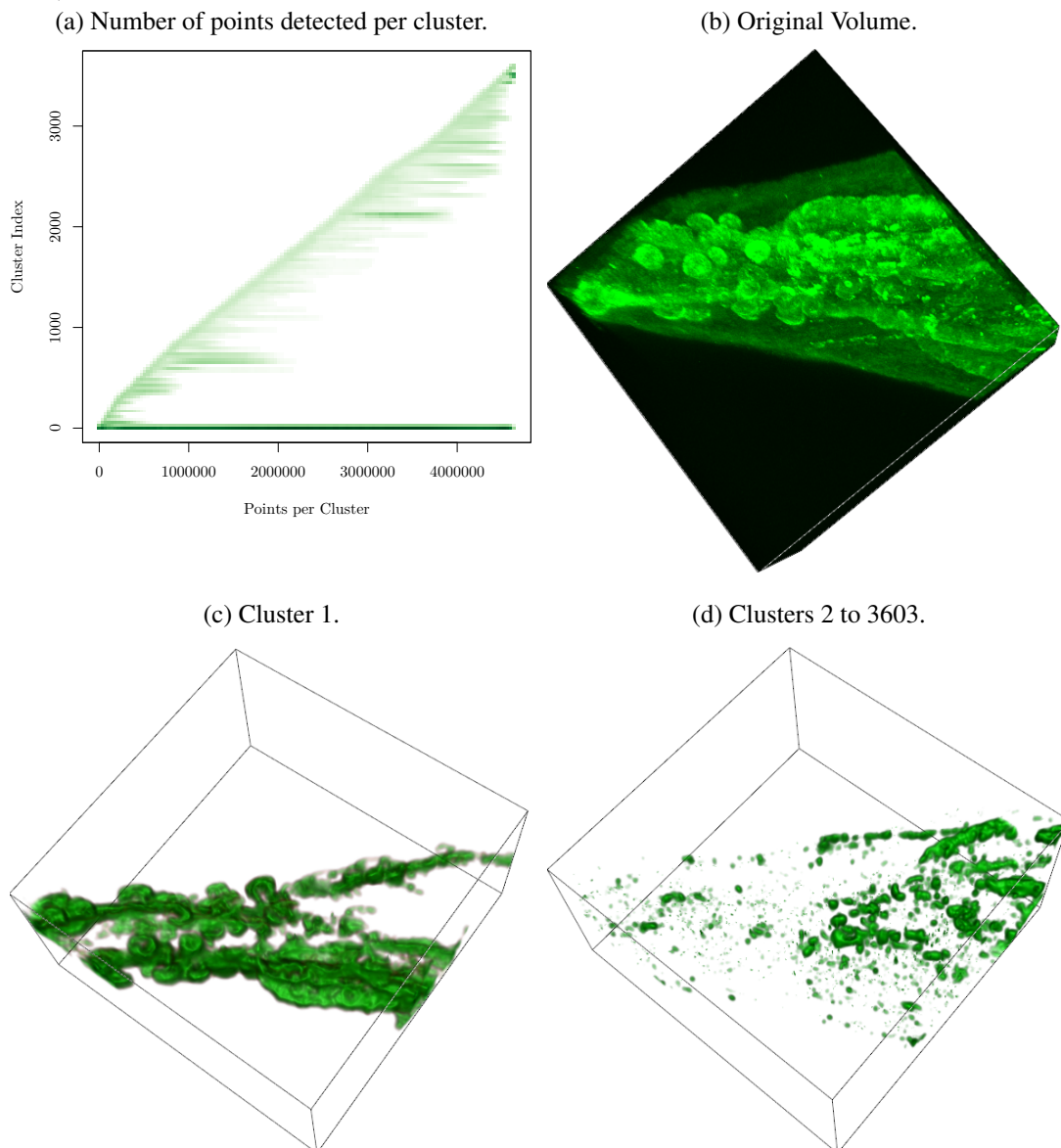
Source: the author

6 CHARACTERIZATION OF BILE DUCTS STRUCTURES

In this chapter, we report our last set of analyses aimed at the structural characterization of bile ducts based on sets of confocal microscopy images. Firstly, we report the fractal dimension analysis that we performed with the original datasets (Section 6.1). Then, in Section 6.2, we present a similar analysis of the structures isolated by the previously described segmentation method (Section 5.2). Finally, we explore the multiscale fractal dimension approach (section 6.3), which eventually provided us with initial evidence for the fractal characterization of bile ducts structures.

6.1 Fractal Dimension Analysis of Confocal Images Datasets

From our very first discussions with hepatologists, we learned that non-normal bile ducts presented severe changes in their relevant structures, i.e., microvasculature, duct wall, and peribiliary glands. Then, we hypothesized that shape analysis methods could help in the characterization of normal and non-normal bile ducts.

From the existing shape analysis methods (COSTA; CESAR, 2000) and also based on recent literature (REICHERT et al., 2017), we chose to explore 3D fractal dimension analysis methods to characterize bile ducts structures.

Fractal dimension can assume non-integer values related to the complexity of a fractal object (i.e., a non-Euclidean geometry) (BACKES et al., 2010). This property enables us to quantify objects and shapes in terms of space occupation (COSTA; CESAR, 2000). Several methods are described in the literature to calculate the fractal dimension. However, according to the study by Backes et al. (BACKES; BRUNO, 2005), the Bouligand-Minkowski method presents the most accurate results. The Bouligand-Minkowski method (TRICOT, 1994; COSTA; CESAR, 2000) has shown to be very sensitive to structural changes of the object (PLOTZE et al., 2005; BRUNO et al., 2008; BACKES et al., 2010; REICHERT et al., 2017; BACKES, 2017), which is likely to happen in non-normal bile ducts.

6.1.1 Materials and Methods

In the first analysis we report herein, we defined the pipeline shown in Figure 6.1 to perform fractal dimension analysis on the original confocal datasets, i.e., non-segmented volumetric datasets. Although we say that the datasets are not segmented, we applied a threshold for removing the background voxels that could impact the fractal dimension computation.

The pipeline for the first fractal dimension analysis is described as follows:

Figure 6.1: Overview of the data flow pipeline for our fractal dimension study using the original datasets.
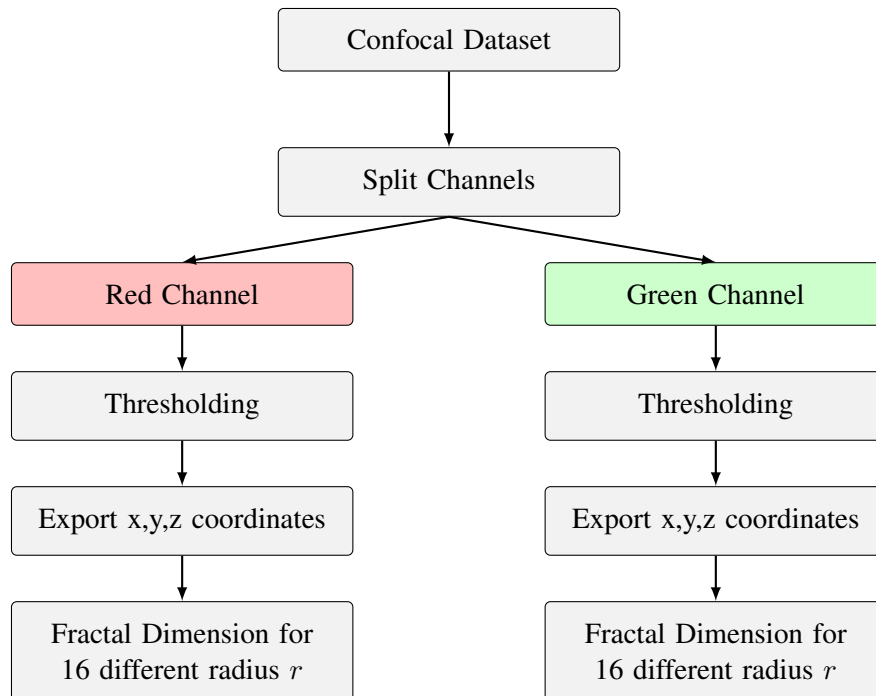
Source: The author.

Table 6.1: Datasets for Fractal Dimension Study

Dataset Id	From	size (pixels)	Threshold for Red Channel	Threshold for Green Channel
mouse1-day5	POA	512x512x85	30-255	30-255
mouse2-day5	POA	512x512x102	30-255	30-255
mouse3-day5	POA	512x512x116	30-255	30-255
mouse2-day7	POA	512x512x140	30-255	30-255
mouse3-day7	POA	512x512x100	30-255	30-255
mouse4-day7	POA	512x512x117	30-255	30-255
Day9CKa-tubulin	USA	512x512x192	100-255	46-255
Day9CKa-tubulin0	USA	512x512x165	100-255	46-255
Day9CKa-tubulin1	USA	512x512x232	100-255	46-255
Day9CKa-tubulin2	USA	512x512x171	100-255	46-255
Day9CKa-tubulin3	USA	512x512x33	100-255	46-255
Day9CKa-tubulin4	USA	512x512x55	100-255	46-255
Day9CKa-tubulin5	USA	512x512x50	100-255	46-255

- *Datasets:*

  We use two datasets groups for our experiments. The first dataset group is composed of six confocal datasets acquired in the Hospital de Clinicas de Porto Alegre. This group is described in Section 5.1.1. The second dataset group is composed of seven confocal datasets acquired in the Cincinnati Children's Hospital, as we described in Section 5.2.1. Although different confocal microscopes were used for image acquisition, both datasets are composed of two fluorescence channels containing bile duct structures. We summarize the datasets in Table 6.1. We refer to *POA* and *USA* the datasets acquired in Porto Alegre and Cincinnati, respectively.

- *Thresholding:*

  For each dataset, images are binarized on several threshold levels. Since we have two fluorescence channels that are labeling juxtaposed microstructures, we split each dataset in Red and Green channels. The threshold criterion was determined by experimental procedures. We report these levels in Table 6.1.

- *Export x,y,z coordinates:* Resulting binarized datasets were exported as x,y,z coordinates for subsequent fractal dimension computation.

- *Fractal Dimension FD:*

  The fractal dimension was computed using the Bouligand-Minkowski method (BACKES et al., 2010) (Equation 2.7). Section 2.5 describes the method to calculate the fractal dimension on 3D shapes.

6.1.2 Results and Analysis

Fractal Dimension FD was computed for each channel, for each bile duct dataset, using Equation 2.7 with 16 different radius (5-20). Thus, for each fluorescence channel, we obtain 16 fractal dimension values for each bile duct image dataset. We will refer to this as 'bile-duct FD descriptor'. In this first experiment, we computed the fractal dimension of the original datasets; the only preprocessing step was the thresholding.

Tables 6.2 and 6.3 summarize the results, presenting the bile-duct FD descriptors for all bile duct images datasets. These 'bile-duct FD descriptors' correspond to the rows of Tables 6.2 and 6.3. Spheres with radius larger than 20 were overlapping too much, and the results could compromise the characterization power we need from the fractal dimension.

Table 6.2: Fractal Dimension values obtained for the original datasets (Red Channel). Each row is a bile-duct FD descriptor.

Dataset Id	From	Threshold for Red Channel	Fractal Dimension														
			r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	r17	r18	r19	r20
mouse1-day5	POA	30 - 255	2.8442	2.8359	2.8260	2.8156	2.8037	2.7924	2.7810	2.7691	2.7571	2.7449	2.7332	2.7216	2.7095	2.6975	2.6859	2.6746
mouse2-day5	POA	30 - 255	2.6377	2.6379	2.6377	2.6365	2.6346	2.6325	2.6296	2.6262	2.6222	2.6176	2.6127	2.6075	2.6016	2.5954	2.5891	2.5827
mouse3-day5	POA	30 - 255	2.7138	2.7136	2.7097	2.7032	2.6937	2.6833	2.6718	2.6589	2.6453	2.6311	2.6168	2.6025	2.5872	2.5717	2.5565	2.5415
mouse2-day7	POA	30 - 255	2.8523	2.8430	2.8332	2.8234	2.8123	2.8020	2.7917	2.7812	2.7706	2.7599	2.7496	2.7395	2.7290	2.7185	2.7085	2.6988
mouse3-day7	POA	30 - 255	2.5814	2.5962	2.6065	2.6122	2.6153	2.6159	2.6142	2.6110	2.6066	2.6011	2.5950	2.5883	2.5808	2.5728	2.5647	2.5564
mouse4-day7	POA	30 - 255	2.7994	2.7915	2.7828	2.7736	2.7630	2.7528	2.7424	2.7316	2.7207	2.7096	2.6989	2.6883	2.6773	2.6662	2.6556	2.6452
Day9CKa-tubulin	USA	100 - 255	2.5656	2.5591	2.5549	2.5511	2.5472	2.5440	2.5407	2.5374	2.5341	2.5307	2.5273	2.5240	2.5204	2.5167	2.5131	2.5095
Day9CKa-tubulin0	USA	100 - 255	2.4387	2.4207	2.4091	2.4005	2.3933	2.3884	2.3846	2.3817	2.3797	2.3783	2.3774	2.3768	2.3765	2.3762	2.3760	2.3757
Day9CKa-tubulin1	USA	100 - 255	2.6472	2.6374	2.6305	2.6249	2.6197	2.6157	2.6121	2.6089	2.6060	2.6033	2.6010	2.5989	2.5968	2.5948	2.5928	2.5908
Day9CKa-tubulin2	USA	100 - 255	2.5649	2.5583	2.5556	2.5546	2.5550	2.5564	2.5580	2.5599	2.5616	2.5630	2.5642	2.5649	2.5653	2.5653	2.5648	2.5639
Day9CKa-tubulin3	USA	100 - 255	2.1341	2.0690	2.0172	1.9751	1.9356	1.9046	1.8783	1.8553	1.8359	1.8198	1.8072	1.7973	1.7893	1.7834	1.7793	1.7768
Day9CKa-tubulin4	USA	100 - 255	2.3236	2.2688	2.2230	2.1846	2.1475	2.1175	2.0915	2.0684	2.0483	2.0310	2.0167	2.0049	1.9948	1.9865	1.9803	1.9758
Day9CKa-tubulin5	USA	100 - 255	2.3791	2.3259	2.2814	2.2446	2.2098	2.1821	2.1585	2.1375	2.1193	2.1035	2.0903	2.0791	2.0691	2.0607	2.0540	2.0486

Table 6.3: Fractal Dimension values obtained for the original datasets (Green Channel). Each row is a bile-duct FD descriptor.

Dataset Id	From	Threshold for Green Channel	Fractal Dimension														
			r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	r17	r18	r19	r20
mouse1-day5	POA	30 - 255	2.8225	2.8090	2.7942	2.7793	2.7625	2.7464	2.7301	2.7128	2.6951	2.6770	2.6592	2.6415	2.6227	2.6037	2.5853	2.5672
mouse2-day5	POA	30 - 255	2.6134	2.6261	2.6326	2.6340	2.6321	2.6279	2.6218	2.6141	2.6054	2.5957	2.5856	2.5752	2.5639	2.5523	2.5409	2.5295
mouse3-day5	POA	30 - 255	2.6188	2.5907	2.5623	2.5346	2.5039	2.4758	2.4488	2.4217	2.3956	2.3702	2.3464	2.3240	2.3014	2.2798	2.2598	2.2412
mouse2-day7	POA	30 - 255	2.7590	2.7496	2.7361	2.7200	2.6997	2.6793	2.6583	2.6363	2.6143	2.5926	2.5723	2.5533	2.5345	2.5171	2.5015	2.4877
mouse3-day7	POA	30 - 255	2.7728	2.7522	2.7301	2.7075	2.6811	2.6557	2.6298	2.6027	2.5753	2.5476	2.5209	2.4948	2.4677	2.4411	2.4158	2.3915
mouse4-day7	POA	30 - 255	2.8034	2.7898	2.7752	2.7598	2.7410	2.7219	2.7019	2.6800	2.6571	2.6332	2.6094	2.5857	2.5605	2.5352	2.5107	2.4870
Day9CKa-tubulin	USA	100 - 255	2.7196	2.7300	2.7379	2.7427	2.7460	2.7475	2.7474	2.7463	2.7443	2.7415	2.7382	2.7344	2.7301	2.7254	2.7204	2.7153
Day9CKa-tubulin0	USA	100 - 255	2.7705	2.7626	2.7552	2.7478	2.7393	2.7312	2.7229	2.7143	2.7055	2.6966	2.6880	2.6794	2.6703	2.6612	2.6523	2.6436
Day9CKa-tubulin1	USA	100 - 255	2.5393	2.5451	2.5511	2.5555	2.5595	2.5628	2.5650	2.5667	2.5679	2.5684	2.5684	2.5678	2.5668	2.5651	2.5631	2.5607
Day9CKa-tubulin2	USA	100 - 255	2.6913	2.6960	2.7002	2.7027	2.7041	2.7045	2.7036	2.7019	2.6995	2.6964	2.6928	2.6889	2.6843	2.6794	2.6743	2.6691
Day9CKa-tubulin3	USA	100 - 255	2.4235	2.4059	2.3918	2.3801	2.3693	2.3610	2.3537	2.3469	2.3403	2.3336	2.3270	2.3204	2.3132	2.3059	2.2985	2.2912
Day9CKa-tubulin4	USA	100 - 255	2.5431	2.5316	2.5217	2.5131	2.5049	2.4984	2.4926	2.4874	2.4829	2.4790	2.4758	2.4729	2.4700	2.4670	2.4641	2.4610
Day9CKa-tubulin5	USA	100 - 255	2.4615	2.4312	2.4053	2.3828	2.3608	2.3431	2.3276	2.3139	2.3017	2.2910	2.2817	2.2736	2.2660	2.2592	2.2533	2.2479

To empirically assess the fractal dimension values presented in Tables 6.2 and 6.3, we plot all fractal dimension values (the bile-duct FD descriptors) for each bile duct as curves. These results also were analyzed with statistical measures such as mean, standard deviation (SD), distribution, and variance.

6.1.2.1 Creating profiles of bile ducts from Fractal Dimensions computed from the original datasets

We investigated the fractal dimensions computed for thirteen datasets (seven datasets from the USA and six datasets from the POA group). Figure 6.2 summarizes the results. Fractal dimensions values computed from each bile duct dataset are depicted as the fractal dimension $FD(r)$ plotted against the dilation radius $[r5 - r20]$. Each curve corresponds to one bile-duct FD descriptor or the 'FD profile' for each bile duct.

Figure 6.2a represents the red channel. This plot shows similar profiles for ten datasets. The other three datasets have a different profile considering both the range of FD values and curve derivative. more spaced at the bottom. These three datasets are from the USA group. Figure 6.2b represents the green channel. In this case, we obtained a "spaghetti" plot with many lines displayed together. Such a plot is hard to read, and thus provides little insight about the data.

Figure 6.2: Fractal Dimension per bile duct computed from the original datasets: a) Red Channel (Table 6.2) b) Green Channel (Table 6.3)

(a) Red Fluorescence Channel (b) Green Fluorescence Channel
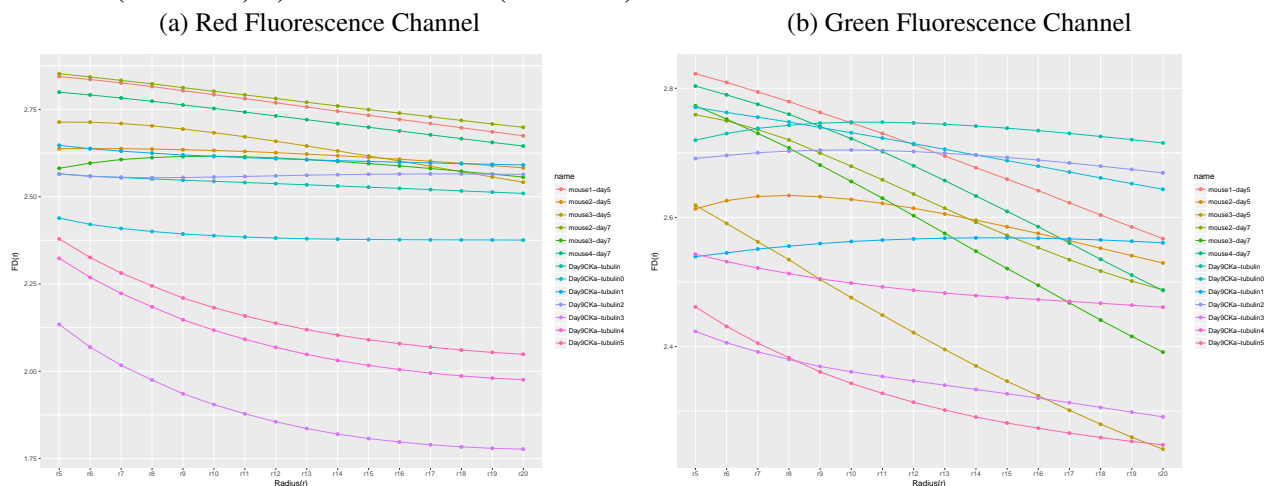
Source: The author.

We also calculated basic statistical measures for each bile duct FD descriptor from Tables 6.2 and 6.3. Figure 6.3 shows the central tendency and variability of fractal dimensions within each bile-duct FD descriptor. The bars represent the fractal dimension mean for each bile duct, and the vertical error bars represent the corresponding SD.

Figure 6.3: Fractal Dimension mean and standard deviation computed for the original datasets: a) Red Channel (Table 6.2) b) Green Channel (Table 6.3).

(a) Red Fluorescence Channel.

(b) Green Fluorescence Channel.
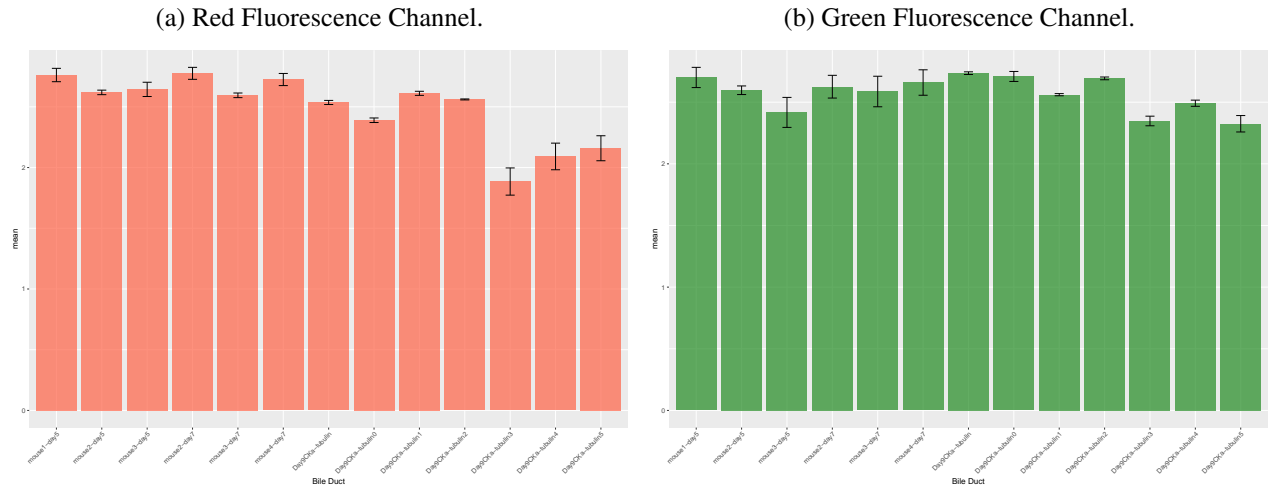
Source: The author.

We noticed that for the red channel (Figure 6.3a), the last three bile-duct FD descriptors have lower mean values, and the variance within each descriptor is larger. These three bile ducts are the same ones that have a different profile in Figure 6.2a. We verify that for these specific datasets, a higher magnification was used during the image acquisition. Then, these datasets contain only a magnified portion of the same bile duct. This fact is not necessarily an acquisition problem and could be a decision of the medical team to get more details from specific bile duct regions.

Visual inspection of results for the green channel (Figure 6.3b) shows that the central tendency within each bile-duct FD descriptor appears to be similar, but the variation seems to be different.

A Shapiro-Wilk test for normality showed that the distributions of the fractal dimensions is not normal in three bile-duct FD descriptors, for the red channel, and in two descriptors, for the green channel. Then, we performed Friedman tests for comparing all the bile-duct FD descriptors, in each channel.

For the red channel, we found significant differences among the bile-duct FD descriptors (Chi-squared=187.6, p<0.0001). Dunn's Multiple Comparison test showed that the six datasets from POA do not show significant differences among them, but they are significantly different from the USA datasets. As for the USA datasets, the three datasets that have a separate FD profile (see Figure 6.2a) show significant differences from the other four ones.

Regarding the green channel, Friedman test also showed that there are significant differences among the bile-duct FD descriptors (Chi-squared=164.5, p<0.0001). Dunn's

post hoc test showed significant differences among several datasets, as can be seen from the FD profiles curves in Figure 6.2b. Only the three datasets that have separated FD profiles are similar.

6.1.2.2 Finding the best radius values for computing the Fractal Dimension of bile ducts

As seen in the previous section, the bile-duct FD descriptor is actually a feature vector composed of fractal dimensions computed from 16 radius values. As part of a more detailed analysis, we aimed to verify which radius value could be used for computing a single fractal dimension value for each bile duct images dataset. We created box-plots of the fractal dimensions separating the two groups (POA-USA) to compare the central tendency and fractal dimension distribution within each group for each radius value. That is, for this analysis we used the 'radius-based FD datasets', since we group the FDs computed for each radius value, for each bile duct image datasets. These 'radius-based FD datasets' correspond to the columns of Tables 6.2 and 6.3. We plot the results for each channel (Red, Green), for each group (POA-USA), in Figures 6.4 and 6.5, respectively.

In the red channel, the median for POA datasets is similar for all radius, and we can see that the fractal dimension distribution is more concentrated with the increase of the radius (Figure 6.4a). Tested for normality using Shapiro-Wilk, this group of radius-based FD datasets shows normal distribution. An ANOVA (Analysis of variance) test resulted that there are statistically significant differences ($F(5,15)=23.38$, $p=0.0034$). A pos-hoc Tukey's multiple comparison test shows that there are differences between the FDs computed from the radius larger than 7 units.

As for the USA group, the behavior is different. The values are more sparse with the increase of the radius. However, the median is similar for all radius values (Figure 6.4b). Similar analysis (Shapiro-Wilk followed by ANOVA) showed that there are no significant differences between the radius-based FD datasets ($F(6,15)=457.2$, $p<0.0001$).

We verified the variance for each radius value and found that, for POA datasets, the radius $r = r19$ shows the smallest variance of the fractal dimension (0.0042, Figure 6.4c). As for the USA datasets, radius-based fractal dataset for $r5$ has the lowest variance (0.0308, Figure 6.4d).

Figure 6.4: Fractal Dimension analysis for the Red channel (Table 6.2): a) FD Distribution in POA Datasets: there are statistically significant differences only between the FDs computed from the radius larger than 7 units; b) FD Distribution in USA Datasets: there are no significant differences among the datasets; c) FD Variance for POA Datasets: the radius $r19$ shows the lowest variance of fractal dimension (0.0042). d) FD Variance for USA Datasets: the radius $r = 5$ shows the lowest variance in the fractal dimension distribution (0.0308). e) FD Distribution for all Datasets: there are statistically significant differences between the FDs computed from radius that differ for more than 6 units; f) FD Variance for all datasets: for all the sixteen radius values, the first radius $r = r5$ has the smallest variance of the fractal dimension (0.0452).

(a) FD Distribution for POA Datasets.
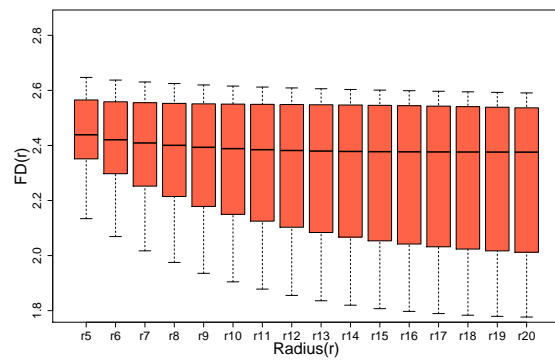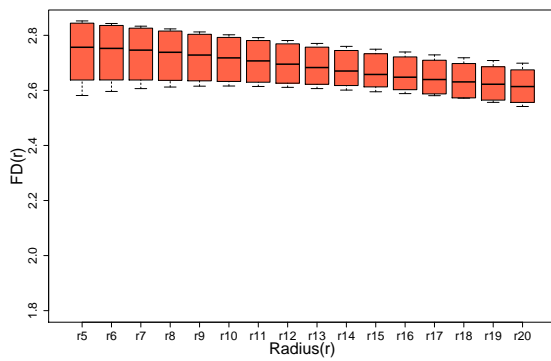
(b) FD Distribution for USA Datasets.

(c) FD Variance for POA Datasets.

(d) FD Variance for USA Datasets.

(e) FD Distribution for all Datasets.

(f) FD Variance for all Datasets.
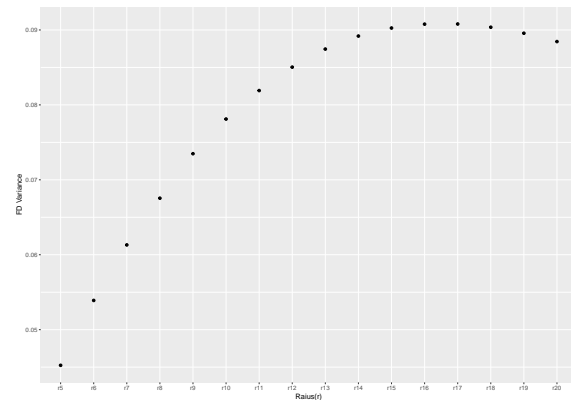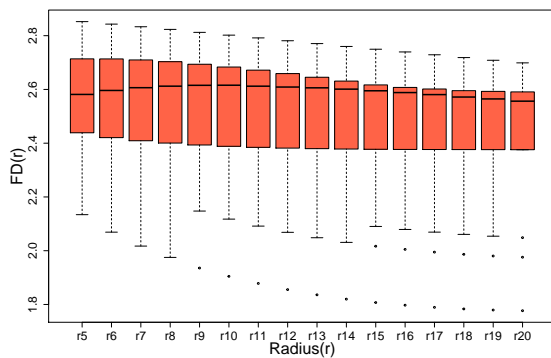
Source: The author.

We analyzed the distributions within the radius-based FD datasets grouping all datasets (POA and USA) for the red channel (Figure 6.4e). A Shapiro-Wilk test for normality showed that the distributions of the fractal dimensions are not normal within the grouped datasets. A Friedman test comparing fractal dimensions obtained from all radius showed that there are significant differences among the FD computed for different radius values (Friedman statistic, Chi-squared=140.3, p<0.0001). Dunn's post hoc test pointed out that there are no significant differences between fractal dimensions computed for radius with a difference up to 6 units.

When we verify the variance of fractal dimensions computed from radius-based FD datasets grouping all datasets (POA and USA, Figure 6.4f) for the red channel, we find the radius $r5$ shows the lowest variance.

In the green channel, for POA datasets, it is not easy to see a pattern for the distribution due to the outliers (Figure 6.5a). On the other hand, the FD distribution in USA Datasets is more sparse with the increase of the radius, and the median is similar (Figure 6.5b). Shapiro-Wilk tests for normality resulted that both groups of radius-based FD datasets have normal distributions. ANOVA performed with POA datasets showed that there are significant differences within the group (F(5,15)=43.08, p=0.0010). A pos-hoc Tukey's multiple comparison test shows that there are differences between the FDs computed from the radius larger than 10 units. ANOVA for USA datasets also showed that there are statistically significant differences within the group (F(6,15)=6.17, p=0.0445). The pos-hoc Tukey's multiple comparison test shows that there are differences only for FDs computed from radius larger than 14 units.

We plot the variance for each radius in the two groups (POA-USA) and verified that the radius $r7$ has the smallest variance of the fractal dimension for POA datasets (0.0080, Figure 6.5c), and for USA datasets, $r5$ shows the smallest variance (0.0180, Figure 6.5d).

We also analyzed the distributions within the radius-based FD datasets grouping all datasets (POA and USA) for the green channel (Figure 6.5e). As occurred with the red channel, the distributions are not normal. The Friedman test showed significant differences (Chi-squared=129.0, p<0.0001), and Dunn's post hoc test pointed out that there are no significant differences between fractal dimensions computed for radius with a difference up to 10 units.

When we verify the variance of fractal dimensions computed from radius-based FD datasets grouping all datasets (POA and USA, Figure 6.5f) for the green channel, we

find that the radius $r5$ shows the lowest variance.

Figure 6.5: Fractal Dimension analysis for the Green channel (Table 6.3): a) FD Distribution for POA Datasets: there are statistically significant differences between the FDs computed from radius values larger than 5 units. b) FD Distribution for USA Datasets: there are statistically significant differences between the FDs computed from radius values larger than 14 units. c) FD Variance for POA Datasets: the radius $r = r7$ has the lowest variance of the fractal dimension (0.0080). d) FD Variance for USA Datasets: the radius $r = r5$ has the lowest variance of the fractal dimension (0.0180). e) FD Distribution for all Datasets: statistically significant differences appear between fractal dimensions computed for radius with a difference larger than 10 units. f) FD Variance: for all the sixteen radius values, the radius ($r = r5$) has the smallest variance of the fractal dimension (0.0177).

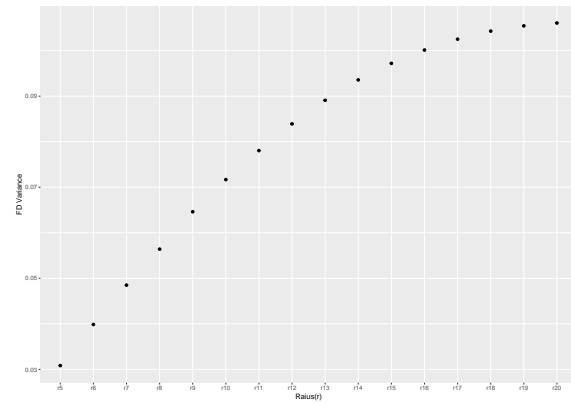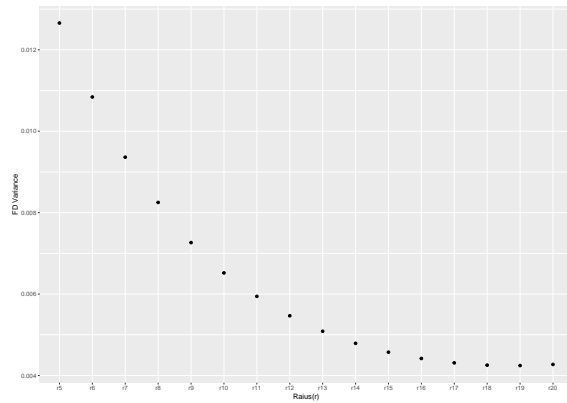(a) FD Distribution for POA Datasets.
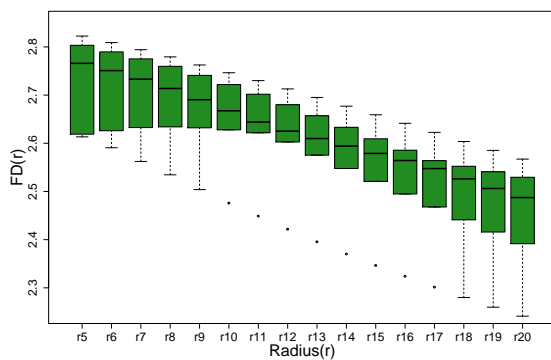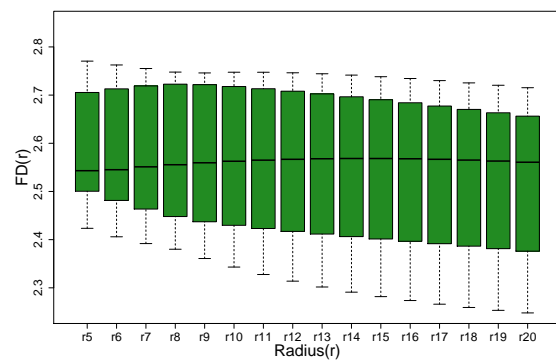
(b) FD Distribution for USA Datasets.

(c) FD Variance for POA Datasets.
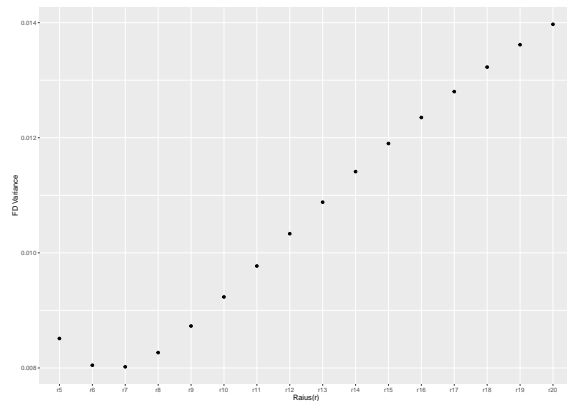
(d) FD Variance for USA Datasets.
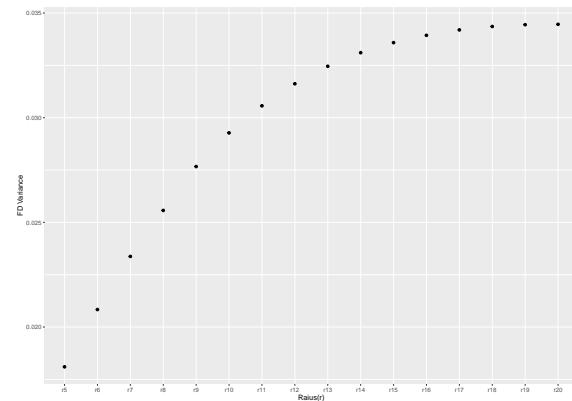
(e) FD Distribution for all Datasets.
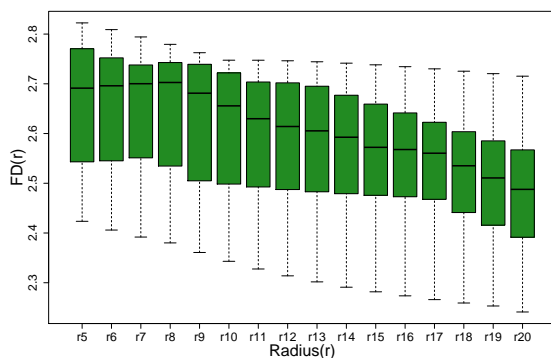
(f) FD Variance for all Datasets.
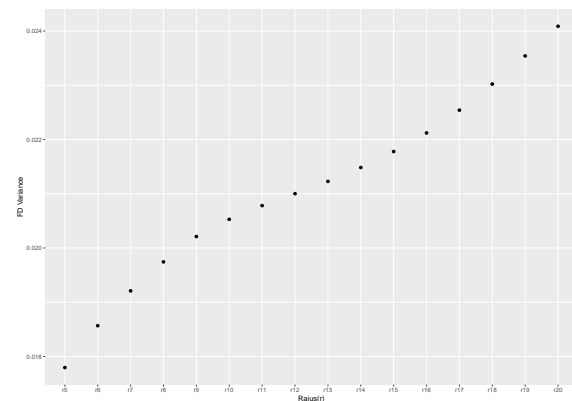
Source: The author.

6.1.3 Final Comments

The results obtained with these analyses showed that the fractal dimension mean (computed from the FD for each radius, $r = 5 - 20$) do not represent the similarity we need for characterizing normal bile ducts. Moreover, we also found that FDs computed from some radius values show significant differences.

We use the variance as a statistical measure to assess the dispersion of the fractal dimension results. So, we took the radius values where the radius-based FDs show the lowest variance.

Figures 6.4c-d and 6.5c-d show the variance of the fractal dimension for each radius value for the POA and USA datasets, for the red and green channels, respectively. Analyses within specific groups indicated some differences in the fractal dimension distribution between the POA and USA datasets. In some cases, the first radius ($r = 5$) has the lowest variance in the fractal dimension distribution. For other cases, we found other radius values with the lowest variance ($r = 7, r = 19$). Then, we can not assume that there is a single radius value with the fractal dimension distribution that best represents the data. A possible explanation for this discrepancy is the input data and the fact that we use the original data, and the only preprocessing step was the thresholding. That could be a problem as a dataset may not contain a representative view of all relevant structures for the fractal dimension computation.

As a final analysis with these datasets, we compared the two groups (POA and USA) using the radius-based FD datasets with the lowest variance. For the red channel, radius $r19$ and $r5$ showed the lowest variances for POA and USA datasets, respectively. Distributions tested as not normal. Significant differences were found between the two groups derived from an unpaired t-test with Welch'correction (t=0.02911, $p < 0.05$). As for the green channel, radius $r5$ showed the lowest variance, and significant differences were also found (t-test = 0.1101, $p < 0.05$).

6.2 Characterization of Bile Ducts Based on Fractal Dimensions Computed from the most Representative Clusters

In this second analysis, we modified the pipeline to include a new step to perform clustering based on the DBSCAN method before fractal dimension computation. With this pipeline modification, we aim to improve the input data for the fractal dimension method. Then, we calculate the fractal dimension from the most representative cluster for each dataset. As will be noticed, we have designed the fractal dimensions analyses of these new datasets based on conclusions from the previous analyses, described in Section 6.1.

Figure 6.6: Overview of the data flow pipeline for our fractal dimension study over the most representative cluster.
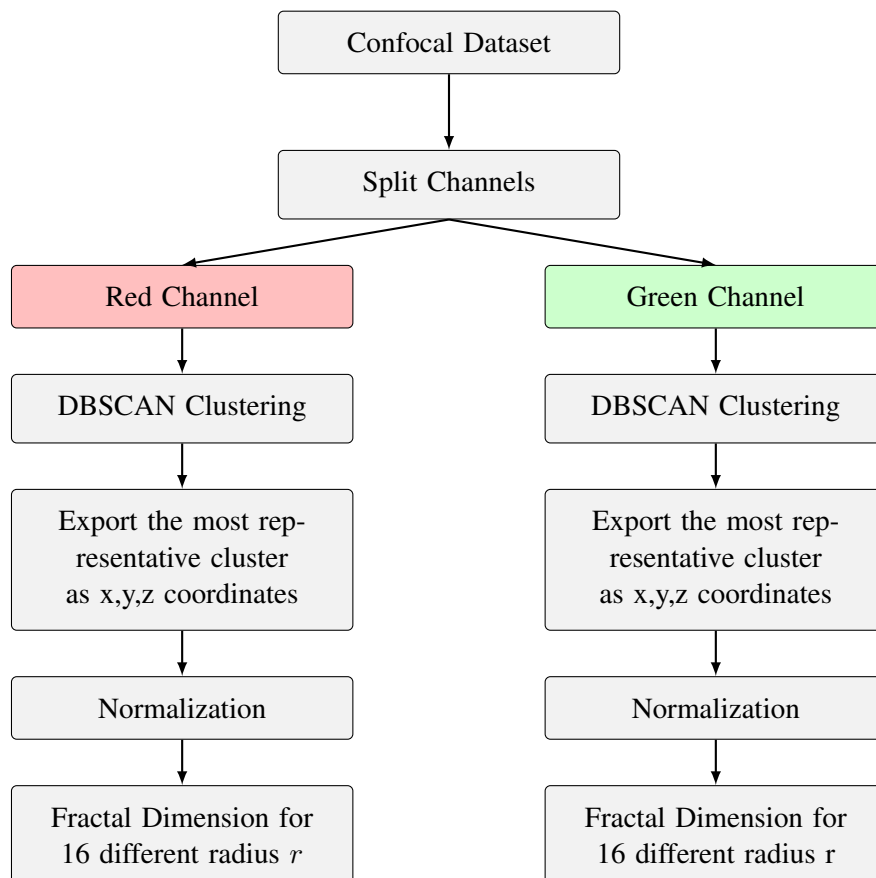
Source: The author.

6.2.1 Material and Methods

We use the same datasets described in Table 6.1. In this way, we can compare the new fractal dimension results with the previous ones reported in Section 6.1. Figure 6.6 shows our data flow pipeline for fractal dimension analyses using DBSCAN clustering. The clusters were obtained as described in Section 5.2. Since DBSCAN results several clusters, we then explore them to select the cluster that best contains the structure of interest.

From the mathematical definition, the fractal dimension is invariant to scale. However, the Bouligand-Minkowski method is based on the idea of a volume of influence being dilated over the object. Then, two identical objects, but with different sizes, tend to have different fractal dimensions because their initial volume is different.

In Section 6.1, we noticed that we have some datasets with different magnification. Then, for the current analyses, we apply a normalization step to avoid that datasets with different magnification sizes, result in inconsistent fractal dimensions.

The normalization is applied to the most representative clusters considering the longest distance between any two points in the dataset. After the normalization step, all fractal dimensions' values are recalculated, resulting Tables 6.4 and 6.5.

6.2.2 Results and Analysis

For each dataset, we extract the most representative cluster. The inspection of the several clusters produced by DBSCAN showed that, in most cases, cluster number 1 contains the segmented structure, while in other cases, we had to explore the clusters until the cluster that contains the structure of interest was found. As in the previous analyses, the Fractal Dimension descriptor FD was computed for the most representative cluster using Equation 2.7 with 16 different radius, for each bile duct. Tables 6.4 and 6.5 summarize the results.

Table 6.4: Fractal Dimension computed from the Most Representative Cluster-(Red Channel)

Dataset Id	From	Extracted Cluster	Fractal Dimension															
			r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	r17	r18	r19	r20
mouse1-day5	POA	2	2.6862	2.6710	2.6580	2.6469	2.6362	2.6272	2.6187	2.6104	2.6022	2.5942	2.5865	2.5791	2.5713	2.5636	2.5560	2.5486
mouse2-day5	POA	1	2.5554	2.5396	2.5266	2.5148	2.5025	2.4919	2.4817	2.4717	2.4620	2.4527	2.4440	2.4357	2.4274	2.4193	2.4117	2.4044
mouse3-day5	POA	1	2.5472	2.5159	2.4891	2.4659	2.4429	2.4235	2.4059	2.3893	2.3740	2.3598	2.3471	2.3357	2.3247	2.3146	2.3054	2.2969
mouse2-day7	POA	1	2.6532	2.6449	2.6389	2.6335	2.6278	2.6225	2.6170	2.6113	2.6053	2.5991	2.5929	2.5867	2.5802	2.5737	2.5674	2.5612
mouse3-day7	POA	1	2.6076	2.5806	2.5567	2.5352	2.5130	2.4936	2.4752	2.4572	2.4399	2.4231	2.4073	2.3924	2.3771	2.3623	2.3484	2.3351
mouse4-day7	POA	1	2.6793	2.6633	2.6492	2.6363	2.6231	2.6114	2.6002	2.5891	2.5783	2.5678	2.5581	2.5488	2.5394	2.5303	2.5217	2.5135
Day9CKa-tubulin	USA	1	2.5020	2.4748	2.4529	2.4340	2.4150	2.3989	2.3839	2.3695	2.3558	2.3428	2.3309	2.3198	2.3088	2.2983	2.2887	2.2799
Day9CKa-tubulin0	USA	2	2.3453	2.3077	2.2783	2.2533	2.2289	2.2090	2.1912	2.1751	2.1607	2.1478	2.1367	2.1269	2.1177	2.1093	2.1020	2.0954
Day9CKa-tubulin1	USA	1	2.5592	2.5374	2.5199	2.5045	2.4889	2.4757	2.4636	2.4521	2.4415	2.4318	2.4234	2.4160	2.4091	2.4029	2.3975	2.3925
Day9CKa-tubulin2	USA	1	2.5171	2.4947	2.4770	2.4618	2.4469	2.4346	2.4234	2.4126	2.4025	2.3929	2.3839	2.3756	2.3675	2.3600	2.3533	2.3475
Day9CKa-tubulin3	USA	39	2.3621	2.3179	2.2796	2.2462	2.2132	2.1851	2.1594	2.1350	2.1121	2.0906	2.0711	2.0531	2.0353	2.0185	2.0030	1.9886
Day9CKa-tubulin4	USA	39	2.5257	2.4968	2.4691	2.4438	2.4181	2.3958	2.3750	2.3549	2.3357	2.3175	2.3007	2.2851	2.2695	2.2547	2.2412	2.2287
Day9CKa-tubulin5	USA	1	2.3594	2.3251	2.2947	2.2690	2.2463	2.2288	2.2141	2.2018	2.1918	2.1841	2.1787	2.1750	2.1727	2.1714	2.1708	2.1706

Table 6.5: Fractal Dimension computed from the Most Representative Cluster-(Green Channel)

Dataset Id	From	Extracted Cluster	Fractal Dimension															
			r5	r6	r7	r8	r9	r10	r11	r12	r13	r14	r15	r16	r17	r18	r19	r20
mouse1-day5	POA	3	2.6708	2.6439	2.6193	2.5970	2.5743	2.5545	2.5360	2.5181	2.5010	2.4847	2.4696	2.4555	2.4413	2.4278	2.4151	2.4032
mouse2-day5	POA	6	2.7119	2.6888	2.6679	2.6492	2.6299	2.6131	2.5975	2.5823	2.5679	2.5543	2.5420	2.5304	2.5188	2.5076	2.4971	2.4871
mouse3-day5	POA	4	2.6854	2.6810	2.6742	2.6661	2.6563	2.6466	2.6365	2.6259	2.6152	2.6044	2.5939	2.5836	2.5729	2.5622	2.5518	2.5417
mouse2-day7	POA	1	2.7170	2.6960	2.6772	2.6600	2.6422	2.6267	2.6122	2.5980	2.5845	2.5714	2.5594	2.5480	2.5365	2.5254	2.5150	2.5052
mouse3-day7	POA	1	2.7212	2.6966	2.6738	2.6528	2.6306	2.6108	2.5920	2.5733	2.5553	2.5377	2.5213	2.5057	2.4898	2.4744	2.4598	2.4459
mouse4-day7	POA	1	2.6535	2.6265	2.6023	2.5809	2.5593	2.5410	2.5242	2.5083	2.4933	2.4791	2.4661	2.4539	2.4416	2.4298	2.4188	2.4084
Day9CKa-tubulin	USA	1	2.5729	2.5418	2.5163	2.4943	2.4725	2.4543	2.4378	2.4223	2.4080	2.3946	2.3823	2.3709	2.3594	2.3482	2.3376	2.3274
Day9CKa-tubulin0	USA	1	2.5570	2.5317	2.5090	2.4884	2.4671	2.4489	2.4320	2.4158	2.4006	2.3861	2.3727	2.3602	2.3475	2.3352	2.3237	2.3127
Day9CKa-tubulin1	USA	1	2.6612	2.6398	2.6233	2.6098	2.5973	2.5872	2.5781	2.5696	2.5615	2.5537	2.5464	2.5392	2.5319	2.5245	2.5173	2.5102
Day9CKa-tubulin2	USA	1	2.5759	2.5440	2.5183	2.4969	2.4763	2.4595	2.4443	2.4300	2.4167	2.4041	2.3924	2.3814	2.3701	2.3591	2.3486	2.3383
Day9CKa-tubulin3	USA	1	2.4873	2.4438	2.4052	2.3717	2.3403	2.3157	2.2950	2.2767	2.2605	2.2457	2.2324	2.2201	2.2077	2.1958	2.1845	2.1736
Day9CKa-tubulin4	USA	1	2.5028	2.4642	2.4304	2.4012	2.3727	2.3489	2.3277	2.3077	2.2893	2.2722	2.2567	2.2424	2.2282	2.2148	2.2023	2.1906
Day9CKa-tubulin5	USA	1	2.4953	2.4654	2.4371	2.4109	2.3840	2.3605	2.3387	2.3176	2.2975	2.2784	2.2607	2.2442	2.2275	2.2115	2.1966	2.1825

To empirically assess the results presented in Tables 6.4 and 6.5, we plot all fractal dimension values (the 'bile-duct FD descriptors') for each bile duct, for each channel, as curves (Figure 6.7). We also present statistical measures such as mean, SD, distribution, and variance for these FD values.

6.2.2.1 Creating Profiles of Bile Ducts based on Fractal Dimensions computed from the most representative clusters

Similarly to the analysis we reported in Section 6.1, we examined the fractal dimensions computed from thirteen datasets (six datasets from the POA and seven datasets from the USA group). Figure 6.7 summarizes the results for all bile ducts datasets. Bile duct profiles are depicted as curves, relating the fractal dimension $FD(r)$ to each dilation radius $[r5 - r20]$. Each curve corresponds to the FD descriptor of a bile duct. Figures 6.7a and 6.7b present the red and the green channels, respectively. Comparing with previous results shown in Figure 6.2, we can notice that the new curves show similar profiles for all bile ducts. This pattern is similar for the two fluorescence channels.

Figure 6.7: Fractal Dimension profiles for each bile duct computed from the most representative cluster: a)Red Channel (Table 6.4) b)Green Channel (Table 6.5).

(a) Red Fluorescence Channel.   (b) Green Fluorescence Channel.
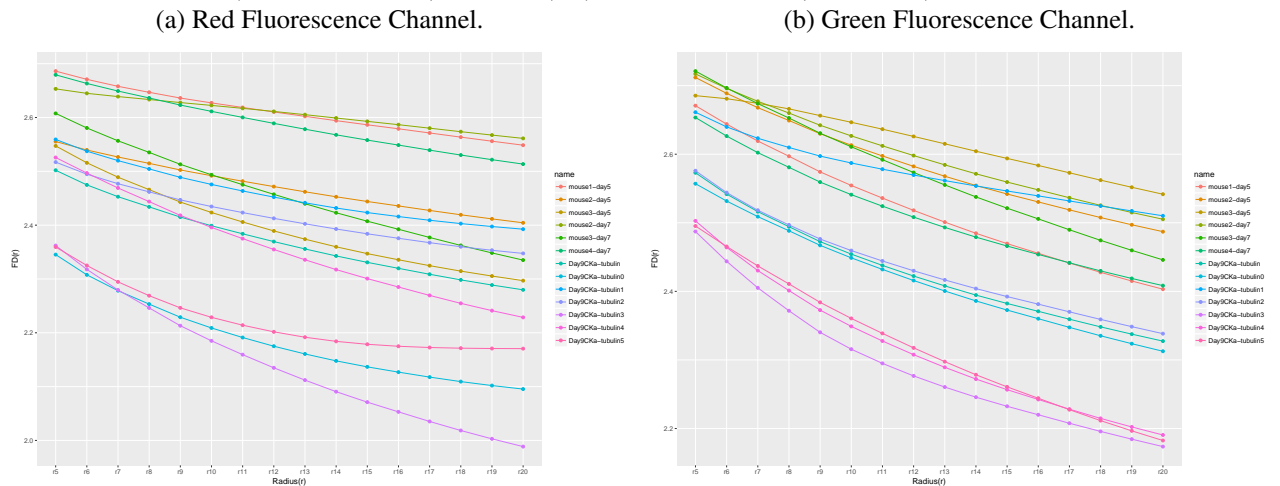
Source: The author.

We also calculated the mean FD and SD for each bile duct FD dataset (i.e., the mean value and SD for each row in Tables 6.4 and 6.5) and plotted the results, as can be seen in Figure 6.8.

From the analyses performed with the original datasets and reported in Section 6.1, we learned that radius values with the lowest FD variance could be used for characterizing similarity between bile ducts. So, we decided to proceed with the assessment of the variances of radius-based FD datasets, for both the red and green channels, and POA and

USA datasets.

Figure 6.8: Fractal Dimension Mean and Standard Deviation of each bile duct FD dataset computed from the Most Representative Cluster: a)Red Channel (Table 6.4) b)Green Channel (Table 6.5).

(a) Red Fluorescence Channel.   (b) Green Fluorescence Channel.
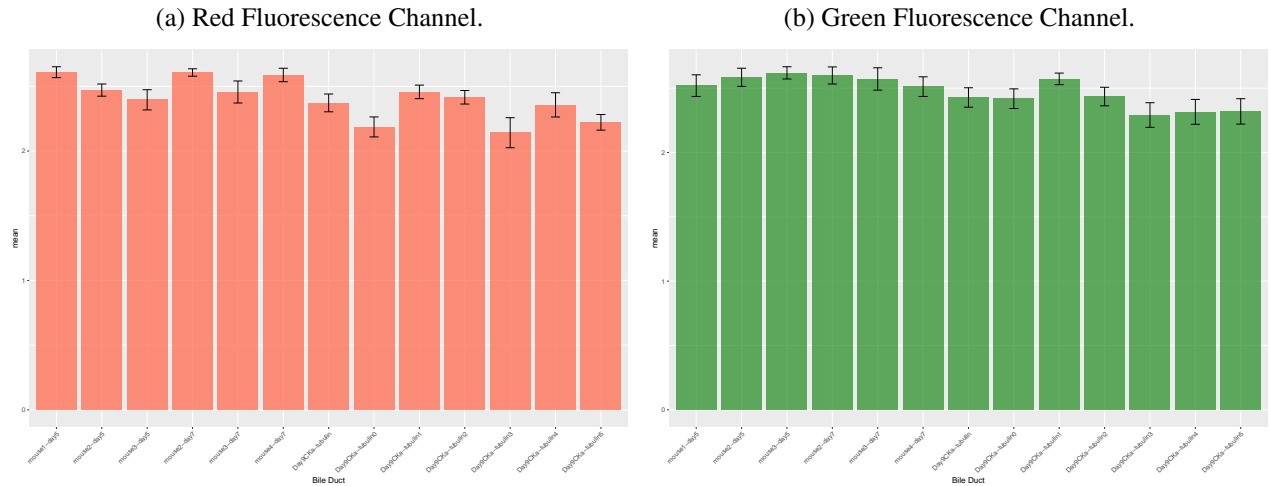
Source: The author.

6.2.2.2 Finding the best radius values for computing the Fractal Dimension of bile ducts

As we did for the original datasets, we analyzed the radius-based FDs (the columns of Tables 6.4 and 6.5), to find the radius values that would give the most representative FD for each bile duct. Figures 6.9 and 6.10 summarize the distribution of the radius-based FD datasets for each fluorescence channel. These plots are useful to observe the distribution of the fractal dimension within each radius.

Comparing Figures 6.9 and 6.10 with previous results reported in Figures 6.4 and 6.5), one notices that the distributions of FDs computed from the representative clusters, and normalized, show similar trends to the ones computed from the original datasets. The same pattern was found when we analyze both groups together, POA and USA. These patterns are related to the variance of the radius-based FD datasets.

We use the variance as a statistical measure to understand the dispersion of the fractal dimension over the most representative clusters. We noticed that the first radius value $(r = 5)$ represents the lowest variation in the fractal dimension distribution (Figures 6.9f and 6.10f). As an interesting result, we can see that the variance has the same behavior in the two fluorescence channels. If we compute FDs for the specific groups (POA and USA datasets), radius-based FDs for $r = 5$ also have the same pattern.

These analyses indicated that the first radius values show the lowest variance of the fractal dimension in all datasets. Then, we can assume that the radius with the lowest

variance is the radius with the fractal dimension distribution that best represents the data. Then, this value ($r = 5$) is chosen for the subsequent statistical tests.

Figure 6.9: Fractal Dimension Analysis of the Red channel datasets (Table 6.4): a) FD Distribution for POA Datasets. b) FD Distribution for USA Datasets c) FD Variance for POA Datasets. d) FD Variance for USA Datasets e) FD Distribution for all Datasets. f) FD Variance for all Datasets: radius $r5$ has the smallest variance among all.

(a) FD Distribution for POA Datasets.

(b) FD Distribution for USA Datasets.
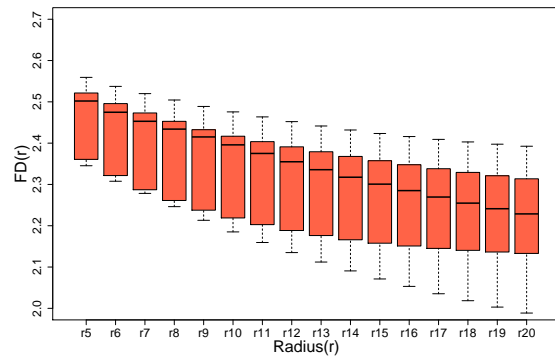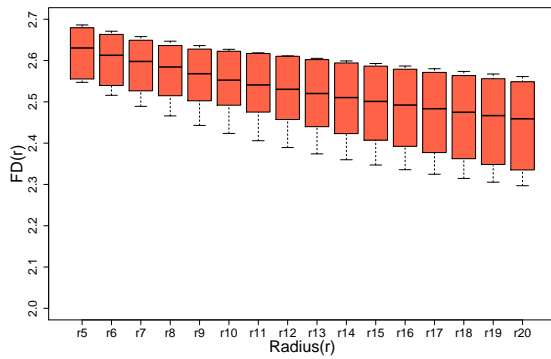
(c) FD Variance for POA Datasets.

(d) FD Variance for USA Datasets.
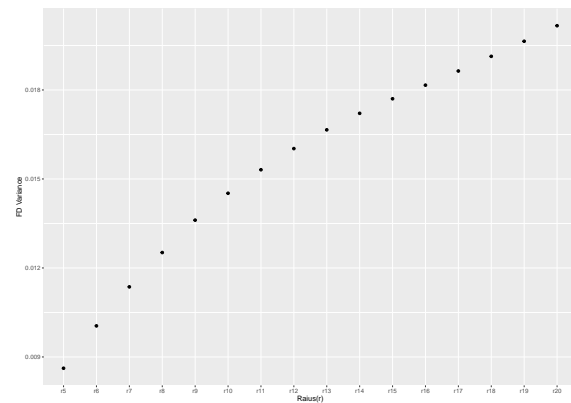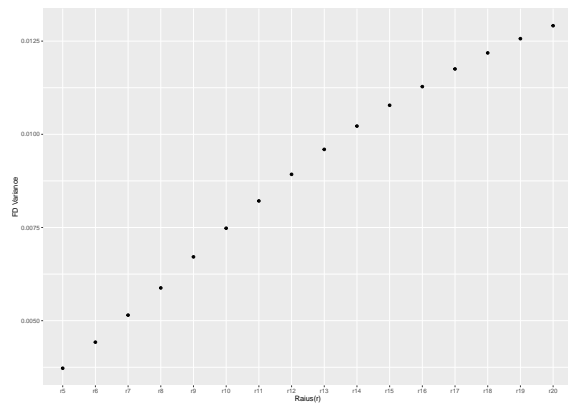
(e) FD Distribution for all Datasets.

(f) FD Variance for all Datasets.
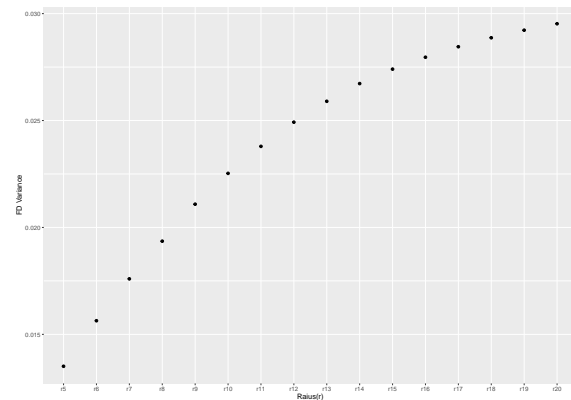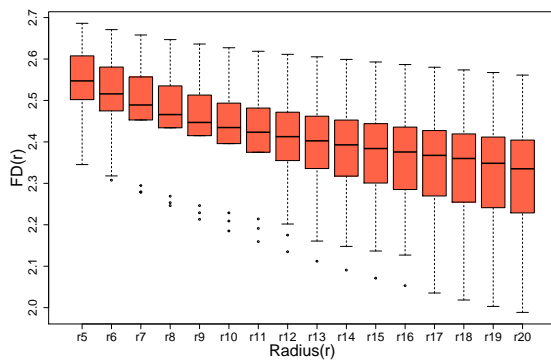
Source: The author.

Figure 6.10: Fractal Dimension Analysis of the Green Channel (Table 6.5): a) FD Distribution for POA Datasets. b) FD Distribution for USA Datasets. c) FD Variance for POA Datasets. d) FD Variance for USA Datasets. e) FD Distribution for all Datasets. f) FD Variance for all Datasets: radius $r5$ has the smallest variance among all.

(a) FD Distribution for POA Datasets.

(b) FD Distribution for USA Datasets.
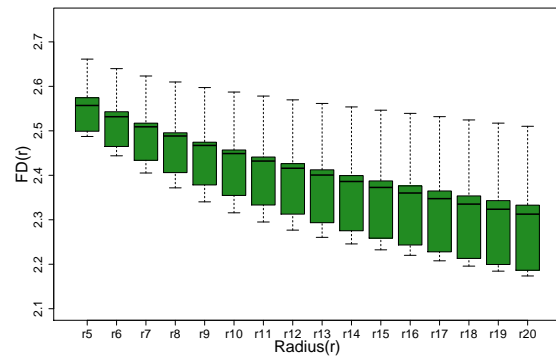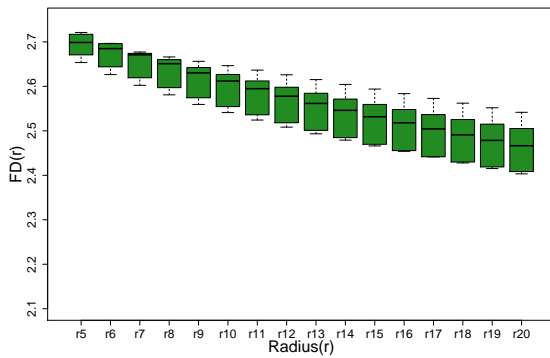
(c) FD Variance for POA Datasets.

(d) FD Variance for USA Datasets.
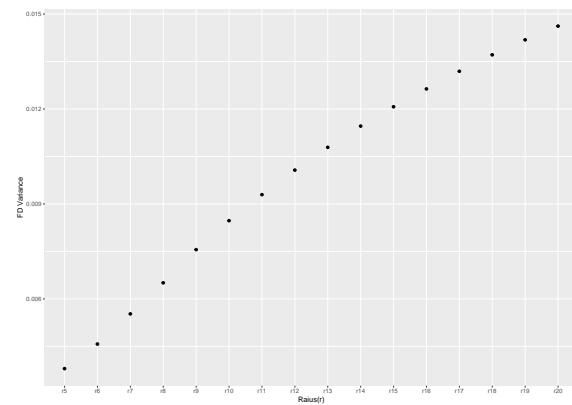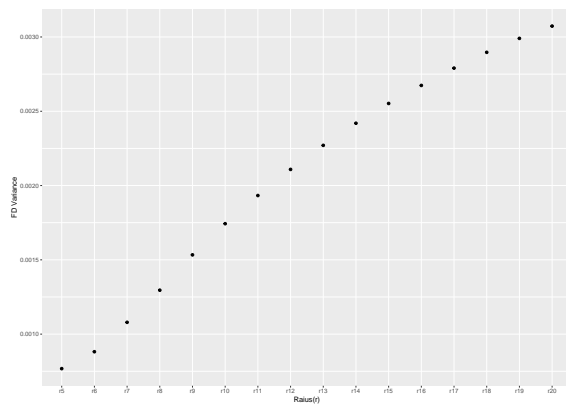
(e) FD Distribution for all Datasets.

(f) FD Variance for all Datasets.
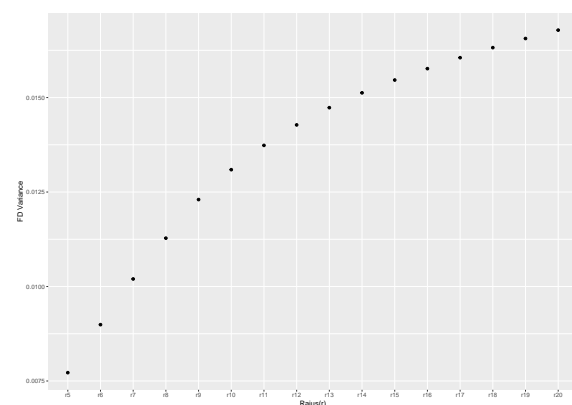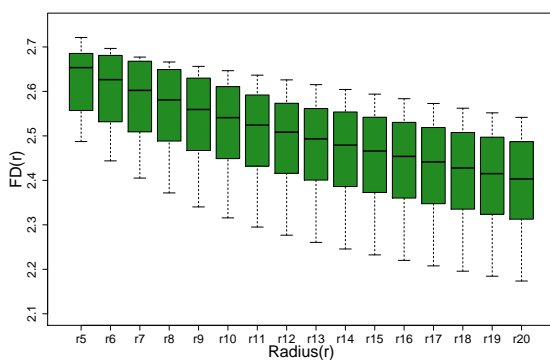
Source: The author.

To support our decision regarding the choice between the Fractal Dimensions computed from the original datasets and the ones computed from the most representative clusters, we performed a comparison between both results.

We compared the radius-based FD datasets for $r5$ from Table 6.2 with the corre-

sponding ones from Table 6.4, for the red channel. We did the same for the 'radius-5-based' FD datasets from Table 6.3 with those from Table 6.5, for the green channel.

Shapiro-Wilk test for normality showed that the distributions of the fractal dimensions are normal for $r = 5$, in both channels and all four tables. Then, we performed two paired t-tests.

For the red channel, t-test showed that there are no significant differences among the radius-based FD descriptors for the original datasets and the most representative clusters (Paired t-test = 0.2393, p < 0.05). Regarding the green channel, t-test also showed that there are no significant differences among the radius-based FD descriptors of the original datasets and the most representative clusters (Paired t-test = 0.2084, p < 0.05).

Based on these statistical criteria, we can select any of the two radius-based FD datasets for quantification (either the original datasets or the most representative clusters). Since the most representative clusters correspond to a better segmentation of the structures while allowing for enhanced visualization, we decided to adopt the most representative clusters to perform our further analyses towards a characterization measure of bile ducts.

As a final analysis considering only the most representative clusters, we compared the two groups (POA and USA) using the radius-based FD datasets with the lowest variance, (radius $r5$). Significant differences were found between groups POA-USA derived from unpaired t-test (p < 0.05): POA-USA, red channel (Two Sample t-test = 0.003006, p < 0.05), and green channel (Two Sample t-test = 0.0002836, p < 0.05). The analyses between these specific groups indicated that there are differences in the fractal dimension means between the POA and USA datasets in the two fluorescence channels.

6.2.3 Final Comments

The results reported in this section with fractal dimensions computed using the radius with the lowest FD variance from the most representative clusters showed that the range of FDs remained between 2.55 and 2.68 for POA datasets, and 2.34 and 2.56 for USA datasets, for the red channel. For the green channel, we found that FDs range from 2.65 to 2.72 for POA datasets and from 2.48 to 2.66 for USA datasets.

One can explain the difference between the ranges based on the fact that bile ducts from POA datasets were obtained from samples of mice 5- and 7-days old, while bile ducts from USA datasets were from mice 9-days old. Although normalized within the groups, fractal dimensions computed with the methods employed herein retain the size

feature. Also, the different development phases might explain the structural difference that can impact shape and, thus change the fractal dimension.

In the next section we report results from further analyses aimed at clarifying the similarities and differences between the datasets.

6.3 Characterization of Bile Ducts Based on Multiscale Fractal Dimension Computed from the most Representative Clusters

As mentioned in Chapter 2 (Sect. 2.5.2), Multiscale Fractal Dimension (MFD) provides a richer information about the shape of complex objects than a single FD value (FLORINDO; CASTRO; BRUNO, 2011).

Aiming at a further analysis of the bile ducts' confocal images datasets, we performed a study based on the Multiscale Fractal Dimension computed from the most representative clusters obtained as reported in Section 6.2.

In this section, we restrain ourselves to present the results and analysis because the only difference from the previous section is the computation of Multiscale Fractal Dimensions for the representative cluster of each bile duct instead of the computation of 16 Fractal Dimensions (one for each radius).

6.3.1 Multiscale Fractal Dimension (MFD) Descriptors

MFDs were computed for each bile duct as described in Section 2.5.2, considering the radius values up to 20. Only values for radius starting at 4.25 were computed because for smaller radius values, there were not enough points in the dataset to compute the FD. From the log-log curve, derivatives at 275 points resulted in MFD descriptors of 275 components for each bile duct. Since performing analyses with such long descriptors would be cumbersome, we analyzed the variances of each radius-based MFD datasets.

We found out that we could analyze the most representative clusters considering two cases:

- **Radius values with the lowest variance:** Preliminary analyses of the MFD results showed that the components between $r[4.25 - 7.36]$ delivered the fractal dimensions with the lowest variance (Figures 6.11a and 6.12a). Then, for the red and green channels, we selected the range of values $r[4, 25 - 7, 36]$ for our analyses.

- **Radius values with the highest variance:** Preliminary analyses of the MFD results showed that the components between $r[17.82 - 18.70]$ delivered the fractal dimensions with the highest variance (Figures 6.11b and 6.12b). Then, for the red and green channels, we also selected the range of values $r[17.82 - 18.70]$ for our analyses.

The resulting MFD descriptors for each bile duct are presented in Figures 6.11 and 6.12, for the red and green channel, respectively. These curves are MFD profiles of each bile duct considering different radius ranges: radius between 4.25 and 7.36, and radius between 17.82 to 18.70.

For the red channel, Figure 6.11a represents the MFD profiles plotted for the radius values with the lowest variance. In this case, we can see a similar trend of the fractal dimension values for all datasets. On the other hand, Figure 6.11b represents the MFD profiles plotted for the radius values with the highest variance. In this case, we can not see a pattern in the trend of values.

Figure 6.11: Multiscale Fractal Dimension computed from the Most Representative Clusters for each bile duct(Red Channel): a) MFD descriptor plotted for the radius values with the lowest variance $r[4.25 - 7.36]$. b) MFD descriptor plotted for the radius values with the highest variance $r[17.82 - 18.70]$.

(a) MFD descriptor (lowest variance) (b) MFD descriptor (highest variance).
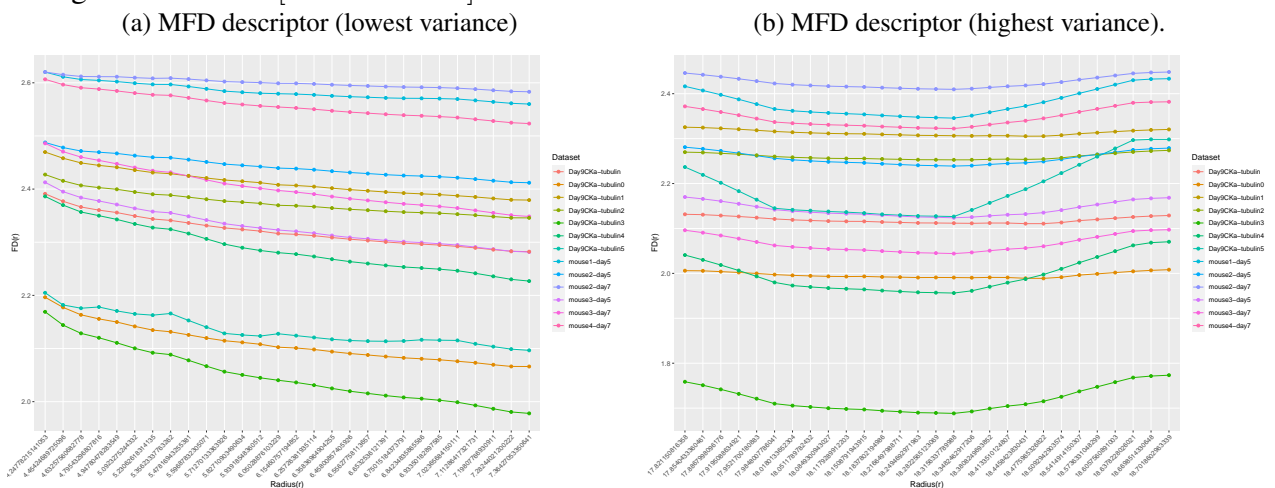
Source: The author.

As for the green channel, we used the same ranges of radius values and plotted Figures 6.12a and 6.12b that represents the MFD profiles of each bile duct. Figure 6.12a shows the bile ducts' MFD profiles considering the range that showed the lowest variance, while Figure 6.12b presents the same profiles but computed for the radius values with the highest variance.

6.3.2 Using MFD Descriptors to Find Similarities and Differences between Bile Ducts

Since the Multiscale Fractal Dimension is a multivariate approach, the analytical approach should also be multivariate. In this way, we chose to perform a Principal Component Analysis (PCA) to investigate how similar are the bile ducts, considering the

Figure 6.12: Multiscale Fractal Dimension descriptors computed from the Most Representative Clusters for each bile duct (Green Channel): a) MFD plotted for the radius values with the lowest variance $r[4.25 - 7.36]$. b) MFD plotted for the radius values with the highest variance $r[17.82 - 18.70]$.

(a) MFD descriptor (lowest variance)   (b) MFD descriptor(highest variance)
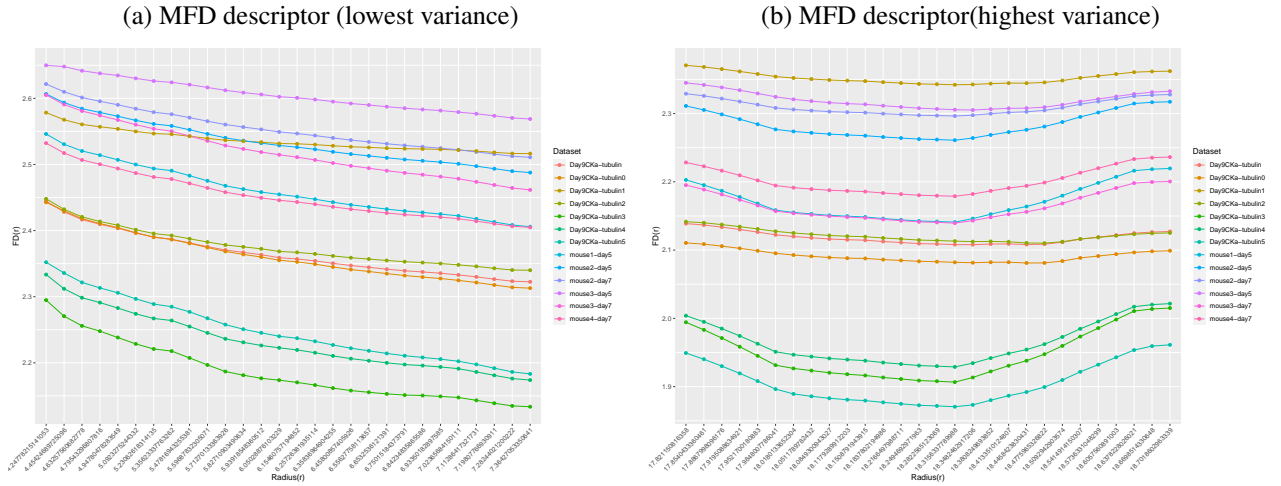
Source: The author.

fluorescence channels separately.

In the previous analyses, we have separated the datasets into two groups, POA and USA datasets. However, since we hypothesize that the differences between them might be caused by the size or age of the mice that were sample, we decided to analyze the bile ducts' MFD profiles based on three categories: poaDay5, poaDay7, and usaDay9, corresponding to the ages: mice at 3, 5, and 9 days of age.

Figure 6.13 shows the PCA of the 13 MFD descriptors for the red channel. We should recall that the red channel represents the bile duct microvasculature. Figures 6.13a and 6.13b present the PCA considering the same radius ranges we used for depicting the bile ducts' MFD profiles, i.e., $r[4.25 - 7.36]$ and $r[17.82 - 18.70]$, intervals of radius values that showed the lowest variance and highest variance, respectively.

We observe that in the three cases (Figures 6.13a, 6.13b and 6.13c), the group usaDay9 is more separated from the poaDay5 and poaDay7 groups. When we observe the bile ducts for these two groups, they form overlapping clusters. That situation confirms the results we obtained from the statistical analyses, where we found out that these bile ducts, for some radius values, do not show significant differences. Also, it should be noticed that PC1 "explains" more than 99% of the differences between all the bile ducts.

In Figure 6.13c, we present the PCA for the MFD descriptors computed from radius 4.25 to 4.92. We observe that PC1 explain 100% of the differences between all the bile ducts, and overlapping of groups is more evident.

Figure 6.14 shows the PCA for the same descriptors, but for the green channel. We

Figure 6.13: PCA based on Multiscale Fractal Dimension (Red Channel)): a) PCA using MFD descriptors computed from radius 4.25-7.36. b) PCA using MFD descriptors computed from radius 17.82-18.70. c) PCA considering the radius values between 4.25 and 4.92

(a) PCA of MFDs from radius 4.25-7.36 (b) PCA of MFDs from radius 17.82-18.70.
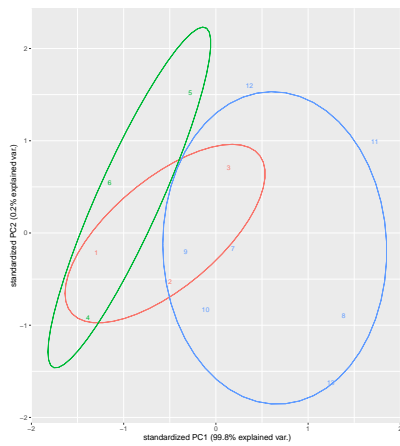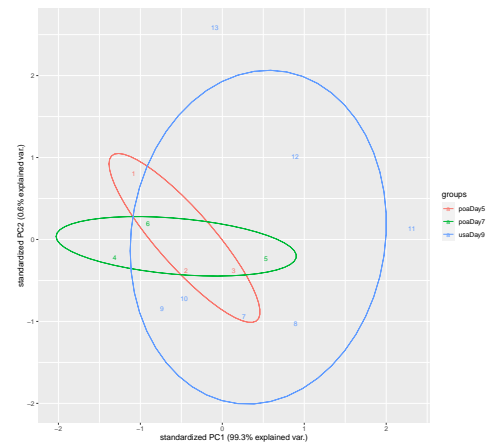
(c) PCA of MFDs from radius 4.25-4.92
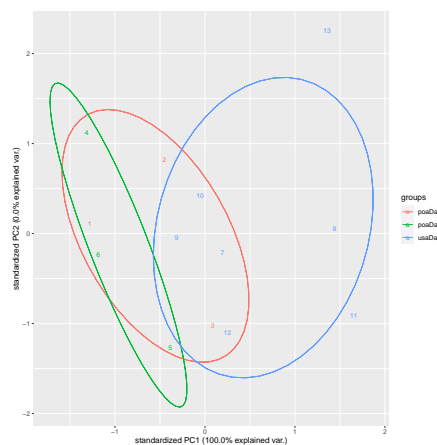
Source: The author.

should recall that the Green channel represents the bile duct wall containing the peribiliary glands.

We observe that the bile ducts belonging to the usaDay9 group form a well separated cluster for radius values with the lowest variance $[4.25 - 7.36]$ (Figure 6.14a), while for the highest variance, the usaDay9 cluster is closer to the other two groups. Bile ducts from the poaDay5 and poaDay7 groups are in two overlapping clusters in both analyses, although for radius between 4.25 and 7.36, the similarity given by proximity is more evident (Figures 6.14a and 6.14b). Also here, the results confirm what we obtained from the statistical analyses, where we found out that the group from USA Datasets are different from the POA datasets, and these, for some radius values, do not show significant differences. Also, it should be noticed that PC1 "explains" more than 99% of the differences

between all the bile ducts.

The PCA analysis for the more constrained range of radius with lowest variance [4.25-4.92] (Figure 6.14c) shows bile ducts from the usaDay9 group well separated, while bile ducts belonging to the poaDay5 and poaDay7 groups are clustered together. This fact also makes sense, as all bile ducts from poaDay5 and poaDay7 come from mice with a small difference in age, and were sampled in the same experiment.

Figure 6.14: PCA based on Multiscale Fractal Dimension (Green Channel)): a) PCA using MFD descriptors computed from radius $r[4.25-7.36]$. b) PCA using MFD descriptors computed from radius $r[17.82 - 18.70]$. c) PCA considering the radius values between 4.25 and 4.92

(a) PCA considering the radius 4.25-7.36 (b) PCA considering the radius 17.85-18.73
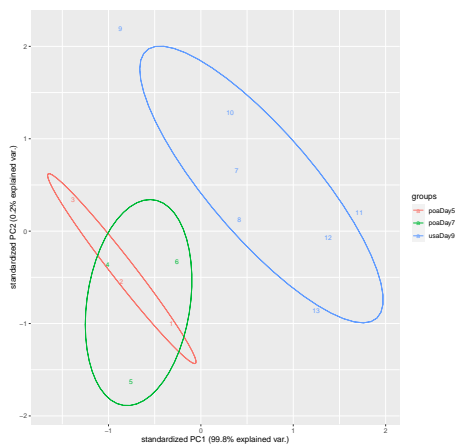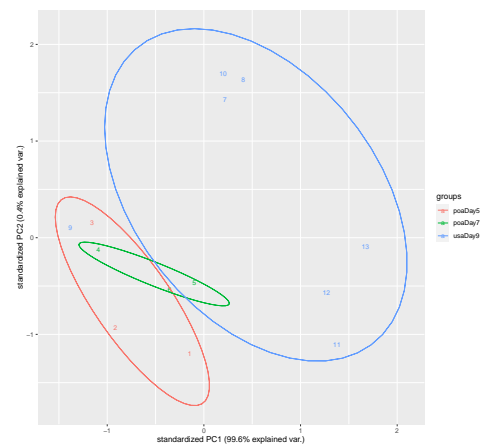
(c) PCA considering the radius 4.25-4.92
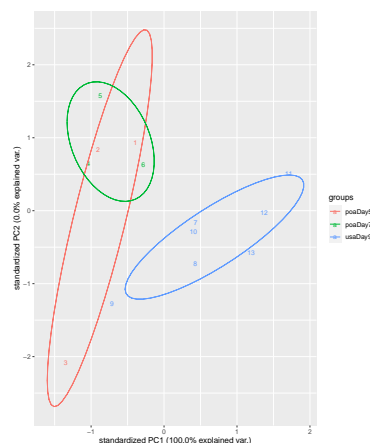
Source: The author.

We also performed a Principal Component Analysis (PCA) to investigate how similar are the bile ducts considering the two fluorescence channels together (microvasculature and bile duct wall together for the radius range $4.25 - 4.92$). We present that result in Figure 6.15. In this case, we can see that all three groups, poaDay5, poaDay7, and usaDay9, form a large cluster. However, we also observe that some bile ducts belonging

to usaDay9 are more apart from the others.

Figure 6.15: PCA based on Multiscale Fractal Dimension over MFD red + MFD green: considering the radius values between $r[4.25 - 4.92]$.
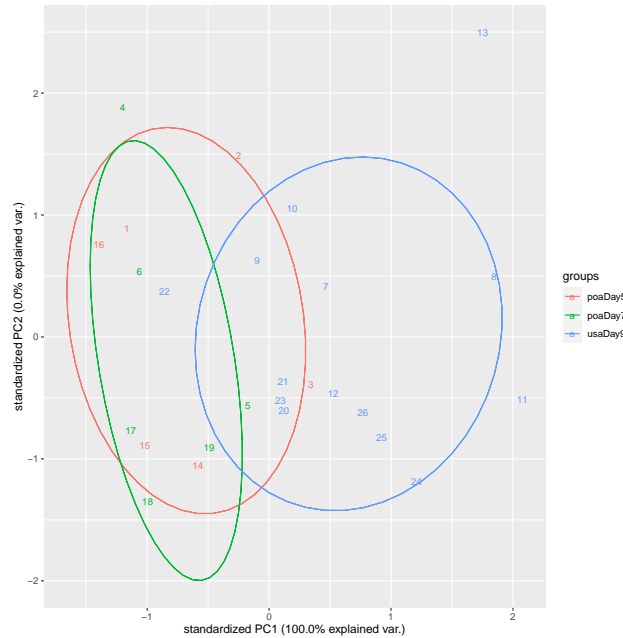
Source: The author.

6.3.3 Final Comments

From the analyses performed with the most representative clusters reported in Section 6.2, we learned that the radius-based FD datasets computed using radius value equals to 5 could be used for characterizing similarity between bile ducts.

When we analyzed the results from multiscale fractal dimension computation, we noticed that the smallest radius values were also those where radius-based FD datasets showed the lowest variance of FDs. We used this information to support our decision regarding the choice of the interval of radius values for computing the bile ducts' MFD descriptors, i.e., the radius values between 4.25 and 4.92.

The results reported in this section with multiescale fractal dimensions computed using the radius with the lowest FD variance from the most representative clusters showed that the range of MFDs remained between 2.37 and 2.62, for POA datasets, and 2.11 and 2.47 for USA datasets, for the red channel. For the green channel, we found that FDs range from 2.49 to 2.65, in POA datasets, and from 2.24 to 2.58 for USA datasets.

The splitting of the bile ducts in groups based on the age of the mice sampled for the study (5, 7, and 9-days old) allowed us to observe how similar or different the bile

ducts are when analyzed through a classical technique for multivariate analysis. PCA has made more evident that the shape of all bile ducts varies depending on age. Moreover, when we compare the proportion of overlapping of the clusters in the red and green channels, we observe that the bile duct walls (green channel) are more similar than the microvasculature (red channel). One might suggest that this is because the development of the microvasculature is more related to shape modification, while bile duct walls change more in size, which is not captured by the MFDs due to normalization of the datasets.

7 CONCLUSIONS AND FUTURE WORK

In this thesis, we have proposed methods based on the analysis of confocal microscopy images with the final goal of characterizing bile ducts. The study of bile ducts and associated micro-anatomy has been essential in hepatology research, especially in the detection of biliary atresia, a disease that leads to transplant or death during childhood.

We started our studies motivated and in collaboration with hepatologists from the Hospital de Clínicas de Porto Alegre and Cincinatti Children's Hospital, which were studying that disease and were trying to come up with an animal model of its development. The initial step of such study from the computational point of view is the characterization of normal bile ducts. In this chapter we review our work and draw comments on future work.

7.1 Summary of our work and contributions

Our approach to characterize bile ducts from confocal images datasets is actually a three-stage process.

The first stage was motivated by the fact that the acquired confocal images are affected by several sources of noise. To improve the quality of the acquired images, we developed a method based on anisotropic diffusion. We applied the method in several slices and compared their image quality using quantitative measures. Qualitative and quantitative analysis has shown that our results so far provide a better context for the visual study of bile ducts' microvasculature.

The significant result in this first stage was the enhanced volumetric visualization of the bile duct microanatomy, which allowed the visualization of details that are hardly seen in the original data. We reported this contribution in a first publication associated to the thesis and reproduced in Appendix A. In this way, we validate our first hypothesis (*Regarding the 3D visualization of confocal images datasets, it is possible to achieve quality by enhancing structures using a pre-processing step with appropriate techniques to deal with the noise*).

In the second stage, we explored the application of unsupervised machine learning to extract relevant structures from confocal microscopy datasets. The input for this stage was the dataset already pre-processed with the voxels to be clustered. We have explored the gradient magnitude as a feature that allowed us to extract relevant information

from the density-based spatial clustering. As a result, we obtained a better visualization of the most prominent vessels and internal structures. We reported this contribution in the second publication associated to the thesis and reproduced in Appendix B. In this way, we validate our second hypothesis (*Regarding the 3D visualization of confocal images datasets, it is possible to improve the distinction of the relevant structures using an unsupervised image segmentation method before rendering.*).

These two contributions can be considered as an answer to our first research question (*How can we improve the visualization of multichannel confocal datasets to provide a better distinction of the structures of interest?*)

Regarding the last stage, we characterize the peribiliary glands and the more prominent vessels (obtained in stage two) using fractal dimension analysis and extended the characterization with multiscale fractal dimension analysis, which allowed us to start understanding how the shape of the bile ducts structures seem to evolve. Although we had few datasets, our analyses based on fractal dimensions and multiscale fractal analyses give us some evidence that the fractal dimension is a measure that can be used for quantification and characterization of bile ducts. Moreover, to the best of our knowledge, our work is the first to analyze 3D confocal images datasets using a fractal dimension analyses approach. We reported this contribution in Chapter 6. In this way, we validate our third hypothesis (*Fractal dimension analysis can be used for the quantification and characterization of structures in confocal images datasets*).

With this final stage, we have accomplished our plans towards investigating measurements for the characterization of bile ducts obtained from confocal microscopy answering the last research question (*Which measurements can we use for quantifying and characterizing 3D structures in confocal images datasets?*).

However, as can be learned from recent literature (JONKMAN et al., 2020), quantitative analysis of confocal images datasets are still a challenge because the whole process, since the acquisition, has many issues that could compromise the quality of the acquired data. These authors concluded that *"Is 'quantitative confocal microscopy' an oxymoron? The more experience you have with confocal imaging, the more you realize just how many things can go wrong. Indeed, several experts in the field have opined that it is nearly impossible to obtain rigorous measurements of intensities in a confocal experiment. Yet, no reviewer will accept qualitative comparisons between microscopy images: you will most likely be asked to quantify them!"*. So, we can affirm that our work is a step towards a quantitative characterization of structures in confocal images datasets.

7.2 Future Work

From the application domain point of view, the envisioned future work is related to deepening our analyses of confocal datasets from bile ducts. The first study we would like to conduct as future work is the analyses of other confocal images datasets sampled from mice older than those we have analyzed here. Such work would support with more strength our conclusion regarding how the bile ducts fractal characteristic change depending on age, and what is the range of FDs for normal bile ducts. A second study related to this is the characterization of non-normal bile ducts through the analyses of how changes in the shape of bile duct walls and microvasculature affect the fractal descriptors. Such study would contribute to the research on biliary atresia.

From the computational point of view, future work can tackle issues we could not pursue during this project due to several reasons. Confocal images datasets are noisy datasets, and the staining of samples also introduces noise, which usually makes the distinction of structures harder. Image processing techniques can be improved for this kind of image. Moreover, research on segmentation is challenging in these images due to the staining process, which although introducing a "natural" segmentation, can also change parameters that could be used for a better separation of the structures of interest. We used DBSCAN for segmentation. Straightforward future work is the study of this algorithm to investigate other measurements that can be used for guiding the clusterization. As for visualization, we have adopted the conventional ray casting algorithm for volumetric visualization and have not investigated multidimensional transfer functions for improving visualization. Such techniques could also be the focus of future work. Finally, regarding the fractal analysis of volumetric datasets and 3D objects in general, there are many subjects that one could explore in future work. Instead of using PCA as we did in this work, other multivariate analyses or dimensional reduction techniques could be investigated to verify if they reveal other facts that PCA might be missing to represent.

REFERENCES

ANNADHASON, A. Fractal geometry in image processing. **International Journal of Research in Management and Technology**, v. 2, n. 1, p. 110–114, 2012.

ARAUJO, A. D. A. et al. Enhancing microscope biological images with dip techniques. In: **Computer Graphics and Image Processing, 2000. Proceedings XIII Brazilian Symposium on**. [S.l.: s.n.], 2000. p. 349–. ISSN 1530-1834.

BABU, C. S. R.; SHARMA, M. Biliary tract anatomy and its relationship with venous drainage. **Journal of clinical and experimental hepatology**, Elsevier, v. 4, p. S18–S26, 2014.

BACKES, A. R. Upper and lower volumetric fractal descriptors for texture classification. **Pattern Recognition Letters**, Elsevier, v. 92, p. 9–16, 2017.

BACKES, A. R.; BRUNO, O. M. Técnicas de estimativa da dimensão fractal: um estudo comparativo. **INFOCOMP**, v. 4, n. 3, p. 50–58, 2005.

BACKES, A. R.; BRUNO, O. M. Medical image retrieval based on complexity analysis. **Machine Vision and Applications**, Springer, v. 21, n. 3, p. 217–227, 2010.

BACKES, A. R.; CASANOVA, D.; BRUNO, O. M. Plant leaf identification based on volumetric fractal dimension. **International Journal of Pattern Recognition and Artificial Intelligence**, World Scientific, v. 23, n. 06, p. 1145–1160, 2009.

BACKES, A. R. et al. Characterizing 3d shapes using fractal dimension. In: SPRINGER. **Iberoamerican Congress on Pattern Recognition**. [S.l.], 2010. p. 14–21.

BELTRAN, L. A. et al. Enhancing the visualization of the microvasculature of extrahepatic bile ducts obtained from confocal microscopy images. In: IEEE. **Graphics, Patterns and Images (SIBGRAPI), 2016 29th SIBGRAPI Conference on**. [S.l.], 2016. p. 25–31.

BELTRAN, L. A. C. et al. Visualizing structures in confocal microscopy datasets through clusterization: A case study on bile ducts. In: IEEE. **2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.], 2019. p. 405–410.

BEYER, J. et al. Connectomeexplorer: Query-guided visual analysis of large volumetric neuroscience data. **IEEE Transactions on Visualization and Computer Graphics**, IEEE Educational Activities Department, Piscataway, NJ, USA, v. 19, n. 12, p. 2868–2877, dec. 2013. ISSN 1077-2626.

BOARD, P. A. T. E. Bile duct cancer (cholangiocarcinoma) treatment (pdq®). In: **PDQ Cancer Information Summaries [Internet]**. [S.l.]: National Cancer Institute (US), 2019.

BRUNO, O. M. et al. Fractal dimension applied to plant identification. **Information Sciences**, Elsevier, v. 178, n. 12, p. 2722–2733, 2008.

CALAPEZ, A.; ROSA, A. A statistical pixel intensity model for segmentation of confocal laser scanning microscopy images. **IEEE Transactions on Image Processing**, IEEE, v. 19, n. 9, p. 2408–2418, 2010.

CELEBI, M. E.; ASLANDOGAN, Y. A.; BERGSTRESSER, P. R. Mining biomedical images with density-based clustering. v. 1, p. 163–168, 2005.

CHAN, P.; CHENG, S. H.; POON, T.-C. Automated segmentation in confocal images using a density clustering method. **Journal of Electronic Imaging**, International Society for Optics and Photonics, v. 16, n. 4, p. 043003, 2007.

CHEN, D.; SMID, M.; XU, B. Geometric algorithms for density-based data clustering. **International Journal of Computational Geometry & Applications**, v. 15, n. 03, p. 239–260, 2005. Available from Internet: <https://doi.org/10.1142/S0218195905001683>.

CHEN, Y.-C.; CHEN, Y.-C.; CHIANG, A.-S. Template-driven segmentation of confocal microscopy images. **Computer Methods and Programs in Biomedicine**, v. 89, n. 3, p. 239 – 247, 2008. ISSN 0169-2607.

CLAXTON, N. S.; FELLERS, T. J.; DAVIDSON, M. W. Laser scanning confocal microscopy. **Department of Optical Microscopy and Digital Imaging, National High Magnetic Field Laboratory, The Florida State University**, 2006.

COSTA, L. d. F. D.; CESAR, R. M. **Shape analysis and classification: theory and practice**. [S.l.]: CRC Press, Inc., 2000.

DIPAOLA, F. et al. Identification of intramural epithelial networks linked to peribiliary glands that express progenitor cell markers and proliferate after injury in mice. **Hepatology**, Wiley Online Library, v. 58, n. 4, p. 1486–1496, 2013.

DREBIN, R. A.; CARPENTER, L.; HANRAHAN, P. Volume rendering. **SIGGRAPH Comput. Graph.**, ACM, New York, NY, USA, v. 22, n. 4, p. 65–74, jun. 1988. ISSN 0097-8930.

EMERSON, C. W. Multi-scale fractal analysis of image texture and pattern. 1998.

ESTER, M. et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: **Proceedings of the Second International Conference on Knowledge Discovery and Data Mining**. [S.l.]: AAAI Press, 1996. v. 96, p. 226–231.

FLORINDO, J. B.; CASTRO, M. D.; BRUNO, O. M. Enhancing volumetric bouligand–minkowski fractal descriptors by using functional data analysis. **International Journal of Modern Physics C**, World Scientific, v. 22, n. 09, p. 929–952, 2011.

FORMAGGIA, L.; QUARTERONI, A.; VENEZIANI, A. **Cardiovascular Mathematics: Modeling and simulation of the circulatory system**. [S.l.]: Springer Science & Business Media, 2010.

FRANGAKIS, A. S.; HEGERL, R. Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion. **Journal of structural biology**, Elsevier, v. 135, n. 3, p. 239–250, 2001.

GAMARRA, M. et al. A study of image analysis algorithms for segmentation, feature extraction and classification of cells. **Journal of Information Systems Engineering & Management**, v. 2, n. 4, p. 20, 2017.

GOULD, D. J. et al. Multifractal and lacunarity analysis of microvascular morphology and remodeling. **Microcirculation**, Wiley Online Library, v. 18, n. 2, p. 136–151, 2011.

HADWIGER, M. et al. **Real-time Volume Graphics**. Natick, MA, USA: A. K. Peters, Ltd., 2006. ISBN 1568812663.

HAHSLER, M.; PIEKENBROCK, M.; DORAN, D. dbscan: Fast density-based clustering with r. **Journal of Statistical Software**, Springer-Verlag, v. 25, p. 409–416, 2017.

HAMMAD, S. et al. Protocols for staining of bile canalicular and sinusoidal networks of human, mouse and pig livers, three-dimensional reconstruction and quantification of tissue microarchitecture by image processing and analysis. **Archives of toxicology**, Springer, v. 88, n. 5, p. 1161–1183, 2014.

HINNEBURG, A.; KEIM, D. A. An efficient approach to clustering in large multimedia databases with noise. In: **Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining**. AAAI Press, 1998. (KDD'98), p. 58–65. Available from Internet: <http://dl.acm.org/citation.cfm?id=3000292.3000302>.

HLADUVKA, J.; KÖNIG, A.; GRÖLLER, E. Curvature-based transfer functions for direct volume rendering. In: **Spring Conference on Computer Graphics**. [S.l.: s.n.], 2000. v. 16, p. 58–65.

HOHEISEL, W. et al. Confocal microscopy: Applications in materials science. **Macromolecular Materials and Engineering**, v. 286, 11 2001.

HONG, M. Advantages of laser confocal microscopes for semiconductor inspection. **Manufacturing News**, Gross Publ., n. April, 2019.

HUI, Y.; LIU, Y. Volumetric data exploration with machine learning-aided visualization in neutron science. **arXiv preprint arXiv:1710.05994**, 2017.

HUI, Y.; LIU, Y.; PARK, B.-H. Discovering features in sr _{14} cu _{24} o _{41} neutron single crystal diffraction data by cluster analysis. **arXiv preprint arXiv:1809.05039**, 2018.

IEVA, A. D. et al. **The fractal geometry of the brain**. [S.l.]: Springer, 2016.

JONKMAN, J. et al. Tutorial: guidance for quantitative confocal microscopy. **Nature protocols**, Nature Publishing Group, p. 1–27, 2020.

KELCH, I. D. et al. Organ-wide 3d-imaging and topological analysis of the continuous microvascular network in a murine lymph node. **Scientific reports**, Nature Publishing Group, v. 5, p. 16534, 2015.

KIM, H. S. et al. Multichannel transfer function with dimensionality reduction. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Visualization and Data Analysis 2010**. [S.l.], 2010. v. 7530, p. 75300A.

KINDLMANN, G. et al. Curvature-based transfer functions for direct volume rendering: Methods and applications. In: **Proceedings of IEEE Visualization**. [S.l.: s.n.], 2003. p. 513–520.

KNISS, J.; KINDLMANN, G.; HANSEN, C. Multidimensional transfer functions for interactive volume rendering. **IEEE Transactions on Visualization and Computer Graphics**, v. 8, n. 3, p. 270–285, 2002. ISSN 1077-2626.

KUIKKA, J. T. Fractal analysis in medical imaging. **International Journal of Nonlinear Sciences and Numerical Simulation**, v. 3, n. 2, p. 81–88, 2002.

LAMETSCHWANDTNER, A. et al. Histomorphology and microvasculature of extrahepatic bile ducts, extrapancreatic ducts and choledocho-pancreatic duct in the adult african clawed toad, xenopus laevis: Histomorphology and scanning electron microscopy of microvascular corrosion casts. **Transactions of the Royal Society of South Africa**, Taylor & Francis, p. 1–10, 2015.

LANDINI, G. Fractals in microscopy. **Journal of microscopy**, Wiley Online Library, v. 241, n. 1, p. 1–8, 2011.

LEEUW, W. de; VERSCHURE, P.; LIERE, R. van. Visualization and analysis of large data collections: a case study applied to confocal microscopy data. **Visualization and Computer Graphics, IEEE Transactions**, v. 12, n. 5, p. 1251–1258, 2006. ISSN 1077-2626.

LEVOY, M. Display of surfaces from volume data. **IEEE Computer Graphics and Applications**, IEEE Computer Society Press, v. 8, n. 3, p. 29–37, 1988. ISSN 0272-1716.

LI, J.; SUN, C.; DU, Q. A new box-counting method for estimation of image fractal dimension. In: IEEE. **2006 International Conference on Image Processing**. [S.l.], 2006. p. 3029–3032.

LI, Z. et al. An automatic and efficient coronary arteries extraction method in ct angiographies. **Biomedical Signal Processing and Control**, Elsevier, v. 36, p. 221–233, 2017.

LOPES, R.; BETROUNI, N. Fractal and multifractal analysis: a review. **Medical image analysis**, Elsevier, v. 13, n. 4, p. 634–649, 2009.

MADDAH, M.; SOLTANIAN-ZADEH, H.; AFZALI-KUSHA, A. Snake modeling and distance transform approach to vascular centerline extraction and quantification. **Computerized Medical Imaging and Graphics**, Elsevier, v. 27, n. 6, p. 503–512, 2003.

MANDELBROT, B. How long is the coast of britain? statistical self-similarity and fractional dimension. **Science**, American Association for the Advancement of Science, v. 156, n. 3775, p. 636–638, 1967. ISSN 0036-8075. Available from Internet: <https://science.sciencemag.org/content/156/3775/636>.

MANDELBROT, B. **The Fractal Geometry of Nature**. Henry Holt and Company, 1983. (Einaudi paperbacks). ISBN 9780716711865. Available from Internet: <https://books.google.com.br/books?id=SWcPAQAAMAAJ>.

MASTERS, B. R. Handbook of biological confocal microscopy. **Journal of biomedical optics**, International Society for Optics and Photonics, v. 13, n. 2, p. 029902, 2008.

MATSUMOTO, B. **Cell biological applications of confocal microscopy**. [S.l.]: Elsevier, 2003.

MEHMOOD, R. et al. Clustering by fast search and find of density peaks via heat diffusion. **Neurocomputing**, Elsevier, v. 208, p. 210–217, 2016.

MEIJERING, E. Cell segmentation: 50 years down the road [life sciences]. **IEEE Signal Processing Magazine**, IEEE, v. 29, n. 5, p. 140–145, 2012.

MEIJERING, E. et al. Imagining the future of bioimage analysis. **Nature biotechnology**, Nature Publishing Group, v. 34, n. 12, p. 1250, 2016.

METE, M.; KOCKARA, S.; AYDIN, K. Fast density-based lesion detection in dermoscopy images. **Computerized Medical Imaging and Graphics**, Elsevier, v. 35, n. 2, p. 128–136, 2011.

Michael W. Davidson. **Basic Concepts in Fluorescence**. 2014. <http://micro.magnet.fsu.edu/primer/techniques/fluorescence/fluorescenceintro.html>. Accessed on: November 2018.

MORALES-NAVARRETE, H. et al. Automatic recognition and characterization of different non-parenchymal cells in liver tissue. In: IEEE. **Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on**. [S.l.], 2016. p. 536–540.

MORALES-NAVARRETE, H. et al. A versatile pipeline for the multi-scale digital reconstruction and quantitative analysis of 3d tissue architecture. **Elife**, eLife Sciences Publications, Ltd, v. 4, 2015.

MU, J. et al. Segmentation, reconstruction, and analysis of blood thrombus formation in 3d 2-photon microscopy images. **EURASIP Journal on Advances in Signal Processing**, Springer, v. 2010, n. 1, p. 147216, 2009.

PARAZZA, F.; HUMBERT, C.; USSON, Y. Method for 3d volumetric analysis of intranuclear fluorescence distribution in confocal microscopy. **Computerized medical imaging and graphics**, Elsevier, v. 17, n. 3, p. 189–200, 1993.

PAUL, P. et al. Automatic noise quantification for confocal fluorescence microscopy images. **Computerized Medical Imaging and Graphics**, Elsevier, v. 34, n. 6, p. 426–434, 2010.

PERONA, P.; MALIK, J. Scale-space and edge detection using anisotropic diffusion. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, IEEE, v. 12, n. 7, p. 629–639, 1990.

PLOTZE, R. d. O. et al. Leaf shape analysis using the multiscale minkowski fractal dimension, a new morphometric method: a study with passiflora (passifloraceae). **Canadian Journal of Botany**, NRC Research Press, v. 83, n. 3, p. 287–301, 2005.

PRICE, R. L.; JEROME, W. G. **Basic Confocal Microscopy**. New York: Springer, 2011. ISBN 978-1-84800-173-2.

RAMESH, N.; OTSUNA, H.; TASDIZEN, T. Three-dimensional alignment and merging of confocal microscopy stacks. In: **2013 IEEE International Conference on Image Processing**. [S.l.: s.n.], 2013. p. 1447–1450. ISSN 1522-4880.

REICHERT, J. et al. The power of 3d fractal dimensions for comparative shape and structural complexity analyses of irregularly shaped organisms. **Methods in Ecology and Evolution**, Wiley Online Library, v. 8, n. 12, p. 1650–1658, 2017.

RIDEOUT, D. Infrared laser confocal microscopy: Fast, flexible, cost-effective inspection and metrology tool for microelectronic manufacturing. **Microscopy Today**, Cambridge Univ. Press, v. 15, n. 1, p. 36–37, 2007.

SANDER, J. et al. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. **Data mining and knowledge discovery**, Springer, v. 2, n. 2, p. 169–194, 1998.

SMITH, G. D.; SMITH, G. D.; SMITH, G. D. S. **Numerical solution of partial differential equations: finite difference methods**. [S.l.]: Oxford university press, 1985.

ŢĂLU, Ş.; GIOVANZANA, S. Image analysis of the normal human retinal vasculature using fractal geometry. **Human and Veterinary Medicine**, Bioflux SRL, v. 4, n. 1, p. 14–18, 2012.

TENGINAKAI, S.; LEE, J.; MACHIRAJU, R. Salient iso-surface detection with model-independent statistical signatures. In: **Proceedings of IEEE Visualization**. [S.l.: s.n.], 2001. p. 231–238. ISBN 0-7803-7200-X.

TORIWAKI, J.; YOSHIDA, H. **Fundamental of Three Dimensional Digital Image Processing**. New York: Springer, 2009. ISBN 978-0-387-78175-4.

TRAN, T. N. et al. A density-based segmentation for 3d images, an application for x-ray micro-tomography. **Analytica chimica acta**, Elsevier, v. 725, p. 14–21, 2012.

TRICOT, C. **Curves and fractal dimension**. [S.l.]: Springer Science & Business Media, 1994.

TSIOTSIOS, C.; PETROU, M. On the choice of the parameters for anisotropic diffusion in image processing. **Pattern Recognition**, v. 46, n. 5, p. 1369 – 1381, 2013. ISSN 0031-3203.

TSYGANKOV, D. et al. User-friendly tools for quantifying the dynamics of cellular morphology and intracellular protein clusters. In: **Methods in cell biology**. [S.l.]: Elsevier, 2014. v. 123, p. 409–427.

VARTAK, N. et al. Cholestasis-induced adaptive remodeling of interlobular bile ducts. **Hepatology**, Wiley Online Library, 2016.

VOCI, F. et al. Estimating the gradient in the perona-malik equation. **Signal Processing Magazine, IEEE**, IEEE, v. 21, n. 3, p. 39–65, 2004.

WAN, Y. et al. An interactive visualization tool for multi-channel confocal microscopy data in neurobiology research. **Visualization and Computer Graphics, IEEE Transactions on**, IEEE, v. 15, n. 6, p. 1489–1496, 2009.

WAN, Y. et al. Fluorender: an application of 2d image space methods for 3d and 4d confocal microscopy data visualization in neurobiology research. In: IEEE. **Visualization Symposium (PacificVis), 2012 IEEE Pacific**. [S.l.], 2012. p. 201–208.

WEICKERT, J. **Anisotropic diffusion in image processing**. [S.l.]: Teubner Stuttgart, 1998.

WU, Q.; MERCHANT, F.; CASTLEMAN, K. **Microscope image processing**. [S.l.]: Elsevier, 2010.

YUAN, J.; WANG, J. Perona–malik model with a new diffusion coefficient for image denoising. **International Journal of Image and Graphics**, World Scientific, v. 16, n. 02, p. 1650011, 2016.

Zhou, Y.; Liang, J. Fractal features for object recognition. In: **2014 12th International Conference on Signal Processing (ICSP)**. [S.l.: s.n.], 2014. p. 1225–1229. ISSN 2164-523X.

ZIEGLER, U.; BITTERMANN, A.; HOECHLI, M. **Introduction to Confocal Laser Scanning Microscopy (LEICA)**. 2013.

APPENDIX A — ENHANCING THE VISUALIZATION OF THE MICROVASCULATURE OF EXTRAHEPATIC BILE DUCTS OBTAINED FROM CONFOCAL MICROSCOPY IMAGES

SIBGRAPI MAIN TRACK 2016

Authors: BELTRAN, L. A. C., dos SANTOS, J. L., CRUZ, C. U., and FREITAS, C. M. D. S

Abstract: Confocal microscopy is an important tool for visualizing 3D datasets of fluorescent specimens, and has been used to investigate the structure of biological specimens. However, such images are affected by the noise introduced during the specimen preparation and image acquisition processes. Anisotropic diffusion is a non-linear filter that can significantly improve image quality while removing noise without blurring edges. This study investigates the application of anisotropic diffusion in confocal microscopy images by exploring different models for parameters' estimation. Our data consists of several slices of extrahepatic bile ducts containing a network of small vessels named Peribiliary Vascular Plexus (PVP), which are affected by several sources of noise. Experimental results show that anisotropic diffusion improved the volumetric visualization of the PVP. We validated the results using MSE and PSNR quantitative approaches and qualitative description by an expert user.

Keywords: Anisotropic diffusion; Image processing; Confocal microscopic images; Volumetric Visualization.

Enhancing the Visualization of the Microvasculature of Extrahepatic Bile Ducts Obtained from Confocal Microscopy Images

Lizeth A.C Beltran*, Jorge Luiz dos Santost†, Carolina Uribe Cruz‡, Carla M.D.S Freitas*

*Institute of Informatics
Federal University of Rio Grande do Sul UFRGS, Porto Alegre, Brazil
†Centro de Investigação em Ciências da Saúde, Universidade da Beira Interior, Portugal
‡Laboratório Experimental de Hepatologia e Gastroenterologia, Porto Alegre, Brazil
{lacbeltran, carla}@inf.ufrgs.br
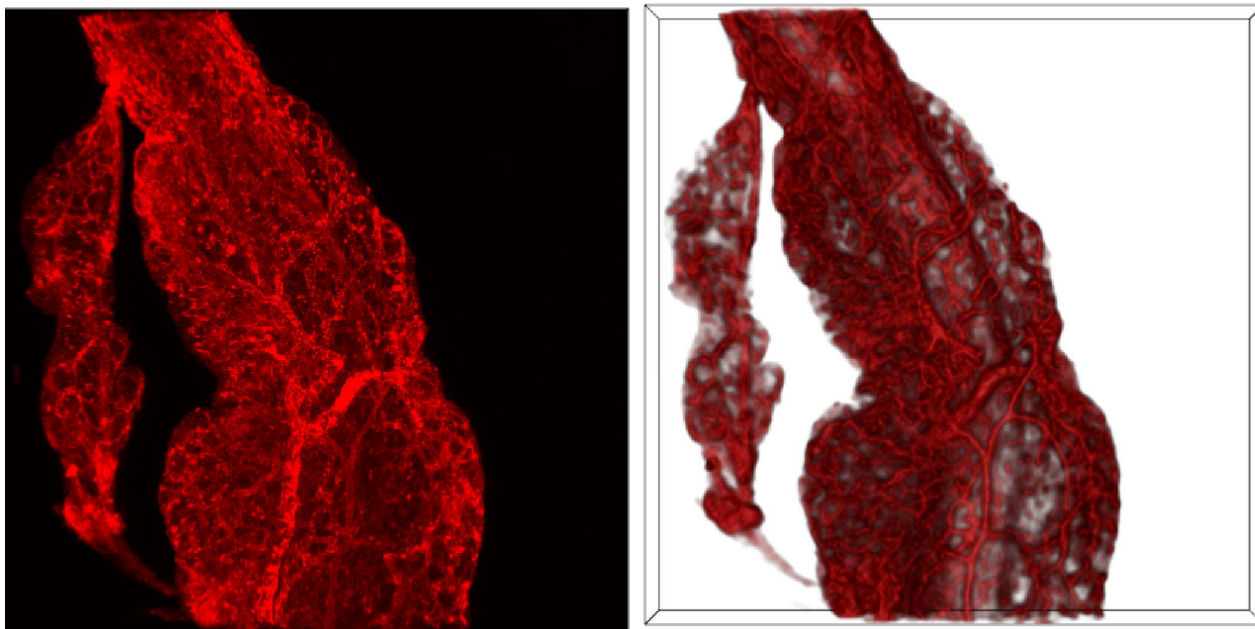
Fig. 1. Peribiliary Vascular Plexus of an extrahepatic mouse bile duct. Left: 3D visualization of the original confocal dataset. Right: Resulting volume after relevant features enhancement.

Abstract—Confocal microscopy is an important tool for visualizing 3D datasets of fluorescent specimens, and has been used to investigate the structure of biological specimens. However, such images are affected by the noise introduced during the specimen preparation and image acquisition processes. Anisotropic diffusion is a non-linear filter that can significantly improve image quality while removing noise without blurring edges. This study investigates the application of anisotropic diffusion in confocal microscopy images by exploring different models for parameters' estimation. Our data consists of several slices of extrahepatic bile ducts containing a network of small vessels named Peribiliary Vascular Plexus (PVP), which are affected by several sources of noise. Experimental results show that anisotropic diffusion improved the volumetric visualization of the PVP. We validated the results using MSE and PSNR quantitative approaches and qualitative description by an expert user.

Keywords-Anisotropic diffusion; Image processing; Confocal microscopic images; Volumetric Visualization.

I. INTRODUCTION

Bile ducts are tubular structures that carry bile from the liver to the gallbladder and duodenum. Bile ducts are classified as intra- and extrahepatic depending on their position in relation to the liver. [1]. Studying the anatomy of these structures is a hot topic in hepatology research [2]–[5]. In particular, microscopic visualization of bile ducts can provide additional information for detection of morphological changes arising

from biliary diseases. Confocal microscopy enables the acquisition of three dimensional image datasets. In recent years, this technique has been used in medical studies for investigating the micro-anatomical structure of extrahepatic bile ducts [5].

However, there are some challenges associated to confocal imaging. The first challenge is about the image post-processing. Although microscopes provide proprietary software for exploring the datasets, there is still a lack of post-processing techniques to show more details about specific structures of interest. Another challenge is associated to the kind of specimen. Confocal microscopy provides a useful tool to study 3D structure of transparent specimens, because in this case the light can pass through it with minimal scattering [6]. However, bile ducts are nontransparent, thus making necessary the design of specific clearing procedures and staining methods before the image acquisition. Consequently, the images are affected by the noise introduced during the specimen preparation process [7]. In addition to the noise associated to the specimen preparation, the Signal-to-Noise-Ratio (SNR) of the slices of the stack obtained using confocal microscopes is reduced with increasing depth [8]. Hence, the quality of the contrast decreases with increasing depth.

Our work aims at enhancing the quality of images of bile ducts and associated vessels, both obtained by confocal microscopy. In this context, image processing techniques are essential to remove noise and enhance the acquired images. In other kinds of medical images, anisotropic diffusion offers an elegant solution for image enhancement [9]–[12]. We propose the use of anisotropic diffusion to enhance the confocal images, and volumetric visualization techniques to create projections of the bile duct samples in a 3D space so the experts can examine the microvasculature distribution and interact with the bile ducts. Fig. 1 illustrates some results obtained with our approach, showing the volume rendering of the original dataset (left) and the resulting volume after the anisotropic diffusion (right). The network of vessels surrounding the bile duct can be clearly observed. These vessels supply blood to the biliary structures and are called peribiliary vascular plexus (PVP) [13]. The microscopic visualization of PVP is essential to analyze the interrelationship between a bile duct and its vascular plexus, in order to understand the development of biliary diseases associated with vascular disorders.

Contributions: The main contribution of this paper is the enhancement of the volumetric visualization of the PVP obtained from confocal microscopy images. We explore two models to estimate appropriate parameters for the anisotropic diffusion equation used to improve the original volume. As results, we enhance details that are hardly visualized in the original data. Moreover, using interactive manipulation like rotation and zooming operations on the resulting volumetric visualization, the hepatologists can have different views of the microvasculature.

The rest of this paper is organized as follows. Next section briefly discusses the noise present in confocal microscopy images and introduces the classical anisotropic diffusion equation. In Section III, we scrutinize the related works in the field

of confocal microscopy images and the estimation of parameters for the anisotropic diffusion equation. Section IV gives details about the bile duct preparation and image acquisition. In Section V we explore two existing models to calculate the k parameter for the anisotropic diffusion equation. The results of our experiments are described in Section VI, and in Section VII we discuss our findings and draw final comments.

II. BACKGROUND

In recent years there has been an explosion in the popularity of confocal microscopy [14], more specifically in cell biology applications using fixed and living cells and tissues. Confocal microscopy data have their own characteristics, which differ from other biomedical data [15]. We describe some operating principles of confocal microscopes to understand the advantages and disadvantages of this technique, which affect directly the data that is being processed. Since, anisotropic diffusion was used to reduce the noise and enhance details of the bile duct images, we also present the basic principles of anisotropic diffusion filtering.

A. Confocal Microscopy Images

In confocal miroscopy the images are acquired point-by-point, using lasers and the principle of fluorescence. Fluorescence is the property of some atoms and molecules to absorb light at a particular wavelength and to subsequently emit light of a longer wavelength after a brief interval [16]. The biological samples can be labeled with several appropriate fluorescent antibodies during the staining process, which allows to mark different tissues or cells. The characteristics of a confocal microscope offer several advantages over conventional optical microscopy such as: the ability of removing out-of-focus light [17], the capability of controlling the depth of field and the capability of collecting several aligned images of the same sample. However, confocal images are normally affected by several artifacts and noise sources:

- Low signal-to-noise ratio: confocal images have a strong decrease in the signal-to-noise ratio over the slices depth [8].
- Diversity of density values: the physical meaning of density values is not limited to image subjects [18]. Confocal images have an inhomogeneous density inherent to the fluorescent staining process [7].
- Visual occluders: structures irrelevant to the analysis may also be labeled through the fluorescent staining process, resulting in visual occluders that obscure the structures to be visualized [19].
- Subtle boundaries: meaningful boundaries may be only subtly presented in the confocal data [19].

Confocal microscopes use proprietary formats. We use datasets acquired by a Leica confocal microscope. These kind of microscopes produce datasets in a format named *LIF* (Leica Image File Format). This specific format encodes the information about the images and the microscope configuration used during acquisition. The information regarding the process of image acquisition is important for further interpretation of the

obtained data by the hepatologist. However, the downside of proprietary formats is that they need proprietary software to decode the stored images [17].

B. Anisotropic Diffusion Filtering

Anisotropic diffusion was introduced by Perona and Malik [20], and has been used as an effective approach in image processing and computer vision for noise removal, edge detection and image restoration [9] [10] The main idea behind this approach is that smoothing should be low on relevant edges and stronger in regions dominated by noise [21]. In the classical formulation [20], the anisotropic diffusion equation is given by the following PDE:

$$\frac{\partial I}{\partial t} = \nabla \cdot c(\nabla I)\nabla I \tag{1}$$

where t is the time parameter, ∇I is the gradient of the image at time t and c is the diffusivity. The diffusivity can be expressed as a decreasing function of the image gradient magnitude, such as:

$$c(x,t) = e^{-\frac{\nabla I^2}{k}}, c(x,t) = \frac{k^2}{k^2 + \nabla I^2} \tag{2}$$

where k is the gradient magnitude threshold parameter that controls the rate of the diffusion and serves as a soft threshold between the image gradients that are attributed to noise and those attributed to edges [22]. The great success of the Perona and Malik's model can be mainly attributed to its excellent performance in edge preservation and noise removal [23]. However, the estimation of the parameters for the anisotropic diffusion equation is not an easy task. In this work we explore two models to estimate the k parameter in confocal images (Section V).

III. Related Work

A. Confocal Microscopy Imaging

Related work on confocal microscopy are mostly devoted to biological studies with small structures such as cells. Examples include neurobiology research using animal models such as Drosophila and Zebrafish [15], microtubule spindles during mitosis [24], profiling gene expression of cells [25]–[27], screening phenotypic data and reconstructing the morphology of neurons [28]–[30].

In our research, differently from those previous studies, we analyze a macro structure (the bile duct). Since cells are thinner than bile ducts, this poses a challenge associated to the data because the noisy sources cited in Section II-A affect the image quality in thick specimens to a greater extent than thinner specimens such as cells.

Regarding the use of confocal microscopy to study bile ducts, DiPaola et al. [5] identified networks of glands residing within the bile duct walls. These structures were identified by the visual exploration of the serial sections using the microscope's proprietary software. However, proprietary software has limited possibilities for enhancing the images. Hammad et al. [3] and Vartak et al. [4] used confocal microscopy to visualize intrahepatic bile ducts that are much smaller than the extrahepatic bile ducts we work on.

B. Image Enhancement in Confocal Images

Usually, the proprietary software that comes with confocal microscopes allows simple contrast enhancement of 2D images. However, image filtering techniques are important when the relevant information is noisy. Median filter is a traditional filter used for noise reduction in confocal images. Parazza et al. [31] used a 3D median filter for noise reduction in confocal microscopy images from cell nuclei. A median filter was also used for noise reduction in images from rat brain [32] In another work, Paul et al. [33] used median filtering to estimate the global noise variance in images from cells. Araujo et al. [34] propose the use of blurring filter, histogram equalization and arithmetic operations to enhance cells from animal nervous system. However, traditional filters, such as the median filter, do not provide significant results to enhance details in our images. That is because they are linear filters performing in all the data, and do not discriminate the important structures we need to improve.

C. Estimating Parameters for the Anisotropic Diffusion Equation

Previous works on anisotropic diffusion indicate that the diffusion process can be improved with the proper choice of parameters of the anisotropic diffusion equation [35] [22]. According to Formaggia et al. [21], the anisotropic filter must be tuned for specific applications in terms of k and t. This implicates that empirical evaluation of the effects of the filter is necessary.

For the estimation of the k parameter, Perona and Malik [20] proposed the use of a noise estimator. This consists in a histogram of the absolute values of the gradient throughout the image and the k is computed as the 90% of its integral. On the other hand, Voci et al. [35] used a morphological approach to estimate k. Several papers discuss aspects for optimizing the anisotropic diffusion. However, none of those proposals are applied to confocal microscopy images.

IV. Image Acquisition

BALB/c mice of 5- and 7-postnatal days were euthanized by overdose of isoflurane and the extrahepatic bile ducts were isolated. The ducts were fixed with Fixer Dent's, rehydrated in decreasing concentrations of MeOH and washed with PBS/BSA/Triton. The preparation methods of bile ducts and related PVP included a technique, developed at the Cincinnati Children's Hospital [5], which is innovative in relation to the mounting of the histologic sample for in situ staining in order to preserve the anatomic structures. This image acquisition was conducted at Hospital de Clinicas de Porto Alegre (HCPA-Brazil). After that, the histologic specimens were labeled with primary antibody: PECAM-1 and CK19 (1:150 Abcam. USA) and secondary antibody Alexa Fluor 647 and 488, respectively (1:150 Abcam. USA). An example of a mouse bile duct after the specimen preparation is shown in Fig. 2.

The handling, caring, and processing of the animals were carried out according to regulations approved by local ethics committee at HCPA (protocol number 11-0190) and complied with the National Guidelines on Animal Care.
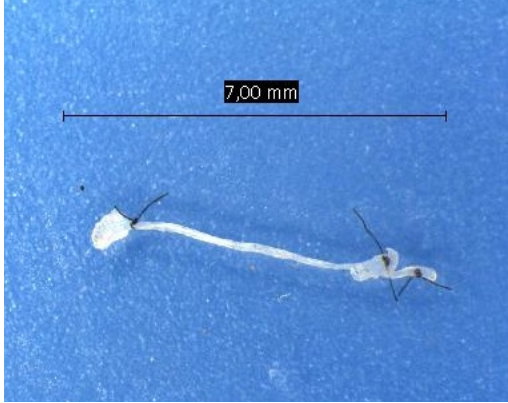

Fig. 2. An extrahepatic mouse bile duct after the clearing and staining process.

Images of all samples were obtained using the confocal microscopy Leica TCS SP5. The resulting datasets had images of 512 x 512 pixels, and the stack was composed by a variable number of slices. Then, the datasets have dimensions *X, Y* and *Z*, where *X* and *Y* are the size of each slice in pixes, and *Z* is the number of slices. Table I illustrates the information about the datasets obtained for each bile duct. The number of slices varies depending on the bile duct size and the microscope settings.

The dataset consists of images with two fluorescence channels. The first one (red channel) represents the PVP stained with PECAM-1, and the second one (green channel) represents the biliary structure stained with CK19. We extracted the red channel because this channel contains the relevant structures that are our focus in this work.

V. ESTIMATING GRADIENT THRESHOLD IN THE CONFOCAL IMAGES DATASET

Anisotropic diffusion only affects parts where the gradient value is below a certain threshold. Therefore, the estimation of this parameter plays an important role in the anisotropic diffusion process. Methods for estimating a suitable value for the *k* parameter are useful in cases in which we have no idea about an appropriate value of the diffusion coefficient and

TABLE I
SIZE OF ACQUIRED DATASETS FROM EXTRAHEPATIC MICE BILE DUCTS.

Dataset Id	Image size (pixels)	# Slices
mouse1-day5	512x512	85
mouse2-day5	512x512	102
mouse3-day5	512x512	116
mouse2-day7	512x512	140
mouse3-day7	512x512	100
mouse4-day7	512x512	117

we would like to perform noise reduction with low loss of details [35]. We compared the two approaches described in related works (Section III-C) to explore the *k* estimation in our images: the first model proposed by Perona and Malik [20] and a second model, proposed by Voci et al. [35]. Then, we use these *k* values as input parameter to calculate the anisotropic diffusion.

A. Estimation of the k parameter using Perona and Malik's model.

The model for the *k* estimation proposed by Perona and Malik [20] is based on a noise estimator using the histogram of the gradient. This noise estimator consists in calculating the histogram of the absolute values of the gradient for every image, and the *k* parameter value is equal to the 90% value of its integral [20].

Fig. 3 shows an example of a slice extracted from our datasets, the gradient histogram calculated for this image and the anisotropic filtering result.

B. Estimation of the k parameter using Voci et al. model.

The model for the *k* estimation proposed by Voci et al. [35] is based on mathematical morphology. The idea of using a morphological approach derives from the fact that morphology can be used for an estimation of noise intensity in the image. Their model is based on opening and closing operations from mathematical morphology. The *k* is given by the following equation:

$$k = \sum_{i,j \in I} \frac{(I(i,j) \circ st)}{(r.c)} - \sum_{i,j \in I} \frac{(I(i,j) \bullet st)}{(r.c)} \qquad (3)$$

where $I(i,j)$ refers to the image consisting of r rows and c columns, a structuring element st (we use a st with size 5x5), and the symbols $\circ$ and $\bullet$ represent the opening and closing operations, respectively.

Fig. 4 compares the anisotropic diffusion results using the two different models presented in Section V-A and Section V-B. As can be observed, the results show very similar filtered images using the two models.

VI. PRELIMINARY EXPERIMENTAL RESULTS

The two models for estimating the *k* parameter were implemented in Python. We used the Visualization Toolkit (VTK) [36] to render the confocal datasets. The VTK class (*vtkLIFReader*) was used to load the confocal dataset codified in the LIF format. This class was developed by Kankaanpää et al. [37], and it can be freely used for academic research [38].

Using the models presented in Sections V-A and V-B, we calculated the *k* parameter in several slices from the same dataset. The parameter *t* that represents the time in the anisotropic diffusion was experimentally established as 10 iterations.

For quantitative analyses, we calculated the Peak Signal-to-Noise Ratio (PSNR) and the Mean Square Error (MSE) for comparing the enhanced images with the original images. As for quality, the measured values of MSE should be small and
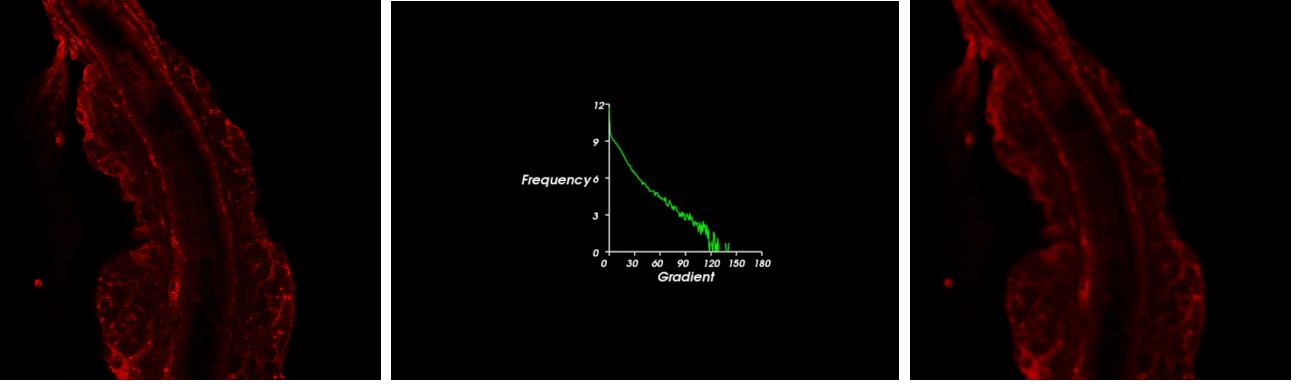
Fig. 3. Estimation of the k parameter using Perona and Malik's model. Left: Original slice. Center: Gradient histogram. Right: Anisotropic filtering result
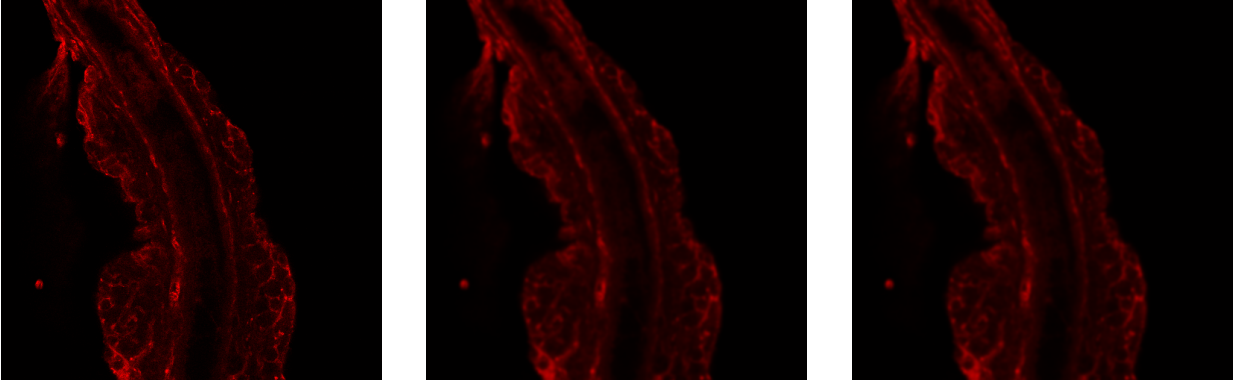
Fig. 4. Comparison of the anisotropic diffusion using the two models for estimating the k parameter. Left: Original slice. Center: Anisotropic filtering result using the Perona and Malik's model presented in Sec. V-A. Right: Anisotropic filtering result applying the Voci et al.'s model presented in Sec. V-B.

PSNR should be large. The PSNR and the MSE are defined by:

$$MSE = \frac{\sum_{i=1}^{r} \sum_{j=1}^{c} \mid I(i,j) - \hat{I}(i,j) \mid^2}{r.c} \qquad (4)$$

where $I(i,j)$ is the original image, $\hat{I}(i,j)$ is the enhanced image, and $r.c$ the size of the image.

$$PSNR = 10 \log_{10}(\frac{MAX_I^2}{MSE}) \qquad (5)$$

where $MAX_I = 2^n - 1$ and n is the number of bits. Since the confocal images are 8-bits depth, n is set to 255.

Table II and Table III show the k values and the respective measures of MSE and PSNR of the enhanced images.

After applying the anisotropic diffusion with both models, we verified that PSNR values are very similar in the resulting images. In terms of image quality this means that the two models are adequate for enhancing our images.

For enhancing the whole confocal dataset using the two models presented in Section V-A and Section V-B, we calculated the average value of k considering all slices. The results for the enhanced volume are shown in Fig. 5.

As for qualitative analyses, we invited a senior hepatologist to describe how he found the enhanced volume in comparison to the original volume. According to the hepatologist, the resulting volumetric visualization solves some problems

TABLE II
k PARAMETER ACCORDING TO PERONA AND MALIK'S MODEL AND MSE, PSNR OF THE ENHANCED IMAGES

mouse2-day7	k parameter	MSE	PSNR
Slice #1	131	25.0112	34.1494
Slice #10	139	25.7968	34.0151
Slice #20	139	25.9635	33.9871
Slice #30	138	26.9295	33.8285
Slice #43	137	29.4593	33.4385
Slice #50	155	30.9956	33.2177
Slice #60	156	32.7181	32.9829
Slice #70	145	34.3088	32.7767
Slice #80	138	34.8454	32.7093
Slice # 90	144	35.4471	32.6349
Slice #100	155	34.242	32.7852
Slice #110	137	31.7267	33.1165
Slice #120	147	27.1433	33.7941
Slice #130	141	22.3949	34.6293
Slice #140	144	18.1274	35.5474

Fig. 5. Comparison of the anisotropic diffusion using the two models for estimating the *k* parameter. Left: Original reconstructed volume. Center: anisotropic filtering results using Perona and Malik's model. Right: anisotropic filtering results using Voci et al.'s model

TABLE III
k PARAMETER ACCORDING TO VOCI ET AL.'S MODEL AND MSE, PSNR
OF THE ENHANCED IMAGES

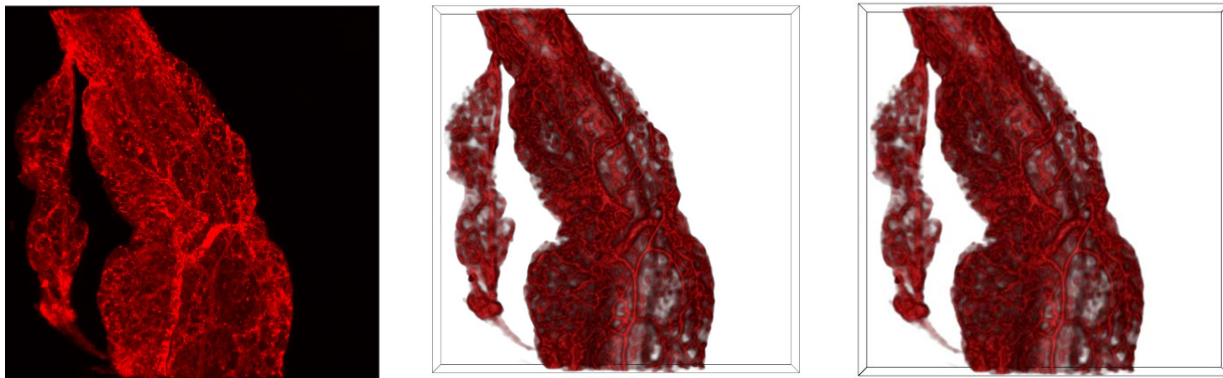| mouse2-day7	*k* parameter	MSE	PSNR
Slice #1	243	25.0208	34.1477
Slice #10	240	25.8104	34.0128
Slice #20	242	25.9736	33.9854
Slice #30	242	26.9542	33.8245
Slice #43	236	29.4543	33.4393
Slice #50	238	30.9819	33.2197
Slice #60	223	32.7058	32.9845
Slice #70	124	34.2854	32.7797
Slice #80	123	34.8371	32.7103
Slice # 90	123	35.4506	32.6345
Slice #100	126	34.2397	32.7854
Slice #110	139	31.7235	33.1169
Slice #120	140	27.1388	33.7948
Slice # 130	159	22.4014	34.628
Slice #140	182	18.141	35.5441

associated to the original data such as noise and superposition of vessels. He also commented that the microvasculature was clearly discernible, which gives a better idea of the 3D distribution of the vessels. This observation is really important because it represents that our method allows hepatologists to evaluate morphological alterations in the bile ducts.

VII. FINAL COMMENTS AND FUTURE WORK

The study of bile ducts and associated vessels is an important goal of current hepatology research. Microscopic visualization is essential to the analyses of the interrelationship between a bile duct and its vascular plexus. The acquired confocal images are affected by several sources of noise. So, the first step in the analyses of the microvasculature is to improve the quality of the acquired images.

In this paper, we presented experimental results of applying anisotropic diffusion to enhance noisy confocal images of bile ducts. We applied anisotropic diffusion in several slices and compared their image quality using quantitative measures such as PSNR and MSE. Qualitative analysis by an expert has shown that our results so far provide a better context for the

visual study of bile ducts' microvasculature.

As for future work, there are several possibilities for helping the understanding of the development of biliary diseases associated with vascular disorders. We intend to give support for quantitative analysis such as measuring the size of the structures, as well as devising shape analysis methods tailored for the microvasculature study.

REFERENCES

[1] C. S. R. Babu and M. Sharma, "Biliary tract anatomy and its relationship with venous drainage," *Journal of clinical and experimental hepatology*, vol. 4, pp. S18–S26, 2014.
[2] A. Lametschwandtner, H. Bartel, C. Radner, and B. Minnich, "Histomorphology and microvasculature of extrahepatic bile ducts, extrapancreatic ducts and choledocho-pancreatic duct in the adult african clawed toad, xenopus laevis: Histomorphology and scanning electron microscopy of microvascular corrosion casts," *Transactions of the Royal Society of South Africa*, pp. 1–10, 2015.
[3] S. Hammad, S. Hoehme, A. Friebel, I. Von Recklinghausen, A. Othman, B. Begher-Tibbe, R. Reif, P. Godoy, T. Johann, A. Vartak *et al.*, "Protocols for staining of bile canalicular and sinusoidal networks of human, mouse and pig livers, three-dimensional reconstruction and quantification of tissue microarchitecture by image processing and analysis," *Archives of toxicology*, vol. 88, no. 5, pp. 1161–1183, 2014.
[4] N. Vartak, A. Damle-Vartak, B. Richter, O. Dirsch, U. Dahmen, S. Hammad, and J. G. Hengstler, "Cholestasis-induced adaptive remodeling of interlobular bile ducts," *Hepatology*, 2016.
[5] F. DiPaola, P. Shivakumar, J. Pfister, S. Walters, G. Sabla, and J. A. Bezerra, "Identification of intramural epithelial networks linked to peribiliary glands that express progenitor cell markers and proliferate after injury in mice," *Hepatology: Official Journal of the American Association for the Study of Liver Diseases*, 2013.

[6] Y.-Y. Fu, C.-W. Lin, G. Enikolopov, E. Sibley, A.-S. Chiang, and S.-C. Tang, "Microtome-free 3-dimensional confocal imaging method for visualization of mouse intestine with subcellular-level resolution," *Gastroenterology*, vol. 137, no. 2, pp. 453–465, 2009.

[7] Y.-C. Chen, Y.-C. Chen, and A.-S. Chiang, "Template-driven segmentation of confocal microscopy images," *Computer Methods and Programs in Biomedicine*, vol. 89, no. 3, pp. 239 – 247, 2008.

[8] N. Ramesh, H. Otsuna, and T. Tasdizen, "Three-dimensional alignment and merging of confocal microscopy stacks," in *2013 IEEE International Conference on Image Processing*, Sept 2013, pp. 1447–1450.

[9] J. Weickert, *Anisotropic diffusion in image processing*. Teubner Stuttgart, 1998, vol. 1.

[10] A. S. Frangakis and R. Hegerl, "Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion," *Journal of structural biology*, vol. 135, no. 3, pp. 239–250, 2001.

[11] G. Gavriloaia, C. Neamtu, M. Gavriloaia, and A. Ghemigan, "Anisotropic diffusion filtering of infrared medical images," in *Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on*, June 2011, pp. 1–4.

[12] V. Ruela Pereira Borges, D. Junqueira dos Santos, B. Popovic, and D. Farias Cordeiro, "Segmentation of blood vessels in retinal images based on nonlinear filtering," in *Computer-Based Medical Systems (CBMS), 2015 IEEE 28th International Symposium on*. IEEE, 2015, pp. 95–96.

[13] K. Washington, P.-A. Clavien, and P. Killenberg, "Peribiliary vascular plexus in primary sclerosing cholangitis and primary biliary cirrhosis," *Human pathology*, vol. 28, no. 7, pp. 791–795, 1997.

[14] N. S. Claxton, T. J. Fellers, and M. W. Davidson, "Laser scanning confocal microscopy," *Department of Optical Microscopy and Digital Imaging, Florida State University, Tallahassee, http://www. olympusconfocal. com/theory/LSCMIntro. pdf*, 2006.

[15] Y. Wan, H. Otsuna, C.-B. Chien, and C. Hansen, "Fluorender: an application of 2d image space methods for 3d and 4d confocal microscopy data visualization in neurobiology research," in *Visualization Symposium (PacificVis), 2012 IEEE Pacific*. IEEE, 2012, pp. 201–208.

[16] Michael W. Davidson., "Basic concepts in fluorescence," http://micro.magnet.fsu.edu/primer/techniques/fluorescence/fluorescenceintro.html, June 2014, accessed on: November 2015.

[17] R. L. Price and W. G. Jerome, *Basic Confocal Microscopy*. New York: Springer, 2011.

[18] J. Toriwaki and H. Yoshida, *Fundamental of Three Dimensional Digital Image Processing*. New York: Springer, 2009.

[19] Y. Wan, H. Otsuna, C.-B. Chien, and C. Hansen, "An interactive visualization tool for multi-channel confocal microscopy data in neurobiology research," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 6, pp. 1489–1496, 2009.

[20] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 7, pp. 629–639, 1990.

[21] L. Formaggia, A. Quarteroni, and A. Veneziani, *Cardiovascular Mathematics: Modeling and simulation of the circulatory system*. Springer Science & Business Media, 2010, vol. 1.

[22] C. Tsiotsios and M. Petrou, "On the choice of the parameters for anisotropic diffusion in image processing," *Pattern Recognition*, vol. 46, no. 5, pp. 1369 – 1381, 2013.

[23] J. Yuan and J. Wang, "Perona–malik model with a new diffusion coefficient for image denoising," *International Journal of Image and Graphics*, vol. 16, no. 02, p. 1650011, 2016.

[24] S. Inoué, "Microtubule dynamics in cell division: exploring living cells with polarized light microscopy," *Annual review of cell and developmental biology*, vol. 24, pp. 1–28, 2008.

[25] O. Rubel, G. H. Weber, M.-Y. Huang, E. W. Bethel, M. D. Biggin, C. C. Fowlkes, C. L. Luengo Hendriks, S. V. Keranen, M. B. Eisen, D. W. Knowles *et al.*, "Integrating data clustering and visualization for the analysis of 3d gene expression data," *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, vol. 7, no. 1, pp. 64–79, 2010.

[26] I. U. Rafalska-Metcalf and S. M. Janicki, "Show and tell: visualizing gene expression in living cells," *Journal of cell science*, vol. 120, no. 14, pp. 2301–2307, 2007.

[27] F. Long, H. Peng, X. Liu, S. K. Kim, and E. Myers, "A 3d digital atlas of c. elegans and its application to single-cell analyses," *Nature methods*, vol. 6, no. 9, pp. 667–672, 2009.

[28] H. Peng, Z. Ruan, D. Atasoy, and S. Sternson, "Automatic reconstruction of 3d neuron structures using a graph-augmented deformable model," *Bioinformatics*, vol. 26, no. 12, pp. i38–i46, 2010.

[29] H. Peng, F. Long, and G. Myers, "Automatic 3d neuron tracing using all-path pruning," *Bioinformatics*, vol. 27, no. 13, pp. i239–i247, 2011.

[30] E. Meijering, M. Jacob, J.-C. Sarria, P. Steiner, H. Hirling, and M. Unser, "Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images," *Cytometry Part A*, vol. 58, no. 2, pp. 167–176, 2004.

[31] F. Parazza, C. Humbert, and Y. Usson, "Method for 3d volumetric analysis of intranuclear fluorescence distribution in confocal microscopy," *Computerized medical imaging and graphics*, vol. 17, no. 3, pp. 189–200, 1993.

[32] M. Maddah, H. Soltanian-Zadeh, and A. Afzali-Kusha, "Snake modeling and distance transform approach to vascular centerline extraction and quantification," *Computerized Medical Imaging and Graphics*, vol. 27, no. 6, pp. 503–512, 2003.

[33] P. Paul, H. Duessmann, T. Bernas, H. Huber, and D. Kalamatianos, "Automatic noise quantification for confocal fluorescence microscopy images," *Computerized Medical Imaging and Graphics*, vol. 34, no. 6, pp. 426–434, 2010.

[34] A. D. A. Araujo, B. M. D. Faria, H. J. Rees, and M. A. R. Silva, "Enhancing microscope biological images with dip techniques," in *Computer Graphics and Image Processing, 2000. Proceedings XIII Brazilian Symposium on*, 2000, pp. 349–.

[35] F. Voci, S. Eiho, N. Sugimoto, and H. Sekibuchi, "Estimating the gradient in the perona-malik equation," *Signal Processing Magazine, IEEE*, vol. 21, no. 3, pp. 39–65, 2004.

[36] *The Visualization Toolkit User's Guide*, Kitware, Inc., January 2003.

[37] P. Kankaanpää, L. Paavolainen, S. Tiitta, M. Karjalainen, J. Päivärinne, J. Nieminen, V. Marjomäki, J. Heino, and D. J. White, "Bioimagexd: an open, general-purpose and high-throughput image-processing platform," *Nature methods*, vol. 9, no. 7, pp. 683–689, 2012.

[38] N. Díaz Rodríguez, P. Kankaanpää, M. M. Saleemi, J. Lilius, and I. Porres, "Programming biomedical smart space applications with bioimagexd and pythonrules," in *4th International Workshop on Semantic Web Applications and Tools for the Life Sciences (SWAT4LS 2011)*, A. Paschke, A. Burger, P. Romano, M. S. Marshall, and A. Splendiani, Eds. ACM, 2011, p. 10–11.

APPENDIX B — VISUALIZING STRUCTURES IN CONFOCAL MICROSCOPY DATASETS THROUGH CLUSTERIZATION: A CASE STUDY ON BILE DUCTS

IEEE CBMS International Symposium on Computer-Based Medical Systems

MAIN TRACK 2019

Authors: BELTRAN, L. A. C., dos SANTOS, J. L., CRUZ, C. U., SHIVAKUMAR, P., BEZERRA, J., and FREITAS, C. M. D. S

Abstract: Three-dimensional datasets from biological tissues have increased with the evolution of confocal microscopy. Hepatology researchers have used confocal microscopy for investigating the microanatomy of bile ducts. Bile ducts are complex tubular tissues consisting of many juxtaposed microstructures with distinct characteristics. Since confocal images are difficult to segment because of the noise introduced during the specimen preparation, traditional quantitative analyses used in medical datasets are difficult to perform on confocal microscopy data and require extensive user intervention. Thus, the visual exploration and analysis of bile ducts pose a challenge in hepatology research, requiring different methods. This paper investigates the application of unsupervised machine learning to extract relevant structures from confocal microscopy datasets representing bile ducts. Our approach consists of pre-processing, clustering, and 3D visualization. For clustering, we explore the density-based spatial clustering for applications with noise (DBSCAN) algorithm, using gradient information for guiding the clustering. We obtained a better visualization of the most prominent vessels and internal structures.

Visualizing Structures in Confocal Microscopy Datasets Through Clusterization: A Case Study on Bile Ducts

Lizeth A. C. Beltran[*], Carolina U. Cruz[†], Jorge Luiz dos Santos[‡], Pranavkumar Shivakumar[§],
Jorge Bezerra[§] and Carla M.D.S. Freitas[*]
[*]*Federal Univ. of Rio Grande do Sul, Porto Alegre, Brazil, Email:[lacbeltran,carla]@inf.ufrgs.br*
[†]*Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil, Email: carolinaurib10@yahoo.com.ar,*
[‡]*Universidade da Beira Interior, Portugal, Email: jlsantos@fcsaude.ubi.pt,*
[§]*Cincinnati Children's Hospital, Cincinnati, USA, Email: [lpranav.shivakumar,jorge.bezerra]@cchmc.org*

Abstract—**Three-dimensional datasets from biological tissues have increased with the evolution of confocal microscopy. Hepatology researchers have used confocal microscopy for investigating the microanatomy of bile ducts. Bile ducts are complex tubular tissues consisting of many juxtaposed microstructures with distinct characteristics. Since confocal images are difficult to segment because of the noise introduced during the specimen preparation, traditional quantitative analyses used in medical datasets are difficult to perform on confocal microscopy data and require extensive user intervention. Thus, the visual exploration and analysis of bile ducts pose a challenge in hepatology research, requiring different methods. This paper investigates the application of unsupervised machine learning to extract relevant structures from confocal microscopy datasets representing bile ducts. Our approach consists of pre-processing, clustering, and 3D visualization. For clustering, we explore the density-based spatial clustering for applications with noise (DBSCAN) algorithm, using gradient information for guiding the clustering. We obtained a better visualization of the most prominent vessels and internal structures.**

Keywords-**confocal microscopy data; image processing; DBSCAN clustering; volumetric visualization**

I. INTRODUCTION

The confocal microscope has the ability to remove out-of-focus light and the capability of controlling the depth of field and collecting several aligned images of the same sample [1]. These characteristics have led to its increasing use for the acquisition of volumetric datasets from biological samples. In this work, we focus on the visualization of the bile ducts structure imaged with confocal fluorescence microscopy.

Bile ducts are thin tubular structures that carry the bile, and studying their microanatomy is a hot topic in hepatology research [2–4]. Usually, confocal microscopy images of bile ducts are studied by analyzing the serial slices obtained from fluorescent samples of mice. However, despite the effort dedicated to the samples preparation and the image acquisition itself, it still is necessary adequate computational support for the analysis and visualization of such confocal microscopy datasets. Bile ducts are mainly composed of two different groups of complex and juxtaposed microstructures (Fig.1): Microvasculature and Peribiliary glands (PBGs). The microvasculature refers to the network of small vessels

that surround the bile ducts [5]. PBGs are clusters of cells that elongate to form complex epithelial networks that course and branch within the bile duct walls [2]. The visual inspection of these structures is decisive to understand the development of biliary diseases associated with vascular disorders. However, there are some challenges to understand the characteristics of the microvasculature as well as PBGs because of their complex morphology, induced by their shapes and overlapping. Furthermore, confocal microscopy images are affected by the noise introduced due to the specimen preparation process, such as the procedure of staining [6].

In this paper, we propose an exploratory approach to detect, identify and visualize clusters of voxels that represent similar structures in bile ducts confocal microscopy datasets. We adopt clustering by the Density-Based Spatial Clustering DBSCAN algorithm, which creates clusters with arbitrary shapes, even in the presence of noise in large spatial databases [7]. To the best of our knowledge, there is no reported application of this technique in the study of data from hepatological samples. Our work aims at adapting the DBSCAN method for extracting structures from confocal images of bile ducts. The main challenge is to find the appropriate similarity features between voxels that allow for differentiating such structures. The main contributions of this work are the use of gradient information as a feature to guide the clustering process and the proposal of a specific pre-processing step that can also be used in other applications involving confocal microscopy images.

II. BACKGROUND AND RELATED WORK

Confocal microscopes produce multichannel fluorescent datasets in which each channel is collected separately [1]. Confocal images are commonly affected by some artifacts and noise, and irrelevant structures may also be labeled through the fluorescent staining process, resulting in visual occluders [8].

Regarding hepatology research, we found a few works using confocal microscopy to study the micro-anatomy of bile ducts. DiPaola et al. [2] identified peribiliary glands (PBGs)

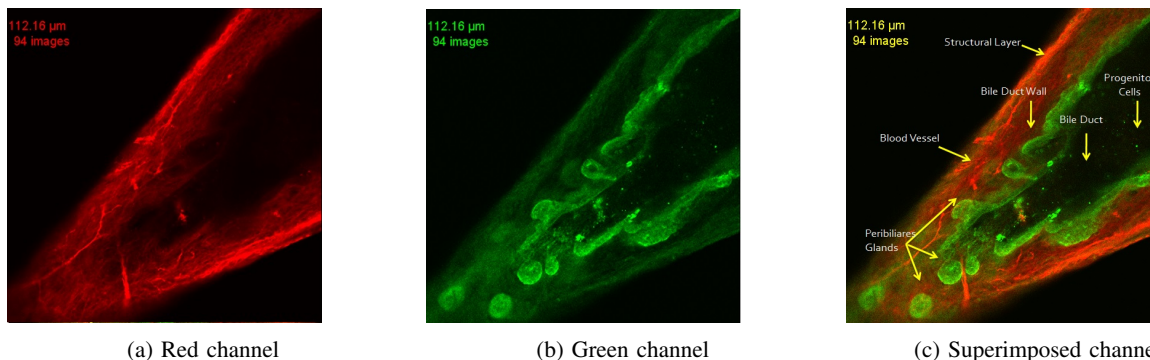(a) Red channel   (b) Green channel   (c) Superimposed channels

Figure 1: View of a single slice from a bile duct dataset: (a) the red channel encodes the microvasculature, while (b) the green one encodes the peribiliary glands. The dataset has 192 slices (512 x 512 image each) $\approx$ 50 millions of points.

residing within the bile duct walls. However, the images were visualized using the confocal microscopy proprietary software, which provided limited features for image post-processing. Hammad et al. [3] and Vartak et al. [4] used confocal microscopy to visualize intrahepatic bile ducts that are much smaller than the extrahepatic bile ducts we work with.

In a previous work [9], we proposed a pipeline to enhance and visualize the microvasculature of bile ducts. The pipeline consists of a non-linear filtering step and direct volume rendering. However, direct volume rendering requires the design of transfer functions, which are difficult to create for noisy data. In our previous work, transfer functions were obtained by a trial and error process.

In this paper, we also adopt direct volume rendering for displaying the clusters containing the structures of interest, with transfer functions based on the voxel values that characterize the clusters. Since the clustering method minimizes noise points, transfer functions are easier to design.

Clustering plays an important role in the fields of knowledge discovery and data mining [10]. Since our approach is exploratory, and we do not know a priori the number of clusters to partitioning the data, we decided to focus on density-based algorithms. The density-based clustering algorithm known as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [7] discovers clusters of arbitrary shapes and is based on two global parameters: *eps*, which is the radius around a pixel for the density calculation, i.e., the size of the *eps* neighborhood, and *minPts* that corresponds to the minimum number of points required to form a cluster.

DBSCAN has been successfully applied in images datasets obtained from different sources for application in distinct domains [11–16], including confocal images [17, 18]. Table I summarizes the main characteristics of these works.

Due to space constraints, we restrain ourselves to give details about those works related to the use of density-

based clustering in images from confocal microscopy. Mu et al. report that the density-based spatial clustering approach is useful for image segmentation of blood thrombus [17]. Mu et al. did not use DBSCAN, but a generalized version of the density-based clustering proposed by Chen et al. [19]. Chan et al. also modified a different density-based clustering method, known as DENCLUE [20], to perform segmentation in confocal images to study gene expression on zebrafish [18]. The original method is based on a set of density distribution functions, which are, in fact, influence functions that model the influence of a given data point in its neighborhood. In Chan et al., the Density-Based Segmentation (DBS) method the density function of each pixel is calculated using the differences of pixel intensity with the neighboring pixels, which is an approximation of the gradient of each pixel like we did in our approach.

From all the surveyed papers that use the DBSCAN method, two features are used to guide the clustering: pixel location and intensity. In our work, in addition to the spatial position and the size of the neighborhood of the voxel, we also use its gradient magnitude to guide the clustering.

III. DISCOVERING STRUCTURES IN BILE DUCTS IMAGES

In hepatology research, the *a priori* labels (ground truth) on the pixels are not available. Creating labels by hand is a hard task due to the complexity of the structures and the high dimensionality of data. Thus, we formulate our problem of extracting structures from these data sets as a clustering problem.

The input datasets that we use in this work were acquired at the Cincinnati Children's Hospital [2]. The mice bile duct was stained with two different fluorescent antibodies, α-tubulin and Cytokeratin CK, to mark different tissues. The resulting dataset consists of two channels: the first one (red channel) represents the microvasculature or blood vessels around the bile duct with (α-tubulin staining) (Fig.1a); the second one (green channel) represents the bile duct wall

Table I: Summary of papers reporting density-based clustering in image datasets.

Ref No.	Year	Image Acquisition	Database	Image Size	3D Stack ?	Clustering Algorithm	Clustering use	Features for clustering
[14]	2005	Dermascopy.	135 color skin lesions images.	256 x 256 pixels.	No	GDBSCAN (Generalized DBSCAN) [21].	Segmentation.	Color and pixel location.
[18]	2007	Confocal Microscopy.	4 images of zebrafish embryos.	Not mentioned.	No	Density-Based Segmentation (DBS).	Segmentation.	Intensity, pixel location.
[17]	2009	Confocal Microscopy.	15 Z-stacks of thrombi (clots). 80 2D slices per stack.	512 x 512 pixels.	Yes	Density-Based Clustering (DBC) algorithm [19].	Segmentation.	Pixel location.
[15]	2011	Dermascopy.	100 RGB color images.	From 577 x 397 to 1921 x 1285 pixels.	No	Boundary driven density-based clustering (BD-DBSCAN).	Edge detection.	Pixel location.
[16]	2012	X-ray micro- tomography.	A synthetic 3D binary image containing 144 rod-like particles.	Not mentioned.	Yes	DBSCAN for binary 3D images XMT-DBSCAN.	Edge detection.	Pixel location.
[13]	2017	CT angiography.	12 datasets of coronary arteries.	Not mentioned.	Yes	Original formulation of DBSCAN [7].	Segmentation.	Pixel location.
[11]	2017	Diffuse Scattering Spectrometer and Tomography.	1 diffuse scattering dataset and 1 neutron tomography dataset.	701 x 701 x 701 and 1997 x 1997 x 1997 pixels.	Yes	Original formulation of DBSCAN [7].	Remove noise and Segmentation.	Intensity and pixel location.
[12]	2018	Neutron Single Crystal Diffraction.	1 dataset.	501 x 501 x 501 pixels.	Yes	Original formulation of DBSCAN[7].	Data Reduction and Segmentation.	Intensity and pixel location.

containing the peribiliary glands with CK staining (Fig.1b). In the remainder of this section, we explain our approach constituted by a pre-processing phase, the clustering to isolate structures and visualization.

A. Pre-processing

We use two operations to normalize the image stacks and prepare the data for the clustering process.

- **Normalization**: We apply contrast stretching to increase the visibility of the structures.
- **Data Reduction**: We remove all points with intensity 0 (background), for eliminating unnecessary points and reducing the amount of data that will undergo the clustering phase.

B. Density-Based Spatial Clustering

The spatial information, i.e., the coordinates (x,y,z) are a typical candidate **clustering feature**. As for images, any kind of pixel (or voxel) attribute can be used as a clustering feature. Confocal images have an inhomogeneous intensity inherent to the fluorescent staining process [22], and the gradient was investigated as a more robust candidate feature. After experimenting with the intensity and gradient values, we found out that the gradient was a richer source of information for distinguishing the regions of interest. Then, we adopted the gradient magnitude to guide the clustering process.

1) Determining the parameters for 3D clustering: In the original DBSCAN algorithm [7], the key idea is that, for each point of a cluster, the neighborhood defined by a given radius *(eps)* around it has to contain at least a minimum number of points *(minPts)*, i.e., the local density in the

neighborhood has to exceed some threshold. Based on some heuristics we determined the appropriate *eps* parameter, and set *minPts* empirically. In the following, we give details about the configuration of DBSCAN for clusterizing our dataset.

The **eps-neighborhood of a point** dictates the maximum distance (radius) between two points for them to reside in the same neighborhood. A general heuristic to establish the value for *eps* is by computing the *k*-nearest neighbor distances. However, in a recent application of DBSCAN [11], a simplified calculation for *eps* was proposed. The author's idea is that the coordinates of the data points in the case of 3D image datasets are uniformly distributed voxels. Then, it is possible to use the Cartesian coordinate system and Euclidean distance to obtain the neighborhood. Values of *eps* in the interval $[1, \sqrt{2}]$ includes the six first nearest neighbors, values in $[\sqrt{2}, \sqrt{3}]$ to include the twelve second nearest neighbors, and so on. Based on this last approach, we fixed the *eps* to $1.7 \approx (\sqrt{7})$. This value means that the local density function uses 18 nearest neighbors of a given point data in the clustering.

minPts denotes the minimum number of points located in an *eps*-neighborhood, and is data dependent. If we select a low *minPts* value, we get more clusters from noise. We have experimented *minPts* values from 50 to 300, and finally set it to 200 points for the green channel and 50 points for the red channel.

The density in a neighborhood is just the sum of the **weights** of the points inside the neighborhood. By default, each data point has weights 1, so the density estimate for the neighborhood is just the number of data points inside the

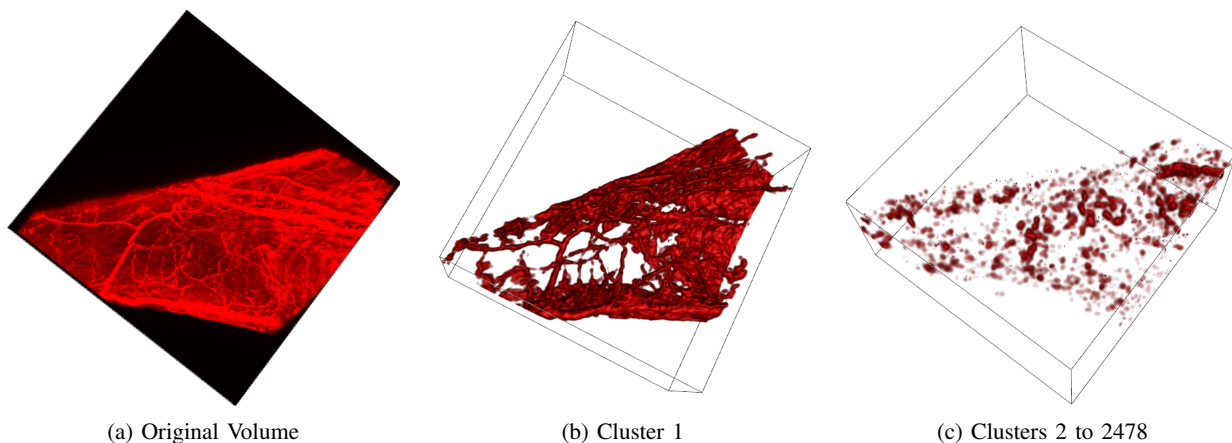(a) Original Volume   (b) Cluster 1   (c) Clusters 2 to 2478

Figure 2: 3D visualization of the microvasculature of a bile duct: points shown in (b) represent the most prominent vessels extracted as cluster 1, and those shown in (c) are considered noise and were detected as clusters 2 to 2478. The color associated to the data points maps the depth of the data points.

neighborhood. We can use the parameter *weight* to change the importance of points [23]. The *weight* is an optional parameter to perform clustering based on a specific feature.

As described before, we have chosen the gradient magnitude as a feature to guide the clustering. We follow the model for the *weight* parameter proposed by [11]. However, we use the gradient magnitude instead of intensity. We calculate the *gradient magnitude* for every point of the dataset considering the x, y, and z dimensions. Then, we take a specific value of gradient magnitude as a *threshold*. We fixed the threshold empirically as 20 for the red and the green channel. Any data point with gradient magnitude less than the threshold will take the weight of 1, while the data points with gradient magnitude greater than the threshold will have their weights assigned to the difference between their gradient magnitude and the threshold.

2) DBSCAN applied to 3D data points using gradient information: We used the DBSCAN R package [23] to perform the clustering on the 3D data points. As mentioned before, we configured the *eps* and *minPts* parameters and feed the algorithm with a list of data points containing their x, y, and z coordinates and the weight obtained from the gradient information. It is important to recall that background voxels are not considered in the clustering phase (refer to Section III-A).

In this way, the clustering method uses both information (gradient and spatial location) to obtain at least the *minPts* data points for each cluster. The output is the list of points labeled with the cluster identification of each point as well as basic numbers about the clusters detected. Then, we use the original volumetric dataset again, and voxels belonging to the cluster of interest form a new volume that is passed to the visualization module.

IV. RESULTS AND DISCUSSION

A. Microvasculature: Red Channel

Figure 2 shows 3D visualizations of selected regions in the dataset that contain the microvasculature (red channel) of the bile duct. We obtained a total of 2478 clusters from the clustering process in the red channel.

Due to the large number of clusters detected by DBSCAN, we summarize the results in the plot shown in Figure 4a, and use it to select the clusters for 3D visualization. Figure 2a shows the original dataset rendered with direct volume rendering. We identified most of the points as belonging to clusters 0 and 1. The cluster 0 is composed by 250,081 noise points, which can be discarded for visualization and analyses purposes. In other words, cluster 0 contains all the points that do not satisfy the conditions to belong to a cluster. Since clusters 1 to 2478 represent the detected objects, and cluster 1 is the largest one among them representing a connected region, it is the one that best represents the microvasculature (Figure 2b). Figure 4a shows that the clusters 2 to 2478 contain a lower quantity of points, and so we can also consider these points as noise (Figure 2c).

B. Peribiliary Glands (PBGs): Green Channel

In the case of the green channel, we obtained a total of 3,603 clusters (more clusters than in the red channel). In this case DBSCAN detected 1,998,026 noise points. Figure 3 shows the 3D visualization of data points belonging to clusters chosen among the ones that were detected in the green channel dataset. As we did in the processing of the red channel, we summarized the DBSCAN result in a plot (Figure 4b) that allowed us to analyze and select for visualization only the relevant clusters. Cluster 0 corresponds to the noise points. For the other clusters, we find a behavior similar to the red channel: cluster 1 is the most prominent

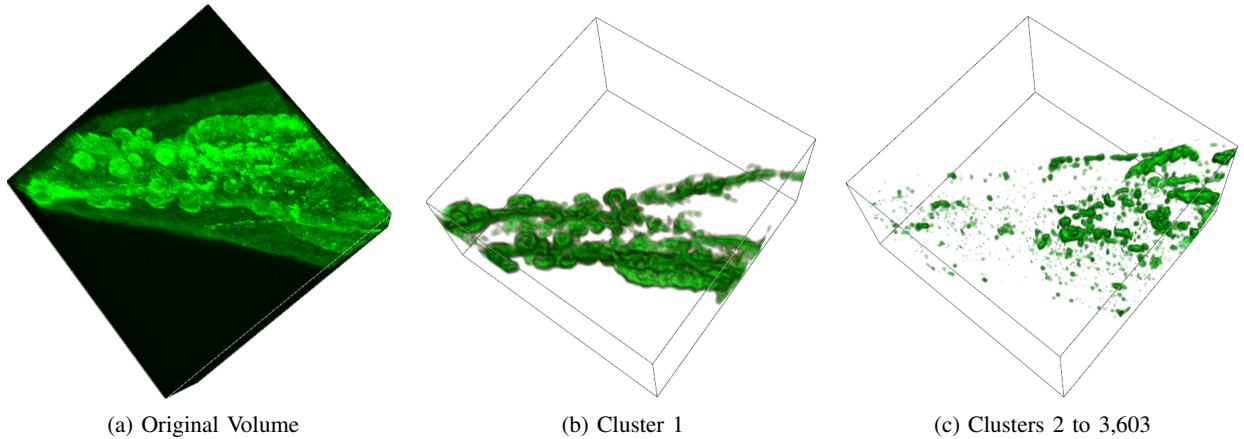(a) Original Volume   (b) Cluster 1   (c) Clusters 2 to 3,603

Figure 3: 3D visualization of the bile duct wall and PBGs: points shown in (b) represent mostly the PBGs identified as cluster 1, and those shown in (c) are also considered noise and were detected as clusters 2 to 3,603.

one, representing a connected region containing the internal bile duct wall and the peribiliary glands. The other clusters, i.e., clusters 2 to 3,603, contain a lower quantity of points, and we can also consider them as noise. While Figure 3a shows the original volume, cluster 1 representing the internal bile duct wall and the PBGs are presented in Figure 3b. Figure 3c present the clusters $[2-3603]$, considered as noise.
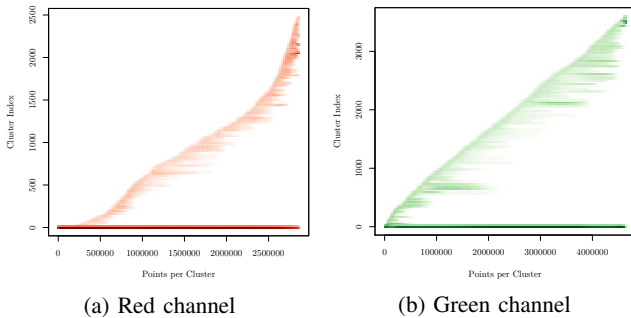

(a) Red channel   (b) Green channel

Figure 4: Number of points detected per cluster.

C. Discussion

When comparing our work to others that adapt DBSCAN for their application domain, we found different approaches. For example, Celebi et al. used the original DBSCAN method for segmenting 2D digital images of skin lesions [14], while Tran et al. presented a version of DBSCAN to process 3D binary images, using the coordinates of the original image data and solving a known instability issue of the original DBSCAN in classifying border points of adjacent objects [16]. Our method is not limited to binary images and also uses the original data points' coordinates.

Regarding the use of additional features to guide clustering with DBSCAN, only two works adopt this approach. Hui and collaborators [11, 12] use the intensity value as a feature for selecting the points during the clustering. In our work, besides the spatial position and the size of the neighborhood, we use the gradient information to select the points during the clustering.

V. FINAL COMMENTS

In this paper, we have studied the density-based spatial clustering method to extract relevant structures from confocal microscopy images of bile ducts. Our confocal microscopy images of bile ducts can be divided into two different datasets, each one representing a separate channel that encodes distinct, but hard to visualize structures: the microvasculature, in the red channel, and the bile duct wall and peribiliary glands, in the green channel.

Aiming at a better result from previous works, we employed some heuristics found in the literature to determine the appropriate parameters for the clustering. We proposed our methodology by adding some steps to be performed before the clustering phase: one step for pre-processing the volumetric dataset and another to analyzing candidate features to guide the clustering. In this latter aspect, we provide an interesting contribution: we have explored the gradient magnitude as a feature that allowed to extract relevant information from the density-based spatial clustering. Besides the fact that DBSCAN allows easy detection of noise points, an interesting result for both datasets was that the first and largest cluster found as significant for the visualization represents the structure of interest. In the red channel, this cluster represents the most prominent vessels, while in the green channel, the peribiliary glands were made more evident.

As future work, we want to explore multidimensional features to continue the search for better discriminating peribiliary glands from the internal bile duct wall. Also, since we are interested in analyzing the peribiliary glands and the

more prominent vessels, we will work on quantitative and qualitative measures for such structures.

REFERENCES

[1] R. L. Price and W. G. Jerome, *Basic Confocal Microscopy*. New York: Springer, 2011.

[2] F. DiPaola, P. Shivakumar, J. Pfister, S. Walters, G. Sabla, and J. A. Bezerra, "Identification of intramural epithelial networks linked to peribiliary glands that express progenitor cell markers and proliferate after injury in mice," *Hepatology*, vol. 58, no. 4, pp. 1486–1496, 2013.

[3] S. Hammad, S. Hoehme, A. Friebel, I. Von Recklinghausen, A. Othman, B. Begher-Tibbe, R. Reif, P. Godoy, T. Johann, A. Vartak *et al.*, "Protocols for staining of bile canalicular and sinusoidal networks of human, mouse and pig livers, three-dimensional reconstruction and quantification of tissue microarchitecture by image processing and analysis," *Archives of toxicology*, vol. 88, no. 5, pp. 1161–1183, 2014.

[4] N. Vartak, A. Damle-Vartak, B. Richter, O. Dirsch, U. Dahmen, S. Hammad, and J. G. Hengstler, "Cholestasis-induced adaptive remodeling of interlobular bile ducts," *Hepatology*, 2016.

[5] K. Washington, P.-A. Clavien, and P. Killenberg, "Peribiliary vascular plexus in primary sclerosing cholangitis and primary biliary cirrhosis," *Human pathology*, vol. 28, no. 7, pp. 791–795, 1997.

[6] Y.-C. Chen, Y.-C. Chen, and A.-S. Chiang, "Template-driven segmentation of confocal microscopy images," *Computer Methods and Programs in Biomedicine*, vol. 89, no. 3, pp. 239 – 247, 2008.

[7] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A density-based algorithm for discovering clusters in large spatial databases with noise." in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, vol. 96. AAAI Press, 1996, pp. 226–231.

[8] Y. Wan, H. Otsuna, C.-B. Chien, and C. Hansen, "An interactive visualization tool for multi-channel confocal microscopy data in neurobiology research," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 15, no. 6, pp. 1489–1496, 2009.

[9] L. A. Beltran, J. L. dos Santos, C. U. Cruzz, and C. M. Freitas, "Enhancing the visualization of the microvasculature of extrahepatic bile ducts obtained from confocal microscopy images," in *Graphics, Patterns and Images (SIBGRAPI), 2016 29th SIBGRAPI Conference on*. IEEE, 2016, pp. 25–31.

[10] R. Mehmood, G. Zhang, R. Bie, H. Dawood, and H. Ahmad, "Clustering by fast search and find of density peaks via heat diffusion," *Neurocomputing*, vol. 208, pp. 210–217, 2016.

[11] Y. Hui and Y. Liu, "Volumetric data exploration with machine learning-aided visualization in neutron science," *arXiv preprint arXiv:1710.05994*, 2017.

[12] Y. Hui, Y. Liu, and B.-H. Park, "Discovering features in sr $\_\{14\}$ cu $\_\{24\}$ o $\_\{41\}$ neutron single crystal diffraction data by cluster analysis," *arXiv preprint arXiv:1809.05039*, 2018.

[13] Z. Li, Y. Zhang, H. Gong, G. Liu, W. Li, and X. Tang, "An automatic and efficient coronary arteries extraction method in ct angiographies," *Biomedical Signal Processing and Control*, vol. 36, pp. 221–233, 2017.

[14] M. E. Celebi, Y. A. Aslandogan, and P. R. Bergstresser, "Mining biomedical images with density-based clustering," in *Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on*, vol. 1. IEEE, 2005, pp. 163–168.

[15] M. Mete, S. Kockara, and K. Aydin, "Fast density-based lesion detection in dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 35, no. 2, pp. 128–136, 2011.

[16] T. N. Tran, T. T. Nguyen, T. A. Willemsz, G. van Kessel, H. W. Frijlink, and K. van der Voort Maarschalk, "A density-based segmentation for 3d images, an application for x-ray micro-tomography," *Analytica chimica acta*, vol. 725, pp. 14–21, 2012.

[17] J. Mu, X. Liu, M. M. Kamocka, Z. Xu, M. S. Alber, E. D. Rosen, and D. Z. Chen, "Segmentation, reconstruction, and analysis of blood thrombus formation in 3d 2-photon microscopy images," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, p. 147216, 2009.

[18] P. Chan, S. H. Cheng, and T.-C. Poon, "Automated segmentation in confocal images using a density clustering method," *Journal of Electronic Imaging*, vol. 16, no. 4, p. 043003, 2007.

[19] D. Chen, M. Smid, and B. Xu, "Geometric algorithms for density-based data clustering," *International Journal of Computational Geometry & Applications*, vol. 15, no. 03, pp. 239–260, 2005.

[20] A. Hinneburg and D. A. Keim, "An efficient approach to clustering in large multimedia databases with noise," in *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, ser. KDD'98. AAAI Press, 1998, pp. 58–65.

[21] J. Sander, M. Ester, H.-P. Kriegel, and X. Xu, "Density-based clustering in spatial databases: The algorithm gdbscan and its applications," *Data mining and knowledge discovery*, vol. 2, no. 2, pp. 169–194, 1998.

[22] J. Toriwaki and H. Yoshida, *Fundamental of Three Dimensional Digital Image Processing*. New York: Springer, 2009.

[23] M. Hahsler, M. Piekenbrock, and D. Doran, "dbscan: Fast density-based clustering with r," *Journal of Statistical Software*, vol. 25, pp. 409–416.

APPENDIX C — RESUMO EXPANDIDO

Introdução

A microscopia confocal é uma ferramenta útil para adquirir dados 3D de amostras fluorescentes. Na hepatologia, pesquisadores vêm usando microscopia confocal para investigar a microanatomia dos dutos biliares. Como as imagens confocais são difíceis de segmentar devido ao ruído introduzido durante a preparação das amostras, as análises quantitativas tradicionais, em geral, são difíceis de serem executadas e requerem extensa intervenção do usuário. Assim, a análise dos dutos biliares representa um desafio na pesquisa em hepatologia, exigindo diferentes métodos.

Nesta tese, são propostos métodos para caracterizar estruturas em imagens confocais de dutos biliares. No estudo de caso motivador, supõe-se que a caracterização dessas estruturas ajudará os hepatologistas a distinguir amostras afetadas por atresia biliar, uma doença que requer transplante de fígado para evitar a morte prematura. Nossos dados consistem em volumes de imagens de dutos biliares de camundongos organizados em dois subconjuntos, um para cada canal de fluorescência. O canal vermelho contém uma rede de pequenos vasos denominados Plexo Vascular Peribiliar (PVP), e o canal verde representa o ducto biliar interno com as Glândulas Peribiliárias (PBGs). Nossa abordagem para caracterizar as estruturas dos dutos biliares inclui um processo de três estágios: um estágio para melhorar a visualização 3D dos dutos biliares, um estágio para extrair estruturas importantes e um estágio para quantificar estruturas específicas de interesse.

Principais Resultados

A primeira etapa do nosso estudo foi motivada pelo fato das imagens confocais serem afetadas por diversas fontes de ruído. Para melhorar a qualidade das imagens adquiridas, desenvolvemos um método baseado na filtragem por difusão anisotrópica. Aplicamos o método no volume e comparamos a qualidade da imagem usando medidas quantitativas. A análise qualitativa e quantitativa mostrou que nossos resultados fornecem um contexto melhor para o estudo visual da microvasculatura dos dutos biliares. O resultado significativo nesta primeira etapa foi a visualização volumétrica aprimorada da microanatomia do ducto biliar, que permitiu a visualização de detalhes que dificilmente são vistos nos dados originais. Relatamos essa contribuição em uma primeira publicação

associada à tese e reproduzida no Apêndice A. Desta forma, validamos nossa primeira hipótese (*Em relação à visualização 3D de conjuntos de dados de imagens confocais, é possível obter uma melhor qualidade de imagem através do realce de estruturas usando uma etapa de pré-processamento com técnicas apropriadas para lidar com o ruído*).

No segundo estágio, exploramos o agrupamento espacial baseado no método conhecido como *Density-based spatial clustering of applications with noise* (DBSCAN), usando, porém, informações de gradiente para orientar o agrupamento. A entrada para esta etapa foi o conjunto de dados já pré-processado contendo os pontos a serem agrupados. Exploramos a magnitude do gradiente como um recurso que nos permitiu extrair informações relevantes do agrupamento espacial baseado em densidade. Como resultado, descobrimos um cluster representativo para cada conjunto de dados. Para o canal vermelho, o cluster contém os vasos mais representativos, e para o canal verde, as estruturas internas. Relatamos essa contribuição na segunda publicação associada à tese e reproduzida no Apêndice B. Desta forma, validamos nossa segunda hipótese (*Em relação à visualização 3D de conjuntos de dados de imagens confocais, é possível melhorar a distinção das estruturas relevantes usando um método de segmentação de imagens não supervisionado antes da renderização.*).

Essas duas contribuições podem ser consideradas como uma resposta à nossa primeira pergunta de pesquisa (*Como podemos melhorar a visualização de conjuntos de dados confocais multicanal para fornecer uma melhor distinção das estruturas de interesse?*)

Por fim, exploramos os conceitos de dimensão fractal e dimensão fractal multiescala aplicados às estruturas obtidas dos agrupamentos, que consideramos úteis para extrair informações quantitativas com o objetivo de caracterizar estruturas relevantes. Nossas análises nos dão algumas evidências de que a dimensão fractal é uma medida que pode ser usada para quantificação e caracterização dos dutos biliares. Em relação à última etapa, caracterizamos as glândulas peribiliares e os vasos mais proeminentes (obtidos na etapa dois) por meio da análise da dimensão fractal e ampliamos a caracterização com a análise da dimensão fractal multiescala, o que nos permitiu começar a entender como é o formato das estruturas dos ductos biliares. Embora tivéssemos poucos conjuntos de dados, nossas análises baseadas em dimensões fractais e análises fractais multiescala nos dão algumas evidências de que a dimensão fractal é uma medida que pode ser usada para quantificação e caracterização de dutos biliares. Além disso, até onde sabemos, nosso trabalho é o primeiro a analisar conjuntos de dados de imagens confocais 3D usando uma

abordagem de análise de dimensão fractal. Relatamos essa contribuição no Capítulo 6. Desta forma, validamos nossa terceira hipótese (*A análise da dimensão fractal pode ser usada para a quantificação e caracterização de estruturas em conjuntos de dados de imagens confocais*).

Com esta etapa final, concluímos nossos planos de investigação de medidas para a caracterização de dutos biliares obtidas por microscopia confocal respondendo à última questão de pesquisa (*Quais medidas podemos usar para quantificar e caracterizar estruturas 3D em conjuntos de dados de imagens confocais?*) .

Considerações Finais

Como pode ser observado na literatura recente (JONKMAN et al., 2020), a análise quantitativa de conjuntos de dados de imagens confocais ainda é um desafio, pois todo o processo, desde a aquisição, apresenta muitos problemas que podem comprometer a qualidade dos dados adquiridos. Esses autores concluíram que *"A 'microscopia confocal quantitativa' é um oxímoro? Quanto mais experiência você tem com imagens confocais, mais você percebe quantas coisas podem dar errado. Na verdade, vários especialistas na área opinaram que é quase impossível obter medições rigorosas de intensidades em um experimento confocal. No entanto, nenhum revisor aceitará comparações qualitativas entre imagens de microscopia: muito provavelmente será solicitado que você as quantifique! "*.

Assim, podemos afirmar que nosso trabalho é um passo na direção de uma caracterização quantitativa de estruturas em conjuntos de dados de imagens confocais.