UNIVERSIDAD DE NAVARRA
ESCUELA SUPERIOR DE INGENIEROS
SAN SEBASTIÁN

Automated Image Analysis for the Characterization of Animal Models of Lung Disease using X-ray micro-Computed Tomography

DISSERTATION

submitted for the Degree of Doctor of Philosophy by

XABIER ARTAECHEVARRIA ARTIEDA

under the supervision of

Arrate Muñoz Barrutia and

Carlos Ortiz de Solórzano Aurusa

San Sebastian, February 2010

Servicio de Publicaciones de la Universidad de Navarra

ISBN 84-8081-147-1

Acknowledgement

I would like to thank the Oncology Division of the Center for Applied Medical Research and Tecnun, the School of Engineering of the University of Navarra, for the opportunity to carry out my PhD project.

This thesis is the result of a large team work. Arrate Muñoz Barrutia and Carlos Ortiz de Solórzano were the demanding but friendly coaches: thanks for your trust from the very first day. Luis Montuenga, Gorka Bastarrika, Javier J. Zulueta and Juan Pablo de Torres formed the board of directors. Despite their ties, they never minded coming down to the court for whatever it was required. Thanks for your invaluable contributions and encouragement.

Daniel Pérez Martín was the perfect team-mate: the word *no* was not in his dictionary. This thesis would have probably been impossible without him, but I learnt much more important things during those endless scanning days. Gabriel "Txiki" de Biurrun and David Blanco were simply essential: brilliant players, incredibly generous, and always with a joke or a smile ready. Even though Thomas Pengo played a different game, I kept bothering him with all sorts of problems. He was always so helpful! Mario Ceresa stepped in once the game had started, but raised all our spirits and improved the team's performance. Carlos Jauquicoa, Beatriz Úcar, Miguel Galarraga and Gabriel Heras were always there for anything. Nacho Arganda Carreras was an intermittent presence, always positive. Nacho Fernández García was lab mate, flat mate and problem mate. Nice to meet you, hope to see you soon.

Foreign players were also important. Thank you Bram van Ginneken for letting me integrate in your team for a while. Eva van Rixkoort and Adriënne Mendrik were my office mates for three months and became two friends for ever. Sascha Münzing was my psychotherapist and I was his. Clarisa Sánchez: thank you, don't ever change. Stefan Klein and Marius Staring, thanks for elastix, help and friendship: great work and exemplary generosity. In Iowa Joe Reinhardt was my host, thank you. I shared work and free time with Dragos Vasilescu, who made me feel a little bit less far

from home. Tarunashree Yavarna, Kai Ding and Panfang Hua were lab mates, always busy but always smiling.

No puedo olvidarme de las *chicas del tupper*: Laura, Elena R., Úrsula, Elena S., y otras que pasaron por aquí. Gracias por hacerme un hueco en la mesa, y hacerme olvidar papers, revisores y malos resultados durante unos minutos cada día.

Unai, Izko, Sanfo, Millán, Ibarro, Javi eta guztiak, eskerrikasko hain fan sutsuak izateagatik. Tere eta Luis, Imanol eta Mertxe, Fernando eta Xoko, Amaia eta June, nigan sinesteagatik. Atta, ama dena zor dizuet. E. G., beroa eta argiagatik.

Xabier Artaechevarria
Pamplona, February 23, 2010

Summary

Lung cancer and chronic obstructive pulmonary disease (COPD) are among the deadliest diseases worldwide, and animal models play a key role in understanding the natural history of these diseases, as well as in pre-clinical treatment trials. Different techniques can be used to study animal models of lung disease, such as pulmonary function tests or histology. X-ray micro-computed tomography (micro-CT) represents a very convenient technology to obtain three-dimensional images of the lungs with minimum invasiveness.

Multiple preliminary studies have shown the use of micro-CT to assess the progress of mouse models of lung disease. In this thesis, we set up a generic protocol for image acquisition which is of use even for heavily diseased animals. The protocol includes endotracheal intubation, pulmonary function tests and iso-pressure breath holds for movement artifact reduction. Chest micro-CT image segmentation and analysis methods have been developed to quantify the effects of disease. These methods allow for quantitative measurements on the lungs and the airways separately, which can be used to monitor disease development. Moreover, significant contributions have been made to the field of atlas-based segmentation, with applications in multiple image modalities and segmentation problems.

Developed methods have been applied to characterize the dynamic evolution of three relevant mouse models of lung disease: elastase-induced emphysema, silica-induced chronic pulmonary inflammation and urethane-induced lung cancer combined with emphysema. Apart from micro-CT, other techniques have also been used to complement the data.

Results show the use of micro-CT and automated image analysis to quantify the effect of different pulmonary diseases on small animal models. Methods presented in this thesis will be of use to characterize other models of lung disease, as well as for treatment testing.

Contents

List of Figures

List of Tables

I hesitate ... For, usually and fitly, the presence of an introduction is held to imply that there is something of consequence and importance to be introduced.

Arthur Machen, writer

1 Introduction

1.1 BACKGROUND

The lung is the basic organ for oxygen and carbon dioxide exchange in all air-breathing animals, including humans. Diseases that affect the lung are likely to impair this gas exchange and therefore compromise physiological respiration. The deadliest lung disease is chronic obstructive pulmonary disease (COPD), being the fourth leading cause of death in the developed world [1]. COPD is characterized by airflow limitation that is not fully reversible, and can be related to chronic bronchitis, emphysema, or both [2]. The disease has an important inflammatory component and is generally progressive. Lung cancer is another severe lung disease, being the most common cancer in terms of both incidence and mortality worldwide [3]. Although advances have been made in the detection and treatment of other cancer types, lung cancer still presents very high mortality. One of the main reasons for this is late detection, due to the lack of early symptoms.

Extensive research efforts are being put into both COPD and lung cancer [4, 5]. As on any other biomedical research, two different types can be distinguished depending on the environment and subject employed for the research. *In vitro* research is done in a controlled environment outside a liv-

ing organism, for instance in a test tube or a Petri dish. On the other hand, *in vivo* research is done on living organisms. These can be humans, which are generally the final subject of research. However, biomedical research on human subjects raises multiple practical and ethical issues. These can be partially overcome by using animals, especially small animals such as mice and rats. Therefore, animal models of lung disease are widely used to gain insight into the pathological processes, as well as for drug testing. Models exist for the most common diseases, including asthma, emphysema, and lung cancer [6–8].

In 1959, William Russell and Rex Burch proposed the "3 Rs" principle for animal research. Their point was that, when experimenting with animals, every effort should be made to Replace them with non-sentient alternatives, to Reduce to a minimum the number of animals used, and to Refine experiments which used animals so that they caused the minimum pain and distress [9]. The "3 Rs" principle represents not only an ethical demand, but also an incentive for high quality research.

Currently available non-invasive imaging techniques, such X-ray micro-computerized tomography (micro-CT), positron emission microtomography (micro-PET) or microscopic magnetic resonance imaging (micro-MRI), offer the possibility to go further with the "3 Rs". In particular, the number of animals can be reduced by performing longitudinal studies, and non-invasiveness allows for considerable experiment refinement. In the case of lung imaging in small animals, micro-CT is particularly well suited because the air that fills the lungs has very low X-ray absorption, compared to the tissue in the lungs. Therefore, high contrast can be achieved between tissue and air within the lungs [10]. An important drawback of this techique is the presence of respiratory motion, which results in blurred micro-CT chest images if it is not avoided or compensated for.

Once the micro-CT images have been acquired, they must be interpreted. This can be done by mere visual assessment: the extent of emphysema or inflammation can be qualitatively analyzed, the distribution and size of tumors can be compared between different animal groups, etc. Nevertheless, much more information can be obtained by quantitative image analysis, which can be done manualy (by the user), semi-automatically (with interactive methods) or fully automatically.

Multiple previous studies have shown the use of micro-CT to monitor disease development in small animal models of lung disease, such as em-

physema, lung cancer, fibrosis and inflammation [11–14]. Artifacts derived from respiratory motion being one of the main drawbacks, different solutions have been proposed and adopted to overcome this problem. They include minimizing the movement by banding the chest area with nonelastic paper tape [12], synchronizing image acquisition with different respiratory phases (gating) [15, 16] and forced iso-pressure breath holds on an endotracheally intubated animal [17]. The latter results in the best image quality, but animal recovery can be problematic, as well as its application to diseased animals.

Regarding automated image quantification, the analysis of radiological images of human subjects is an area of active research, with dozens of papers being published every year in journals such as IEEE Transactions on Medical Imaging, Medical Image Analysis or the International Journal of Biomedical Imaging. However, due to the morphological differences between humans and rodents (lung shape, airway structure), and the differences caused by the imaging modality (artifacts, image noise), methods that are valid in CT images of humans can not generally be directly translated to micro-CT images of rodents. Therefore, specific tools need to be developed and validated. The application of automatic tools to animal studies, and in particular to chest micro-CT images, has been very limited, and most published works have made use of manual or interactive tools [13, 18].

1.2 CONTRIBUTIONS

This thesis was developed in the context of two coordinated projects funded by the Spanish Ministry of Health (grants PI070751 and PI070792). The general aim of these projects is to study the relationship between COPD and lung cancer, based on the observation that smokers that suffer from COPD are more likely to develop lung cancer than the rest [19]. One project dealt with the relationship between both diseases in human samples. The second project approached the problem using animal models of the disease. Different animal models had to be thoroughly characterized. The animal models studied during the project development were:

- Elastase-induced emphysema
- Urethane-induced lung cancer
- Silica-induced inflammation

- Combined elastase and urethane model
- Combined silica and urethane model

The latter model is out of the scope of this thesis, since the lack of contrast makes it virtually impossible to detect lung nodules surrounded by dense inflammation in micro-CT images without contrast. These animal models are very different, thus the analysis and characterization has been done according to their particularities and estimated interest of the scientific community. To this aim, several methods and techniques had to be used, including micro-CT, pulmonary functional tests [20], histological analysis and cytokine measurements.

One of the main contributions of this thesis is the protocol for data acquisition, which allows obtaining lung function parameters and high quality micro-CT images *in vivo* in a considerably short time, with very low mortality. Relevant advances have also been made in the field of image analysis. Automatic methods for lung segmentation have been developed and validated, which are of use not only on healthy but also on animals presenting both low density (emphysema, for instance) or high density diseases (inflammation or fibrosis, for example). Based on the automatic segmentations, measurements that quantitatively characterize disease development can be obtained. An algorithm for airway segmentation was also developed. This algorithm is based on previously published methods for human airway segmentation on CT images, but substantial changes were required to adapt the method to micro-CT images of mice, and to include the possibility of segmenting lungs affected by disease. The devised acquisition and segmentation methods were used to characterize the animal models detailed above, thus contributing to a better understanding of the similarities and differences between the models and the actual disease. Significant contributions have also been made to the field of atlas-based segmentation, which can be applied to a wide variety of segmentation tasks. The advanced multi-atlas methods proposed in the corresponding section have not been required in our experiments, but they are present in the thesis for their generic interest.

1.3 OUTLINE

The rest of this thesis is divided in three main parts. Chapter 2 summarizes the methods that have been developed or used to acquire and analyze data for all animal models. This includes a new protocol for micro-CT

image acquisition, pulmonary function tests and various image segmentation and quantification techniques, specifically devised and validated for the analysis of small animal models of lung disease using micro-CT.

Chapter 3 separately describes the experiments and results for each animal model, which were held making use of the methods presented in Chapter 2. If a given method was used only in one of the models, it is explained in the corresponding subsection. Each section of this chapter also contains a discussion about the results of that particular model.

Finally, Chapter 4 includes a brief general discussion with the main implications of this thesis. In Appendix A a method to perform spatially variant convolution using scaled B-splines is presented. This method has many applications in image processing and was developed during the research period. It must be noted that a large part of the contents presented in this thesis have been or are being considered to be published in international conferences and journals. A detailed list of publications can be found in the appendices.

So here we have pi squared, which an engineer would call 10.

Frank King, cartoonist

2

Material and Methods

This chapter summarizes the main materials and methods employed in this thesis. These include the methods for disease generation in mice, micro-computed X-ray tomography (micro-CT) and pulmonary function data acquisition protocols, and lung and airway segmentation algorithms for the micro-CT images. A final section on multi-atlas segmentation introduces significant contributions to the field of atlas-based segmentation.

2.1 ANIMAL MODELS

In this section, the protocols followed for disease generation in each animal model will be detailed. All animal procedures were approved by the University of Navarra Animal Experimentation Ethics Committee. The studied models were:

- Elastase-induced emphysema: Animals from this group aspired 6 Units per 30 g porcine pancreatic elastase (PPE, EC134GI, EPC, MI, USA) following a previously published protocol [21]. Briefly, animals were anesthetized using a mixture of oxygen and isoflurane and laid on their backs on a metal board slanted at a 60°angle. The tongue was held with forzeps to avoid swallowing and nostrils were kept closed. Then, the aspiration fluid was dropped in the mouth and the animals

were kept in the same position for a few seconds. Animals recovered in 1 to 2 minutes from this procedure.

- Silica-induced inflammation: Inflammation was produced by a single oropharyngeal aspiration of crystalline silica (9 mg of 99% pure alpha-quartz, $< 5\mu m$ particle size, in 90 μl of saline; Min-U-Sil 5, US Silica Co., Berkeley Springs, WV, USA). Oropharyngeal aspirations were performed following the method described in the previous subsection for elastase administration.

- Urethane-induced lung cancer: Animals were injected intraperitoneally with urethane (1 mg per gram of body, Urethane U2500 Sigma, Saint Louis, Missouri, USA) at the beginning of the experiment.

- Combined elastase and urethane model: The combined model was created by first administering porcine pancreatic elastase by aspiration and injecting urethane one week later (same quantity as in the simple model).

- Control animals: To have a reference of animals without disease, control mice aspired 90 l of saline.

These models have very different appearance in micro-CT images, as shown in Figure 2.1.

2.2 MICRO-CT IMAGE ACQUISITION

Micro-CT is a widely used non-invasive imaging technique for assessment of animal models of lung disease, including lung cancer, emphysema and fibrosis [12,13,18,22]. The principles of micro-CT have been explained elsewhere [23, 24]. Briefly, an X-ray source and a detector rotate around the specimen. Projection data is acquired from different angles and it is used to reconstruct a three-dimensional image. Image acquisition protocols depend on various parameters that include the desired image quality, the delivered radiation and the X-ray absorption characteristics of the specimen or organ to be imaged. Generally, a trade-off exists between image quality and dose [25].

Apart from the scanner settings, chest micro-CT images are affected by respiratory motion, which causes image blurring. As pointed out in the introductory chapter, different techniques have been proposed to minimize

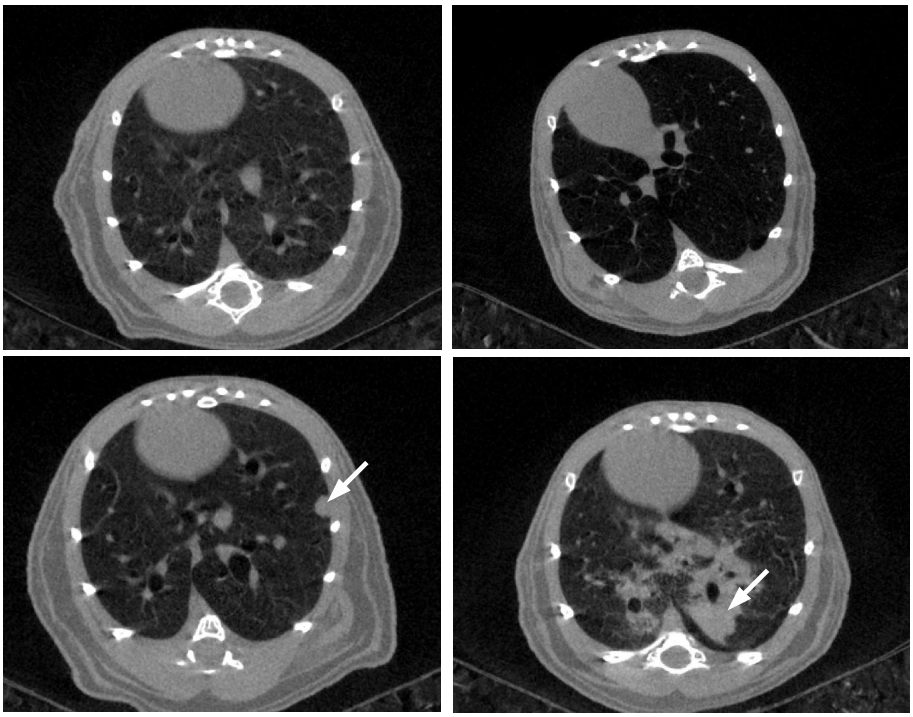


Figure 2.1: *Sample transversal micro-CT slices of different mouse models of lung disease. A healthy control subject (top left) is displayed for comparison. The elastase-induced emphysema model (top right) results in a loss of lung tissue, and the lungs appear darker and more inflated than the healthy lungs. Urethane causes spherical nodules (bottom left, white arrow). Silica aspiration results in heavy inflammation in central areas of the lungs (bottom right, white arrow).*

this problem. Solutions range from limiting the chest movement with a nonelastic tape [12], to prospective and retrospective gating [15, 26]. To achieve minimally blurred images with limited radiation, our protocol uses artificial ventilation, and the micro-CT projections are acquired during ventilator-induced breath holds.

We defined our protocol starting from the work by Namati *et al.* [17]. Several preliminary experiments were made to find the optimum workflow and scanning parameters. Every aspect was optimized to end up with a protocol with high image quality, high animal survival rate and non-detectable harm to the lungs.

Animals were anesthetized with an intraperitoneal injection of 90 mg/kg ketamine and 10 mg/kg xylacine. Endotracheal intubation was performed on anesthesized animals using the BioLite system (Biotex, Houston, Texas), to illuminate the trachea with a fiber optic stylet. After intubation, animals were connected to a Flexivent rodent ventilator (Scireq, Montreal, Canada) at a rate of 200 breaths/min and a tidal volume of 10 ml/kg. Animals were kept breathing isoflurane at 2% concentration until complete relaxation was achieved. 0.5% isoflurane was maintained during the rest of the experiment. No paralyzer was required in order to prevent mice from trying to breath on their own. Before starting the micro-CT image acquisition, lung function measurements were made (see section 2.3 for details).

Seven hundred (700) micro-CT projections were acquired during iso-pressure breath holds at 12 cm H_2O, which represents a physiological pressure and minimizes the probability of ventilator induced lung injury (VILI) [27]. Breath hold duration was 650 ms and normal breathing was induced for two inspiration-expiration cycles between breath holds. A total lung capacity (TLC) perturbation was performed every 20 breath holds. This inflates the lung to the maximum capacity (30 cm H_2O) for 3 s and contributes to prevent atelectasis [28]. No projections were acquired during the TLC perturbations.

Images were acquired with a Micro-CAT II scanner (Siemens Pre-Clinical Solutions, Knoxville, Tennesee), with a source voltage of 80 kVp and a current of 500 μA. The exposure time per projection was 450 ms and each projection was acquired during the 650 ms iso-pressure breath holds at peak inspiration (12 cm H_2O). This resulted in images of 640 slices with 1024×1024 voxels per slice, and an isotropic voxel size of 46 μm. A water phantom was used to calibrate the images to Hounsfield Units (HU).The resolution estimated by the Resolution Estimator software (Siemens Pre-Clinical Solutions, Knoxville, Tennessee) was 43.74 μm. Scan time was approximately 30 minutes and the dosage computed by the Dose Calculator software (Siemens Pre-Clinical Solutions, Knoxville, Tennessee), was 71.6 cGy per scan. To estimate image noise, we scanned a cylindrical phan-

tom filled with water and computed the standard deviation of the water region (mean 0.97 HU), which resulted in 74.55 HU.

To avoid a dangerous reduction in animals' body temperature, a plastic tube was placed around the head of the animal. This protected the mice from the air flow caused by the ventilators inside the scanner. A cotton cloth placed around the body further contributed to keep the animal warm. After the scan, animals were intraperitoneally injected with warm saline (10 μl per 1 g) to aid recovery from anesthesia and artificial ventilation.

2.3 LUNG FUNCTION MEASUREMENTS

In humans, parameters of lung function such as forced volume vital capacity (FVC) or forced expiratory volume in 1 second (FEV_1) can be measured with patient's cooperation. This is not possible with animals, thus alternative approaches must be taken. Small animals, such as mice, pose an extra difficulty due to their size [29]. There exists a trade-off between invasiveness and accuracy: the less invasive the technique, the less precise the measurement [20].

Unrestrained plethysmography lies on the least invasive side. It involves placing the subject into a small closed box and measuring the pressure changes within the box that occur as the animal breathes. The animal is conscious and unrestrained. Changes of air pressure in the lung are measured, and this data is used to compute a parameter called enhanced pause (Penh). This technique is still in use due to its simplicity, but it has been severely criticised for its lack of precision and physiological meaning [30]. On the other hand, input impedance measurement on intubated mice represents an invasive and precise method. This method consists in measuring the response to a known input pressure signal (perturbation), and deriving the respiratory system input impedance. Then, parameters of a previously defined model for the respiratory system can be computed minimizing the distance between the measured curves and what the model predicts.

As stated above, in our studies lung function measurements were done once the animals were connected to the ventilator and did not show any reflex or attempt to breath on their own, and before the micro-CT image acquisition. To obtain measurements of respiratory mechanics, two different perturbations were used. The Single-Frequency Forced Oscillation provides the estimation of the Single Compartment Model parameters [20]:

resistance (R) and compliance (C). This model assumes that the lung is an elastic compartment with a given compliance or elastance value ($E = 1/C$), served by a conduit that poses a certain resistance to air flow. The equation of the system is then given by:

$$P(t) = R\frac{dV(t)}{dt} + \frac{V(t)}{C} + P_0, \tag{2.1}$$

where $P(t)$ is the pressure at the entrance to the model, $\frac{dV(t)}{dt}$ is the flow of gas into the model, $V(t)$ is volume of gas in the elastic compartment, P_0 is the resting applied pressure (e.g., positive end-expiratory pressure), and t is time

Additionally, Broadband-Frequency Forced Oscillations allow for estimation of the Constant Phase Model parameters [31]: airway resistance (Raw), airway inertance (I), tissue damping (G), and tissue elastance (H). The input impedance of the system is expressed as:

$$Z_{in} = R + i2\pi fI + \frac{G - iH}{(2\pi f)^{\alpha}}, \tag{2.2}$$

where R is a Newtonian resistance, I is an inertance essentially equal to that of the gas in the central airways, G characterizes viscous dissipation of energy in the respiratory tissues, H characterizes energy storage in the tissues and f is frequency. G and H are coupled via the equation:

$$\alpha = \frac{2}{\pi}\arctan\frac{H}{G}. \tag{2.3}$$

This model separates the resistive and reactive components of main airways and lung tissue, which in principle allows for differentially evaluating the contribution of each of these two components to overall lung function. An important limitation of this model is that it assumes a homogeneous lung tissue.

In our experiments, each perturbation was performed three times and the results averaged, to minimize variability. Before each pair of perturbations, a total lung capacity (TLC) maneuver was performed, which inflates the lungs to the maximum capacity (30 cm H2O) for 3 s. The TLC was

performed to open small airways and to set comparable initial conditions. The entire lung function parameter acquisition took around 5 minutes.

2.4 LUNG FIXATION

The same standard protocol was used for lung fixation in all animal models. Mice were anesthetized with ketamin and xylacin and then sacrificed by exsanguination. A 22G cannula was inserted in the trachea and secured with a thread. The lungs and the heart were removed *en bloc* and then fixed for one hour with 4% formaldehyde introduced in the lungs through the cannula at a constant pressure of 20 cm H_2O. The lungs were immersed in fresh 4% formaldehyde for 24 hours and then in 70% ethanol for 24 hours, before being included in paraffin.

2.5 LUNG SEGMENTATION IN MOUSE MODELS OF LUNG DISEASE

2.5.1 PROBLEM STATEMENT AND PREVIOUS WORK

Segmenting the lungs is a basic preliminary step to quantify the effect of any pulmonary disease in micro-CT images of the mouse chest. Healthy lungs can be segmented using semi-automatic [32] or automatic methods similar to the ones used in humans [33]. Emphysematous lungs show a decreased voxel intensity, which does not require different algorithms. However, lungs showing high intensity areas, for instance due to inflammation or fibrosis, pose a considerable challenge (see Figure 2.1).

Automatically segmenting pulmonary inflammation areas in micro-CT images resembles in some respects the segmentation of computed tomography (CT) images of pathologies such as severe consolidation or extended fibrosis. Sluimer *et al.* worked on algorithms to segment scans with dense pathology [34] by combining image registration and voxel classification. More recently, Lee *et al.* used level sets to segment images of patients with diffuse interstitial lung disease (DILD) [35]. Their results were generally correct although the method failed in the most severe cases.

In this section, we explain the segmentation algorithms employed in this thesis as a preliminary step for the analysis of mouse models of lung disease. The method to segment healthy and emphysematous lungs is based

on a previously reported algorithm [33], and the method for the inflamed lungs was specifically devised for this work. Both methods are validated by comparing computer-generated segmentations to manually-defined segmentations. The rest of the section is organized as follows. In subsections 2.5.2 and 2.5.3 the algorithms for lung segmentation are explained and evaluated. In subsection 2.5.4 some implications of the methods and results are discussed.

2.5.2 SEGMENTATION OF HEALTHY AND EMPHYSEMATOUS LUNGS

2.5.2.1 Method

The segmentation algorithm used to extract the healthy and emphysematous lungs is largely based on the method described by Hu *et al.* [33]. In particular, the implementation by Carlos Jauquicoa for his Master Thesis was used [36]. The basic steps of the algorithm are stated below, and we refer to the mentioned two works for further details:

1. *Automatic thresholding*: a threshold is iteratively computed to separate air and body tissue in the scan.
2. *Background removal*: air regions connected to the borders are removed.
3. *Small volume removal*: small air volumes, which are likely to correspond to air blobs in the digestive tract, are removed.
4. *Hole filling*: possible holes within the lung are filled.
5. *Trachea extraction*: the trachea and the main airways are deleted from the volume.
6. *Small vessel inclusion*: a 3D closing operator is used to include small vessels in the lung segmentation.

2.5.2.2 Evaluation

Six images of control and emphysematous mice were semi-automatically segmented using the Amira software (Visage Imaging, Fürth, Germany)

Table 2.1: *SI for the six manually segmented images of control and emphysematous mice*

Image number and group	Similarity Index
Image 1 (control, week 0)	0.99
Image 2 (control, week 4)	0.98
Image 3 (control, week 14)	0.99
Image 4 (emphysema, week 4)	0.99
Image 5 (emphysema, week 14)	0.99
Image 6 (emphysema, week 34)	0.98

and were used as reference for the evaluation of the automatic segmentation method. A seeded region growing tool was used for segmentation, and boundaries were corrected in case of inaccuracies.

The SI between two segmentations, S_a and S_b, of the same object, is defined as:

$$SI = \frac{2|S_a \bigcap S_b|}{|S_a| + |S_b|}, \tag{2.4}$$

where $\bigcap$ indicates the overlapping voxels between the two segmentations, and $|S_a|$ indicates the number of voxels of the corresponding segmentation [37]. SI has value 1 when there is a perfect match between labels and 0 when there is no overlap.

Resulting SI values are shown in Table 2.1. The average SI is 0.99.

2.5.3 SEGMENTATION OF LUNGS AFFECTED BY CHRONIC INFLAMMATION

2.5.3.1 Method

Due to the difficulty of the task, our lung segmentation algorithm combines three different segmentation strategies: threshold-based segmentation, atlas-based segmentation and geodesic active contours. This subsection details the implementation of each technique separately, concluding with the method used to combine the three segmentation results.

2.5.3.1.1 Threshold-based segmentation

We first segmented the healthy, i.e., low density, parts of the inflamed lungs using the algorithm used for healthy and emphysematous lung segmentation (see subsection 2.5.2). With this method the largest connected area of the lung with low density is segmented, but the inflamed areas remain unsegmented.

2.5.3.1.2 Atlas-based segmentation

This step is essential for the correct inclusion of the diseased lung areas in the final segmentation. Furthermore, the result of this step will be used for the initialization of the final level set refinement. Atlas-based segmentation uses registration between a reference image, the atlas image, and the image to be segmented, the target image, to transform the segmentation of the atlas image into a segmentation of the target image [38]. Further details on atlas-based methods can be found in section 2.7.

In this case, a randomly chosen silicotic lung was manually segmented to serve as an atlas for the segmentation of the rest. Instead of directly registering the atlas and the target images, binary images containing rib segmentations were used. The reason is that there is a large potential variability in inflammation location, extension and appearance, which can lead to errors in the registration. Although inflamed areas appear mainly near the main airways in the upper part of the lung, they can occasionally be in the lower parts too. Therefore, in all images ribs were segmented by thresholding and the resulting binary images were registered. A Gaussian blurring was applied on the binarized images to help registration. The alignment process consisted of two steps: affine registration and B-Spline non-rigid registration. Mutual information was used as metric together with a stochastic gradient descent optimizer, which allows for fast and accurate registrations [39].

2.5.3.1.3 Geodesic Active Contours for refinement

The result of the atlas-based step is generally accurate at the ribs, has the expected lung shape in the upper, inflamed areas of the lung, but is inaccurate around the diaphragm and the heart. To refine the segmentation

results in those areas we used geodesic active contours [40] . The geodesic active contours, which were implemented in ITK, follow the equation [41]:

$$\frac{d\psi}{dt} = -\alpha \mathbf{A}(\mathbf{x}) \cdot \nabla\psi - \beta P(\mathbf{x}) \cdot |\nabla\psi| + \gamma Z(\mathbf{x})\kappa \cdot |\nabla\psi| \tag{2.5}$$

where ψ is the level set function; α, β and γ are the weights for the advection, propagation and curvature terms, respectively; $\mathbf{A}(\mathbf{x})$ is an advection term; $P(\mathbf{x})$ is a propagation term; and $Z(\mathbf{x})$ is a spatial modifier for the curvature term κ.

To limit the refinement to the areas of interest near the diaphragm, we applied the active contour segmentation only to the lower two thirds of the lung in the coronal direction. The weights were set to 0.1 for propagation, 10 for curvature and 20 for advection. The high curvature and advection terms lead to smooth segmentations, which at the same time are strongly attracted to the edges.

2.5.3.1.4 Segmentation combination

The three previous segmentation results must be combined into a single final segmentation. In this fusion process, two basic assumptions are made:

- All voxels segmented by the threshold-based segmentation method or the level sets are correct.
- All voxels segmented by the atlas-based segmentation method are correct, as long as there is another lung voxel with the same x-y coordinates between the current point and the diaphragm, segmented by any of the other two methods. An x-y plane is defined as an axial or transverse slice of the micro-CT image.

With these assumptions, no extra voxel is added below the diaphragm, which is assumed to be well delineated by the combination of the threshold-based and the level sets segmentations. The atlas-based segmentation is used to add pathological areas int the upper area of the lung.

Figure 2.2 shows an example of the segmentation process.

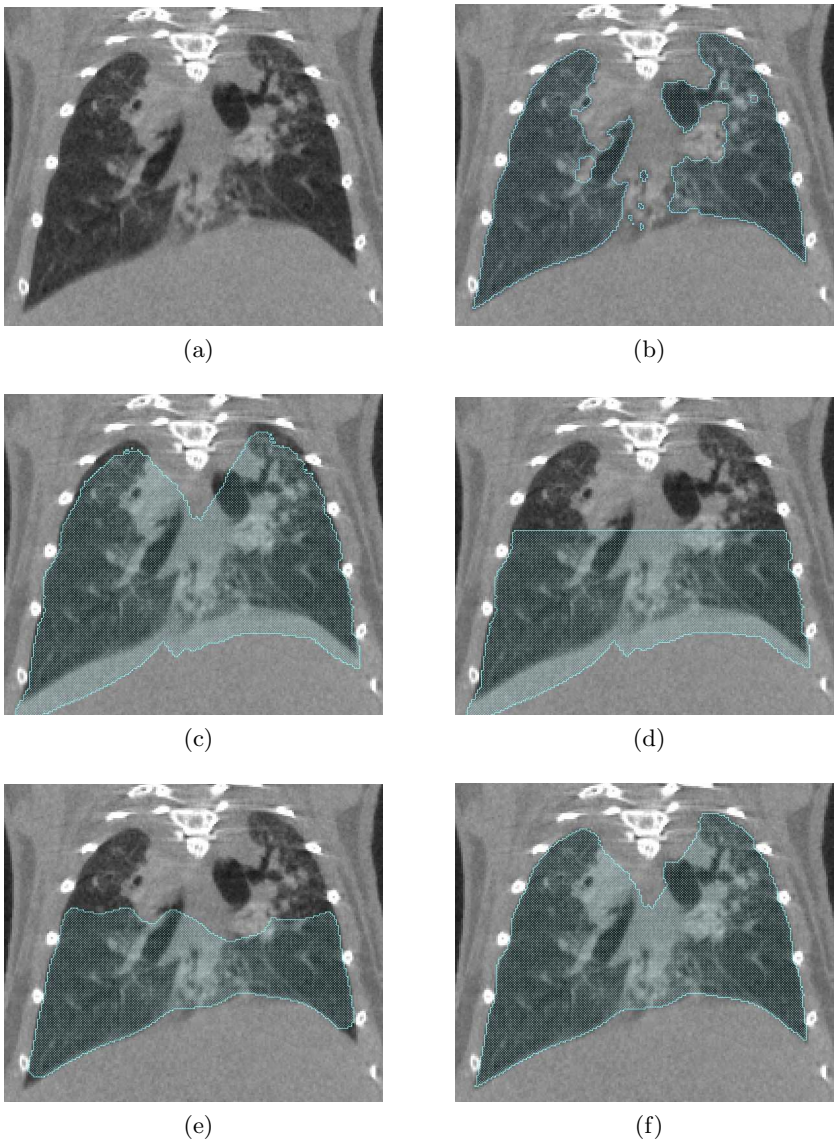


Figure 2.2: *(a) Micro-CT image of a silicotic lung; (b) Threshold-based segmentation; (c) Atlas-based segmentation; (d) Cropped atlas-based segmentation; (e) Cropped atlas-based segmentation refined using level sets; (f) Final combined segmentation.*

Table 2.2: *SI for the six manually segmented images of the chronic inflammation study*

Image number and group	Similarity Index
Image 1 (control, week 0)	0.97
Image 2 (silica, week 0)	0.97
Image 3 (silica, week 4)	0.94
Image 4 (silica, week 4)	0.96
Image 5 (silica, week 14)	0.94
Image 6 (silica, week 14)	0.93

An airway segmentation algorithm similar to the one reported in section 2.6 of this thesis was used to remove the airways from the final lung segmentation.

2.5.3.2 Evaluation

Six different scans were manually traced using the Amira software (Visage Imaging, Fürth, Germany) and were used as reference for the evaluation of the automatic segmentation method. The similarity index (SI) between the manual and the automatic segmentations was computed (see section 2.5.2.2).

Table 2.2 shows the SI for the six manually segmented images. The average SI is 0.95.

Automatic segmentations can be used to obtain a three-dimensional view of the lung volume. In Figure 2.3, the evolution of silica-induced inflammation in a mouse can be seen.

2.5.4 DISCUSSION AND CONCLUSION

The quantitative study of animal models of lung disease using micro-CT requires accurate segmentations of the lungs. Automatic segmentation methods such as the ones described in this section are essential, because manual segmentations are very time consuming, requiring as much as 5 hours in the silica-treated cases. Results show that our automatic segmentation methods yield accurate delineations.

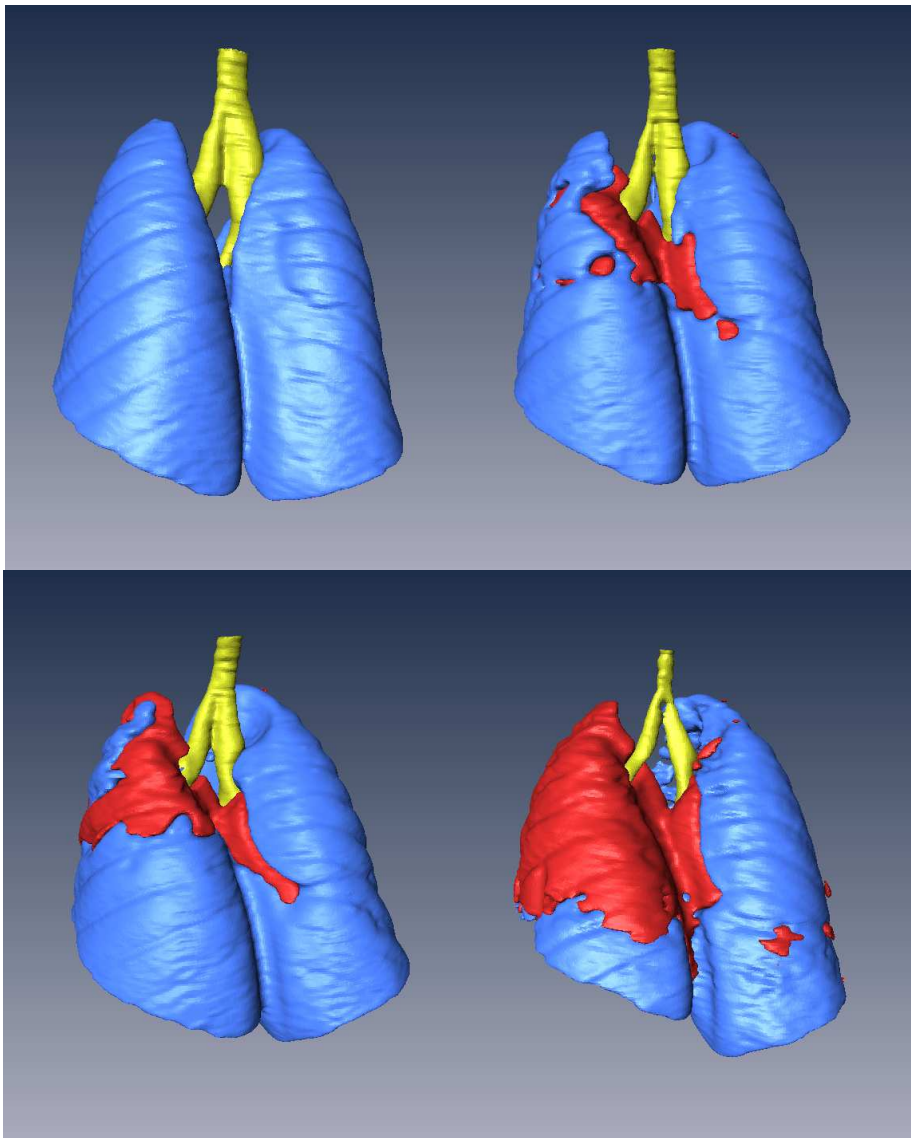


Figure 2.3: *Temporal evolution of silica-induced inflammation on the same mouse. From top-left to bottom-right, time points are baseline, 4 weeks, 14 weeks and 34 weeks after silica aspiration. The blue volume indicates lung with normal appearence, the red volume indicates the inflamed volume and airways are represented in yellow. The increase of the inflamed area can be observed.*

Automatic segmentation of human lungs without dense pathology in CT images can be said to be a solved problem [33, 34]. Despite the differences between the image modalities and imaged subjects, very similar techniques can be used to segment the same kind of lungs in micro-CT images of mice chests. In fact, the method proposed and evaluated here is an implementation of the multi-step method initially proposed by Hu *et al.* [33]. Evaluation results reveal that virtually perfect segmentations are achieved. However, the same technique can not be used in presence of dense pathologies, because it relies on the contrast between the air in the lungs and tissue. Therefore, a novel method which combines multiple steps had to be devised and evaluated.

Atlas-based segmentation is one of the steps of the devised automatic segmentation method for lungs with inflammation. A randomly chosen image of the silicosis group was chosen as the atlas image. Segmentation accuracy results show that this approach does not lead to large errors, but overall accuracy could improve with the use of a more generic atlas (see section 2.7 for possible atlas selection and combination strategies).

In conclusion, we have used a well-known method for the segmentation of healthy and emphysematous lungs on micro-CT images. However, this method was not valid to segment mouse lungs affected by chronic inflammation. The intrinsic complexity of the latter segmentation task has produced the need for a custom segmentation method, within which several well-known methods are combined taking the characteristics of the expected disease into account.

2.6 AIRWAY SEGMENTATION IN MOUSE MODELS OF LUNG DISEASE

2.6.1 PROBLEM STATEMENT AND PREVIOUS WORK

Segmenting the airways is required by most automatic lung analysis algorithms, especially in mice micro-CT images, due to the relatively large volume occupied by the airways. It is well known that the functionality of central airways and lung peripheral areas varies considerably. Thus, the morphological effects of most lung pathologies are generally different in the airways and the parenchyma. Separating airways from the rest of the lung is thus of utmost importance when analyzing a disease (like emphysema, chronic inflammation or cancer) which affects lung density differentially. Moreover, the analysis of segmented airways may be interesting in itself to study airway-specific morphological changes such as stenosis, bronchiectasis, etc. Finally, the airways can be used as a reference for image registration in follow-up studies or for atlas-based segmentation of the lungs, lobes and pulmonary segments.

To the best of our knowledge, only two airway segmentation methods in micro-CT images of mice lungs have been reported. Chaturvedi *et al.* [42] segmented silicon casts of excised mice lungs using an interactive region growing algorithm. More relevant to the present work, Shi *et al.* [43] reported a fully automatic algorithm to segment *in vivo* images of healthy mice, acquired with an iso-pressure breath hold protocol [17]. This algorithm works by searching airway lumens in 2-D transversal slices that are then used to reconstruct the full 3-D tree. The main drawback of this method is that it is based on the search of candidates in 2-D, which may lead to problems due to the high variablity of the shape of the airway lumens as seen in transversal slices. Moreover, reported computation times were fairly high (about 30 minutes for each scan).

Due to the scarce previous work on mouse airway segmentation, it is worth reviewing the most relevant approaches that exist for segmenting the human airways in CT images. Schlathoelter *et al.* [44] introduced an algorithm for simultaneous segmentation and reconstruction of the airways. This algorithm was based on a propagating front that divides the tree into branches during segmentation. The framework was extended and generalized in a subsequent work by Bülow *et al.* [45]. More recently, van Ginneken *et al.* [46] also used the framework for human airway segmentation with a

multi-threshold approach to increase robustness. Kiraly *et al.* [47] compared two different methods: an adaptive region growing algorithm and an algorithm that combined region growing and mathematical morphology. Their conclusion was that the region growing method was faster than the hybrid method, but also slightly less accurate. Aykac *et al.* [48] used a two step approach. In a first step, candidate airways were identified on transversal slices using grayscale morphological reconstruction. In a second step, valid candidates were connected to build a 3-D airway tree. Fetita *et al.* [49] also employed mathematical morphology to obtain a first approximation of the airways. The morphological operator worked in 3-D and was specifically built for this purpose. An energy-minimizing reconstruction algorithm was used to build the final airway tree. The work by Tschirren *et al.* [50] was based on fuzzy connectivity. They made use of small adaptive regions of interest around the already segmented airway areas. Thus, the algorithm adapted to local image characteristics, leaks were detected early and the computation time was reduced. In a recent work, Graham *et al.* [51] present a method that can be summarized in three main steps. In the first step, a conservative segmentation of the major airways is obtained via 3-D region growing on heavily smoothed data. Then, possible branch segments are detected and connected to each other by nonlinear filtering and surface interpolation. The final segmentation is obtained using a global graph partitioning algorithm, which connects the valid branch segments to the major airways.

The main reason that prevents a direct translation of these algorithms to mice micro-CT data is the low signal to noise ratio (SNR) of the images, caused by the required small voxel size and the limitation on radiation dose imposed by the *in vivo* studies. In particular, following the simple model assumptions from reference [52], a reduction of one order of magnitude in voxel size (0.5 mm in CT to 50 μm in micro-CT) implies an increase of two orders of magnitude in the variance of the linear attenuation coefficient, if the rest of parameters remain unchanged.

Taking this into account, we decided to develop on the flexible segmentation and reconstruction framework first reported by Schlathoelter *et al.* [44]. This framework has multiple advantages. First, it allows for simultaneous segmentation and reconstruction of tree-like structures. Second, the topological and morphological information from the segmented tree can be used to guide the segmentation of the remaining branches. Third, its modular configuration allows for easy introduction of application-specific

segmentation rules. In fact, one of the major contributions of our work is the use of new features when compared to those in the previously reported applications of the framework, to adapt to the low SNR and the special morphology of the mice airways.

To complete the morphological information provided by the image-based measurements, we use respiratory system input impedance measurements and the constant phase model parameters, which have been widely used to assess respiratory mechanics in multiple animal models ([31], [53], [54]). The combination of micro-CT imaging and constant phase model parameters has also been used to analyze animal models of lung disease before. In particular, Lundblad *et al.* [55] qualitatively analyzed *ex vivo* micro-CT images in a mouse model of allergical inflammation and combined it with measurements of tissue elastance. In this work, we propose to use quantitative airway measurements, since this approach is likely to provide valuable information to better understand the morphology and function of lungs affected by diseases like emphysema and inflammation.

The rest of the section is structured as follows. In subsection 2.6.2, we briefly summarize the airway segmentation framework and give details about the new features we have introduced. In subsection 2.6.3 the airway segmentation validation experiments are detailed. Results are presented in subsection 2.6.4. A final discussion (subsection 2.6.5) concludes the section.

2.6.2 METHODS

2.6.2.1 Prefiltering

Micro-CT images contain high levels of noise. This justifies the need of a filtering step before the analysis. To this end, we used a 3-D grayscale closing by reconstruction filter with a spherical structuring element of radius one voxel [56]. The 2-D version of this filter was reported to yield positive results in the previous work by Shi *et al.* [43]. This filter increases the contrast of the darkest regions of the image while preserving the shape of the structures. The radius was selected because it represented a good trade-off between noise removal and contour information preservation.

2.6.2.2 Airway tree segmentation and reconstruction

The adopted framework has been explained in detail in previous works by Schlathoelter *et al.* and Bülow *et al.* [44, 45]. A block diagram summarizing the main execution flow is shown in Figure 2.4. A key concept is the *segment*. A segment is a set of contiguous points that has been accepted as valid by a growing wavefront without bifurcations. The growing wavefront is initialized in one seed point, and grows into neighbouring voxels that fulfill certain voxel acceptance criteria. After every wavefront propagation step, several conditions are checked to prevent leaks. When the wavefront bifurcates, new segments are initialized and added to the rest of pending segments in a *segment queue*. This queue is a *first-in, first-out* (FIFO) list, thus ensuring that branches from upper airways are processed first. When a segment is finished, its correctness is checked in a segment evaluation step.

It was necessary to develop new voxel acceptance, propagation evaluation and segment evaluation criteria, due to the particularities of our segmentation task. Details about these new features are given in this subsection.

2.6.2.2.1 Initialization

The initial wavefront consists of a seed point located inside the trachea detected on the first transversal slice. To find the trachea, the first image slice is automatically thresholded using the method described by Hu *et al.* [33]. This threshold value separates air from tissue and it is iteratively computed for each image. Once this separation is done, two main objects appear in the thresholded binary image: the trachea and the background around the animal body. Other smaller objects can appear, for instance between the animal body and the bed. To select the trachea, the following process is used. In a first step, very small objects -smaller than 100 pixels- are removed. Next, very large objects -larger than 5000 pixels- are discarded. This step removes the background. The last step consists in computing the center of mass of the remaining objects. The most centered object is considered to be the trachea. The center of mass of the trachea in the first slice region is set as initial seed for the wavefront propagation. The Insight Toolkit [41] was used to implement the initialization and also the rest of the algorithm.

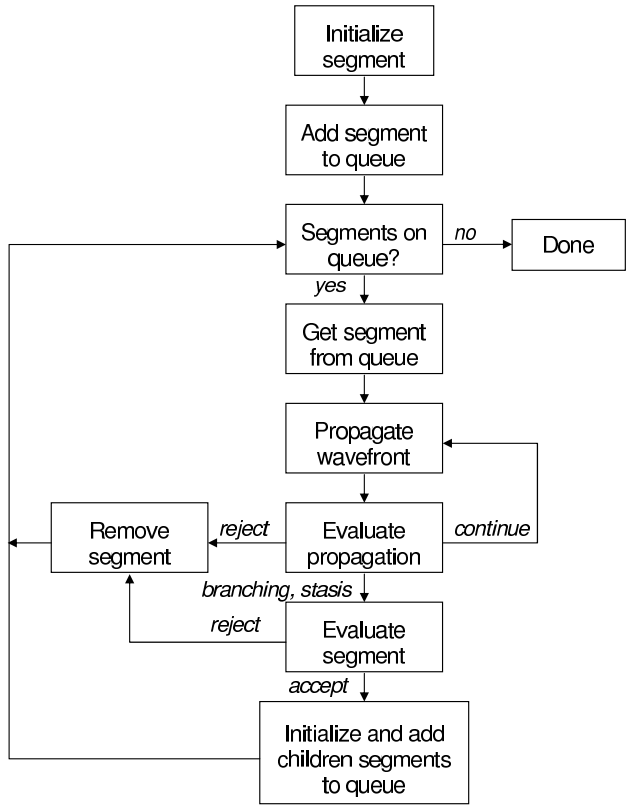


Figure 2.4: *Block diagram of general tree segmentation and reconstruction framework.*

2.6.2.2.2 Wavefront propagation

To propagate the wavefront, we use a 3-D Fast Marching level set algorithm. In particular, our implementation is based on the algorithm described by Deschamps *et al.* [57] and we refer to their work for details. An important aspect is that, since the intensity threshold is variable -as it will be detailed in subsection 2.6.2.2.3- it is not guaranteed that the potential will be monotonically increasing as the wavefront propagates. Since the Fast Marching algorithm assumes this, the wavefront stops as soon as a

negative growth of the potential is detected. An important advantage of a Fast Marching front compared to region growing is the dome-shape of the wavefront, which is very helpful for correctly detecting bifurcations.

Bifurcation detection is a critical aspect of wavefront propagation. To detect a bifurcation, the connectivity of the wavefront is checked after every propagation step. If the wavefront is divided in two or more parts, a bifurcation exists. Sensitivity to detect a bifurcation depends on the connectivity size parameter. The connectivity size defines how many voxels around a given voxels are considered as neighbors. A neighbor connectivity of radius one means that only the 26 voxels around a given voxel are considered neighbors. If this size is increased to two, all voxels in the $5 \times 5 \times 5$ cube around a given voxel will be considered neighbors. Therefore, a low connectivity is very sensitive to bifurcations and will detect them immediately, although it is also sensitive to noise and will probably detect non-existent bifurcations caused by noise. To combine the advantages of high and low connectivities, we have implemented a variable connectivity scheme. Namely, a high connectivity size is used by default -cubical neighborhood of radius three- and when rapid growth of the wavefront in succesive steps is detected, the parameter is reduced to radius one, in order to detect the bifurcation more rapidly. The connectivity returns to the default value of three once the segment stops growing.

2.6.2.2.3 Voxel acceptance criteria

The viscosity function of the Fast Marching algorithm determines whether a voxel is to be added to the current segment or not. If a certain voxel fulfills the acceptance criteria, its viscosity value is set to one. Otherwise it is set to infinity. There are two criteria that a voxel must fulfill before being accepted: the similar intensity condition and the low gradient condition. The similar intensity condition is given by a variable upper threshold, which accounts for the large variability in image intensities observable in micro-CT images. In other words, the low SNR and the strong artifacts (beam hardening, circular artifacts) cause the airway branches to have different intensities depending on their location. Therefore, it becomes necessary to have a threshold that varies within the same image. We define the variable threshold considering the intensity distribution of the current segment as well as the intensities of its ancestor segments:

$$T_i = \mu_l + \alpha \cdot max(\sigma_p, \sigma_{gp}), \tag{2.6}$$

where μ_l is the mean intensity of the current segment (l stands for *local*), α is a tunable parameter, and σ_p and σ_{gp} are the standard deviations of the intensity values of the voxels in the parent and grandparent segments of the current one, respectively. The largest of both is taken, to avoid very small σ-s in case of a short or very homogeneous segment, which would lead to a too low threshold T_i. Intensities are averaged in a $3 \times 3 \times 3$ neighborhood around the voxel of interest, to minimize the effect of noise. In our experience with different imaging protocols, α values between 1.2 and 2 work generally well. If the image is very noisy α must be rather small (near 1.2), while less noisy images require an α value closer to 2.

For the low gradient condition, the three-dimensional Sobel gradient is computed and voxels are rejected if their Sobel gradient value exceeds a given threshold T_s. This way segmentation leakage from the airways and into the lungs is avoided.

2.6.2.2.4 Propagation evaluation

There are two criteria to accept or reject a propagation step. The first one controls the current size of the wavefront, and prevents it from growing in excess. This is implemented by not allowing wavefronts to be larger than the wavefronts in their parent segments. To account for noise and anatomical irregularities, the propagation stops if

$$r > \gamma \cdot \frac{min(r_{anc}) + mean(r_{anc_{min-r}})}{2}, \tag{2.7}$$

where r is the current wavefront radius, γ is a tunable parameter that allows certain growth, $min(r_{anc})$ is the minimum radius among all ancestors of the current segment and $mean(r_{anc_{min-r}})$ is the mean value of wavefront radius of the ancestor containing the minimum radius. The latter is averaged with the smallest radius because the smallest radius alone resulted in a too noisy measure, which limited propagation in some cases. The value of γ indicates the tolerance to the growth of the airway diameter as the wavefront propagates inside the tree. A large value allows segments to be wider than their ancestors, but can also lead to accept leaks as correct airways. Given that mouse airway segments can be considerably wider than their ancestors - especially the right mainstem bronchus can be wider than the trachea-, values between 1.5 and 2 should be considered.

The second propagation evaluation rule refers to the number of neighbor segments. This number is limited to two, because the number of neighbors is computed before the current segment has stopped its propagation and its children segments have been initialized. That is, only a parent and a *brother* segment can be neighbors. A number greater than two generally indicates that several small segments are growing next to each other, which is a common indicator of leakage.

2.6.2.2.5 Segment evaluation

Three different parameters are considered to accept or reject a fully grown segment. A segment is considered to be fully grown when a bifurcation of the wavefront is detected or it can not grow further.

To evaluate if the wavefront has propagated uniformly, the growth rate (GR) is measured, which is given by:

$$\mathrm{GR} = \frac{1}{N} \sum_{i=1}^{N} \frac{|W_i|}{|W_{i-1}|}, \tag{2.8}$$

where N is the number of propagation steps in the segment and $|W_i|$ is the number of voxels of the wavefront at propagation step i. Thus, a GR larger than one indicates that the segment has grown during propagation. A threshold T_{GR} slightly larger than one allows a slight growth of the segment.

Then the discrete compactness (C) is computed, as defined by Bribiesca *et al.* [58] for a solid volume of n voxels in 3-D:

$$\mathrm{C} = \frac{n - A/6}{n - (\sqrt[3]{n})^2}, \tag{2.9}$$

where A is the area of the enclosing surface of the segment, and corresponds to the sum of the areas of the external plane polygons of the voxels which form the visible faces of the solid. Correct airways tend to be more compact than leakages. A threshold T_C is defined to separate correct and incorrect segments.

Finally, the difference between the sizes of the last and the first wavefront is also computed and compared to a threshold T_W, because a large difference is a typical sign of leakage.

These three parameters are combined through voting, as detailed in Algorithm 1. Two of the three parameters must be above the thresholds in order to reject a segment.

Algorithm 1 Segment evaluation

```
votes ⇐ 0
if GR > T_GR then
  votes++
end if
if C < T_C then
  votes++
end if
if |W_last| − |W_first| > T_W then
  votes++
end if
if votes > 1 then
  reject segment
else
  accept segment
end if
```

2.6.3 EXPERIMENTS

2.6.3.1 Image dataset

We tested our algorithm on three different groups of 10 A/J male mice each (Harlan UK Limited, Oxon, UK), scanned at 12 weeks of age and with a mean weight of 21.1 gr (variance 1.45). The three groups received a different treatment 4 weeks before the scan. The groups were control, emphysema and chronic inflammation (see section 2.1).

Animals were imaged following the protocol detailed in section 2.2.

2.6.3.2 Airway segmentation evaluation measures

Automatic airway segmentations were compared to reference manual segmentations. Only every fifth transversal slice was considered for evaluation, due to the long time required to manually segmentat the images. The Amira software (Visage Imaging, Fürth, Germany) was used to perform

the manual segmentations. To this end, a seed was placed inside every visible airway and a 2-D region growing algorithm was applied to segment the airway. The upper threshold for region growing was modified to adjust to local airway characteristics in each case. Whenever this approach failed (mainly due to leakage into the parenchyma) the airways were delineated manually. The entire manual delineation process took approximately 90 minutes per scan. To assess inter-observer variability, a second expert segmented 5 scans from each of the three groups.

To assess the accuracy of our segmentation method, we computed the True Positive Volume Fraction (TPVF) and the False Positive Volume Fraction (FPVF) as defined in [43]:

$$\text{TPVF} = \frac{|S_A \cap S_M|}{|S_M|},\ \text{FPVF} = \frac{|S_A - S_M|}{|S_M|}, \tag{2.10}$$

where S_A is the automatic segmentation, S_M is the manual segmentation, $|S_A \cap S_M|$ represents the cardinality of the intersection between the two segmentations and $|S_A - S_M|$ represents the number of pixels that were marked as airway by the automatic algorithm but not by the manual segmentation.

2.6.3.3 Test parameter sets

Our airway segmentation algorithm has a number of parameters that must be set. For this purpose, we used two extra images from each group, which were excluded from the validation study. The tunable parameters can be divided in two different groups: wavefront propagation and segment evaluation. The wavefront propagation parameters are α (equation (2.6)), T_s and γ (equation (2.7)). The parameter α, (see subsection 2.6.2.2.2) acts on the threshold of the propagation: a high α value leads to a high threshold. T_s is the threshold for the Sobel gradient and, similarly, the higher the threshold the less restrictive the propagation. The last parameter related to propagation is γ (equation (2.7)). A smaller value implies more restrictions to radius growth in subsequent wavefronts. Three thresholds (T_{GR}, T_C and T_W) comprise the segment evaluation parameters. A more lenient set generally leads to more accepted segments, with the risk of more invalid segments being accepted.

We used a simple methodology to set the tunable parameters. The wavefront propagation parameters were selected first, by visually observing

Table 2.3: *Two test parameter sets, with different values for the tunable parameters of the airway segmentation and reconstruction algorithm. Values in Set 1 were selected using 2 images of the control group and 2 images of the silica group. 2 images of the elastase group were used for Set 2. The symbol $|S|$ indicates the number of voxels of the evaluated segment.*

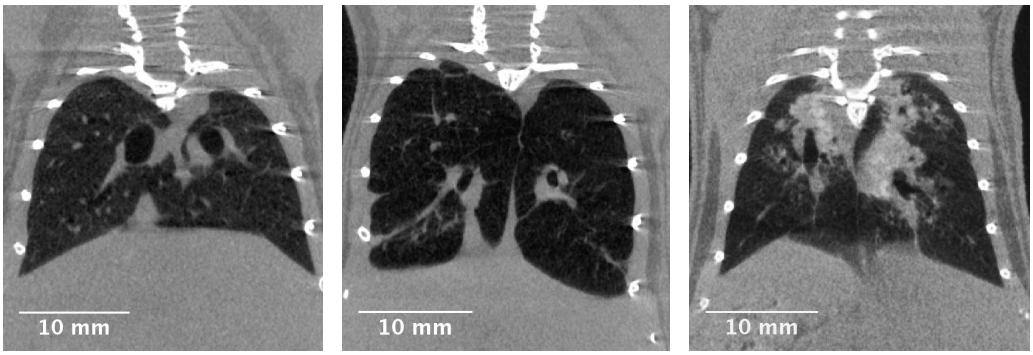
	Set 1	**Set 2**
$\boldsymbol{\alpha}$	1.4	1.3
$\boldsymbol{T_s}$	*none*	4250
$\boldsymbol{\gamma}$	2.5	2.5
$\boldsymbol{T_{GR}}$	1.2	1.1
$\boldsymbol{T_C}$	0.7	$5 \cdot 10^{-5} \cdot \lvert S \rvert$
$\boldsymbol{T_{W_d}}$	*none*	$3.5 \cdot 10^{-2} \cdot \lvert S \rvert$

the results in two images of each group. These parameters were set so that all the main bronchi and most of the small bronchi were segmented, without excessive leakage into the parenchyma. To select T_{GR}, T_C and T_W, the growth rate, compactness and difference between last and first wavefront parameters of a few correct and incorrect segments were observed, and thresholds that best separated those two groups were chosen.

Due to the very different image characteristics of the three groups, two different parameter sets resulted from the tuning process. The first set, aimed at control and silica-treated mice, has more lenient rules for both propagation and segment evaluation than the second set. The reason is that airway walls tend to be worse defined in elastase-treated animals and this results in a much higher leakage probability. Precise values for parameters are given in Table 2.3.

2.6.3.4 Robustness to image noise

To test how the algorithm behaves on images of different characteristics, we scanned three A/J mice with three different X-ray doses. Two mice were 11 months old and the third one was 16 months old. The imaging protocol was in all cases virtually identical to the one described in subsection 3.2. The only difference was the tube voltage, which was set to 80 kV_p for the



(a) Control (b) Emphysema (c) Inflammation

Figure 2.5: *Sample coronal micro-CT slices of control, emphysematous and inflammed mice lungs. Differences in lung texture and shape are clearly visible.*

normal dose images, to 65 kV_p for the low dose images and to 40 kV_p for the very low dose images. The standard deviation of the water volume in a water phantom used to estimate the noise was 78.32 HU, 86.64 HU and 94.42 HU respectively.

2.6.4 RESULTS

Micro-CT sample images from the three groups are shown in Figure 2.5. Compared to the control group, emphysematous mice show considerably darker lungs, due to the loss of parenchymal tissue [22]. This, together with a stronger tendency to movement-related artifacts, leads to poor intensity contrast between airways and parenchyma. Airway diameter is also generally smaller and lungs appear more inflated. In contrast, the silica group shows large inflammed central areas which display with high intensity in the image. Some lung lobes and airways leading to them can appear collapsed.

Table 2.4 shows the segmentation accuracy results for the three different groups using the two different parameter sets presented in Table 2.3. As expected, the first parameter set yields good results in the control and

Table 2.4: *Results of the automatic segmentation on the three different image datasets. Mean TPVF and FPVF are indicated, together with the standard deviation in parentheses. The parameter set for each group is marked in bold.*

Dataset	Parameter Set	TPVF (%)	FPVF (%)
Control	1	**89.44 (3.26)**	**7.97 (5.53)**
	2	73.14 (22.88)	2.82 (2.18)
Elastase	1	53.28 (11.41)	53.76 (31.69)
	2	**81.20 (3.58)**	**12.99 (8.17)**
Silica	1	**87.86 (3.84)**	**6.52 (3.01)**
	2	72.04 (21.97)	2.98 (2.10)

Table 2.5: *Inter-observer variability. Mean TPVF and FPVF of the second observer's segmentations are indicated, together with the standard deviation in parentheses.*

Dataset	TPVF (%)	FPVF (%)
Control	96.06 (1.41)	8.81 (1.07)
Elastase	95.59 (0.86)	14.3 (1.82)
Silica	90.16 (4.36)	6.76 (3.51)

silica groups, while a more restrictive set is required for the elastase-treated emphysema group.

The result of manual segmentations depends up to a certain point on the subjective criteria used by the manual segmenter. These effects can be partially quantified by looking at inter-observer variability. In this case, it was estimated by comparing the manual segmentations of two independent observers on 5 randomly chosen images of each group. The first observer's segmentations were taken as reference to compute the TPVF and FPVF of the second observer. Results are shown in Table 2.5, where relatively high values of TPVF and FPVF can be observed. Visual comparison of the two observers' segmentations revealed that the second observer drew larger contours around the airway lumens, which explains the high FPVF values.

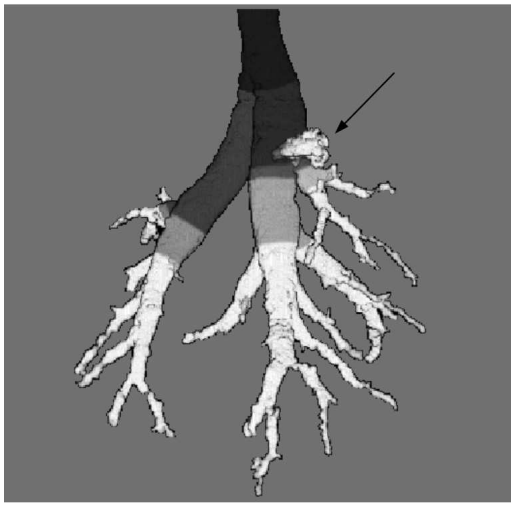


Figure 2.6: *3-D rendering of a segmentation of a control mouse airway. The black arrow points at an undetected leakage in the upper right side.*

Most oversegmentation errors of our automatic method occured due to late detection of leaks, as shown in the 3-D rendering of Figure 2.6. A more restrictive set of segment acceptance rules would reduce these kind of leaks, but it would also affect the TPVF, as already seen in Table 2.4.

Computation time varies from image to image depending on particular image characteristics (airway wall thickness, movement artifact, etc.) and lies between one and three minutes, with non-optimized C++ code on an Intel Xeon 3.20 GHz processor.

Our algorithm not only segments the airway tree, but also provides a reconstruction of the bronchial tree, by means of a list of segments hierarchically related. The length, volume and mean wavefront size of each segment are also given. The mean wavefront size can be seen as a surrogate of airway area, since the front propagates in normal direction to the airway lumen. Figure 2.7 shows three renderings of images of the three groups, with main segments displayed in different gray-levels. It must be noted that segments do not necessarily correspond to airway branches, because they only have algorithmic meaning, as explained in subsection 2.6.2.2.

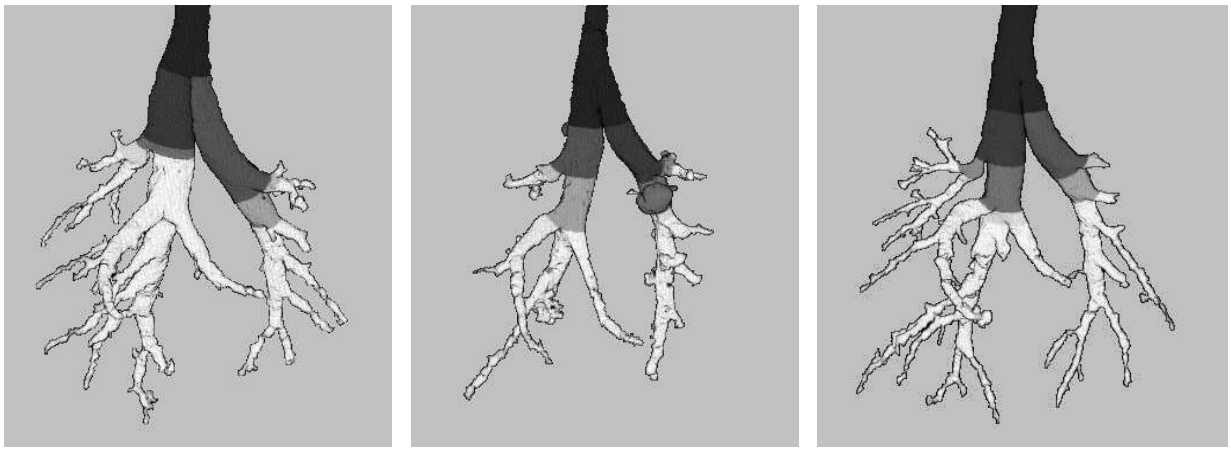
Table 2.6: *Mean radius of the right and left mainstem bronchus (MB) in μm. (*) and (**) indicate $p < 0.05$ and $p < 0.01$ when compared to the control group using the Wilcoxon or Mann-Whitney U test.*

	Right MB	**Left MB**
Control	760.38	529.92
Elastase	613.64**	475.18*
Silica	636.64**	449.88*

To gain a better understanding of the effect of the diseases on the lung physiology, we first compared the radii of the two mainstem bronchi of the three different mice groups. To have a more accurate measurement and to have segments that correspond to real bronchi, the algorithm was run again on the already segmented trees. Table 2.6 shows the mean radii of the three different groups. It can be seen that, as expected, the right mainstem bronchus is larger than the left, and that the elastase and silica groups have smaller radii than the control group. The difference was statistically significant according to the Wilcoxon or Mann-Whitney U test [59].

Regarding the constant phase model parameters, Raw, I and G did not show any difference between the control and the diseased mice. H was slightly lower in silica (22.26$\pm$2.94 cmH_2O/ml) compared to control mice (24.84$\pm$3.31 cmH_2O/ml), and significantly lower in elastase-treated mice (12.59$\pm$1.42 cmH_2O/ml). If mean mainstem bronchi radii and H are plotted jointly, clear differences between the 3 groups can be seen (see Figure 2.8). Silicotic mice have smaller bronchi radii than controls, and virtually the same tissue elastance. On the other hand, emphysematous mice also show smaller bronchi radii, but a significantly lower tissue elastance. This suggests that the causes of mainstem bronchi shrinking are different. In the inflammation case, the inflammation around the main airways is the most probable cause of radius reduction. In contrast, reduced airway diameter in emphysematous mice is likely to be due to the increased compliance of their lung tissue, revealed by the low H value.

Table 2.7 shows the mean TPVF and FPVF values for the images acquired at different X-ray doses (see subsection 2.6.3.4). It can be seen that, as expected, decreasing dose leads to worse segmentations. However, the



(a) Control (b) Emphysema (c) Inflammation

Figure 2.7: *Sample renderings of segmented airway trees of the three groups. In emphysema, less small branches are segmented due to the lack of contrast between airway and parenchyma in those areas, resulting in a less detailed segmentation compared to the other groups. A similar effect took place in manual segmentations. To allow a clear distinction between the main segments in the renderings, only main segments are shown in different gray levels, while smaller branches are all displayed in white.*

Table 2.7: *Effect of decreasing dose on the automatic segmentation. Mean TPVF and FPVF of the automatic segmentations with different levels of X-ray dose, with the standard deviation in parentheses.*

	Dose 69.1 cGy	**Dose 48.3 cGy**	**Dose 30.2 cGy**
TPVF (%)	86.59 (6.02)	82.39 (11.01)	84.15 (2.88)
FPVF (%)	9.19 (2.84)	15.29 (8.61)	16.94 (7.88)

decrease in accuracy is not dramatic and, depending on the application, the algorithm might be of use even with very low dose images.

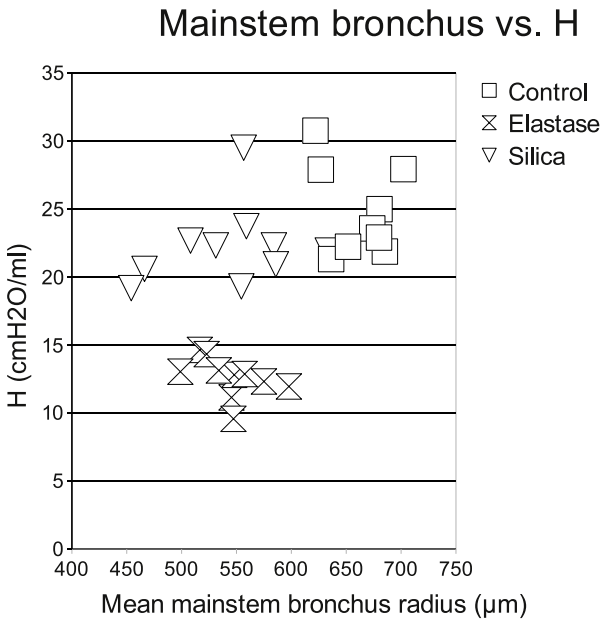


Figure 2.8: *Mean radius of left and right mainstem bronchi plotted along with tissue elastance (H).*

2.6.5 DISCUSSION AND CONCLUSION

Accurately segmenting mice airways in *in vivo* micro-CT images is a challenging task. On the one hand, the imaging technology imposes certain image quality constraints, as mentioned above (see subsection 2.6.1). On the other hand, the position of lung structures can vary during the scanning process, which takes several minutes. Despite artificial ventilation, iso-pressure apneae and deep anesthesia, the way in which lungs respond to repeated mechanical ventilation does not remain constant during the acquisition process. We observed that this was especially significant in emphysematous specimens, which lead to a generally lower image quality due to movement artifacts.

To overcome the difficulty imposed by the low SNR and movement artifacts, a robust segmentation algorithm is required. We built on a previously published framework for tubular structure segmentation and reconstruction. Previous works have centered on the framework itself and its applications [44–46]. The framework provides with generic concepts for the segmentation and reconstruction of tree-like structures, such as the propagating wavefront and the *segment* definition. Here, we have used these basic concepts of the framework and we have devised new features in its different components, such as the variable threshold and the voting segment evaluation, to adapt to the particularities of thoracic micro-CT images. These extra components have certain tunable parameters that must be fixed. In our experience, it is useful to divide these parameters in wavefront propagation and segment evaluation parameters to better understand their global effect. To fine tune the parameters, a few images of the dataset should be used as a guide, until an acceptable trade-off between sensitivity and specificity is achieved.

We tested our algorithm on normal mice and on two mouse models of lung disease. The three groups showed very different image characteristics. Particularly, the elastase group displayed very low voxel intensity in the lungs, which made both manual and automatic segmentations more difficult. Moreover, movement artifacts were stronger in this group. Therefore, we used a different parameter set for this image dataset. The main characteristic of this parameter set was that it was more restrictive in both wavefront propagation and segment acceptance.

The validation was performed by comparing automatic segmentations with manual ones. In an attempt to quantify the existing human observer errors, a second expert segmented 50% of the evaluated images. Results show that a considerable variability exists between the two observers, suggesting the difficulty of precisely delineating the airways.

The automated image analysis tool that we have developed may be of use for a variety of studies related to lung physiology and pathology. A number of potential applications of murine airway segmentation have been outlined in the introductory section. This can be used to quantify both emphysema and inflammation, in the context of chronic-inflammation related lung carcinogenesis (see chapter 3 in this thesis). An interesting topic for future work would be the measurement of airways walls, since they are also affected in COPD.

We showed that differences in airway diameter among different groups can be detected with this segmentation method. However, the question of how early in a disease process changes can be quantified has not been addressed. A different dataset showing disease progression would be required. The factors that will affect the technique's sensitivity include image resolution, signal to noise ratio and the presence of motion artifacts in the images.

We observed that, despite the segment evaluation rules, some leaks remained undetected. Depending on the application, an interactive tool for incorrect segment removal might be of use. The relatively large standard deviations in the TPVF and FPVF values reveal that the algorithm performs better in some images than in others. This could be improved by modifying the tunable parameters separately for each image, which should be done automatically if a fully automatic algorithm is desired. Variables such as the number of segmented branches or the intensity values within and outside the already segmented airways could be used for an iterative tuning process. Another limitation of the algorithm that it is based on a propagating wavefront. Any interruption in the growth of an airway branch due to noise or disease-related airway obstruction stops the propagation. Additional features would be required to detect and avoid these interruptions.

Segment hierarchical information is given together with the segmentation. Thus, the topology of the tree can be easily traced, because the parent and children of each branch are identified. We believe that this is a

major advantage of the technique when compared to other algorithms that simply segment the tree, for two reasons. First, topological information can be used to help the segmentation, for instance by comparing the current segment with its ancestors' radius or by checking the number of neighbor segments. Second, it eliminates the need for an extra skeletonization step. Moreover, our algorithm provides with measurements of the length, volume and mean wavefront size of each airway (which is a surrogate of the area). We showed an application of this by comparing the mean radius of the left and right mainstem bronchi among the control, emphysema and inflammation groups. Combining data from the constant phase model parameters and the airway measurements, we were able to detect different patterns in the three groups studied. These differences could not be clearly seen when only input impedance parameters or mainstem bronchi measurements were analyzed.

In conclusion, we have presented a fast and robust algorithm for murine airway segmentation and reconstruction. By adjusting the algorithm parameters to the particular characteristics of different models of lung disease, the airways were segmented with high sensitivity and specificity values. We have shown that measurements derived from these segmentations can be combined with pulmonary input impedance measurements to gain more insight into the changes in normal lung physiology caused by different diseases.

2.7 MULTI-ATLAS SEGMENTATION

2.7.1 PRINCIPLES AND PREVIOUS WORK

The principles of atlas-based segmentation have been successfully applied to a wide variety of image modalities and segmentation tasks [34, 60–63]. This approach has a major advantage when compared to other segmentation algorithms, such as level sets [64] or watersheds [65]. Namely, it allows introducing *a-priori* knowledge about the shape and the distribution of the segmented structures in a simple way, by using a pre-segmented image as a reference that guides the segmentation.

In principle, a single atlas image can be used for segmenting new images. Indeed, in section 2.5.3 of this thesis, we used a single atlas in one of the steps of a multi-step segmentation method for lungs with severe inflammation. However, it has been recently shown that using multiple atlas images can yield better results [38, 66–68]. Information from several reference images can be combined into an average atlas [69, 70] or, if probability values for each particular location are included, into a so-called *probabilistic atlas* [34, 60, 62, 67]. In doing so, these atlases try to comprise all the variability of a given population. However, it has been recently suggested that, to gain full advantage of having multiple atlas images at hand, they must be registered to the target image independently and the resulting segmentations combined [38].

An analogy can be made between the combination of segmentations derived independently from multiple atlas images and the combination of multiple independent classifiers in a generic classification problem [71]. In this analogy, each transformed atlas image can be regarded as a classifier, which assigns a label value to each voxel of the target image. The training process can be assimilated to the registration between the atlas image and the target image. It has been widely proven in the pattern recognition field that combining multiple classifiers can yield more robust and accurate results than using single classifiers [72, 73], this fact being the main motivation for multi-atlas approaches.

The most widely used combination strategy in the literature is majority voting, also named majority rule, decision fusion or label voting. This approach weights each candidate segmentation equally and assigns to each voxel the label that most segmentations agree on [38, 68]. Another popular

approach is called Simultaneous Truth And Performance Level Estimation (STAPLE), which uses an expectation-maximization (EM) approach to reach the best possible final segmentation [74, 75]. STAPLE estimates the performance of each classifier iteratively and weights it accordingly. The two different methods presented in [75] are extensions of the one in [74] for images with multiple segmented structures. Shape-based averaging represents another way of combining segmentations [76] which is based on Euclidean distance maps computed for all structures in each candidate segmentation. The method was shown to keep structure regularity and contiguity better than majority voting. Another possibility is atlas selection: instead of combining segmentations, methods can be devised to select atlases *a-priori* -before registration- or *a-posteriori*- after registration [38]. In [77], a number of atlases is selected for combination based on mutual information [78]. In [79], atlas selection is done on a structure basis for the segmentation of brain MR images. The atlas image with highest local mutual information in each structure is selected.

Recent works have contributed significantly to the field of atlas-based medical image segmentation, but we believe that some aspects still remain unexplored. Namely, we think that in order to achieve the highest possible overall segmentation accuracy, a better understanding of the different combination strategies is required. In this paper, we propose a scheme that divides the combination strategies in two major groups: Global strategies, which estimate segmentation accuracy with a single value for the whole image, and local strategies, which evaluate the segmentations at each voxel and perform weighted voting accordingly.

We test several global and local combination methods on digital phantoms and on publicly available MR images of human brains, to conclude that no combination algorithm is better than the others consistently for all images and regions within the images. We study in which kind of structures local strategies do better than global methods, and conclude that the optimum solution is approached by selecting the combination method according to the gray level contrast between each particular region and its neighbors.

The outline of this section is the following. In the methods subsection, different segmentation combination strategies are examined. First, global weighting approaches are listed. Then, we show their main limitations and introduce the local combination strategies. In the third and fourth

subsections, results from experiments on phantom images and on brain MR data with multiple segmented structures are shown, comparing different combination strategies. This section on multi-atlas image segmentation ends with a discussion and a final conclusion.

2.7.2 COMBINATION METHODS

2.7.2.1 Notation

For simplicity and uniformity, we adopt the basic notation used in [75]. Consider a segmentation of an image with potentially L different classes that belong to a label set $\Lambda = \{1, 2, ..., L\}$. A 3D atlas image A_k is a mapping from coordinates to labels $A_k : \mathbb{R}^3 \rightarrow \Lambda$. An atlas-based classifier is defined by a set of atlas images, $A_k, k = 1, ..., K$ and coordinate transformations that map coordinates from the target image to the atlas images $F_{T \rightarrow A_k} : \mathbb{R}^3 \rightarrow \mathbb{R}^3, k = 1, ..., K$. Using a given transformation $F_{T \rightarrow A_k}$ to transform the segmentation of the atlas image A_k^s, an estimated segmentation of the target image $\hat{T}_k^s$ is obtained:

$$\hat{T}_k^s = A_k^s \circ F_{T \rightarrow A_k} \tag{2.11}$$

Each $\hat{T}_k^s, k = 1, ..., K$ is a candidate segmentation which must be combined into a final estimated segmentation $\hat{T}^s$. Following [75], we now regard the segmentation task as the classification of N unordered samples. Thus, when candidate k assigns sample x to class i, we write the output of the classifier as

$$e_k(x) = i. \tag{2.12}$$

Note that in this context the spatial location of the voxels is irrelevant. The set of all samples that belong to class i is denoted by C_i. The outputs from L independent classifiers can be combined to generate a single ensemble response, $E(x)$, which is the output of the combination strategy. The aim when building an ensemble classifier is to achieve a higher probability of correctly classifying the voxels of the image than that obtained by using an individual classifier. The combined classifier output $E(x)$ for a sample x should be the class that maximizes the probability, given all classifier decisions $e_1(x)$ through $e_K(x)$ and some arbitrary classifier performance model $\mathbf{P}$ [75]

$$E(x) = \underset{i}{\operatorname{argmax}}\, P(x \in C_i | e_1(x), ..., e_K(x), \mathbf{P}). \tag{2.13}$$

2.7.2.2 Global combination strategies

This group of combination strategies assigns a global weight w_k to each segmentation of the target image $\hat{T}_k^s$. The weight is calculated by a global evaluation of segmentation accuracy, and its value is used to ponderate the whole target image segmentation derived from the atlas image. Note that although the weight is global for each candidate segmentation, decisions on the combination of the segmentations are taken voxel by voxel.

2.7.2.2.1 Majority Voting

This is the most simple combination method. It is generally used when there is no a-priori knowledge about the accuracy of each classifier. It assigns to each voxel the label that most segmentations agree on. Thus, the ensemble response $E_{MV}(x)$ of majority voting can be expressed as

$$E_{MV}(x) = \max[f_1(x), ..., f_L(x)] \tag{2.14}$$

where $f_i(x) = \sum_{k=1}^{K} w_{k,i}(x)$ for $i = 1, ..., L$ and

$$w_{k,i}(x) = \begin{cases} 1, & \text{if } i = e_k(x) \\ 0, & \text{otherwise.} \end{cases} \tag{2.15}$$

This technique labels a voxel correctly if a majority among the K segmentations agree on the correct value.

2.7.2.2.2 Weighted Voting

The expression for majority voting, given by (2.14), can be generalized to assign arbitrary weights to each candidate segmentation. Weighted voting is commonly used in pattern recognition, as it allows the maximization of classification accuracy when the accuracy of individual the classifiers is known [73]. However, these results can not be directly translated into the multi-atlas segmentation field, because segmentation accuracies can not be known *a-priori*. Thus, a different approach must be adopted. The ensemble response $E_{WV}(x)$ would respond to equation (2.14), but weights would have to be set differently. One option is to set them as:

$$w_{k,i}(x) = \begin{cases} m^p, & \text{if } i = e_k(x) \\ 0, & \text{otherwise} \end{cases} \tag{2.16}$$

where m is a similarity measure between the registered atlas image and the target image and p is the associated gain. If the similarity measure is not sensitive enough, the differences between the weights might not be relevant. In those cases, it might be necessary to increase the value of p.

The selection of a similarity measure m is not straightforward. For instance, in [80], we proposed using the mutual information [78] between the target image and the registered atlas image as weight, i.e., $m^p = [I(A_i^s \circ F_{T \to A_i}, T)]^p$. The rationale behind that choice was that high mutual information values normally imply a better registration, and using this information for the weighting should yield better results. It must be noted that, generally, no strong correlation between mutual information (or any other image similarity measure) and registration accuracy exists [81]. Nevertheless, our tests on MR images of mouse brains resulted in higher segmented region overlap and smaller distances between segmented surfaces than majority voting [80]. It must also be noted that, although statistically significant, differences were small.

More generally, in their work [82], Roche et al. study the assumptions that are implicitly made when different similarity measures are used for image registration. They show that the mean square distance can be used when no relative intensity difference between images is expected, while variants of the correlation coefficient should be preferred in the case of an affine relationship between intensities. Finally, if only a statistical relationship can be assumed, mutual information is the best option. These three similarity measures are tested and compared in our experiments, since they are likely to be good predictors of accuracy in registration.

2.7.2.3 Limitation of global combination strategies

Global fusion strategies have shown their potential and perform generally better than single atlas approaches. Nevertheless, they have a major limiting drawback. Namely, weights or performance estimates are the same for all the voxels of the segmentation. We will show how this can negatively affect overall performance using a simple example. Let us have a star-shaped target image and two different atlas images, numbered 1 and

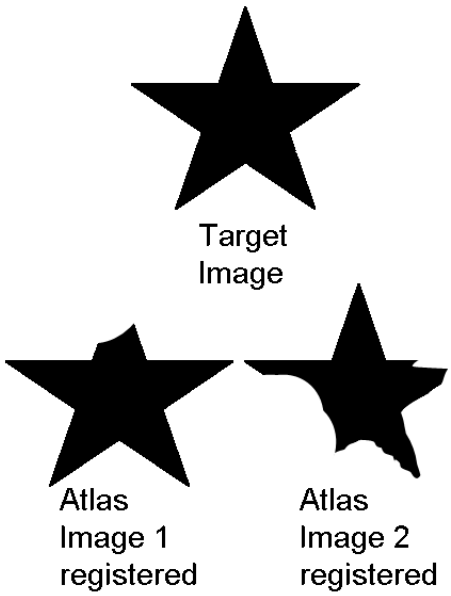


Figure 2.9: *Example that shows the limitations of global candidate segmentation combination strategies. Atlas images* 1 *and* 2 *have been registered to the target image. Image* 1 *is generally better registered, except for the upper arm of the star. However, global strategies can not evaluate registration performance locally. Therefore, they can not take advantage of this fact to obtain a better fused segmentation.*

2, as shown in Fig. 2.9. After registration, the first atlas image fits the target image almost perfectly in all parts, except for one of the arms (upper arm). In contrast, the second atlas image fits the target image correctly in that arm, while registration fails in the rest of the image.

As weights are assigned globally, with these approaches it is impossible to take advantage of the fact that registration between atlas image two and the target image succeeded in a particular point, even if it was inaccurate in the rest of the image. A combination strategy that would account for local registration failures could therefore achieve higher segmentation accuracies.

2.7.2.4 Local combination strategies

2.7.2.4.1 Generalized Local Weighted Voting

To overcome the limitation shown in the previous subsection, a logical solution is to adapt weights locally. Instead of assigning the weight to all the voxels of the segmentation, each voxel might have a different weight value.

In this context, we propose a local segmentation fusion method that we denominate *Generalized Local Weighted Voting*. The general expression for the weights is given by

$$w_{k,i}(x) = \begin{cases} [m(s,r)]^p, & \text{if } i = e_k(x) \\ 0, & \text{otherwise} \end{cases} \tag{2.17}$$

where $m(s,r)$ can be any local similarity measure that assigns a distance between two regions of the registered atlas image and the target image. The distance is measured over a neighborhood region of shape s, radius r with a gain p. The neighborhood shape can also be arbitrary, for example cubical or spherical.

As pointed out in subsection 2.7.2.2.2, it must be noted that similarity measures do not correlate perfectly with registration accuracy. For instance, high information content areas, such as edges or high contrast areas, are likely to result in regions with high mutual information, regardless of the registration accuracy.

Assigning a different weight to each voxel significantly increases the number of degrees of freedom of the problem. This arises the need for a regularization method to ensure a smooth variation of the weights, ignoring the image noise. When local weights are computed, an *intrinsic regularization* occurs thanks to the voxel-by-voxel shifting of the neighborhood region. For instance, assuming a $2D$ image and a square region of 10×10 pixels, two neighboring pixels share 180 out of the 200 pixels (100 from the target image, 100 from the registered atlas image) used to compute the local weight. Sharing this large percentage of voxels between neighboring pixels causes a smooth variation of the weights along the image, regardless of the particular similarity measure employed in each case. Increasing the radius of the local neighborhood further ensures the smoothness of the

weights, but such computed weights are less local and this generally results in worse perfomance. More detailed effects of increasing the neighborhood radius r are explored in sections 2.7.3 and 2.7.4.

On the other hand, the computational burden can be reduced by limiting the number of voxels in which segmentations must be evaluated. If, for a given voxel, all candidate segmentations agree on a certain label, this label is directly assigned without computing the weights. This approach was adopted in all experiments shown in this paper.

2.7.2.4.2 STAPLE

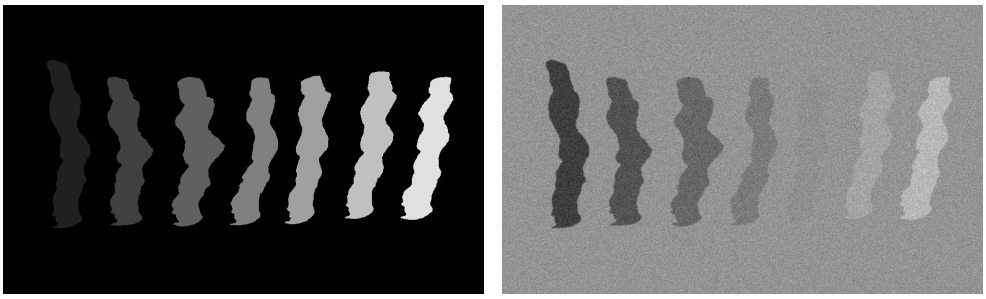
This fusion strategy weights each voxel according to the estimated performance of the disputing labels at that point, using an iterative expectation-maximization algorithm. Three different implementations of the method exist. The binary classification method was first proposed by Warfield *et al.* [74], and Rohlfing *et al.* presented two multilabel generalizations of the method [75]. In the latter work, STAPLE was shown to outperform majority voting on confocal microscopy images of bee brains.

2.7.3 EXPERIMENTS ON DIGITAL PHANTOMS

2.7.3.1 Data

To study the performance of different atlas combination strategies, we built an artificial atlas of 18 images. A manually delineated elongated shape was copied 7 times on a 2D image of size 1000×600. On this first image, random elastic deformations were applied using an online available ImageJ plugin [1]. Each transformation was calculated by deforming a B-spline grid on the image, with a random noise factor. The grid spacing was one pixel. We generated 18 different mask images, each of them containing 7 regions. From each of these mask images, an atlas image was generated in three steps. First, image intensities were modified, so that there were regions that showed different contrast levels with the background. Second, a very slight Gaussian smoothing with $\sigma = 1$ was applied. Third, image intensities were modified to follow a Rician distribution, as MRI magnitude images do [83]. One of the phantom atlas images and the corresponding label mask can be seen in Fig. 2.10. The label mask assigns the label 0 to the background,

[1] http://biocomp.cnb.uam.es/~iarganda/SplineDeformationGenerator/



(a) Phantom mask (b) Phantom

Figure 2.10: *Sample label mask and corresponding phantom with modified intensities and added noise.*

and labels with increasing ordinal values to the regions from left to right, i.e., the leftmost region has label 1 and the rightmost label 7. In the rest of the paper, we assimilate the label to the name of the region. One of the 18 images was used to set the parameters of the different segmentation algorithms, while the other 17 were used for algorithm evaluation.

2.7.3.2 Registration

The registrations required for the multi-atlas segmentation were done in two steps. First, an affine transform was used to account for translations, rotations, anisotropic scaling and shearing. Then, an elastic B-spline registration was performed, using an isotropic grid spacing of 8.0 pixels [84]. Mutual information was maximized in both cases [78]. The reason for using a much less dense grid than that employed for the deformation generation was to obtain non-ideal registrations, as occurs in real complex 3D images. Elastix, an open-source software for elastic image registration was used for all registrations [85]. To accelerate computation time, registration was done in a multi-resolution fashion, with three resolution levels. A stochastic gradient descent optimizer was used, because it provides a good trade-off between precision and computation time [39].

2.7.3.3 Evaluation measures

The Similarity Index (SI) is the most widely used measure to evaluate the performance of a segmentation algorithm. Its definition is given in equation . As pointed out before, SI has value 1 when there is a perfect match between labels and 0 when there is no overlap.

Apart from SI, the mean average surface distance (MASD) between the segmented structures was also evaluated [34, 86]. The MASD is the symmetric distance between the surfaces of the respective segmentations:

$$MASD(S_a, S_b) = \frac{1}{2}[d(S_a, S_b) + d(S_b, S_a)], \tag{2.18}$$

where $d(S_a, S_b)$ is the mean distance from the points of segmentation S_a to segmentation S_b.

For both measures, mask images were considered the ground truth.

2.7.3.4 Combination algorithms

The combination algorithms were implemented based on the Insight Toolkit (ITK) [41], an online available toolkit for image processing. The gain (p) and region size (r) parameters were set independently for each combination method, using a randomly chosen single image from the set. For the global voting methods, the p between 1 and 8 that maximized segmentation accuracy for the selected image was chosen. Selection was also considered as a possible value for p, meaning that the image -or pixel, in case of local methods- with the highest weight is selected, without any further voting. For the local methods, both p and r had to be set. The parameter tuning process consisted in setting r to an initial value of 10 and then varying p from 1 to 8, including selection, to obtain the optimal value. Then, r was varied from 5 to 25 with a 5 pixel step, to find the optimum value.

We tested the following combination algorithms:

- *Majority Voting*

- *Global Weighted Voting based on Normalized Cross-Correlation* (GWV-NCC): Normalized cross-correlation (NCC) between two images is defined as:

$$NCC = \frac{Cov(I_1, I_2)}{\sqrt{Var(I_1)} \cdot \sqrt{Var(I_2)}} \tag{2.19}$$

  where $Cov(I_1, I_2)$ is the covariance of the images and $Var(I_1)$ and $Var(I_2)$ are the variances of each of the images. The gain factor p was set to 4.

- *Global Weighted Voting based on Mutual Information* (GWV-MI) with $p = 8$. The power of eight was required to amplify differences between the weights. Mutual information was computed as in [78], using the ITK implementation.

- *Global Weighted Voting based on Mean Square Distance* (GWV-MSD): This is a particular case that does not require any gain factor, because the difference between the images is already a very large value. Therefore, gain was simply set to -1 to account for the inverse relationship between mean square distance and image overlap after registration.

- *STAPLE*: We used an implementation by Rohlfing available online [2]. We did not limit the number of iterations and the termination threshold for the EM iterations was set to 10^{-5}.

- *Local Weighted Voting based on Normalized Cross-Correlation* (LWV-NCC): The similarity measure employed in this local combination method was the same as in GWV-NCC, but computed in a square local neighborhood around each pixel of size $r = 10$. Best results were achieved with $p = 5$.

- *Local Weighted Voting based on Mutual Information* (LWV-MI): For this local fusion method, we particularized the *Generalized Local Weighted Voting* method with normalized mutual information as similarity measure, defined as:

$$NMI = \frac{H(I_1) + H(I_2)}{H(I_1, I_2)} \tag{2.20}$$

[2] http://www.stanford.edu/~rohlfing/software/index.html

where $H(I_1)$, $H(I_2)$ and $H(I_1, I_2)$ are the entropies of images 1, 2, and 1 and 2 jointly, respectively. The entropy of an image can be computed from its histogram $h(x)$ as:

$$H(I_1) = -\sum_{i=1}^{N} h(x_i) \log_2 h(x_i), \tag{2.21}$$

where N is the number of bins and x_i is the centroid of the i-th histogram bin. The similarity measure was computed in a two-dimensional square neighborhood of isotropic size $r = 15$, and p was set to 8.

- *Local Weighted Voting based on Mean Square Distance* (LWV-MSD): In this case, the chosen similarity measure was the mean squared distance computed in a square region around the voxel of interest of size 10 in each dimension. The power p was set to -6 to account for the inverse relationship between distance and similarity between images.

2.7.3.5 Results

Evaluation measures resulting from the use of different combination strategies on the phantom images are summarized in Tables 2.8 and 2.9. The Wilcoxon matched-pairs signed-ranks test [59] was applied on the results on a region basis, to look for the best method for each region. If we look at SI (Table 2.8), LWV-MSD is generally the best combination method, except for region 5. The particularity of this region is its extremely low contrast with the background (see Fig. 2.10b). In terms of MASD (Table 2.9), it is noteworthy that the performance difference between LWV-MSD and the next best method is generally larger in regions with greater contrast (regions 1, 2 and 3 are the ones with largest contrast, as shown in Fig. 2.10b).

It is noteworthy that, in exceptional cases, a method can be significantly better than the other, even though the mean evaluation measure is worse. This is due to the fact that the employed statistical test works with the magnitude of the differences between the pairs of measures involved in the test. This is the case of region 7 in Table 2.9, where a single region in one image degrades the mean MASD of LWV-MSD considerably, making it worse than the mean MASD of GWV-MI. However, the statistical test is

Table 2.8: *Average SI for different phantom regions with different combination strategies. (*) indicates $p < 0.05$ according to the Wilcoxon matched-pairs signed-ranks test, when compared one-to-one to all the other columns. (**) indicates $p < 0.001$. If two columns are marked, it means that the difference between them is not statistically significant, but it is with the rest of the columns.*

Region	Combination Strategy							
	MV	STAPLE	GWV-MI	GWV-NCC	GWV-MSD	LWV-MI	LWV-NCC	LWV-MSD
1	0.97	0.95	0.98	0.97	0.96	0.98	0.97	**0.99 ****
2	0.97	0.96	0.98	0.97	0.97	0.98	0.97	**0.99 ****
3	0.97	0.95	0.97	0.97	0.96	0.97	0.97	**0.99 ****
4	0.92	0.88	0.93	0.92	0.87	0.92	0.92	**0.97 ****
5	0.83	0.79	**0.84 ***	0.83	0.78	0.83	0.83	**0.84 ***
6	0.91	0.85	0.91	0.91	0.85	0.91	0.91	**0.96 ****
7	0.95	0.91	0.95	0.95	0.89	0.95	0.93	**0.98 ****

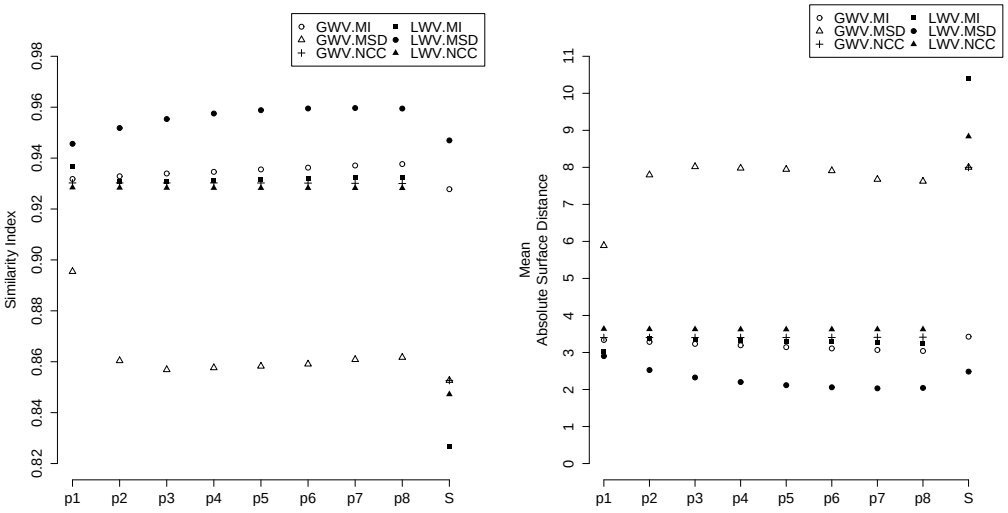
Table 2.9: *Average MASD in pixels for different phantom regions with different combination strategies. (*) indicates $p < 0.05$ according to the Wilcoxon matched-pairs signed-ranks test, when compared one-to-one to all the other columns. (**) indicates $p < 0.001$.*

Region	Combination Strategy							
	MV	STAPLE	GWV-MI	GWV-NCC	GWV-MSD	LWV-MI	LWV-NCC	LWV-MSD
1	1.60	2.37	1.18	1.61	2.06	1.12	1.55	**0.43 ****
2	1.35	2.12	0.99	1.37	1.62	1.02	1.35	**0.54 ****
3	1.58	2.35	1.36	1.58	2.26	1.44	1.58	**0.67 ****
4	3.91	5.95	3.52	3.89	5.85	3.87	3.89	**1.57 ****
5	8.00	11.29	7.59	7.97	11.42	7.97	7.96	**6.66 ****
6	4.61	7.58	4.31	4.65	9.46	4.61	4.63	**2.10 ****
7	2.73	4.44	2.35	2.76	8.53	2.46	4.41	**2.45 ***

able to detect that LWV-MSD performs consistently better (with $p < 0.05$) than the rest of the combination methods, including GWV-MI.

2.7.3.6 Influence of tunable parameter selection

We used a single image to set the tunable parameters, both for global and local combining methods. This arises the question how much the final result can be influenced by a change in the selection of these parameters.



(a) SI for varying gain factor p.

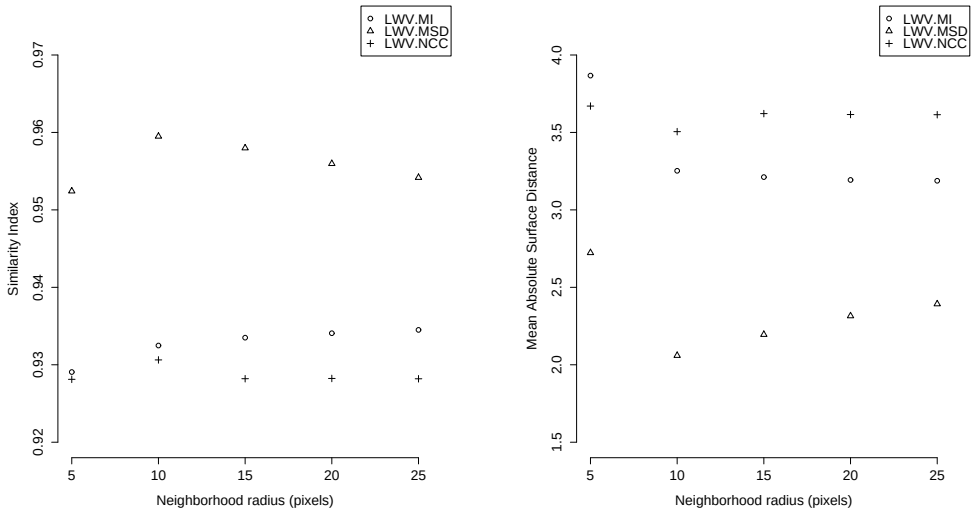
(b) MASD for varying gain factor p.

Figure 2.11: *Plots showing the effect of varying the gain factor p on mean SI and MASD over all regions for different combination strategies on phantom images. p is varied from* 1 *to* 8. *S indicates* selection, *that is, the voxel with the highest weight is selected, without any further voting process. r was set to* 10 *for all LWV methods.*

To evaluate this, we looked at SI and MASD as p varies for the different combination methods, as the well as at the results of varying r. Results are shown in Fig. 2.11 and Fig. 2.12 respectively.

The variation of p shows no clear peaks for any method, except for GWV-MSD. This could be expected because weight values are already very large with this metric, and amplifying them before the voting process is not required. The absence of peaks and the very slow variation between neighboring points on the rest of methods suggests that the selection of the gain factor can lead to slightly better or worse overall results, but it is not critical.

In the case of the neighborhood radius, it can be seen that if the smallest radius, i.e. $r = 5$, is neglected, the accuracies resulting from the different



(a) SI for varying neighborhood region radius r. (b) MASD for varying neighborhood region radius r.

Figure 2.12: *Plots showing the effect of varying the neighborhood region radius r on mean SI and MASD over all regions for different combination strategies on phantom images. For all methods, p was the set to the value employed in the rest of experiments.*

methods do not overlap with each other. That means that, on this dataset, the selection of r does not heavily influence the performance of the local combination methods.

2.7.4 EXPERIMENTS WITH REAL DATA

2.7.4.1 Data

To test the performance of the different combination strategies on real images, we employed 18 T1 MR brain images from the Internet Brain Segmentation Repository (IBSR) [87]. One of the images was used to tune the segmentation parameters, while the other 17 were employed for the evaluation of the algorithms. We studied the segmentation of 18 different structures, 14 of which were paired (left and right side). More structures

Table 2.10: *Gain parameter p for each combination algorithm on the IBSR database.*

	GWV-MI	GWV-NCC	GWV-MSD	LWV-MI	LWV-NCC	LWV-MSD
Gain parameter p	4	6	-1	8	2	-1

were actually segmented on the images, but the segmentation failed almost completely on them due to their small size. All images had been normalized into the Talairach position and processed with biasfield correction routines. Image size is $256 \times 256 \times 128$. Resolutions are variable: In the $x - y$ dimensions they range from 0.837 to 1.0 mm and in the z dimension it is 1.5 mm in all cases.

2.7.4.2 Registration

The registration strategy used with the phantom images was extended to three-dimensional images. A similar approach has been previously used to register brain MR images [68, 88]. The voxel spacing for the B-spline grid was 8.0 voxels.

2.7.4.3 Combination algorithms

We used the same combination algorithms that we compared on the phantoms, modified for 3D images when required. Neighborhood shapes for the local combination methods were in this case cubical. The optimum gain parameters p were generally different, due to the different intensity distribution of the images. One image was used for parameter tuning. Table 2.10 shows the chosen p for each combination method. The process to select the p and r values was the same as the one followed for the phantom images. Gain factors between 1 and 8 were considered, as well as selection. The tested r values were between 2 and 14 with a varying step of 3. The optimum value was 5 for all local combination strategies.

2.7.4.4 Evaluation measures

As in the phantom experiments, SI and MASD were evaluated. In this case, manual segmentations were considered to be the ground truth.

Table 2.11: *Average SI for different brain structures with different combination strategies. In the case of paired structures (left-right), the first number indicates mean SI of the left structure and the second refers to the right structure. The highest value for each structure is highlighted in bold. (*) indicates $p < 0.05$ according to the Wilcoxon matched-pairs signed-ranks test, when compared one-to-one to all the other columns. (**) indicates $p < 0.001$. If two columns are marked, it means that the difference between them is not statistically significant, but it is with the rest of the columns.*

Brain Structure	Combination Strategy							
	MV	STAPLE	GWV-MI	GWV-NCC	GWV-MSD	LWV-MI	LWV-NCC	LWV-MSD
Thalamus	0.86;0.86	0.85;0.85	0.87;0.87	0.86;0.86	0.87;0.87	0.87;0.87	0.86;0.86	**0.87;0.88**
Caudate	0.78;0.78	0.66;0.69	0.79;0.79	0.78;0.78	0.79;0.80	0.81;0.80	0.79;0.78	**0.83;0.83** *
Putamen	0.86;0.85	0.81;0.79	0.86;0.86	0.86;0.85	0.86;0.86	**0.87;0.86**	0.86;0.85	0.86;0.86
Pallidum	0.80;0.80	0.69;0.70	0.80;0.81	0.80;0.80	0.80;0.80	**0.81;0.81** *	0.80;0.80	0.78;0.79
Hippocampus	0.73;0.75	0.52;0.51	0.74;0.75	0.73;0.75	0.73;0.75	**0.74;0.76** *	0.73;0.75	**0.74;0.76** *
Amygdala	0.71;0.71	0.65;0.64	0.72;0.72	0.71;0.71	0.72;0.72	0.72;0.72	0.71;0.71	0.72;0.72
Accumbens area	0.66;0.66	0.51;0.47	0.67;0.68	0.66;0.66	0.66;0.66	**0.68;0.68**	0.66;0.66	0.67;0.67
Ventral DC	0.81;0.82	0.79;0.77	0.82;0.82	0.82;0.82	0.82;0.82	0.82;0.82	0.82;0.82	0.82;0.82
Cerebral WM	0.75;0.75	0.59;0.54	0.75;0.75	0.75;0.75	0.75;0.75	0.76;0.76	0.75;0.75	**0.78;0.78** **
Cerebral Cortex	0.78;0.78	0.56;0.47	0.79;0.79	0.79;0.78	0.79;0.78	0.79;0.79	0.78;0.78	**0.81;0.81** **
Lateral Ventricle	0.77;0.75	0.51;0.50	0.78;0.76	0.77;0.75	0.77;0.76	0.81;0.79	0.78;0.76	**0.83;0.82** **
Inferior Lat Vent	0.16;0.15	0.15;0.12	0.22;0.21	0.18;0.17	0.20;0.18	0.22;0.20	0.18;0.17	**0.22;0.22**
Cerebellum Cortex	0.84;0.84	0.73;0.73	0.85;0.86	0.84;0.85	0.85;0.85	0.85;0.85	0.84;0.84	**0.86;0.86**
Cerebellum WM	0.78;0.78	0.70;0.70	0.79;0.79	0.78;0.78	0.78;0.78	**0.80;0.80** *	0.78;0.78	0.79;0.79
Third Ventricle	0.71	0.70	0.71	0.71	0.70	**0.73** *	0.71	**0.74** *
Fourth Ventricle	0.75	0.54	0.75	0.75	0.75	**0.77** **	0.75	**0.77** **
Brain Stem	0.89	0.85	0.90	0.90	0.90	0.90	0.90	**0.91**
CSF	0.57	0.55	0.60	0.58	0.55	**0.61**	0.58	**0.61**

2.7.4.5 Results

Table 2.11 shows the average SI for different brain regions using several combination methods. Table 2.12 shows the corresponding results for the MASD. In general, local strategies rendered better accuracies than global methods, but performance varies from region to region. No combination method is better than the rest for all regions.

Out of the studied 18 structures, LWV-MSD was among the methods with highest average similarity index in 12 cases. This number was 8 for LWV-MI. If MASD was evaluated, results followed the same trend. LWV-MSD was the best method for 15 structures (including ties), and LWV-MI

Table 2.12: *Average MASD for different brain structures with different combination strategies. In the case of paired structures (left-right), the first number indicates mean MASD of the left structure and the second refers to the right structure. The highest value for each structure is highlighted in bold. (*) indicates $p < 0.05$ according to the Wilcoxon matched-pairs signed-ranks test, when compared one-to-one to all the other columns. (**) indicates $p < 0.001$. If two columns are marked, it means that the difference between them is not statistically significant, but it is with the rest of the columns.*

Brain Structure	Combination Strategy							
	MV	STAPLE	GWV-MI	GWV-NCC	GWV-MSD	LWV-MI	LWV-NCC	LWV-MSD
Thalamus	0.85;0.89	0.94;0.98	0.81;0.81	0.84;0.88	0.84;0.88	**0.79;0.80** *	0.81;0.84	**0.76;0.74** *
Caudate	0.78;0.81	1.74;1.53	0.75;0.75	0.78;0.80	0.75;0.75	0.69;0.71	0.76;0.79	**0.63;0.64** *
Putamen	0.70;0.72	0.94;1.14	0.67;0.67	0.65;0.72	0.68;0.68	0.66;0.68	0.70;0.72	0.66;0.67
Pallidum	0.70;0.69	1.45;1.27	0.70;0.67	0.70;0.75	0.70;0.68	**0.68;0.66** *	0.70;0.69	0.73;0.70
Hippocampus	0.88;0.79	2.98;4.48	0.85;0.78	0.87;0.75	0.86;0.79	**0.83;0.75** *	0.87;0.79	**0.82;0.76** *
Amygdala	0.91;0.92	1.30;1.36	0.87;0.89	0.91;0.91	0.87;0.89	0.88;0.90	0.91;0.92	**0.84;0.86**
Accumbens area	0.74;0.71	3.14;4.33	0.71;0.0.68	0.74;0.71	0.77;0.74	0.70;0.68	0.74;0.71	**0.68;0.67**
Ventral DC	0.80;0.78	0.96;1.09	0.77;0.74	0.80;0.77	0.77;0.75	0.77;0.75	0.80;0.78	**0.74;0.73**
Cerebral WM	1.31;1.29	1.61;1.78	1.24;1.22	1.30;1.29	1.25;1.23	1.22;1.21	1.30;1.28	**1.09;1.07** **
Cerebral Cortex	1.34;1.31	2.64;3.85	1.24;1.20	1.34;1.31	1.26;1.23	1.29;1.25	1.35;1.32	**1.13;1.10** **
Lateral Ventricle	0.88;1.01	3.06;3.10	0.89;0.90	0.87;0.96	0.89;0.91	0.76;0.80	0.85;0.94	**0.67;0.70** *
Inferior Lat Vent	1.64;1.55	2.62;3.53	1.44;1.43	1.55;1.48	1.45;1.44	**1.40;1.46**	1.55;1.50	1.48;1.45
Cerebellum Cortex	1.35;1.35	1.76;1.74	1.23;1.19	1.34;1.32	1.28;1.25	1.26;1.24	1.35;1.33	**1.18;1.14** *
Cerebellum WM	1.19;1.20	1.61;1.65	1.09;1.08	1.17;1.19	1.12;1.12	1.10;1.08	1.18;1.19	**1.04;1.01** *
Third Ventricle	0.68	0.72	0.65	0.68	0.68	**0.62** *	0.67	**0.59** *
Fourth Ventricle	0.68	2.54	0.68	0.68	0.67	**0.61** **	0.67	**0.62** **
Brain Stem	0.80	1.08	0.75	0.79	0.74	0.74	0.79	**0.69**
CSF	1.22	1.30	1.12	1.20	1.08	1.14	1.21	**1.05**

outperformed the others in 6 cases. It must be noted that ties occurred, as it can can be seen in Tables 2.11 and 2.12.

However, more interesting than the absolute number of regions in which one method was better than others, is the study of the particularities of those regions. Specifically, and for illustrative purposes, we will look at two particular structures: the lateral ventricles and the pallidum. The two lateral ventricles conform the largest part of the ventricular system of the brain and appear as a dark region on T1 MR images (see Fig. 2.13). They are mainly surrounded by cerebral white matter, which shows a much brighter gray level. The caudate, a nucleus within the basal ganglia, is also in contact with the lateral ventricles and, even though the intensity difference is smaller, it is still clearly distinguishable. It is in this kind of

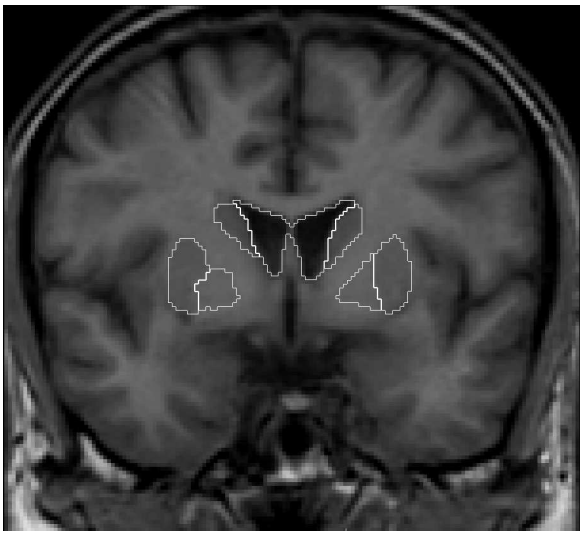


Figure 2.13: *T1 MR slice of human brain, with different anatomical structures delineated manually. Pallidum and putamen are the two paired structures below. The putamen is in the external part and the pallidum is in the interior. The lateral ventricles are the darkest structures delineated above, with the caudates next to them. The lack of gray level difference between pallidum, putamen and cerebral white matter voxels can be observed, together with the large contrast between the ventricles and the surrounding structures.*

structures with large intensity contrast in which the advantage of using local weighting approaches is larger. The average SI of the two-side lateral ventricles using LWV-MSD is 0.825, in contrast with 0.760 for the Majority Voting. This difference was very significant, according to the Wilcoxon matched-pairs signed-ranks test ($p \leq 10^{-7}$) [59]. An illustrative slice with the corresponding segmentation can be seen in Fig. 2.14. The cerebral and cerebellum cortex, the cerebral white matter and the caudate belong to the group of structures with this kind of characteristics.

The pallidum or globus pallidus is part of the basal ganglia. It is surrounded by many regions, mainly cerebral white matter and the putamen,

Table 2.13: *Average SI and MASD for Majority Voting and the best segmentation combination method for each structure, selected a-posteriori according to its SI value.*

Evaluation Measure	Combination Strategy	
	Majority Voting	Best Combination Method
SI	0.733	**0.758**
MASD	0.967	**0.851**

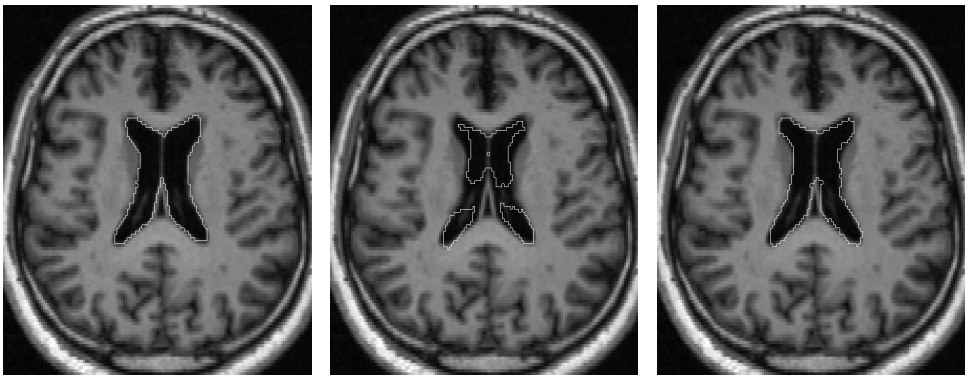
but also the caudate, the amygdala and other close structures. The intensity contrast with most of them is very low (see Fig. 2.13). In this case, LWV-MSD performs worse than Majority Voting ($p \leq 0.03$ with Wilcoxon matched-pairs signed-ranks test), and GWV-MI achieves second best SI, after LWV-MI ($p \leq 0.02$). A slice comparing LWV-MSD and GWV-MI is shown in Fig. 2.15. A comparable results arises for other regions as the putamen or the cerebellum white matter, which also show low contrast with neighboring structures.

If, for each brain structure, the best fusion rule is selected according to its SI value, the mean values for SI and MASD shown in Table 2.13 result. For comparison, the same parameters for Majority Voting are also presented. There is a 3.41% gain in SI and 12.00% reduction in the average distance between surfaces with respect to Majority Voting.

With our current implementation and on an Intel Xeon 3.20 GHz processor, approximate average computation times for segmentation combination in each brain scan were: 10 seconds for Majority Voting, 1 hour for STAPLE, 2 minutes for GWV-MI, 4 minutes for GWV-NCC, 4 minutes for GWV-MSD, 5 minutes for LWV-MSD, 10 minutes for LWV-NCC and 2 hours for LWV-MI. The long time required by LWV-MI is due to the high computational burden of estimating the entropies from the histograms for each voxel in which local weighted voting is required.

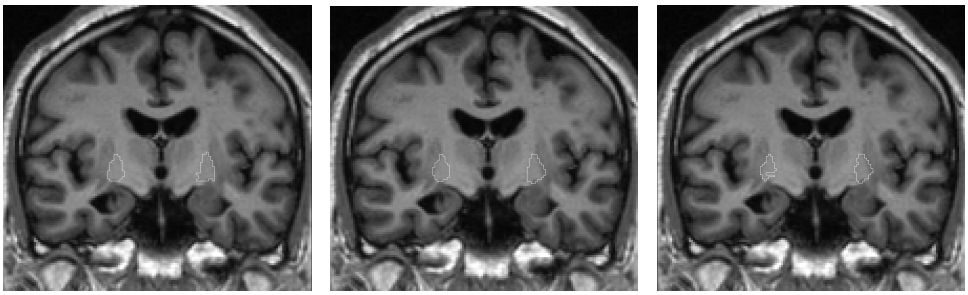
2.7.4.6 Influence of tunable parameter selection

As it was done in the phantom images, we also studied the effect of varying both the gain (p) and neighborhood radius (r) parameters on the



(a) Manual Segmentation (b) Majority Voting (c) Local WV-MSD

Figure 2.14: *Axial slice showing different segmentations of the lateral ventricles. The Local Weighted Voting based on the Mean Square Distance approaches the manual segmentation better than the Majority Voting.*



(a) Manual Segmentation (b) Global WV-MI (c) Local WV-MSD

Figure 2.15: *Coronal slice showing different segmentations of the pallidum. The Local Weighted Voting based on Mean Square Distance results in a more noisy delineation than the Global Weighted Voting based on Mutual Information.*

performance of the different combination methods. Results are summed up in Fig. 2.16 and Fig. 2.17.

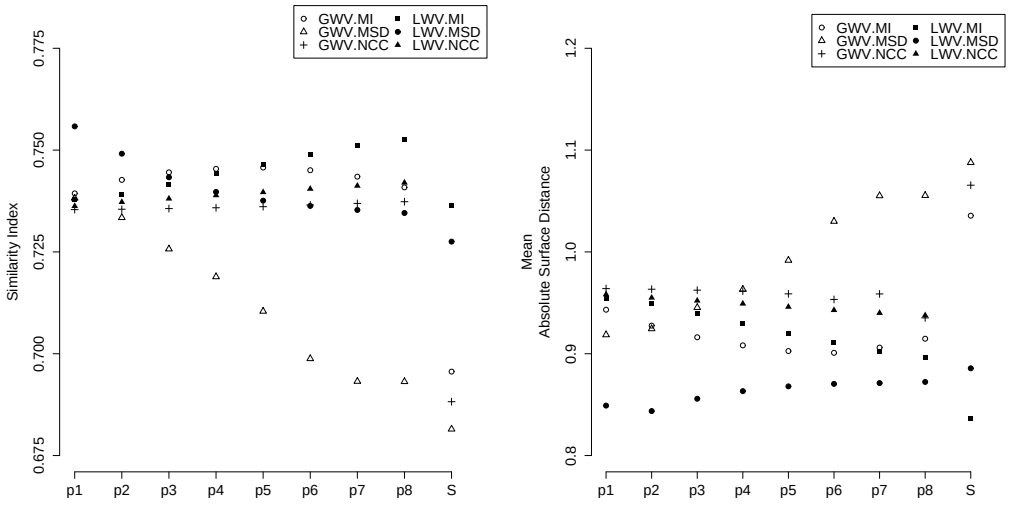
In the case of varying p, results are in principle not as clear as in the phantom dataset. As expected, GWV-MSD shows the greatest variability, $p = 1$ being the optimum solution. Both GWV-NCC and LWV-NCC vary very slightly through all values, implying that gain selection is not relevant. GWV-MI also varies smoothly and shows a maximum at $p = 5$. LWV-MI shows a continuous improvement from $p = 1$ to $p = 8$, in contrast with, LWV-MSD which degrades its performance in that direction. In all cases the proposed method was able to detect the optimum p value except for LWV-MI, where the second best was found.

For local combination strategies, varying the neighborhood radius r had varying effects on the different combination techniques. For LWV-MSD, $r = 5$ proved the optimum global radius with a considerably sharp peak. In the case of LWV-MI and LWV-NCC the radius value does not greatly influence the overall performance, except for the $r = 2$ value for LWV-MI which clearly causes worse results.

2.7.5 DISCUSSION AND CONCLUSION

Results from experiments with phantom images and real brain MR images differ slightly. While for the phantom images LWV-MSD performs better or equal as all the other methods for all regions with a single exception, this is not the case with the brain images. On the phantom images, LWV-MSD is in general the best combination method, followed by LWV-MI and GWV-MI. The difference between them is lower in regions with low contrast with the background. We believe that this is due to the lack of contrasted edges. For each voxel, each segmentation is weighted according to a local similarity measure between the target image and the registered atlas image. If a high contrast between neighboring regions exists at that point, the similarity measure will be able to distinguish between accurate and inaccurate segmentations, because it will be sensitive to the overlap between the different regions. However, if no intensity difference exists, the similarity measure will not be able to assess the accuracy of the registration at that point. Moreover, the weights derived from the local mean square distance might be influenced by noise.

On the IBSR dataset, though, LWV-MSD is not the best choice for all regions. As pointed out in the results section, in regions that show low contrast with neighboring structures, other methods should be preferred. Even global voting strategies perform better than LWV-MSD in those cases.

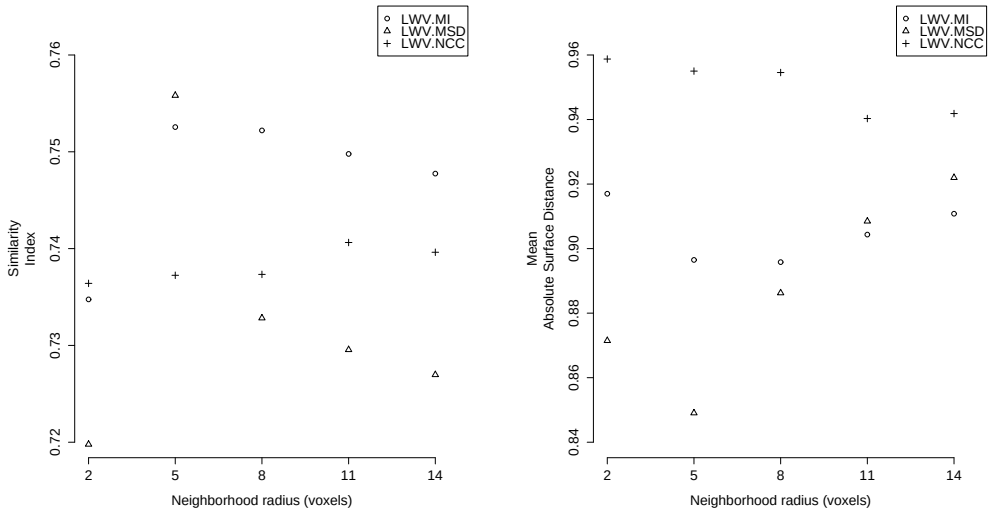


(a) SI for varying gain factor p.

(b) MASD for varying gain factor p.

Figure 2.16: *Plots showing the effect of varying the gain factor p on mean SI and MASD over all regions for different combination strategies on IBSR images. p is varied from* 1 *to* 8. *S indicates* selection, *that is, the voxel with the highest weight is selected, without any further voting process. $r = 5$ was set for all LWV methods.*

One of the reasons for global approaches to perform as good as or even better than local ones in low-contrast regions has been explained above. Another one, which applies only when a manual segmentation is used as the ground truth, is related to the accuracy of manual segmentations in low-contrast regions. It is obvious that it is hard to manually delineate a structure such as the pallidum in an MR image, due to its lack of contrast. As a result, manual segmentations might be less accurate than those from other better defined regions. In this kind of regions, features as shape or relative position play a much more important role than actual voxel intensity. Then, global voting approaches have an advantage over local methods, for they favour typical shapes and relative positions rather than voxel intensity distributions.



(a) SI for varying neighborhood region radius r. (b) MASD for varying neighborhood region radius r.

Figure 2.17: *Plots showing the effect of varying the neighborhood region radius r on mean SI and MASD over all regions for different combination strategies on IBSR images. For all methods, p was set to the value employed in the rest of experiments.*

If LWV-MI and LWV-MSD are compared, the latter is better for more regions but the former proved less sensitive to noise. This means that in low-contrast regions, LWV-MSD can yield worse accuracies than Majority Voting, for example in the case of the putamen or the pallidum. If differences in voxel gray values are only due to noise, as it is the case when there are no intrinsic intensity differences, MSD as local similarity measure might lead to incorrect weights in the voting process. A possible solution would be to increase the neighborhood radius r in order to increase the intrinsic regularization. However, our experiments have shown that this can reduce the mean SI and MASD. It must be noted that LWV-MSD could be used on this image dataset because voxel gray levels had been previously normalized, and the basic assumption of intensity conservation among the images was fulfilled [82]. We believe this is the reason why LWV-MSD outperformed LWV-NCC.

We proposed using a representative image from the dataset to set the two variable parameters -gain factor and neighborhood radius- of the local combination algorithms. The simple assumption behind this approach is that the parameters that are optimum for a single image should not be far from the globally optimum parameters, given that all images have approximately the same intensity distribution. In our two test datasets, parameter selection proved generally not critical and a wide range of parameters yielded comparable results. However, higher robustness could be obtained by employing more images for the parameter selection.

The fact that STAPLE yields worse SI values than the rest of the methods is an unexpected outcome, for it departs from results reported in [75]. Reasons might be multiple. First, in the referred publication recognition rate is used for evaluation, instead of SI and MASD. Second, STAPLE was tested on a completely different image dataset. Third, the lack of any *a priori* probabilities may affect the algorithm negatively, and performance could improve significantly if performance estimates could be initialized within a reasonable range [74]. Moreover, in [77] majority voting and STAPLE showed mostly not statistically significant differences when combining segmentations of MR image of the prostate, and in some atlas sets majority voting performed better.

Many publications on segmentation present a new method for a particular task, compare it with previous techniques and show the advantages of the novel contribution. Results from our experiments suggest that, when combining multiple segmentations in atlas-based segmentation, the principle of the 'no-panacea theorem' applies [89]. No method is better than others always, for all regions and images. One must select the best strategy among the existing (Majority Voting, Global Weighted Voting, any kind of Local Weighted Voting, etc.) according to the particular characteristics of the images and the regions. This can lead to substantial gain compared to majority voting, as results from Table 2.13 show. We believe that a major contribution of our work is the study of the conditions in which local weighting methods perform consistently better than global methods. As it was explained, regions that have large intensity contrast with neighboring tissue, such as the ventricles in T1 MR images, benefit specially from local strategies. We believe that the same concept can be applied to other imaging modalities and segmentation tasks that show similar contrast characteristics, such as lungs in Computed Tomography images. In contrast, there might be little or no benefit from using local approaches in regions

that show similar intensities as their surrounding structures, and global methods should be preferred.

The main limitation of voting strategies is that they can not correct incorrect segmentations when all candidate segmentations have failed in a certain area. An extensive atlas which is representative of the whole population and a good registration algorithm can greatly limit this problem, by maximizing the probability that at least one of the candidate segmentations will be correct in all locations. Then, the local similarity measure would need to be sensitive enough to detect that correct registration and discard the incorrect ones.

The advantage of using local approaches being mainly in high-contrast edges, suggests that it might be worth adding information on the edge strength when locally selecting a combination method. That is, a given structure might have different kind of borders: It might have a high-contrast border with one structure and a low-contrast border with another. Then, if local strategies were applied only on the high-contrast borders, overall accuracy might be better. This could be applied in lungs affected by inflammation: borders with normal intensity contrast would benefit from local combination strategies, while areas without contrast -i.e., high intensity areas in the lungs- would require global strategies.

In conclusion, we have addressed the issue of combining segmentations to achieve the highest possible accuracy in multi-atlas medical image segmentation. We proposed a general local weighted voting method and showed how it can be applied with different similarity measures on different image datasets. We studied the performance of global and local weighted voting strategies for multi-atlas segmentation combination, and concluded that local methods should be preferred in regions that show high contrast with neighbor areas. To achieve optimum overall results, the best fusion method for each region must be found.

The plural of anecdote is not data.

Roger Brinner, economist

3 Experiments

In this chapter, we detail the experiments carried out to characterize different models of lung disease in mice, mainly using micro-CT as a non-invasive imaging technique. The studied models are: silica-induced chronic pulmonary inflammation, elastase-induced emphysema and a combined model of emphysema and urethane-induced lung cancer. Results and limitations of each study are discussed at the end of each section.

3.1 SILICA-INDUCED CHRONIC PULMONARY INFLAMMATION

3.1.1 INTRODUCTION

Silica-induced pulmonary inflammation in mice is a good model for chronic pulmonary inflammation and fibrosis, which are processes related to multiple diseases such as interstitial lung disease and pneumonitis [90]. Due to the significantly different X-ray absorption properties existing between the air and lung parenchyma, high resolution X-ray computed tomography (micro-CT) is very well suited for non-invasively imaging animal models of lung disease [10]. In particular, micro-CT is appropriate in longitudinal studies, aimed at understanding the dynamics of lung diseases. Indeed,

micro-CT has been recently used to image in vivo animal models of lung cancer, emphysema and fibrosis [12, 13, 18, 22].

The use of micro-CT to assess the presence and extent of lung inflammation and fibrosis in rodents has been reported in a few recent works. Jobse *et al.* studied a model of allergic inflammation in rats [14]. They scanned the rats at baseline and longitudinally at different time points, using retrospective gating. Semi-automatic methods for segmentation and quantification of the images were used. Ask *et al.* looked at a model of pulmonary fibrosis without any respiratory gating and with semiautomatic tools for quantification [13]. They complemented their results with pulmonary function tests. Cavanaugh *et al.* showed correlation between semiautomatic quantifications of micro-CT images of bleomycin-induced lung injury in mice, acquired with respiratory gating at end inspiration [18]. Lee *et al.* employed the same animal model, and tried to detect different disease patterns on the micro-CT images [91]. Very good results were achieved in ex vivo scans, but they were considerably worse in *in vivo* scans, due to lack of respiratory gating. In fact, in the same work a preliminary study with gating on intubated mice was reported, which was not applied in the rest of the work due to a mortality rate of 80%.

Therefore, different options for respiratory gating have been reported and applied in assessment of fibrosis and inflammation models in rodents. Nevertheless, a gating technique that offers high quality images applied successfully to the longitudinal study of a lung disease model has not been reported and would be very useful. Moreover, fibrosis and inflammation measurements on micro-CT images have always been done manually or semi-automatically, which requires long operator times and introduces operator variability. In this work, we present a quantitative micro-CT based longitudinal study of a silica-induced mouse model of chronic pulmonary inflammation. To improve the sensitivity of the detection, we reduce movement-related artifacts caused by respiratory motion, by using an iso-pressure breath hold technique, following a carefully devised protocol to reduce mortality. Lungs on micro-CT images are automatically segmented and the extent of damage tissue is computed. Results are validated by comparison with histomorphometry.

Along with histological analysis, we compare our image-based measurements with several parameters of pulmonary function. Several methods exist to measure lung function in murine models of pulmonary disease, as

explained in section 2.3 of this thesis. In this study, we employ both the single compartment model, which consists of a flow resistive conduit (with resistance R) serving a single elastic compartment (with compliance C) [20], and the constant phase model, which separates the contribution of the main airways and the lung tissue, resulting in four different parameters: airway resistance (Raw), inertance (I), tissue damping (G) and tissue elastance (H) [31]. Both models require animal intubation to carry out the measurements, which can be easily done with commercially available equipment. One of the aims of this study was to compare the information given by quantitative micro-CT with lung functional parameters.

In summary, this study was performed to evaluate the feasibility and use of iso-pressure breath hold gated micro-CT to capture the dynamics of a mouse model of chronic pulmonary inflammation, and to compare quantitative micro-CT derived lung damage parameters with the results of pulmonary function tests.

3.1.2 MATERIAL AND METHODS

A total of 49 A/J 8-week-old mice (Harlan UK Limited, Oxon, UK) were used in the study. Inflammation was produced by a single oropharyngeal aspiration of crystalline silica as explained in section 2.1. Control mice aspired 90 μl of saline instead. Fourteen animals were used in the longitudinal study, divided into the inflammation (7 mice) and control groups (7 mice). These animals were imaged at baseline and then 4, 14 and 34 weeks after the first scan. At each time point, 5 additional control and treated animals were imaged and then sacrificed for histological analysis.

Parameters of pulmonary function were computed for all animals in the study following the method described in section 2.3 of this thesis. In particular, six different parameters were obtained: lung resistance (R), lung compliance (C), airway resistance (Raw), airway inertance (I), tissue damping (G) and tissue elastance (H). The protocol for micro-CT image acquisition is described in section 2.5.

The lungs in micro-CT images were automatically segmented from the rest of the chest using the method described in subsection 2.5.3. The airways were segmented and the radii of the right and left bronchi were measured, using the method explained in section 2.6.2. Two parameters were computed from the segmentations: the mean lung voxel intensity (MLVI)

and the percentage of damaged lung volume (DLVF). As damaged volume, we considered all voxels of the lung that displayed intensity above the threshold that we used to find the dark areas of the lung.

Lungs were fixated as explained in section 2.4. One section per lobe (5 lobes per mice) was stained with hematoxylin and eosin and used for qualitative histological analysis. Mosaic images of the two sections corresponding to the left lobe were acquired with a Zeiss Axioplan 2ie microscope (Carl Zeiss, Jena, Germany) with a 12.5x magnification, and the damaged area fraction (DAF), or percentage of damaged area, was computed semi-automatically using custom software, which detects damaged tissue based on the staining differences with respect to normal healthy lung tissue.

Another section per lobe was stained with picro-sirius red. To quantify alveolar fibrosis, two fields per mouse containing mainly alveoli were randomly selected, and a fibrosis score was computed as the ratio between the area occupied by collagen and the total area occupied by tissue. The used magnification was 200x.

For statistical analysis, the mean and standard deviation of each micro-CT and pulmonary function test parameters were computed separately for each animal group (control and inflammation) and time point (baseline, and weeks 4, 14 and 34 after treatment). For each time point and parameter, the values of the two groups were compared using the Mann Whitney U tests (online tool available from the Institute of Phonetic Sciences, University of Amsterdam). A p value below 0.01 was considered to yield statistically significant results. For the parameters that resulted in statistically significant differences between the control and inflammation groups at least in one time point, a receiver operating characteristic (ROC) curve was generated. Animals of the chronic inflammation group were considered positive results, while control animals were considered negative results. If each of the computed parameters is regarded as a classifier, the ROC curve displays the true positive rate versus the false positive rate. Then, the area under the ROC curve (AUC) was computed, to evaluate the ability of each parameter to discriminate between healthy and diseased subjects [92]. Both ROC and AUC were computed with the GNU Octave software [93].

3.1.3 RESULTS

3.1.3.1 Image analysis

Figure 3.1 shows two representative examples of transversal micro-CT slices of a healthy mouse lung and the lung of a silica-treated animal 14 weeks after the silica aspiration. In Figure 3.2, three-dimensional reconstructions at consecutive time points are shown, together with micro-CT slices and histological samples. The evolution of disease can be observed.

At all time points, the MLVI, DLVF and radii of the left and right bronchi were computed for the control and the inflammation groups, and differences between the groups were evaluated.

The MLVI was significantly different in the control versus the inflammation groups at all time points except at baseline ($p < 0.01$). In particular, the inflammation group showed an increase in MLVI of 19.15% at week 4, 23.75% at week 14 and 48.80% at week 34, when compared to the control group. Regarding the evolution of the absolute values, in control mice MLVI decreased from an average of -550 HU at baseline to -631 HU at week 34 (14.72%, $p < 0.01$), while in the silica-treated group it increased from a mean of -560 HU to -323 HU (42.32%, $p < 0.01$).

The difference in DLVF between the control and inflammation group was 171.83% at week 4, 314.92% at week 14 and 642.23% at week 34, statistically significant in all cases. The DLVF value remained constant with time in the control group (5.23% at baseline, 5.66% at week 34, $p = 0.21$), but it increased steadily from 5.25% at baseline to 42.01% at week 34 (800.19%, $p < 0.01$) in the silica-treated group.

The radii of the right and left mainstem bronchi were smaller in the silica-treated group than in the control group at all time points ($p < 0.01$). The right bronchus of the treated mice was 16.38% smaller than the controls at week 4, 20.08% at week 14 and 37.60% at week 34. These differences in the left bronchus were 14.92%, 17.57% and 31.14%. An evolution could also be seen in both radii in the diseased group. During the 34 weeks, the radius of the right bronchus decreased in average from 763.83 μm to 447.16 μm (41.46%, $p < 0.01$), and the left bronchus from 555.01 μm to 349.31 μm (37.06%, $p < 0.01$).

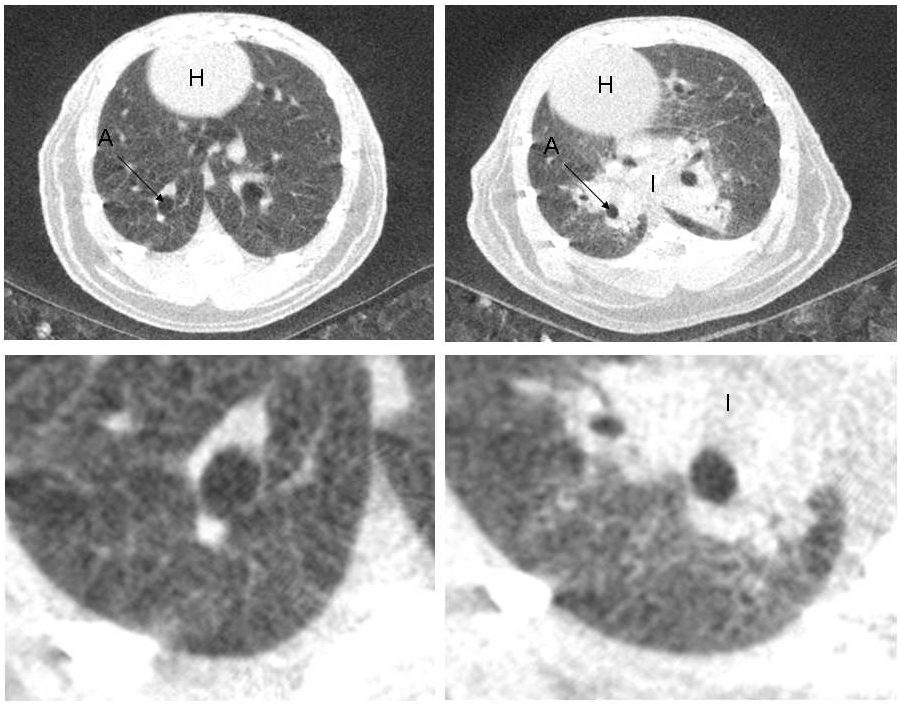


Figure 3.1: *Sample transversal micro-CT slices of a healthy mouse lung (top left), and a lung affected by silica-induced chronic inflammation 14 weeks after silica aspiration (top right). Below, detailed view of the same slices. Note inflammatory changes affecting predominantly central broncovascular structures. The heart (H), an airway (A) and an inflamed area (I).*

Figure 3.3 shows the MLVI and DLVF for each animal group and time point. In Figure 3.4 the evolution of the mainstem bronchi radii can be seen. Figure 3.5 shows the ROC curves for MLVI, DLVF and the bronchi radii. The AUC values were 0.967 for MLVI, 0.999 for DLVF, 0.939 for the right mainstem bronchus radius and 0.892 for the left mainstem bronchus radius.

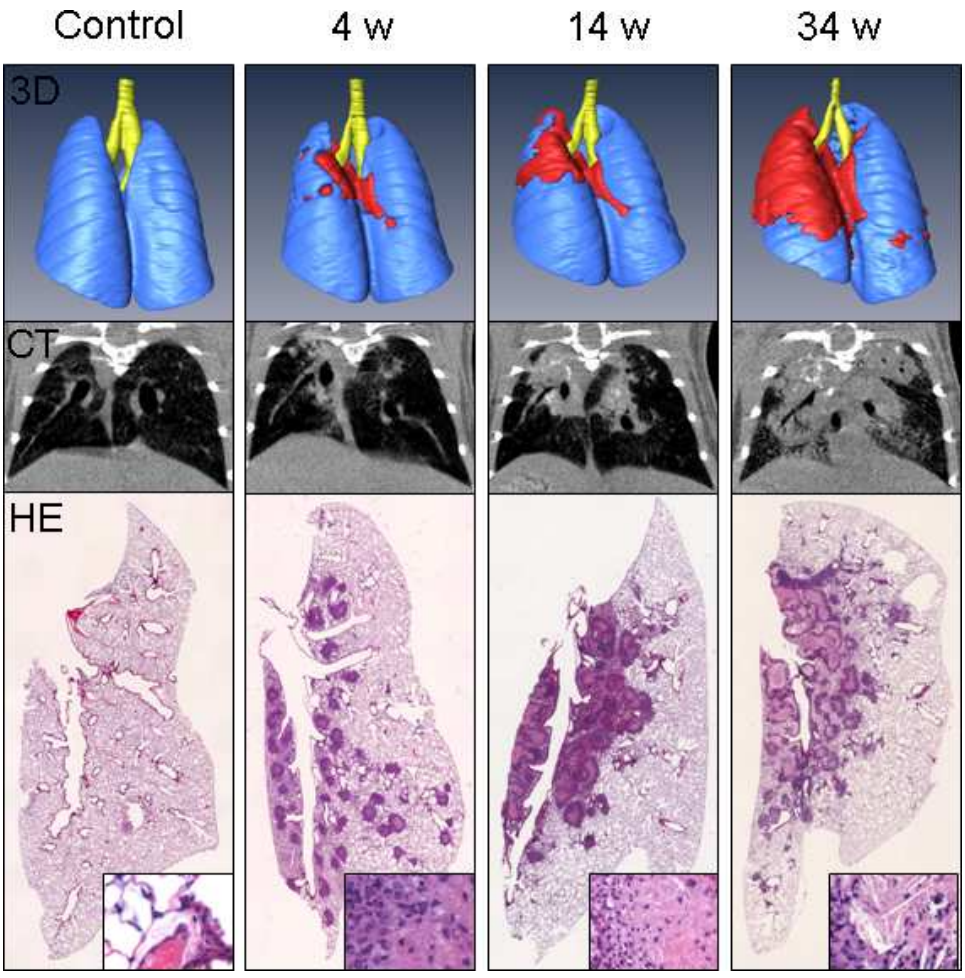


Figure 3.2: *3D reconstruction from the automatic micro-CT image segmentations (normal tissue in blue, lung injury in red, airways in yellow), representative coronal slices and hematoxilin-eosin staining sections of lung in control and silica-treated mice after 4, 14 and 34 weeks. Original magnifications for HE: x9. Inset magnification: x145.*

3.1.3.2 Pulmonary function tests

Six parameters of lung function were measured: resistance (R), compliance (C), airway resistance (Raw), inertance (I), tissue damping (G) and

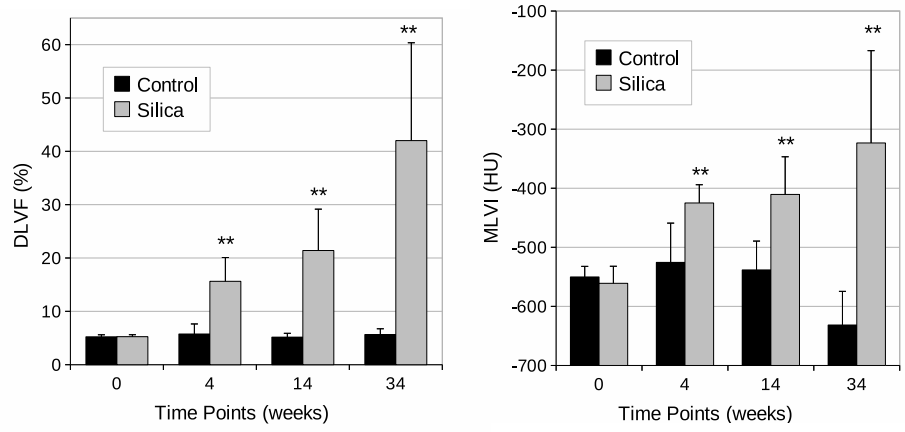


Figure 3.3: *Mean Lung Volume Intensity (MLVI, right) and Damaged Lung Volume Fraction in percentage (DLVF, left) for the control and silica groups at the different time points. Both show the time dependent evolution of lung damage as well as the clear difference with the control untreated group in every time point after baseline. ** indicates statistical significance ($p < 0.01$).*

tissue elastance (H). R, C, Raw and G showed statistically significant differences between the control and the inflammation groups at some time point of the study, while H and I did not yield any statistically significant difference at any time point.

In particular, the inflammation group showed a statistically significant increase in Raw at weeks 14 (54.76%, $p < 0.01$) and 34 (96.61% , $p < 0.01$) but not at week 4. Differences in R, G and C are statistically significant only at week 34. R increased 155.08% in the inflammation group, G increased 98.52% and C decreased 42.37%, all $p < 0.01$. Figures 3.6 and 3.7 show the mean values of these parameters separately for each group. ROC curves for R, C, Raw and G are shown in Figure 3.5. The AUC values were 0.642 for R, 0.547 for C, 0.657 for Raw and 0.627 for G.

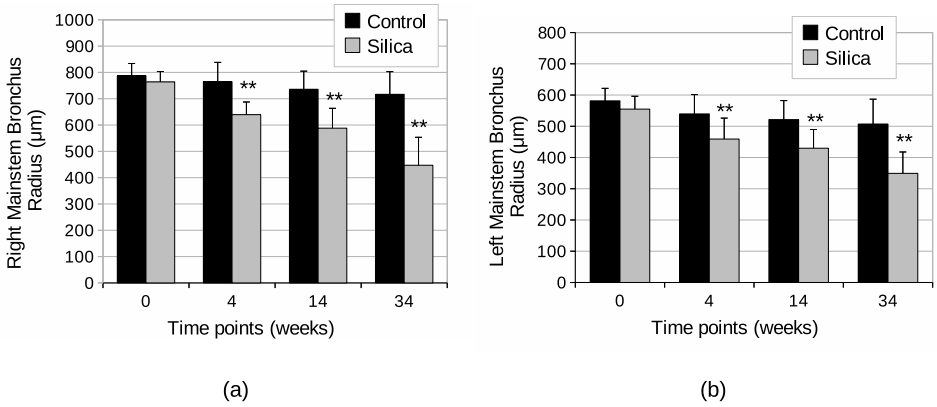


Figure 3.4: *Radii of the right and left mainstem bronchi for the control and silica-treated mice in the different time points. A statistically significant difference ($p < 0.01$) is seen between the two groups at all time points and for both bronchi.*

3.1.3.3 Histological examination

The Pearson correlation coefficient was $R^2 = 0.62$ between MLVI and DAF, and $R^2 = 0.59$ between DLVF and DAF. The relationship between the parameters can be seen in Figure 3.8. DLVF seems to correlate better with low levels of damage, while MLVI correlates better with histomorphometry in the cases of mild and severe lung damage.

There was no difference in the fibrosis score between control and silica-treated animals at week 4, but at weeks 14 and 34 silica-treated animals showed a significantly higher collagen content (7.10%±3.59% vs. 1.05%±0.77% at week 14, and 8.86%±5.48% vs. 1.71%±0.88%, at week 34, p¡0.01 in both cases).

Qualitative analysis of histological sections revealed chronic inflammation at all stages and lung lobes analyzed, and a time related progression of the inflammation and damage was observed (see Figure 3.2). At 4 weeks, silicotic granulomas were observed in all the lung lobes, located mainly in central areas close to branching and terminal bronchioles. These granulomas were often nodular, discrete and prominent. They were composed

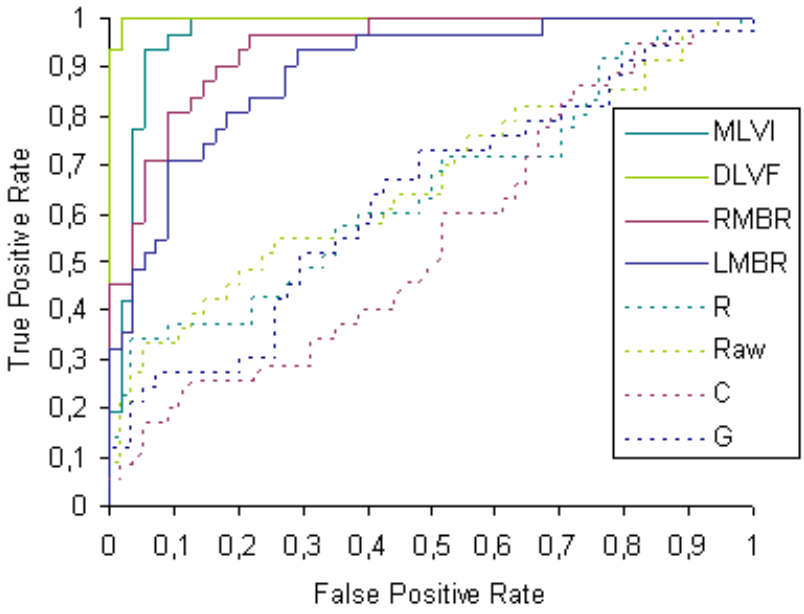


Figure 3.5: *ROC curves for the MLVI, DLVF and right and left mainstem bronchi radius (RMBR and LMBR, respectively), and the functional parameters R, Raw, C and G. The former four measurements were derived from the micro-CT images of the control and silica-treated mice. DLVF shows the best performance, because highest true positive rates can be achieved for the same false positive rate.*

of macrophages and some fibroblasts showing a central area with dust deposits, cellular debris and some fibrosis. Epithelial hyperplasia was observed in lung parenchyma regions, near or around the granulomas. Moreover, a slight perivascular lymphocytic reaction and focal areas of alveolar proteinosis were observed. As previously published, this shows that

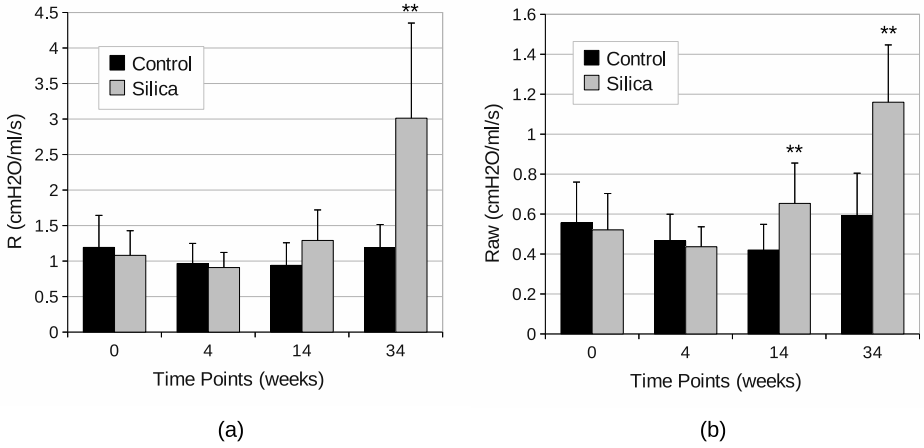


Figure 3.6: *Total lung Resistance (R) and airway resistance (Raw) in different time points for the two animal groups. ** indicates statistical significance between the control and the disease groups ($p < 0.01$).*

macrophages are the first defense mechanism against silica particles that reach the terminal airways and alveolar spaces [94].

At 14 weeks, many silicotic nodules could also be seen, most of them clustered. The inflamed region was larger than in the previous stage, and was located in the same areas of the lung lobes. At this time point, granulomas showed large acellular centers with extensive cellular debris and crystalline silica. The parenchyma between the silicotic nodules showed sparse alveolar proteinosis, thickened septa with interstitial macrophages, and groups of living and necrotic macrophages in alveolar spaces. Moreover, moderate bronchiolar hyperplasia, and hyperplastic type II pneumocytes were also observed surrounding the silicotic lesions. The perivascular lymphoid reaction was increased at this time.

Finally, at 34 weeks, silicotic granulomas remained similar in size to those observed at week 14, and showed larger central necrotic areas with lipid clefts and moderate fibrosis in the periphery. Parenchyma between silicotic nodules showed increased alveolar proteinosis and thickened alveolar septa. Bronchiolar hyperplasia remained near and between the granulomas,

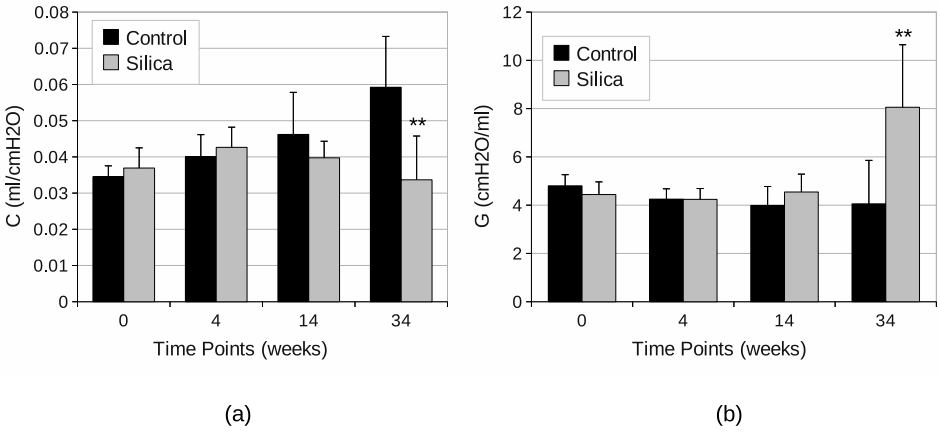


Figure 3.7: *Compliance (C) and tissue damping (G) in different time points for the two animal groups . ** indicates statistical significance between the control and the disease groups ($p < 0.01$)*

but areas with hyperplastic type II pneumocytes were reduced as compared to the previous time point. Interestingly, perivascular and peribronchiolar inflammatory reaction was more remarkable than at week 14. The inflammatory cells observed were predominantly lymphocytes and macrophages.

In summary, the inflammatory lesion induced by the silica insult was initially an acute granulomatous reaction that grew in size over the first 14 weeks. The overall damaged tissue area increased from week 14 to week 34. The morphological appearance and cellular composition of the granulomas and the alterations of the lung parenchyma surrounding them also showed time-related modifications.

3.1.4 DISCUSSION AND CONCLUSION

High-resolution micro-CT is an optimal tool for the in vivo study of the dynamics of animal models of pulmonary disease [52], and previous studies have shown a good correlation between micro-CT and histological findings in animal models of lung inflammation and fibrosis [13, 14, 18, 91].

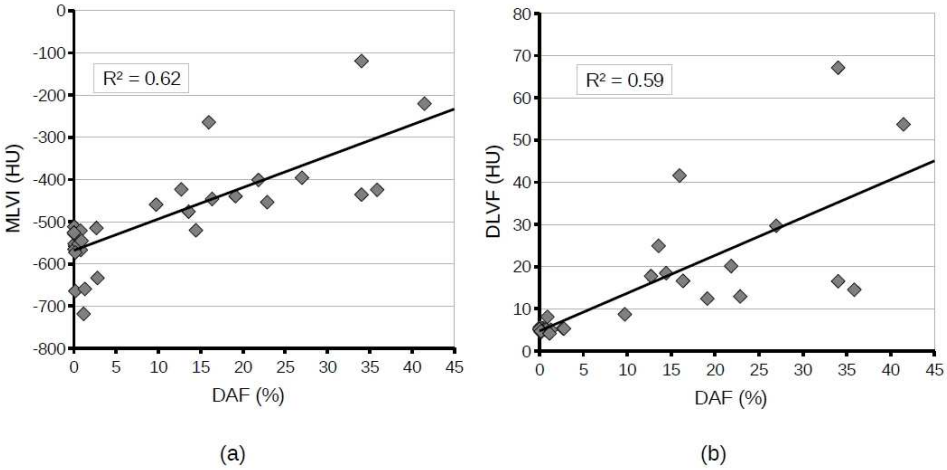


Figure 3.8: *(a) Correlation between Mean Lung Volume Intensity (MLVI) automatically measured from micro-CT images and damaged area fraction (DAF), measured from histological sections. Pearsons correlation coefficient is indicated. (b) Correlation between Damaged Lung Volume Fraction (DLVF) automatically measured from micro-CT images and DAF, measured from histological sections. Pearsons correlation coefficient is indicated. Medium correlation scores are achieved in both cases.*

We have studied the evolution of a mouse model of chronic inflammation using micro-CT, pulmonary function tests and histology. We took quantitative image-based measurements and contrasted them with the results of tests of pulmonary function as well as with histomorphometry.

Regarding the acquisition protocol, an important contribution of this work is the use of an iso-pressure breath hold gating technique, which results in increased image quality and highly accurate and repetitive quantification [17]. To the best of our knowledge, no study to date has evaluated the use of this technique to quantify and follow up lung damage in a longitudinal study. Precise anesthetic dosage and the use of a saline warm injection before disconnecting the animal from the ventilator proved helpful to obtain very high survival rates and rapid recovery. Breath hold gating has many advantages compared to other gating techniques. In contrast with

prospective gating on free breathing animals, lungs are kept filled with at a constant air pressure, virtually removing movement artifacts and increasing tissue contrast at the same time. The use of retrospective gating can be considered if a flat panel detector is available, since the long exposure times required by other kinds of detectors would result in very blurry images. As flat panel detectors have lower resolution, retrospective gating inevitably results in images with larger voxel sizes. The main drawback of breath hold gating is the need to intubate the animals and the potential damage resulting from artificial ventilation. In this work, we showed that following a carefully defined protocol, this method can be used even on heavily diseased animals.

Another technical contribution of this study is the use of automatic tools for airway measurement and lung disease quantification in micro-CT images. These methods are being actively developed and used in CT research [33, 34, 50, 95], but their application to animal studies has been very limited. We believe that these automatic algorithms are indeed key tools, because they provide reliable and reproducible means of quantifying the disease, while reducing drastically the required operator time. In particular, we have evaluated the performance of two parameters based on the image intensities (MLVI and DLVF) along with the radii of the two mainstem bronchi in detecting the presence of disease. We found that all parameters offer very high sensitivity and specificity, DLVF being a practically perfect classifier (AUC of 0.999). We also analyzed the correlation between the micro-CT measurements and histomorphometry. Modest correlation levels were achieved, which were in any case higher than the ones previously reported with different gating and quantification techniques [13, 18, 91]. It must be noted that histological measurement of disease extent was limited to two representative sections of the left lobe, while MLVI and DLVF consider the whole lung volume. Thus, it was unlikely to reach strong correlation levels.

We also calculated parameters of pulmonary function, which are generally used to study the respiratory mechanics in animal models [96]. Our results suggest that histopathologically evident changes in the lung do not necessarily affect the pulmonary function parameters measured by these tests. Airway resistance (Raw) was the parameter that discriminated best between diseased and healthy animals, although with a rather low AUC value (0.657). However, we believe that PFT parameters are of great value to complement micro-CT findings, since they provide relevant physiological

information on lung function. In fact, R, Raw, G and C do reveal the presence of disease at weeks 14 and 34. The increases in R and Raw are likely to be caused by the decrease in airway radius measured in the micro-CT images. It seems that at week 4 the difference in radii is not large enough to cause a significant increase in air flow obstruction. The changes in C may well be caused by the large extent of inflammation, which results in increased lung stiffness. As explained in [97], severe airway constriction and less lung tissue being sampled can also contribute to reduce C. The same reason, together with increased lung parenchyma inhomogeneity, can explain the increase in G seen at week 34 [97].

The X-ray dose delivered to the animals is far from lethal levels, but can not be dismissed. Intubation and artificial ventilation can also affect the normal functioning of the lungs. The control group can be used to analyze those effects. In the longitudinal study, we observed slight inflammation on only one of the control animals at week 4, which disappeared at 14 weeks. On histological evaluation at all time points, no evident changes in normal lung morphology were detected.

This work has several limitations. An exhaustive ROC curve analysis would require the computation of the variance of the curve. This requires multiple datasets, for which the ROC curve should be independently computed and the results averaged. However, the available dataset was limited and dividing the data in different groups would have resulted in too small groups. Nevertheless, differences in ROC curves and AUC values between micro-CT parameters and lung function parameters are very large and we believe they are sufficient to support our conclusions. The small sample size is particularly relevant in the measurements of pulmonary function, because their variability depends on factors which are hard to control, such as the relative position of the cannula in the animal trachea. This variability was limited by performing the measurements three times in each animal. A third limitation is related to the animal model. As any animal model of human disease, the silica-induced chronic inflammation in mice does not reproduce exactly any disease occurring naturally on humans. However, relevant information can be gained on pulmonary diseases that show fibrosis and inflammation.

In conclusion, iso-pressure breath hold gated micro-CT enables detecting inflammatory changes in a mouse model of chronic lung inflammation, and disease evolution can be followed in a longitudinal study. Quantitative

measurements obtained from breath hold gated micro-CT images show differences between the control and the diseased animals before function test parameters, and discriminate the groups with higher sensitivity and specificity. The methods and results presented here may be the basis for further studies involving other lung diseases, such as emphysema or asthma.

3.2 ELASTASE-INDUCED EMPHYSEMA

3.2.1 INTRODUCTION

Emphysema and chronic bronchitis are the two possible forms of chronic obstructive pulmonary disease (COPD). COPD is currently a major health problem in the developed world, and it is predicted to be the third leading cause of death worldwide by 2030 [98]. Emphysema is defined pathologically as the permanent enlargement of the airspaces distal to the terminal bronchioles, accompanied by destruction of their walls, without obvious fibrosis [99, 100]. The main cause of emphysema in humans is tobacco cigarette smoke. Many studies in animals and humans have looked at the causes of the initiation and progression of emphysema, being a protease-antiprotease imbalance the most accepted hypothesis [7]. Briefly, continuous smoking induces a chronic inflammatory process which results in abnormal secretion of proteases, subsequent chemical damage to structural proteins and potentially continuous collagen failure and tissue breakdown due to mechanical forces [101–104].

In order to better understand emphysema, several animal models have been developed and serve as important experimental tools [7]. For instance, cigarette smoke exposure is known to cause emphysema in guinea pigs, as well as in some mice strains [105]. Genetically modified mice can also be used to study the effects of particular genes [53]. However, probably the most widely used model is the elastase-induced emphysema. Administering porcine pancreatic elastase consistently causes airspace enlargement in a large variety of animals, and it is relatively inexpensive and simple to perform [7].

Several methods exist to quantify emphysema and related processes in mice, which are essential for a correct detection and characterization of the disease, as well as to assess the response in pharmacological studies. We propose to characterize and compare a variety of currently available methods, in order to analyze the potential and limitations of each technique, as well as the possible synergies between them.

Histomorphometry is the most accurate method for emphysema quantification [106], but it is an ex-vivo technique and cannot be used in follow up studies. Spirometry is clinically used to diagnose COPD. Although spirometry is not easily translated into animals, there are several invasive

and noninvasive pulmonary function tests that can provide physiological data about lung function in animals [20]. Micro-computed X-ray tomography (micro-CT) can also be used to detect emphysema in rodents in vivo with minimal or no invasiveness [11]. Finally, some studies have looked at the role of inflammatory cytokines in COPD and emphysema [107], suggesting that cytokine measurements could become relevant for both diagnosis and treatment planning. In mice, the inflammatory response of the lungs can be evaluated looking at protein or RNA cytokine levels, using blood plasma or lung tissue extracts, respectively.

All these techniques can provide quantitative measurements of emphysema and related processes. In this study we compare and characterize the type of information provided by each of them and put that information in the framework of the current standard i.e. histomorphometry-. To this end, we have studied the initiation and progression of elastase-induced emphysema in A/J mice using all five technical approaches, providing an exhaustive characterization of this model, which is of great interest due to its low cost and rapid development [7, 108].

3.2.2 MATERIAL AND METHODS

Two studies were done, one short-term and one long-term. 60 A/J mice, 11 weeks old, were used in the short-term experiment. They were divided into treatment (n=30) and control (n=30) groups. Animals of the treatment group were intratracheally instilled with porcine pancreatic elastase as described in section 2.1.

Five treated animals and five control animals were then sampled 1, 6, 12, 24 hours, 7 and 17 days after elastase administration. At each time point, animals were subject to micro-CT thoracic imaging and pulmonary function tests, and then were sacrificed while collecting samples for histomorphometry and cytokine measurements (RNA and protein). 30 additional A/J mice were used in a long-term experiment to confirm the trends seen in the short time group. These mice were 8 weeks old and were equally divided into control and treatment groups. Five treated and five control animals were studied 4 weeks, 14 weeks and 34 weeks after elastase treatment. On these animals, micro-CT imaging, pulmonary function tests and cytokine concentration measurements in blood plasma were performed.

Pulmonary function tests were performed following the method presented in section 2.3. For the sake of simplicity, only the results for the single compartment model will be shown.

Micro-CT image acquisition was done according to the previosuly reported protocol (section 2.2). Lung segmentation was done according to the method in section 2.5.2, and two parameters were computed in the lung volumes: mean lung voxel intensity (MLVI) and relative volume below -900Hounsfield Units (RVB -900 HU). This threshold was selected because intensity values below -900 HU are not generally present in scans of healthy mice (the RVB -900 HU is less than 5% in all healthy animals of any age).

After micro-CT imaging, lungs were excised and fixed as explained in section 2.4. Three different paraffin blocks were built from each lung, dividing the lobes in the following way: block 1 contained right lobes 1 and 2; block 2 contained right lobe 3 divided in two; and block 3 contained the left lobe divided in two. The right lobe 4 was reserved for RNA extraction. Three non-consecutive sections per paraffin block were stained with hematoxylin and eosin, resulting in 9 slides per mouse, each containing two lobe pieces.

All lobe sections were then imaged at 10x magnification using a Zeiss Axioplan 2ie microscope (Carl Zeiss, Jena, Germany). Lobular walls were automatically segmented and analyzed in order to detect and remove vessels and alveolar walls. This step was necessary because these structures would otherwise affect the quantification of emphysema. Finally the Lm and D2 parameters were automatically computed [106]. D2 represents the mean area of the alveoli sections corrected with the standard deviation and the skew to account for variability in alveolar sizes. In order to reduce computation time, the analysis was executed in parallel on 5 machines. The algorithms were programmed in hybrid Python/C++ code and made use of the ITK library.

RNA cytokine expression was measured from the right accesory lobe. The lobe was dissected and frozen in isopentane previously cooled in liquid nitrogen. Tissue was then stored at -80°C for further analysis. Lysate was obtained by physic homogenization of the frozen tissue with a steel mortar and a mixer mill (Retsch MM301). Total RNA isolation of the lysate was performed using standard QUIAGEN RNeasy Micro KIT (including the DNase step) following manufacturer's instruction. RNA concentrations

Table 3.1: *Annealling temperature, primers and product size for the genes used in the qRT-PCR process for measurement of inflammatory cytokine RNA expression.*

GENE	Annealing Temperature	Sense $5' \rightarrow 3'$	Antisense $3' \rightarrow 5'$	Product size (bp)
IL1-β	59°C	GGATGAGGACATGAGCACCT	TAATGGGAACGTCACACACC	108
IL6	60°C	TCTCTGGGAAATCGTGGAAA	TTCTGCAAGTGCATCATCGT	83
IP10	60°C	AATCATCCCTGCGAGCCTAT	TTTTTGGCTAAACGCTTTCATT	131
KC	58°C	TGGGATTCACCTCAAGAACA	TTTCTGAACCAAGGGAGCTT	137
MCP1	60°C	AGGTCCCTGTCATGCTTCTG	GGGATCATCTTGCTGGTGAA	128
MIP1-α	60°C	CTGCCTGCTGCTTCTCCTAC	CCCAGGTCTCTTTGGAGTCA	148
TNF-α	58°C	GCCTCTTCTCATTCCTGCTT	AGGGTCTGGGCCATAGAACT	134
B2M	60°C	ACCCTGGTCTTTCTGGTGCT	ATGTTCGGCTTCCCATTCTC	111

and the A260/A280 ratio were measured with a a NanoDrop ND-1000 (NanoDrop Technologies, Montchanin, DE, USA). Absence of DNA contamination was checked by running samples on 2% agarose gels. 1 μg purified RNA was reverse transcribed. Before transcription, RNA was denatured for 5 min at 65°C followed by cooling on ice. First strand cDNA synthesis was carried out with SuperScript III Reverse Transcriptase (Invitrogen) and random primers (Invitrogen) in a total volume of 20 μl. Reverse transcription was performed at 50°C for 50 minutes followed by 70°C for 15 min. Finally, RNase H was added to the reaction mixture followed by incubation for 20 min at 37°C. cDNA was stored at -80°C until RT-PCR analysis. Each RNA sample was controlled for genomic DNA contamination by a reaction mix without reverse transcriptase addition. All cDNAs were diluted 1:10 before being used as PCR template. qRT-PCR was performed with an Applied Biosystems 7900HT Fast Real-time PCR System.

A set of seven inflammatory cytokines was selected based on a literature search on cytokines present in chronic obstructive pulmonary disease (COPD). The selected cytokines were: interleukin 1-β (IL1-β), interleukin 6 (IL6), inmune protein 10 (IP10), keratinocyte chemoattractant (KC), monocyte chemoattractant protein 1 (MCP1), macrophage inflammatory protein 1 α (MIP1-α), and tumor necrosis factor α (TNF-α). In addition, Beta-2 microglobulin (B2M) was used as an endogenous control for each sample, because it provided homogeneous results at different melting temperatures. The primers for each gene are specified in Table 3.1.

The $2^{-\Delta\Delta CT}$ method was used to quantify relative changes in gene expression between control and treated animals, where:

$$\Delta\Delta CT = (C_{T,Target\ gene} - C_{T,B2M})_{Treated} - (C_{T,Target\ gene} - C_{T,B2M})_{Control} \quad (3.1)$$

We calculated the median and interquartile ranges (IQR) of each parameter obtained with the different techniques, at all time points. The control and elastase-treated groups were compared at each time point using the Mann-Whitney U test. To measure progression, the same test was performed on the elastase-treated animals at successive time points. p values below 0.05 were considered statistically significant and are indicated in the figures with an asterisk (*). The relationship between histomorphometric and micro-CT measurements was analyzed by linear regression. The R language and environment for statistical computing [109] and Open Office Calc software were used for statistical analysis.

3.2.3 RESULTS

3.2.3.1 Micro-CT imaging

Figure 3.9 shows, at four time points, images of representative micro-CT slices and three-dimensional reconstructions obtained from entire micro-CT data sets, together with corresponding histological images. The increase in emphysema extent can be clearly appreciated in the microCT images, and confirmed by histological analysis. Histology, besides emphysema, reveals recruitment of inflammatory cells and the presence of hemorrhage starting 6 hours after elastase aspiration and disappearing after day 7.

Figure 3.10 shows the temporal evolution of MLVI and RVB -900 HU, automatically calculated from the micro-CT images. As can be seen from the short-term experiment (Figure 3.10a), the MLVI values of elastase-treated animals start being significantly lower than the corresponding control values 7 days after treatment. The long-term experiment (Figure 3.10b) reveals that the MLVI values stabilize, while remaining significantly lower than the controls in the last time points (14 and 34 weeks after treatment).

Regarding the RVB -900 HU values, significantly higher values can be detected in the elastase-treated compared to the control animals 6 hours after treatment (Figure 3.10c). The destructive effect of elastase in the lungs -as revealed by the RVB -900 HU values- peaks at 14 weeks and remains constant after that time point (Figure 3.10d). In summary, both

measurements agree in detecting a maximum of lung destruction 14 weeks after treatment, but RVB -900 HU is able to detect changes in the lungs earlier than MLVI.

3.2.3.2 Pulmonary function tests

Lung resistance was not significantly affected in the animals treated with elastase at any time point (data not shown). Figure 3.11 shows the evolution of lung compliance (C) for the short- term (Figure 3.11a) and the long-term (Figure 3.11b) experimental groups. Within the first 24 hours, elastase-treated mice show lower compliance than the controls, implying that elastase administration increases lung stiffness temporarily. At days 7 and 17 no statistically significant difference was found between the groups. In contrast, elastase-treated mice show higher compliance from the fourth week onwards, with an increment in the value depending on time of progression. These data suggest that compliance measurements reveal not only emphysema development, but other processes involved in this particular model.

3.2.3.3 Histological analysis

The mean linear intercept (Lm) and a weighted mean measure of alveolar size (D2) were computed from histological samples to estimate the extent of emphysema. The evolution of both parameters throughout the experimental period is shown in Figure 3.12. From the first experimental hour, both parameters suffer a steady increase, stopping after the first week. Although both parameters yield similar results, D2 is slightly more sensitive to detect differences between the emphysematous mice and the controls, as it has been previously reported [106].

3.2.3.4 Cytokine RNA expression

Cytokine RNA expression was measured as an indicator of inflammation. Figure 3.13a shows the RNA expression of inflammatory cytokines in the elastase-treated mice compared to controls for the short-term experimental group. During the first 24 hours, all the inflammatory cytokines measured in the treated animals display significantly higher expression levels than those measured in control animals. At day 7, only IL-6, MCP-1 and MIP1-α were overexpressed in the elastase-treated group. Interest-

ingly, at day 17, the RNA expression levels of this set of cytokines returned to control levels.

3.2.3.5 Cytokine concentration in plasma

Figure 3.13b shows the protein concentration of the inflammatory cytokines in the plasma of treated and control mice. A marked, statistically significant increase in the plasma concentration of IL-6, IP-10, KC, MCP1 and TNF-α was detected at 6 hours. At 12 hours the plasma levels of these cytokines returned to control levels. The results for the long term experiment are omitted because no differences with the controls were observed at any time point.

3.2.3.6 Micro-CT/Histomorphometry correlation

The best correlation between image-based and histomorphometric emphysema measurements was found between RVB -900 HU and D2, with a coefficient of determination R^2 of 0.63 (Figure 3.14). However, it can be seen that mice with very high RVB -900 HU values do not have correspondingly high D2 values, suggesting that D2 might saturate for very high levels of emphysema. Generally, a tendency towards the upper right side (larger D2, larger RVB -900 HU) can be seen as time progresses.

Finally, Figure 3.15 presents a summary graph containing the four most representative parameters measured: RVB -900 HU from the micro-CT images, D2 from histological section, compliance (C) from the pulmonary function tests and IL6 RNA expression from the cytokine measurements. The evolution of each parameter can be observed.

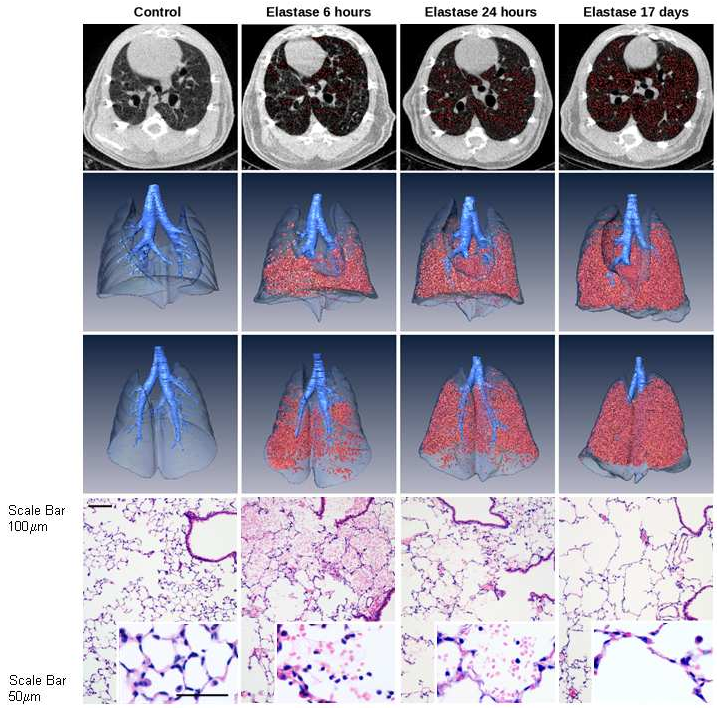


Figure 3.9: *In the upper row, sample transversal micro-CT slices with areas of density lower than -900 HU in red, from different time points. The second and third row show front and back views of three-dimensional reconstructions, with main airways in solid blue, lungs in transparent blue and low density areas in red. The evolution of emphysema can be seen, as the concentration and area covered by the red volume increases (RVB -900 HU is 8.18% at 6 hours, 13.97% at 24 hours and 28.95% at 17 days, in these images). In the lower row, samples of histological sections stained with H& E can be seen. At 6 hours airspace size increases (D2 = 119.19 μm), but there is also inflammation and hemorrhage in the alveolar spaces. The latter reduce significantly at 24 hours, while airspaces keep growing (D2 = 183.67 μm). Finally, at 17 days there is virtually no inflammation visible and airspaces are larger than in previous time points (D2 = 187.42 μm).*

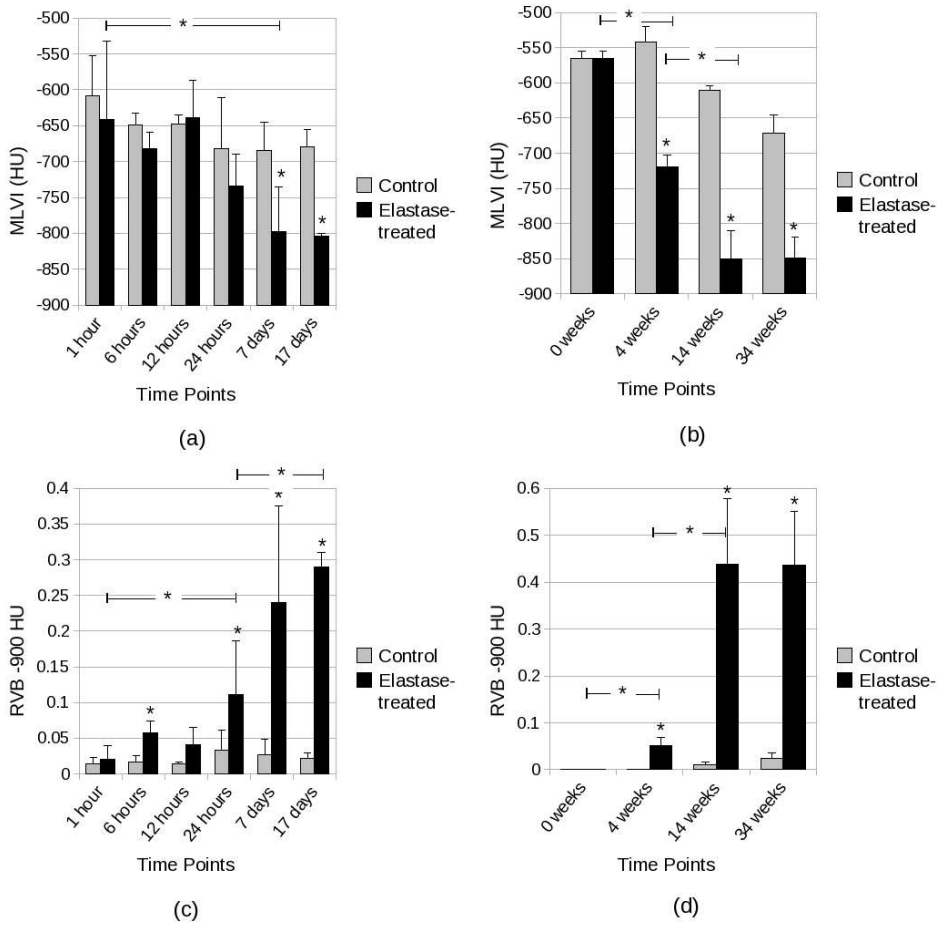


Figure 3.10: *Evolution of automatic measurements on micro-CT images of the mice chest after elastase aspiration. The MLVI for the short time (a) and long time (b) experimental groups and the RVB -900 HU for the same groups (c and d, respectively). Statistically significant differences between the control and the elastase-treated groups are indicated, as well as differences between successive time points in the elastase group ($p < 0.05$, *). First differences in MLVI appear at day 7 days. Differences in RBV -900 HU are evident already 6 hours after treatment and the parameter keeps growing until the 14^{th} week. The large differences in RBV -900 HU between the short time and the long time groups are due to the age and size difference between the mice in the two experiments.*

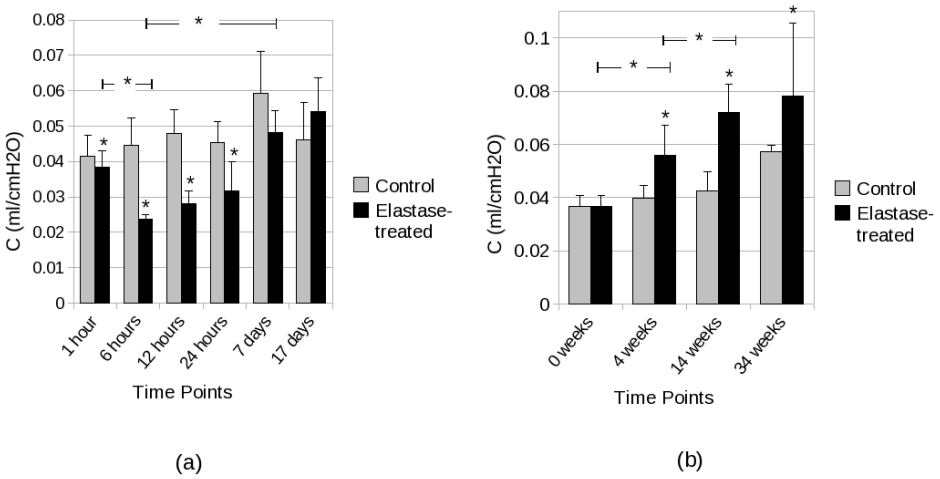


Figure 3.11: *Evolution of lung compliance (C) after elastase aspiration for the short time (a) and long time (b) experimental groups. Statistically significant differences between the control and the elastase-treated groups are indicated, as well as differences between successive time points in the elastase group ($p < 0.05$, *). During the first 24 hours, the elastase treatment causes a reduction in C, which disappears after the first week. From 4 weeks onwards, elastase-treated mice show higher C than controls.*

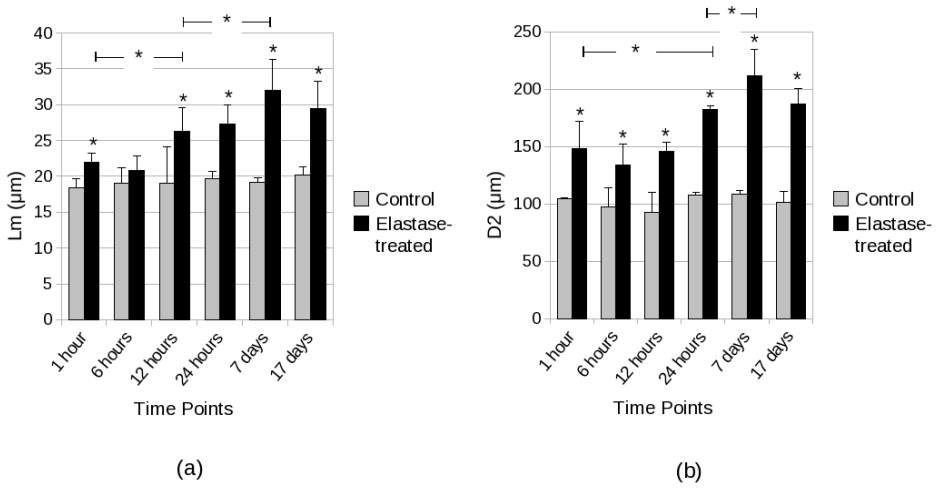


Figure 3.12: *Evolution of Lm (a) and D2 (b) parameters derived from histology at different times after elastase aspiration. Statistically significant differences between the control and the elastase-treated groups are indicated, as well as differences between successive time points in the elastase group ($p < 0.05$, *). Lm shows significant differences at all time points except 6 hours, while for D2 there are significant differences in all time points. A tendency towards growing airspaces through time can be observed, which stops after day 7.*

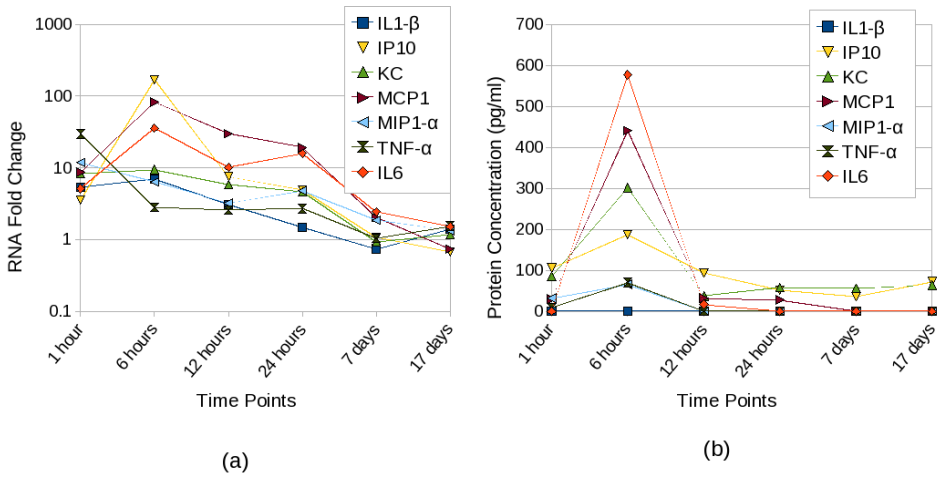


Figure 3.13: *(a) Cytokine RNA expression fold change in the elastase short time group. Controls at paired time points were taken as reference. Each cytokine shows a different time course, but generally a high expression can be seen in the first hours, which decreases to normal levels from 7 to 17 days, depending on the cytokine. (b) Cytokine protein concentration in plasma change in the elastase short time group. High protein levels were detected 6 hours after elastase aspiration. The concentration levels after 24 hours are similar to the control group (data omitted for the sake of clarity).*

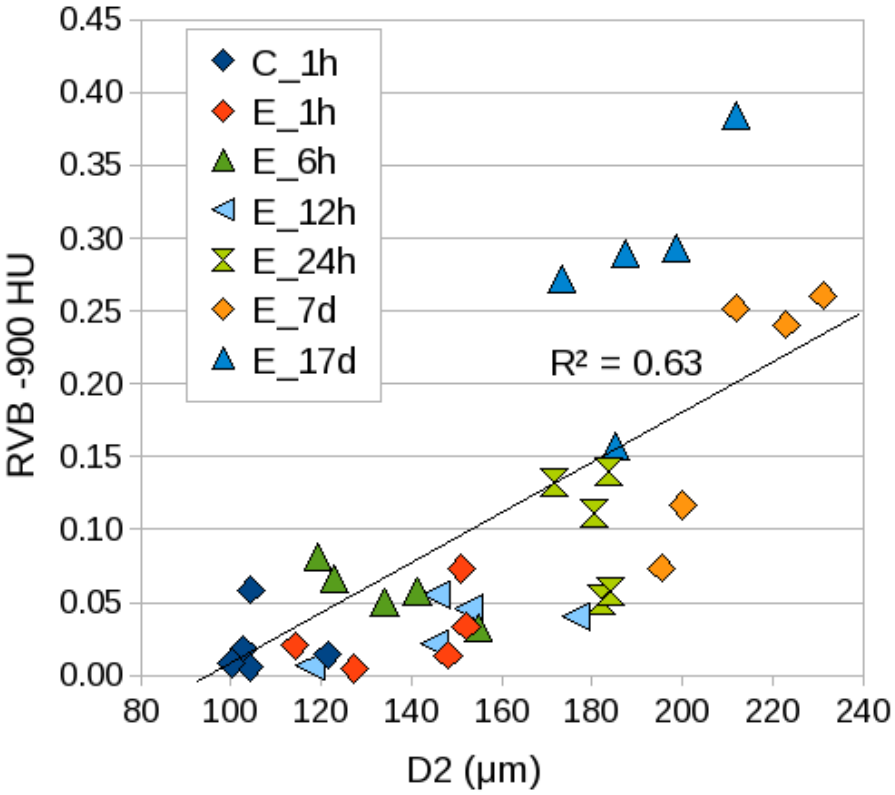


Figure 3.14: *Relative Volume -900 HU measured from micro-CT images plotted versus D2 measurement computed from histological sections. A coefficient of determination R^2 of 0.63 was computed between the two variables. The observations are separated according to their time point. C stands for control and E for elastase, and time points are indicated after the group. Most control mice have been removed for the sake of clarity. A tendency towards the upper right side (larger D2, larger RVB -900 HU) can be seen as time progresses.*

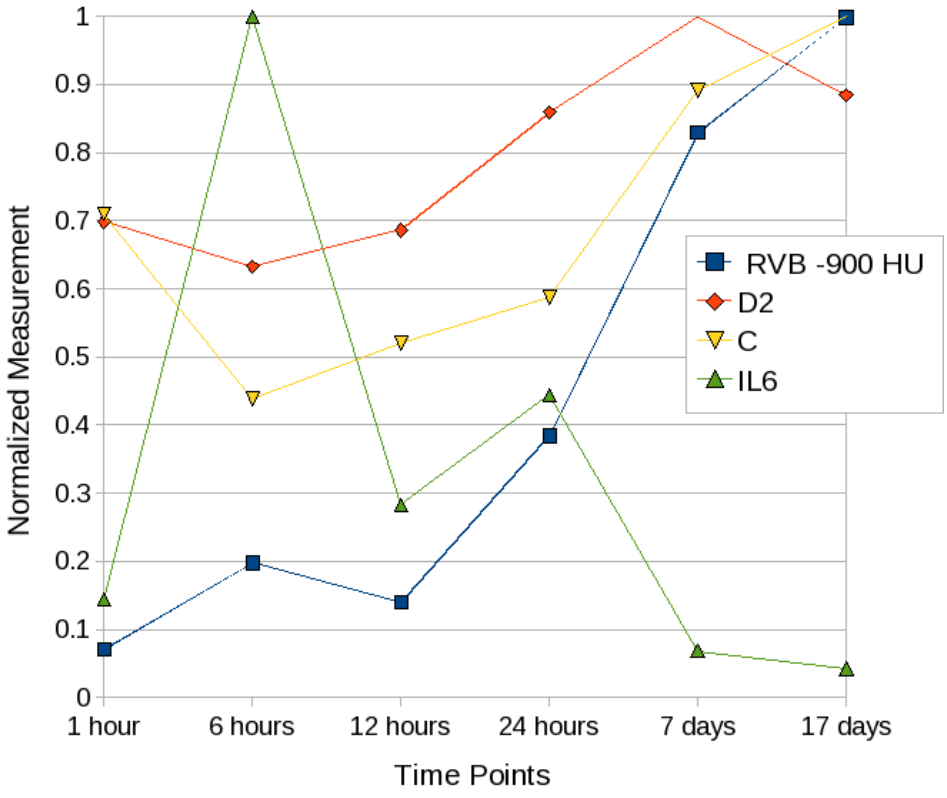


Figure 3.15: *Graph combining the results of a micro-CT measurement (Relative Volume Below -900 HU), a histomorphometric measurement (D2), a lung function measurement (Compliance, C) and a RNA cytokine measurements (Interleukin 6, IL6) for the elastase group. Each measurement has been divided its the maximum value, so that the y axis has a maximum value of 1 for all curves and it is dimensionless. The very different trend of each measurement can be observed.*

3.2.4 DISCUSSION AND CONCLUSION

Our aim was to evaluate and compare the information provided by five methods that can be used to assess emphysema in animal models. To this end, we have performed a detailed multiplatform study with a large animal number at multiple time points, on the elastase-induced emphysema model in mice. By studying our results globally, new light can be shed on the advantages and drawbacks of each technique for emphysema assessment in mice.

Histological analysis is the most used technique to quantify the extent of emphysema in animal models [106]. Recently, several groups have shown the validity of micro-CT for studying these models, showing good correlation between micro-CT image measurements and histology [11,22,110,111]. The main advantage of micro-CT based emphysema quantification, especially if compared to histological analysis, is that animals do not need to be sacrificed, thus allowing longitudinal studies. Moreover, three-dimensional information on the emphysema extent and distribution can be gained with relatively simple image analysis tools.

Compared to previous studies, our histomorphometry values were obtained based on a considerably larger sample size. Furthermore, since we concentrated most of the observation time points in the initial phase of disease development, our results show that histomorphometry is more sensitive than micro-CT at initial stages. This may be due in part to the histologically detected inflammation, which causes important cellular reaction, and thus increased X-ray absorption values in micro-CT images, leading to lower intensity-based emphysema values. Among the micro-CT measurements used, RVB -900 HU proved more sensitive than MLVI in detecting emphysema. It must be noted that previous studies, which did not use breath hold gated image acquisition, employed higher thresholds (around -600 HU) [11, 22, 110]. This means that, as in humans, the thresholds for emphysema quantification in small animals depend on the acquisition protocol and scanner settings.

Regarding the functional tests, different methods exist to obtain lung function parameters in mice, with a general trade-off between invasiveness and accuracy [20]. The forced oscillation techniques can be performed in a relatively short time with commercially available equipment, and provide the most accurate results. The main drawback is the need to perform

tracheotomies or endotracheal intubation. We employed endotracheal intubation, which might reduce measurement accuracy but allows longitudinal studies. As expected, lung resistance (R) did not change in the elastase-treated animals, since the mouse airways which contribute the most to the resistive component of the lungs- are not largely affected by the elastase treatment. However, lung compliance (C) did show remarkable changes in our model. C reflects the ease with which the lungs can be expanded or, in other words, the level of lung stiffness. Our results show that the initial tissue loss does not cause an increase in C. Quite the opposite, C decreases during the first 24 hours. It is only at week 4, when the emphysema extent revealed by histology and micro-CT is already very remarkable, that C starts to increase. Previous studies have underlined that elastase-induced emphysema is associated with an increase in lung compliance [112, 113], and others have pointed at the lack of correlation between emphysema and compliance in other models [114]. However, to the best of our knowledge, this shift from early phase to late phase emphysema has not been yet reported. To explain this shift, we hypothesize that it may be related to the acute inflammatory reaction detected at very early stages: Confirming this point, prominent presence of inflammatory cells and intraalveolar hemorrhage was histologically detected at 6 hours, remained at lower levels at 24 hours, and completely disappeared by day 7. This increased density may compensate for the loss of alveolar walls, thus contributing to a decrease in lung compliance. Accordingly, the C value reaches its minimum at 6 hours, and starts recovering only after 12 hours, when the inflammatory recruitment and hemorrhage levels decrease. Therefore, lung compliance must be carefully used when assessing emphysema in animal models, especially if concomitant inflammation exists.

Inflammatory cytokines can be regarded as markers of inflammatory processes, and they can be measured either at RNA level or at protein level. Measurement of cytokine RNA concentration on RNA extracted from lung samples requires sacrifice. In contrast, blood extracted from living mice can be used to measure cytokine protein concentration in plasma. In agreement with histological observations, our cytokine data reveals an acute inflammatory process starting immediately after treatment and lasting approximately 24 hours. This observation also supports the decrease in lung compliance (C) observed shortly after elastase treatment. A set of 7 cytokines was chosen, which represent a wide variety of inflammatory mechanisms involving different cell types. At days 7 and 17, only very faint signs of inflammation remain. Changes of cytokine concentrations in

plasma were much less evident than in RNA expression in the lungs, suggesting that the elastase-induced emphysema model has a limited systemic effect. To the best of our knowledge, this is the first study in which the cytokine concentration in plasma was measured for this animal model. We confirmed that the elastase-induced model is characterized by a brief initial acute inflammatory response, but shows few or absent signs of chronic inflammation. Therefore, at least in this model, cytokine expression is not directly related to emphysema severity.

In the elastase mouse model, mice develop emphysema very rapidly after elastase instillation or aspiration. Then, according to our data and to previous reports, emphysema development slows down progressively, until practically disappearing after the first month [7]. The model does not require very specific equipment or highly sophisticated expertise, which makes it very attractive for the study of the disease and the evaluation of candidate treatments. Using a diverse set of quantification techniques, we thoroughly characterized this model, adding dynamic, temporal and correlative values that were not addressed in previous studies. Based on our experimental results, direct elastase administration causes a severe alveolar wall breakdown, inducing an acute inflammatory response. The latter may also contribute to the initial protease-antiprotease imbalance caused by elastase. However, once the elastase is cleared from the lungs, which occurs at around 24 hours [115], inflammation clearly decreases, as well as the tissue destruction rate. In this second phase, morphological changes of the lung are progressively attenuated, until alveolar wall breakdown virtually disappears, and we were not able to detect increased cytokine concentration neither in the lungs nor in plasma. Thus, the main difference between the elastase mouse model and the cigarette smoke-induced emphysema in humans may be the early resolution of the acute inflammatory process and the low degree of chronic inflammatory development.

One of the limitations of this study is the distribution of the sampling time points. There is a considerable leap between the 24 hours and the 7 days points, as well as between the 7 days and the 17 days points. The analysis of the biological events taking place on days one to seven after elastase aspiration would have given extra information on each technique, as well as on the disease development. However, instead we chose to cover very thoroughly the first 24 hours, which are very relevant to assess the sensitivity of the quantification techniques that we have studied.

In conclusion, both histomorphometry and micro-CT give reliable measurements of emphysema. However, micro-CT allows longitudinal studies, thus reducing the required number of animals and providing more relevant information on the progression of the disease. The variables automatically computed from micro-CT images can be used as measurements of the extent of emphysema in preclinical trials. Lung compliance and cytokine expression provide different information, which can be related to the presence of emphysema and concomitant inflammation but does not reveal its extent. These measurements can be of interest depending on the animal model and the particular processes under study. Our results confirm that the elastase-induced model of emphysema is characterized by an acute inflammation within the first hours and a subsequent decreasing rate of tissue loss.

3.3 COMBINED MODEL: ELASTASE-INDUCED EMPHYSEMA AND URETHANE-INDUCED CARCINOGENESIS

3.3.1 INTRODUCTION

It has been recently shown that both components of COPD, emphysema and airflow obstruction, are related to lung cancer in high risk populations [116,117]. It seems that patients with COPD are more likely to suffer lung cancer, compared to other patients with similar smoking history and age but without COPD. This observation raises the question of what is the biological link between these two apparently unrelated diseases. In Section 3.2, the inflammatory component of COPD and the potentially resulting protease-antiprotease imbalance was underlined, and the link between inflammation and cancer has been discussed since as early as 1863 [118]. Therefore, the currently most accepted hypothesis is that inflammation triggered by cigarette smoke can be a common origin for both COPD and lung cancer [19]. Although the precise mechanisms are unknown, it has been suggested that bronchoalveolar stem cells (BASCs) involved in tissue repair might be partially responsible for carcinogenesis.

Houghton *et al.* underlined the relevance of animal models to understand the link between COPD and lung cancer [19]. Following that intuition, we have developed and analyzed an animal model which combines emphysema and lung cancer. In particular, we have combined two well-known models into a single model: elastase-induced emphysema and urethane-induced lung cancer. Urethane induces lung hyperplasia, adenoma, and adenocarcinoma in a time-dependent manner in A/J mice [119], and the elastase-induced emphysema model has been analyzed in section 3.2 of this thesis.

We analyzed the hypothesis that emphysema might alter tumorigenesis and tumor growth by comparing the elastase-urethane combined model with the simple urethane model. The technique chosen to study the presence and size of nodules was micro-CT, which has been previously used for this purpose [12, 120].

3.3.2 MATERIAL AND METHODS

A total of 33 A/J mice initially aged 8 weeks were used, divided in two groups: a urethane (simple) model and an elastase-urethane (combined)

model. Animals from the simple model were injected with urethane as explained in section 2.1. The combined model was created by first administering porcine pancreatic elastase by aspiration and injecting urethane one week later (see section 2.1).

In the simple model group, 3 animals were scanned twice, at 14 and 34 weeks after urethane injection. Other 5 animals were scanned at each time point, and sacrificed for histological analysis. In the combined group, 8 animals were scanned and sacrificed at week 14, and 8 at week 34.

The protocol explained in section 2.2 was applied to all animals, to obtain micro-CT images.

Nodules were detected by consensus between a radiologist and a small animal imaging technician. To this end, a modified version of the PointPicker plugin for ImageJ was used. The viewer was set at a lung window for visualization (level -650 HU, window 1500 HU). Measurements of largest diameter in axial slices were performed by a small animal imaging technician, using the ImageJ tool.

The mean and standard deviations of the nodule number and size were computed, and the Mann-Whitney U test was used to compare the two groups. A p-value of 0.05 was set as threshold for statistical significance. The tumor sizes were plotted in histograms, in order to compare their distributions. To further quantify the possible differences in size distributions, a cumulative distribution function (CDF) was estimated for each group at each time point, directly from the data. The CDF represents the probablity that the random variable X (in this case nodule diameter) takes on a value less than or equal to x:

$$F_X(x) = P(X \leq x). \tag{3.2}$$

The R language and environment for statistical computing [109] and GNU Octave [93] were used for the statistical analysis.

3.3.3 RESULTS

At 14 weeks, animals of the combined model had in average 46.8% fewer nodules than those of the simple model. However, at week 34 there were no statistically significant differences between both models. Neither was any

Table 3.2: *Mean and standard deviation of number of nodules and nodule size for the the combined and the simple lung carcinoma models at 14 and 34 weeks after treatment. Statistically significant differences are indicated with an asterisk (*). Note the large standard deviation of the nodule sizes in the combined model at week 34.*

Time Point	Number of Nodules		Nodule Size	
	Urethane	Urethane-Elastase	Urethane	Urethane-Elastase
14 weeks	**19.75* (5.44)**	10.50 (6.32)	13.62 (4.41)	13.17 (4.02)
34 weeks	29.63 (6.50)	24.75 (6.61)	24.04 (8.59)	25.59 (17.42)

difference in the mean nodule size at week 14 nor at week 34. However, the sizes of the nodules in the combined model at week 34 showed considerably larger variability than the simple model (standard deviation 17.42 vs. 8.59). These results are shown in Table 3.2.

The distribution of the nodule sizes at 14 and 34 weeks is given in Figure 3.16. At week 14 no difference can be seen, but at week 34 a slight tendency towards largest nodules in the combined model can be detected. This can be also seen in Figure 3.17, where the CDF for the two groups at 14 and 34 weeks is plotted. At week 14 no clear tendency can be seen, but at 34 weeks the combined model displays a slower increase of the CDF curve for large diameters.

3.3.4 DISCUSSION AND CONCLUSION

The clinical observation that COPD and lung cancer are related arises the question of what is the basis of this relationship. In this preliminary study, we compared the nodule number and size in two different animal models: a model of urethane-induced lung cancer and a combined model, adding elastase-induced emphysema to the simple model.

Regarding the number of nodules, the fact that the simple model contained fewer nodules than the combined model at 14 weeks was an unexpected outcome. A possible explanation is the reduced lung tissue density in the combined model, caused by emphysema. If we assume that the probablity of developing a nodule is proportional to the tissue volume that can be affected, a reduced tissue volume might be responsible for a reduced

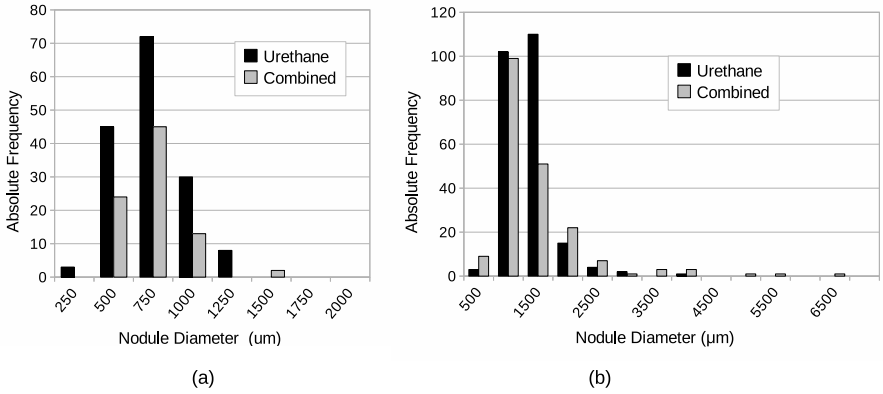


Figure 3.16: *Histograms of nodule diameters for the urethane and combined groups 14 weeks (a) and 34 weeks (b) after initiation. Generally, there are fewer nodules in the combined model. At 14 weeks the distributions are comparable, but at 34 weeks a tendency towards large nodules can be seen in the combined model.*

nodule number. Even though at 34 weeks the difference is not statistically significant, in average animals in the simple model also have more nodules in the lungs.

The nodule size in the two groups was similar at week 14. However, at week 34 a slight tendency was observed towards large nodules in the combined model. Again, this result was not completely expected. Urethane is known to produce hyperplasia detectable in histological sections around 4 weeks after injection. The most extended hypothesis being that it is the inflammatory component of COPD which promotes tumorigenesis [19], and knowing that this emphysema model causes an acute inflammatory response that decreases in time and virtually disappears after the first week (see Section 3.2), we expected to detect differences between the simple and the combined model at earlier times.

We hypothesize that emphysema imposes an initial burden to nodule initiation, probably due to the reduced tissue volume. However, once a nodules appears, it is likely to grow at a faster rate than those caused by urethane only. It is still unclear why this happens, a possibility being that

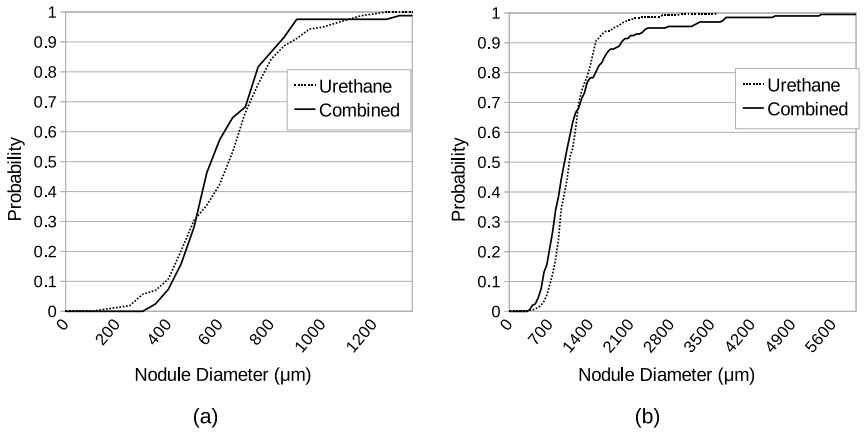


Figure 3.17: *Cumulative density distribution for the urethane and combined groups 14 weeks (a) and 34 weeks (b) after initiation. CDF-s are comparable at week 14, but at 34 weeks the combined model has a higher concentration of nodules larger than 1500 µm.*

the inflammation involved in emphysema may generate some advantages for nodule growth. However, in section 3.2 it was shown that the inflammatory component of this emphysema model is very low. Another possibility is that an emphysematous environment, with less tissue density and degraded cellular matrix, might be beneficial to tumor growth.

Our observations were limited to nodule number and size as detected on a reduced number of micro-CT images. Therefore, care must be taken when deriving conclusions from the data. Before concluding that emphysema increases the probability of developing large nodules, the histological nature of those large nodules has to be studied, to determine whether they indeed correspond to different nodule types. Moreover, detected differences were rather slight, and new experiments would be important to confirm the results.

The chosen animal model combines lung cancer and emphysema. We started by elastase aspiration, and a carcinogen (urethane) was injected one week after. As the acute inflammation caused by elastase aspiration disappears in a few days (see Section 3.2), studying whether a different order

or timing in the process would change the experiment outcome would be of great interest. In particular, it would be interesting to see whether aspirating elastase after urethane injection, or administering smaller elastase doses at regular time intervals would result in a different nodule phenotype. Other models of COPD and lung cancer should also be considered. Previous experiments combining urethane and silica-induced chronic inflammation did not have any influence on tumorigenesis [121], but other models, such as cigarette smoke-induced COPD, might be more appropiate.

In further studies, modifying or adding observation time points should also be considered. Extra nodule measurements between the 14^{th} and the 34^{th} weeks would provide more information on nodule growth rate, which is a key parameter to define their nature. With information on growth rate, more directed histological or genetical studies could be performed.

In summary, we have studied a simple model of lung cancer and a combined model of emphysema and cancer. A slightly increased probability of developing large tumors in the combined model was observed. Further experiments to confirm and interpret the results are required.

The most exciting phrase to hear in science, the one that heralds new discoveries, is not 'Eureka!', but 'That's funny...'

Isaac Asimov, writer and professor of biochemistry

4

Conclusions

In this chapter, the main implications of the thesis are summarized.

COPD and lung cancer are among the deadliest diseases in the developed world. Animal models of both diseases have been developed, which are very useful to study their pathological mechanisms and potential treatments. In this thesis, a set of methods have been developed for the analysis of mouse models of lung disease. These methods have been validated and applied to the characterization of several pulmonary disease models.

A protocol was devised for micro-CT and pulmonary function test data acquisition. The protocol provides high quality images with very high animal survival rate. Advanced image analysis algorithms were also developed, validated and applied to data analysis. In particular, algorithms for airway and lung segmentation and measurement were implemented, and tested in animals suffering from different diseases. Relevant contributions were also made in the field of atlas-based segmentation, which has applications in multiple medical imaging modalities and segmentation problems. Although only basic atlas-based methods were applied in our experiments with mice,

results on other image modalities suggest that specific multi-atlas methods might also be useful.

The methods developed were applied to characterize different models of lung disease. In a silica-induced model of chronic pulmonary inflammation, the information provided by pulmonary function tests and micro-CT image analysis was compared and contrasted with histology. We showed that using breath hold gated micro-CT and automated image quantification, good correlation with histomorphometry can be obtained. Moreover, high animal survival rates were achieved. This implies that longitudinal studies can be performed, which reduce the number of required animals by providing dynamic information on disease evolution. It was also concluded that, in this animal model, micro-CT based analysis detects pathological changes with higher sensitivity than tests of pulmonary function.

The widely used elastase-induced emphysema model was thoroughly studied. Five different techniques were applied on a large number of animals: micro-CT, histomorphometry, pulmonary function tests, RNA cytokine expression and cytokine concentration in plasma. The information provided by each method was evaluated and compared. We concluded that both histomorphometry and micro-CT give reliable measurements of emphysema. However, micro-CT allows longitudinal studies, thus reducing the required number of animals and providing more relevant information on the progression of the disease. Lung compliance and cytokine expression provide different information, which can be related to the presence of emphysema but does not necessarily correlate with its extent. These measurements can be of interest depending on the animal model and the particular processes under study. Our results confirm that elastase-induced emphysema develops rapidly at the beginning along with an acute inflammation, and both disease progress and inflammation are reduced afterwards.

Finally, a mouse model combining emphysema and lung cancer was analyzed and compared to a model of lung cancer without emphysema. We observed a tendency towards larger nodules in the combined model, suggesting that emphysema might increase the probability of having large tumors. However, this observation needs to be confirmed in further studies. These studies should also address the mechanisms through which emphysema and lung cancer are related. Our detailed characterization of the elastase model should be considered when designing new models to test different hypotheses.

In this thesis, methods have been devised and applied for both data acquisition and analysis in mouse models of lung disease. In future work, data acquisition methods might be improved by the use of more modern technology. For instance, most recent fast flat panel micro-CT allows for image acquisition in few seconds, which would ease the process considerably. Regarding the image analysis, the methods presented in this thesis will be of use for basic disease progress assessment in pre-clinical studies. However, more detailed analysis, such as emphysema or nodule localization studies, will require additional tools.

In summary, several methods have been developed and used to study different mouse models of lung disease. Automated analysis of micro-CT images, combined with other techniques, allowed for an extensive characterization of the models. Developed techniques and model characterizations will be key in further studies, involving drug testing or different animal models.

A

Spatially Variant Convolution with Scaled B-splines

We present an efficient algorithm to compute multidimensional spatially variant convolutions -or inner products- between N-dimensional signals and B-splines -or their derivatives- of any order and arbitrary sizes. The multidimensional B-splines are computed as tensor products of $1D$ *B-splines, and the input signal is expressed in a B-spline basis. The convolution is then computed by using an adequate combination of integration and scaled finite differences as to have, for moderate and large scale values, a computational complexity that does not depend on the scaling factor. To show in practice the benefit of using our spatially variant convolution approach, we present an adaptive noise filter that adjusts the kernel size to the local image characteristics and a high sensitivity local ridge detector.*

A.1 INTRODUCTION

The convolution of an image with scaled Gaussian kernels -or their derivatives- is a basic operation in many image processing tasks such as smoothing or feature detection using rotating matched filters [122–125].

Specifically, we are interested in efficiently computing convolutions with kernels whose size vary locally within the image. The common drawback of this type of algorithms is a computational complexity proportional to the size of the convolution kernel, which can be quite high on average.

Unser *et al.* [126] were the first to prove the convergence of B-splines to Gaussians. Following their seminal work, and inspired by the work of Heckbert *et al.* [127], we previously presented a 1D algorithm that computes spatially variant convolutions between spline signals and an approximation of scaled Gaussian kernels based on B-splines [128]. The algorithm is based on a generalized two-scale relation that expresses the equivalence between an arbitrarily expanded B-spline and its unit-scaled counterpart. The algorithm performs the integration of the B-spline coefficients using running sums followed by an inner product with a pre-calculated finite difference filter kernel. The values and locations of the filter kernel depend on the scale value while the number of taps remains constant. This results in an algorithm that has a fixed computational cost per pixel, regardless of the scale. Based on the same principles, Chaudhury *et al.* [129] approximated arbitrarily rotated 2D Gaussian windows as tensor products of four radially uniform zero-degree splines.

In this chapter, we extend to ND the 1D spatially variant, adaptive filter presented in [128]. We choose to approximate the anisotropic Gaussian windows as tensor products of 1D B-splines aligned to the main axes. This way, our algorithm results in a simpler implementation than Chaudhury's, and can be explicitly and straightforwardly extended to ND. Furthermore, since the components of the scale vector are of arbitrary size, our method adapts well to the anisotropy of the sampling rate (e.g., the z-dimension in confocal microscopy data or the temporal dimension in any time-lapse application). In the implementation of the ND algorithm, we take into account that i. the integration of the B-spline coefficients is a separable operation and ii. the subsequent filtering with a FIR differentiation mask cannot be performed in a separable fashion when the kernel is spatially variant.

We illustrate the use of the algorithm with two examples of application: an edge preserving smoothing filter and a ridge detector. Our spatially variant, adaptive filter performs image denoising by convolving with a kernel whose size matches the underlying local image characteristics. Wide kernels are used in smooth regions and narrow ones in edges and highly textured

areas. The filter was first evaluated quantitatively on phantoms and then qualitatively on real image data. Our results improved those obtained using fixed scale B-spline filtering or anisotropic diffusion.

The second example is a ridge detector. Since our algorithm can approximate derivatives of isotropic Gaussians, we use it as a framework to construct steerable filters of the class introduced by Freeman and Adelson [130]. Then, by imposing the flatness criterion introduced in Meijering *et al.* [131], we design a B-spline ridge detector, which is more sensitive to elongated image structures than the standard template (i.e., the second derivative of a B-spline). Moreover, we present a novel, truly spatially variant ridge detector based on a local optimization, which performs comparable to a less efficient classical ridge detector, using the same scale range and resolution. Finally, we show a qualitative example of filament detection on a real fluorescent microscope image.

The chapter is organized as follows. In Section A.2, we review the linear operators and B-spline related expressions that are useful for our work. We start Section A.3 with the derivation of the convolution algorithm of 1D signals with B-splines (or their derivatives). We end the section presenting the extension of the algorithm to N-dimensions. Section A.4 deals with the description of an efficient implementation. In Section A.5, we analyze the propagation of round-off errors and propose a number of error minimization strategies. Section A.6 presents the construction of B-spline ridge detector. Two examples of application are given in Section A.7: Feature adaptive denoising and ridge detection. We end in Section A.8 with some concluding remarks.

A.2 OPERATORS AND DEFINITIONS

In this section, we define operators and present some B-spline expressions that will be useful for the derivation of the ND spatially variant algorithm. As we are working in a Cartesian lattice, we can use tensor products to implement the extension of operators and expressions from 1 to N-dimensions. In this section, we neglect boundary condition issues by assuming signals $f(x), x \in \mathbb{R}$ and discrete sequences s_k, $k \in \mathbb{Z}$ that are specified over the entire line.

A.2.1 FINITE DIFFERENCE OPERATORS

The forward finite difference operator at scale $a \in \mathbb{R}_+$ is defined as $\Delta_a^1(x) = \delta(x) - \delta(x-a)$. In the Fourier domain, we have $\hat{\Delta}_a^1(e^{ja\omega}) = a^{-1}(1 - e^{-ja\omega})$. The n-order finite difference is $\Delta_a^{n+1}(x) = a^{-(n+1)} \sum_{k=0}^{n+1} q(k)\delta(x - ak)$ where $q(k) = \binom{n+1}{k}(-1)^k$ are the modulated binomial coefficients and δ is the Dirac's delta distribution.

The inverse finite difference operator is defined as $\Delta_1^{-1}(x) = \sum_{n\geq 0} \delta(x - n)$. In the Fourier domain, we have $\hat{\Delta}_1^{-1}(e^{j\omega}) = (1 - e^{-j\omega})^{-1}$.

Note that by definition, $\Delta_1^{-(n+1)} * \Delta_1^{n+1} * s = \Delta_1^{n+1} * \Delta_1^{-(n+1)} * s = s, \forall n$.

A.2.2 GENERALIZED TWO SCALE RELATION

The centered B-spline $\beta^n(x/a) = \beta_a^n(x)$ of degree n, at scale $a \in \mathbb{R}_+$ can be related to its unit-scaled counterpart as

$$\beta_a^n(x) = \Delta_1^{-(n+1)}(x) * \Delta_a^{n+1}(x) * \beta_1^n(x + \tau) \tag{A.1}$$

with $\tau = \frac{(a-1)(n+1)}{2}$. See [128] for the proof. Note that this is a generalized version of the two-scale relation encountered in wavelet theory.

A.2.3 DERIVATIVES OF THE SCALED CENTERED B-SPLINE

The derivative of order $d \leq (n+1) \in \mathbb{N}$ of the centered B-spline of degree n and scale $a \in \mathbb{R}_+$ is [122]

$$\frac{d\beta_a^n(x)}{dx^d} = \Delta_a^d(x) * \beta_a^{n-d}\left(x + \frac{ad}{2}\right) \tag{A.2}$$

$$= \Delta_1^{-(n+1-d)}(x) * \Delta_a^{n+1}(x) * \beta_1^{n-d}\left(x + \tau'\right). \tag{A.3}$$

with $\tau' = \frac{(a-1)(n+1)+d}{2}$ with $\beta_1^{-1}(x) = \delta(x)$. See [132] for the proof. To be able to generalize the derivations, we introduce the notation $\beta_{a,d}^n(x)$ to denote the d-derivative of the B-spline $\beta_a^n(x)$. Note that $\beta_{a,0}^n(x) = \beta_a^n(x)$.

A.2.4 SPLINE INTERPOLATION

We assume that the continuous input signal $f_i(x)$ is a spline that interpolates the discrete input samples $f_i(k)$. We can write

$$f_i(x) = (c * \beta_1^m)(x) \tag{A.4}$$

where $c_k = f_i * (b^m)^{-1}$ and $\mathtt{b}^m = \beta_1^m(x)|_{x=k}$ as shown elsewhere [122].

In the rest of the chapter and for the sake of simplicity the unit-scale index will drop off from the finite difference operators and B-splines. Namely, $\beta_1^n = \beta^n$, $\Delta_1^{n+1} = \Delta^{n+1}$ and $\Delta_1^{-(n+1)} = \Delta^{-(n+1)}$.

A.3 DERIVATION OF THE ALGORITHM

A.3.1 1D CONVOLUTION

The discrete output $f_0(b)$ of the convolution of the input function $f_i(x), x \in \mathbb{R}$ with a scaled B-spline (or its d-derivative) $\beta_{a,d}^n$ evaluated at position $b \in \mathbb{Z}$ is

$$f_0(b) = (f_i * \beta_{a,d}^n)(b). \tag{A.5}$$

The expression above can be written as in [128]

$$f_0(b) = (g * \zeta_a)(b) = \sum_{l \in \mathbb{Z}} g(b-l)\zeta_a(l) \tag{A.6}$$

where

$$g = \Delta^{-(n+1-d)} * c \tag{A.7}$$

is an integrated signal and

$$\zeta_a(l) = (\Delta_a^{n+1}(\cdot) * \beta^{m+n+1-d}(\cdot + \tau'))(l) \tag{A.8}$$

is the 1D filter kernel. Note that ζ_a corresponds to the sampled $(n+1)$-finite differences of a shifted B-spline of degree $(m+n+1-d)$. Due to the finite support of the B-splines, the size of the filter kernel is finite. Namely, the filter kernel size is

$$M = m + n + 1 - d + \lceil a(n+1) \rceil. \tag{A.9}$$

Proposition A.3.1. *The convolution Eq. (A.6) can be written as an inner product of the integrated signal g with a filter kernel w_a of scale-dependent size M*

$$f_0(b) = \sum_{k=0}^{M-1} g(b+k+k_0) w_a(k) \tag{A.10}$$

where $b+k_0$ is the first significative index of g for each filter kernel with

$$k_0 = -\lfloor ((a+2)(n+1) + m + 1 - 2d)/2 \rfloor \tag{A.11}$$

and

$$w_a(k) = a^{-(n+1)} \sum_{r=0}^{n+1} q(r) \beta^{m+n+1-d}(-ar - k - k_0 + \tau'). \tag{A.12}$$

The proof of the above proposition follows:

Proof. Inner-Product using a Filter Kernel of Scale-Dependent Size

The support of the filter kernel ζ_a is $[m_L, m_U] = [-(m+n+2-d)/2 - \tau', (m+n+2-d)/2 + a(n+1) - \tau']$. We take advantage of the finite filter kernel size to reduce the number of terms on the sum over l. We make the following change of variable $b - l = k + k_0$, where $k \in [k_L, ..., k_U] = [0, ..., M-1]$, to write the inner product

$$f_0(b) = \sum_{k=0}^{M-1} g(k+k_0) w_{a,b}(k) \tag{A.13}$$

where $k_0 = \lceil b - a(n+1) - (m + +2 - d)/2 + \tau' \rceil$ is the solution of $m_L = b - k_U - k_0$ ($m_U = b - k_L - k_0$) and $w_{a,b}(k) = \zeta_{a,b}(b - k - k_0) = a^{-(n+1)} \sum_{r=0}^{n+1} q(r) \beta^{m+n+1-d}(b - ar - k - k_0 + \tau')$.

In practice, we are only interested in values of b that correspond to the discrete locations of the original samples. Then, we can use the fact that $w_{a,b}(k) = w_{a,0}(k-b) = w_a(k-b)$ if $b \in \mathbb{Z}$, i.e., the value of the weights does not depend on the position b. Therefore, the algorithm is equivalent to a discrete convolution. The expression (A.13) can then be re-written as eq. (A.10). □

Proposition A.3.2. *Alternatively, the convolution Eq. (A.6) can be written as an inner product of the integrated signal g with a filter kernel χ_a of scale-independent size [128]*

$$f_0(b) = \sum_{r=0}^{n+1} \sum_{p=0}^{m+n-d} g(b+p+p_0)\chi_a(r,p) \tag{A.14}$$

where $b+p_0$ is the first index of g for each cluster with

$$p_0 = -\lfloor (2ar - (a-2)(n+1) + (m+1-2d))/2 \rfloor \tag{A.15}$$

and

$$\chi_a(r,p) = a^{-(n+1)} q(r) \beta^{m+n+1-d}(-ar - p - p_0 + \tau'). \tag{A.16}$$

The number of weights in the filter kernel χ_a is $(n+2)$ in the first dimension (the number of coefficients of the finite differences Δ_a^{n+1}) and $(m+n+1-d)$ in the second dimension (the support of the B-spline $\beta^{m+n+1-d}$). This takes into account that the last B-spline sample is always zero. The proof of the proposition is shown below:

Proof. Inner Product using a Filter Kernel of Scale-Independent Size From (A.6) and (A.8) we can derive the following expression $f_0(b) = \sum_{r=0}^{n+1} \sum_{l \in \mathbb{Z}} g(b-l)\chi_{a,b}(r,l)$ where the 2D filter kernel $\chi_{a,b}(r,l)$ is defined as $\chi_{a,b}(r,l) = a^{-(n+1)} q(r) \beta^{m+n+1-d}(b - ar + l + \tau')$.

If we do the change of variable $l = -(p+p_0)$ and take into account the fact that $\chi_{a,b}(r,l) = \chi_a(r, l-b)$, if $b \in \mathbb{Z}$, we have (A.14). □

A.3.2 ND CONVOLUTION

Now, we would like to calculate the ND convolution of the input spline $f_i(\mathbf{x}), \mathbf{x} = (x_1, ..., x_N) \in \mathbb{R}^N$ with a scaled B-spline kernel (or one of its derivatives) $\beta^n_{\mathbf{a},\mathbf{d}}(\mathbf{x}) = \prod_{i=1}^N \beta^n_{a_i,d_i}(x_i)$, being $\mathbf{a} = (a_1, ..., a_N) \in \mathbb{R}^N_+$ and $\mathbf{d} = (d_1, ..., d_N) \in \mathbb{N}^N$. The discrete output $f_0(\mathbf{b})$ evaluated at integer positions $\mathbf{b} = (b_1, ..., b_N) \in \mathbb{Z}^N$ can be expressed as

$$f_0(\mathbf{b}) = \left(f_i * \beta^n_{\mathbf{a},\mathbf{d}}\right)(\mathbf{b}). \tag{A.17}$$

Proposition A.3.3. *The convolution in eq. (A.17) can be written as the inner product of the integrated signal g with a filter kernel $w_{\mathbf{a}}$ of scale dependent size $\mathbf{M}$ as follows*

$$f_0(\mathbf{b}) = \sum_{\mathbf{k}=\mathbf{o}}^{\mathbf{M}-1} g(\mathbf{b}+\mathbf{k}+\mathbf{k}_0) w_{\mathbf{a}}(\mathbf{k}) \tag{A.18}$$

where

- *The matrix g contains the ND integrated spline coefficients where each $1D$ component is given by (A.7).*
- *Each index $k_{i,0}, i = 1, ..., N$, of the vector $\mathbf{k}_0 = (k_{10}, .., k_{N0})$, is given as (A.11).*
- *$\mathbf{o} = (0, \cdots, 0)$ is an all-null vector of length N.*
- *The size of the filter kernel is $\mathbf{M} = M_1 \times M_2 \times \cdots \times M_N$ with $M_i, i = 1, ..., N$, as in (A.9).*
- *The filter kernel of scale-dependent size is calculated as $w_{\mathbf{a}}(\mathbf{k}) = \prod_{i=1}^{N} w_{a_i}(k_i)$ with w_{a_i} defined as in (A.12).*

Proposition A.3.4. *Alternatively, the convolution in eq. (A.17) can be written as an inner product of the integrated signal g with a filter kernel $\chi_{\mathbf{a}}$ of scale-independent size as follows*

$$f_0(\mathbf{b}) = \sum_{\mathbf{r}=\mathbf{o}}^{\mathbf{u}} \sum_{\mathbf{p}=\mathbf{o}}^{\mathbf{u}'} g(\mathbf{b}+\mathbf{p}+\mathbf{p}_0) \chi_{\mathbf{a}}(\mathbf{r}, \mathbf{p}) \tag{A.19}$$

where

- *Each index $p_{0i}, i = 1, ..., N$, of the vector $\mathbf{p}_0 = (p_{01}, ..., p_{0N})$ is given as (A.15).*
- *$\mathbf{u} = (n+1, \cdots, n+1)$ is an all $(n+1)$ vector of length N.*
- *$\mathbf{u}' = (m+n-d_1, ..., m+n-d_N)$ is a vector of length N.*
- *The filter kernel of scale independent size is calculated as $\chi_{\mathbf{a}}(\mathbf{r}, \mathbf{p}) = \prod_{i=1}^{N} \chi_{a_i}(r_i, p_i)$ with χ_{a_i} defined as (A.16).*

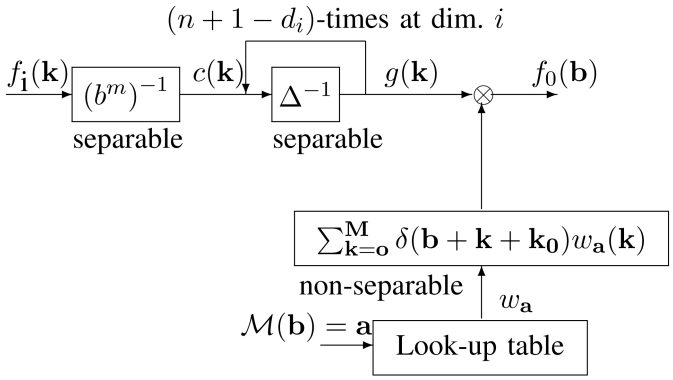


Figure A.1: *Schematic representation of the algorithm that efficiently calculates N D spatially variant convolutions (or inner products) with scaled B-splines (or their derivatives). $\mathcal{M}$ is the scale image that at each pixel* **b** *stores the appropriate scale vector* **a**.

A.4 EFFICIENT IMPLEMENTATION

We now present the implementation of the convolution of the input $f_i(\mathbf{b})$ with a scaled B-spline $\beta^n_{\mathbf{a},\mathbf{d}}(\mathbf{x})$. Figure A.1 shows a schematic representation of the algorithm.

A.4.1 FILTER KERNELS

As explained in Section A.3.B, the ND filter kernels ($w_{\mathbf{a}}$ and $\chi_{\mathbf{a}}$) are efficiently computed as a separable tensor product of the 1D components (w_{a_1} and $\chi_{a_i}, i = 1, ..., N$). In consequence, the ND filter kernels inherit all the convenient properties of their 1D components. Most importantly, the weights of the ND filter kernels are independent of the image to be filtered. Therefore, they can be pre-calculated for each scale and stored in a look-up table which is valid for any image [128]. The ND filter kernel $w_{\mathbf{a}}$ is the spatially explicit version of $\chi_{\mathbf{a}}$. The size of $w_{\mathbf{a}}$ is $(m+n+1-d_i+\lceil a_i(n+1) \rceil)$ per dimension i. The maximum number of non-null weights of $w_{\mathbf{a}}$ equals $(n+2)(m+n+1-d_i)$ per dimension i. The ND filter kernel $\chi_{\mathbf{a}}$ has the

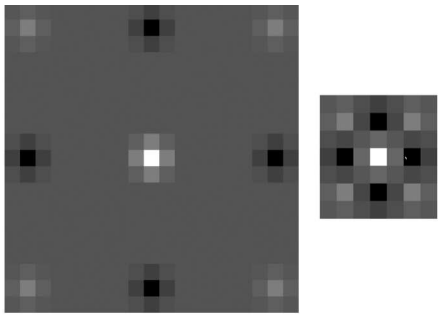


Figure A.2: *2D filter kernel examples for $m = n = 1$ and scale values $\mathbf{a} = (8, 8)$ (Left) and $\mathbf{a} = (2, 2)$ (Right). The filter kernel size per dimension is 19 and 7, respectively. The frequencies of non-null weights are 22.44% and 100%, respectively. The background grey value corresponds to zero, darker grey to negative values and lighter grey to positive values.*

structure of an *a trous* filter [128]. In particular, smoothing is done by filtering the coefficients g with $(n+2)$ clusters of size $(m+n+1-d_i)$ each cluster being separated from its neighbors a distance a_i at each dimension. When the scale a_i is smaller than the cluster size $a_i < (m+n+1-d_i)$ in any of the dimensions, the clusters overlap, and the most efficient approach to compute $f_0(\mathbf{b})$ is to calculate the convolution with the compact filter $w_{\mathbf{a}}$. When filtering at locations with moderate or large scale values, $a_i > m+n+1-d_i, \forall i$, an efficient algorithm will only take into account the non-null filter weights $\chi_{\mathbf{a}}$ at the appropriated coefficient positions. Figure A.2 (left) shows an example of a filter kernel $w_{\mathbf{a}}$ for $m = n = 1$ and $\mathbf{a} = (8, 8)$. The filter kernel size is 361 and the number of non-null weights (which form $\chi_{\mathbf{a}}$) is 81. The reduction in computational complexity when filtering with $\chi_{\mathbf{a}}$ is in this case 77.56 %. See Figure A.2 (right) for an example of a compact filter kernel $w_{\mathbf{a}}$ for $m = n = 1$ and $\mathbf{a} = (2, 2)$.

A.4.2 GLOBAL PREPROCESSING

The spline coefficients and the running sums are calculated over the entire signal at once, with a significant saving in computational complexity

over a point-wise implementation. This is possible because the outputs (c and g) do not depend on the scale vector $\mathbf{a}$ [128].

A.4.2.1 Calculation of the Spline Coefficients

To calculate the spline coefficients c over the entire image, we apply the 1D IIR inverse interpolation filter $(b^m)^{-1}$ successively to each dimension of the data. The filter is implemented as described in [122].

A.4.2.2 Computation of the Inverse Finite Differences

The running sum operator Δ_1^{-1} is applied $(n+1-d_i)$-times successively to each i-dimension of the data. Note that the operator Δ^{-1} can be defined as the running sum filter $y_k = (\Delta^{-1}s)_k = \sum_{n \leq k} s_n$ that can be implemented recursively [132].

A.4.3 SPATIALLY DEPENDENT FILTERING

The output of the filter $f_0(\mathbf{b})$ at each position $\mathbf{b}$ is calculated by non-separable ND filtering with the most appropriate set of weights, $w_{\mathbf{a}}$ or $\chi_{\mathbf{a}}$. The finite difference filter kernel depends on the value of the scale vector $\mathbf{a}$, which can be different at each point. A different filter kernel needs to be applied at every point. In other words, the operation is non-separable.

Note that if a unique scale vector $\mathbf{a}$ is used, the finite difference filter kernel will be the same for the whole image and the scheme reduces to a spatially-invariant filtering. In that case the algorithm can be implemented in a separable fashion with the corresponding reduction in computational complexity.

A.4.4 BOUNDARY CONDITIONS

We extend our 1D signal $\{s_k\}_{k=0,...,L-1}$ using symmetric mirror boundary conditions defined as $s_{-k} = s_k$ and $s_{L-1-k} = s_{L-1+k}$, for $k = 0, ..., L-1$. As proposed in [132], we work with zero-mean signals, i.e., $s = s' - T$. This way we implement the running-sums -inverse finite difference operator- in a way that ensures consistent propagation of the boundary conditions. The mean T is calculated over the period of the extended signal $T = (2L - 2)^{-1} \left(2\sum_{k=0}^{L-1} s'_k - s'_0 - s'_{L-1}\right)$. Thus, assuming that s is zero mean, the

inverse finite difference operator is given by $\left(\Delta_1^{-1} * s\right)_k = \sum_{l=0}^{k \bmod (2L-2)} s_l$. For zero mean signals, the application of the above operator reverses the symmetry of the boundary conditions, as was proven in [132]. That is to say, if the input signal has symmetric boundary conditions, the output will have anti-symmetric boundary conditions defined as $s_k = -s_{-k-1}$ and $s_{L-1+k} = -s_{L-2-k}$, for $k = 0, 1, ..., L-1$. Our implementation of the spatially dependent filtering considers the appropriate boundary conditions resulting from the computation of the running sums.

A.4.5 NORMALIZATION

A.4.5.1 1D Normalization

Since the mean T was subtracted before the inverse finite difference filtering, it must be restored at the final step of the algorithm. Nevertheless, to get the true output f_0 at a given point b, we have to multiply the previously subtracted mean T by a normalization factor F_a that corrects for the intensity scaling caused by the computation of the inner product with the filter kernel, for $a \notin \mathbb{Z}$. The normalization factor F_a is computed as:

- *Inner product using a filter kernel of scale-dependent size*: $F_a = \sum_{l=0}^{M-1}(k+k_0)(l)^{n+1-d}/(n+1-d)! \cdot w_a(l)$.
- *Inner product using a filter kernel of scale-independent size*: $F_a = \sum_{r=0}^{n+1}\sum_{l=0}^{m+n-d}(p+p_0)(r,l)^{n+1-d}/(n+1-d)! \cdot \chi_a(r,l)$.

The normalization factor can be pre-calculated at the same time as the finite difference filter kernel and stored in a look-up table.

As an example, let us consider the input signal to be a delta $f_i(b) = \delta(b)$ for $b = -3:3$, interpolated with a delta basis ($m = -1$) and convolved with the B-spline of degree zero ($n = 0$). The integral signal does not depend on the scale, and the difference between the values is given by $(f_i - T)$, with $T = 0.1667$. For $a = 1$, the filter kernel is $[-1, 1]$, so the inner product recovers the signal $(f_i - T)$ and the addition of T suffices to recover $\beta_1^0(b)$. For $a = 1.1$, the filter kernel is $0.9091[-1, 1]$, and the inner product becomes $0.9091(f_i - T)$. Therefore, to recover the true output $\beta_{1.1}^0(b)$, we need to add $F_{1.1} \cdot T$ with $F_{1.1} = 0.9091$.

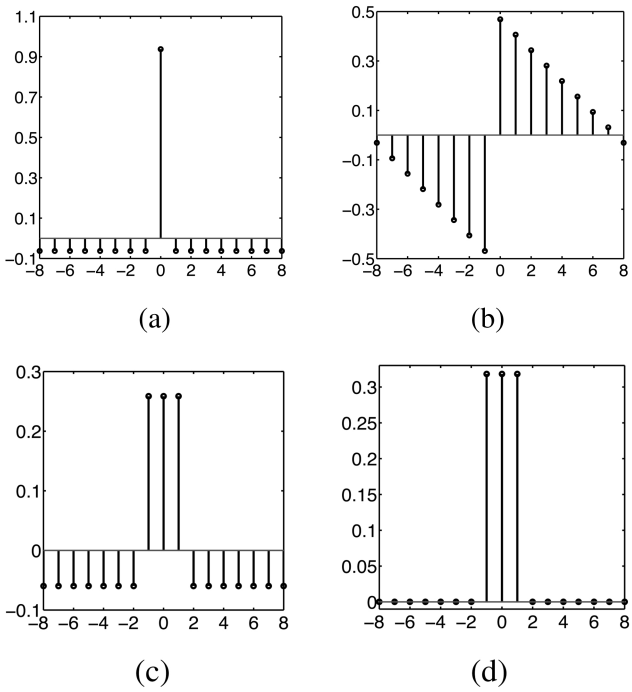


Figure A.3: *Illustration of the implementation of the* 1*D algorithm. A sampled scaled B-spline* $\beta^0_\pi(b)$ *with* $b = -8 : 8$ *is constructed: (a) Input:* $\delta(k) - T$ *with* $k = -8 : 8$ *and* $T = 0.0625$*. (b) Integral (running-sum result) of (a). (c) Result of the inner products given by: eq. (A.10) with* $w_\pi = [-0.3103\ 0\ 0\ 0.3183]$ *and eq. (A.14) with* $\chi_\pi = [-0.3103; 0.3183]$*. (d) Output:* $\beta^0_\pi(b)$ *with* $b = -8 : 8$*. It is computed by adding to (c)* $F_\pi \cdot T$ *with* $F_\pi = 0.9549$*. We choose* $m = -1$*.*

An illustration of the implementation of the 1D algorithm is shown in Figure A.3, where we show the sequence of steps required to construct the sampled scaled B-spline $\beta^0_\pi(b)$. Figure A.3 (a) shows one extended period of the sampled mean subtracted- input signal. In this case, we chose $m = -1$ as to have $f_i(k) = c(k) = \delta(k) - T$ with $k = -8 : 8$ and $T = 0.0625$. Figure A.3 (b) shows the result of the running sum i.e., the

integral coefficients g in eq. (A.7). The integral signal has anti-symmetric boundary conditions and zero mean. Figure A.3 (c) shows the result of the inner products given by eq. (A.10) with $w_\pi = [-0.3103, 0, 0, 0.3183]$ and eq. (A.14) with $\chi_\pi = [-0.3103; 0.3183]$. In simple terms, $f_0(0) = -0.3103 \cdot g(-2) + 0.3183 \cdot g(1)$. For each sample b, the output $f_0(b)$ is computed as a simple linear combination of a number of integral coefficients g. The values of the weights depend on the local scale a. As $n = 0$ the clusters $(..., g(k), g(k-1), ...)$ are of one sample. Note that the distance between the sample of clusters equals three. As $\pi > (m + n + 1) = 1$, it is more efficient to implement (A.14) than (A.10). Figure A.3 (d) shows the sampled scaled B-spline $\beta^0_\pi(b)$ with $b = -8 : 8$. This is computed by adding $f_0(b)$ and $F_\pi \cdot T$ with $F_\pi = 0.9549$.

Finally, we have tested the validity of our implementation by choosing $f_i(b) = \delta(b)$, for $b = -31 : 31$, $n = 0, ..., 3$, and fixing a scale value a. For a range of scale values a from 1 to 5, using a step of 0.1, we can recover $f_0(b) = \beta^n(b/a)$ with an error in the order of the computer precision.

A.4.5.2 ND Normalization

Since the inverse finite difference filter is applied to each dimension of the data, one mean value is extracted from each 1D signal vector in each dimension. For example, in 2D, we have $\mathbf{T}_1 = (T_{1,0}, ..., T_{1,L_1-1})$, a mean vector calculated over the L_1-rows and $\mathbf{T}_2 = (T_{2,0}, ..., T_{2,L_2-1})$ calculated over the L_2-columns.

To obtain the true output f_0 at a given point $\mathbf{b} = (b_1, b_2)$:

- if $\mathcal{M}(\mathbf{b}) = \mathbf{a} \in \mathbb{Z}^N$, add $T_{1,b_1} + T_{2,b_2}$ to $f_0(\mathbf{b})$. $\mathcal{M}$ is an image that stores the scale vector $\mathbf{a}$, for each pixel $\mathbf{b}$ of the input image.

- else, if the scale has non-integer values, calculate the 1D expanded value $\mathbf{R}_j$ of $\mathbf{T}_j, j = 1, 2$, following the algorithm described in the previous section. Then, add the factor $(R_{1,b_1} \cdot NF_{a_2} + R_{2,b_2} \cdot NF_{a_1})$ to $f_0(\mathbf{b})$.

We have tested the validity of the implementation using the same method as for the 1D case.

A.4.6 SPECIAL CASE: INTEGER SCALES

For the special case of using integer scales, there is an efficient implementation of the *spatially-invariant* spline-based filtering, which uses moving sums [133]. Such an implementation also avoids the use of global integration operators which can cause accumulated round-off errors and overflows for large images (see Section A.5).

Namely, when the scale a is an integer, we have the identity [128]

$$\Delta^{-(n+1)}(z)\Delta_a^{n+1}(z) = \left(\frac{1-z^a}{1-z}\right)^{n+1} = (U_a^0(z))^{n+1} \tag{A.20}$$

where $U_a^0(z) = \sum_{k=0}^{a-1} z^k$. The convolution with an integer-scaled B-spline can then be written as

$$f_0(\mathbf{b}) = \left(h * (u_{\mathbf{a}}^0)^{n+1}\right)(\mathbf{b}) \tag{A.21}$$

where h contains the ND spline coefficients, each 1D component is given by $h = f_i * (\mathtt{b}^m)^{-1} * \mathtt{b}^{m+n+1}$ and $(u_{\mathbf{a}}^0)^{n+1} = \prod_{i=1}^{N}(u_{a_i}^0)^{n+1}$ with $r(k) = (u_{a_i}^0 * s)(k) = \sum_{l=k}^{k+a_i-1} s(l)$.

When the convolution is spatially-invariant, $(u_{\mathbf{a}}^0)^{n+1}$ can be implemented by separable filtering. Each filter $(u_{a_i}^0)$ can be computed recursively using a moving-sum method that requires only one addition and one subtraction per point, i.e., $r(k) = r(k-1) + s(k+a_i-1) - s(k-1)$ [133]. In this particular case, the computational complexity is smaller than needed to implement a separable version of eq. (A.19). See [128] for the proof.

When the convolution is spatially variant, $(u_{\mathbf{a}}^0)^{n+1}$ cannot be implemented by separable filtering and therefore, an optimization using moving sums cannot be applied. In other words, implementing eq. (A.21) does not provide any computational advantage with respect to implementing eq. (A.19).

A.5 NUMERICAL ERRORS

We now analyze the constraints that machine accuracy imposes on the computation of the ND convolution with a scaled B-spline, especially for large values of n and for large images ($L_1 \times L_2$), along with the strategies we propose to minimize the errors.

A.5.1 FLOATING POINT REPRESENTATION

Let $s \approx fl(s) = sg \cdot \mathcal{M} \cdot \mathcal{B}^{e-E}$, be the floating point representation of a number $s \in \mathbb{R}$, where sg is the sign, $\mathcal{M}$ is an integer mantissa, e an integer exponent, $\mathcal{B}$ is the base (usually 2) and E is a fixed bias required to represent both negative and positive exponents [134].

We can write $fl(s) = s(1+\kappa)$, for some $|\kappa| \leq \frac{1}{2}\epsilon$, with the machine epsilon defined as $\epsilon = \mathcal{B}^{-e_{max}}$ where e_{max} is the total number of bits used for the mantissa $\mathcal{M}$ in the floating point representation. Then, the maximum possible relative error when representing a real number in floating point format is $\frac{1}{2}\epsilon$. The relative machine precision of s is $eps(s) = \mathcal{B}^{e-e_{max}}$ where e is the integer that fulfills $\mathcal{B}^e \leq s < \mathcal{B}^{e+1}$. A large magnitude of s results in a larger eps. Note that $\epsilon = eps(1)$ and $1 = eps(\epsilon^{-1})$.

To represent floating point numbers, most CPUs follow the IEEE $754-1985$ Standard for Binary Floating Point Arithmetic [135]. The two most widely used formats specified by this standard are *single precision* (32 bits) and *double precision* (64 bits). The latest revision of the standard (unapproved draft) supports the *long double precision* (128 bits). Table A.1 contains the machine epsilon and the number of decimal digits of accuracy for all three IEEE floating point types.

A.5.2 PROPAGATION OF ROUND-OFF ERRORS

A round-off error occurs when a number cannot be exactly represented with the available number of bits. Any operation using floating-point numbers involves round-off errors in the operands, which propagate to the result. In certain circumstances this propagation can lead to a large error that can jeopardize the accuracy of the result. This could happen in two critical steps of our algorithm: the computation of the inverse finite differences computation and the scale dependent filtering.

A.5.2.1 Computation of the Inverse Finite Differences

The 1D inverse finite difference operator Δ^{-1} can be implicitly expressed as $y_k = \sum_{l=0}^{k} s_l$, where each output value is calculated as the addition of the precedent ones. If $|y_k| > B^{e_{max}}$, then $eps(y_k) > 1$ and the number y_k cannot be represented with integer precision. The overflow problem becomes more critical when working with higher values of

Table A.1: *Machine epsilon and decimal digits of accuracy of the IEEE Floating Point types.*

# bits	Epsilon	Digits
32 bits	$2^{-23} \approx 10^{-7}$	6
64 bits	$2^{-52} \approx 10^{-16}$	15
128 bits	$2^{-112} \approx 10^{-34}$	33

Table A.2: *Estimation of the maximum size of 1D signals to have $eps(y_k) \leq 1$ for different IEEE floating point types and n values.*

# bits	$n = 0$	$n = 1$	$n = 2$	$n = 3$
32 bits	$\mathbf{10^{6}}$	$2^{\frac{1}{2}}10^{3}$	$6^{\frac{1}{3}}10^{2}$	$0.24^{\frac{1}{4}}10^{2}$
64 bits	$\mathbf{10^{15}}$	$\mathbf{20^{\frac{1}{2}}10^{7}}$	$\mathbf{6^{\frac{1}{3}}10^{5}}$	$\mathbf{2.4^{\frac{1}{4}}10^{4}}$
128 bits	$\mathbf{10^{33}}$	$\mathbf{20^{\frac{1}{2}}10^{16}}$	$\mathbf{2^{\frac{1}{3}}10^{11}}$	$\mathbf{240^{\frac{1}{4}}10^{8}}$

n and multiple dimensions, as the error growths faster. Table A.2 shows an estimate of the maximum value $\Delta^{-(n+1)}$ of a 1D signal that fulfills $eps(y_k) \leq 1$ for different IEEE floating point types and n values. The increase in magnitude of the inverse finite differences output is polynomial with degree $(n + 1)$. For example, if we consider that $\{s_k = 1\}_{k=0,\cdots,L-1}$, then $y_k = k^{n+1}/(n + 1)!$. For N dimensions, the corresponding increase in magnitude is polynomial with degree $N(n + 1)$.

A.5.2.2 Space Dependent Filtering

Due to the scaled finite differences, the filter kernels (w_a and χ_a) behave as derivative filters (see Figure A.2). The computations defined by eqs. (A.18) and (A.19) involve:

- *the multiplication of large integral values g with small filter weights w_a (or χ_a)* The multiplication is a relatively safe operation but small errors may accumulate in the process.

- *the addition of the results of the multiplication*, i.e. $g \cdot w_a$ (or $g \cdot \chi_a$) Adding large numbers with different signs may result in significant loss of precision leading to an incorrect result. If the two operands differ by less than the lowest of their respective *eps*, a catastrophic cancelation occurs: The result contains no correct digits but only round off errors from previous computations.

A.5.3 ERROR MINIMIZATION STRATEGIES

We implemented our algorithm using the IEEE *long double* floating point data type (128 bits) because it provides the highest significant precision. From our estimates, given in Table A.2, it is safe to work with images of size 1024×1024 and $n = 3$. When applied to larger images, round-off errors and their propagation can be limited by calculating the integrals over independent image blocks so as to limit the maximum magnitude that the integral can reach. We consider that a maximum block size of 1024×1024 remains on the safe side for the typical requirements. For instance, if we have a 1024×1024 2D noise image with a Gaussian intensity distribution of mean 1000, and standard deviation 15, the estimated error bound with the *double* data type is 40, when working with $n = 3$. Using the *long double* type, the error is reduced to $4 \cdot 10^{-16}$. We confirmed these values by filtering the noise image with B-splines of fixed scale $a = 4$ with the minimum block-size -equal to the filter kernel size- and comparing the result when processing the whole image in one block. The difference between the two outputs was zero when represented with 16 bits.

A.6 STEERABLE FILTERING

In this section, we use the filter that we have just developed as a building block of a novel B-spline based steerable filter that detects features of unknown position and orientation.

A.6.1 ROTATED MATCHED FILTERING

The detection of features of unknown position and orientation can be formulated as a rotated matched filtering task [136]. We assume the signal model

$$f(\mathbf{x}) = r \cdot f_t(\mathbf{R}_\theta(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0) + n(\mathbf{x}), \tag{A.22}$$

where f_t is the signal we want to detect; $\mathbf{x}_0$ is its position in the image; r is a weight; $\mathbf{R}_\theta = [\mathbf{r}_\theta\ \mathbf{r}'_\theta]$ is the rotation matrix with $\mathbf{r}_\theta = (\cos\theta, \sin\theta)$, $\mathbf{r}'_\theta = (-\sin\theta, \cos\theta)$, θ is the rotation angle and $n(\mathbf{x})$ is Gaussian white noise. The maximum likelihood estimation algorithm, for this signal model is

$$\begin{aligned} \theta^*(\mathbf{x}_0) &= \arg\max_\theta (f(\mathbf{x}) * h(\mathbf{R}_\theta(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0))) \\ r^*(\mathbf{x}_0) &= f(\mathbf{x}) * h(\mathbf{R}_{\theta^*}(\mathbf{x} - \mathbf{x}_0) + \mathbf{x}_0)), \end{aligned} \tag{A.23}$$

where r^* is the magnitude, θ^* is the orientation of the feature at position $\mathbf{x_0}$ and $h(\mathbf{x}) = f_t(-\mathbf{x})$ is the feature template.

The exhaustive search of the maximum argument in (A.23) is not practical as it requires the computation of inner products with an extensive set of shifted and rotated versions of the feature template. The common practice used to reduce computational complexity is to select the detector within the class of steerable filters introduced by Freeman and Adelson [130]. Namely, the template is chosen such as any rotated version of itself can be computed as a linear combination of a small number of derivatives of an isotropic prototype function. A common choice is for the prototype function to be a Gaussian, i.e. the standard Gaussian steerable template is $h(\mathbf{x}) = g_{\sigma,(2,0)}(\mathbf{x})$ where $g_\sigma(\mathbf{x})$ is a Gaussian function with standard deviation σ. Here, we propose instead to use B-splines.

A.6.2 B-SPLINE STEERABLE FILTERING

We focus our attention on a simple task: ridge detection, and choose the idealized line model $f_t(x_1, x_2) = \delta(x_1)$. Although we particularize the derivations to the ridge detector, this very idea can be extended to any steerable filter of the class introduced in [130].

We start by showing the convergence of the second derivative of a rotated B-spline to a steerable template.

Proposition A.6.1. *Consider the sequence of centered B-splines* $\{\beta^n_a(\mathbf{x})\}_{n\geq 3}$ *of degree* n *at scale* $a \in \mathbb{R}_+$ *with* $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$*. Then, we have the following convergence result* $\lim_{n\to\infty} \beta^n_{a,(2,0)}(\mathbf{R}_\theta \mathbf{x}) = \lim_{m\to\infty}(\cos^2\theta \beta^m_{a,(2,0)}(\mathbf{x}) + 2\cos\theta\sin\theta\beta^m_{a,(1,1)}(\mathbf{x}) + \sin^2\theta\beta^m_{a,(0,2)}(\mathbf{x}))$*, where* $\beta^n_{a,(d_1,d_2)} = \beta^n_{a,d_1}(x_1)\beta^n_{a,d_2}(x_2)$.

Proof. We define $\hat{\beta}_a^n(\mathbf{R}_\theta\omega) = \hat{\beta}_a^n(\mathbf{r}_\theta^T\omega)\hat{\beta}_a^n(\mathbf{r}_\theta'^T\omega)$. In the Fourier domain, we have

$$\hat{\beta}_{a,(d_1,d_2)}^n(\mathbf{R}_\theta\omega) = (j\mathbf{r}_\theta^T\omega)^{d_1}(j\mathbf{r}_\theta'^T\omega)^{d_2}\hat{\beta}_a^n(\mathbf{R}_\theta\omega). \tag{A.24}$$

Thus, we can write $\hat{\beta}_{a,(2,0)}^n(\mathbf{R}_\theta\omega) = (j\mathbf{r}_\theta^T\omega)^2\hat{\beta}_a^n(\mathbf{R}_\theta\omega)$. From the Central Limit Theorem each of the constituent B-spline components converge to a 1D Gaussian as $n \to \infty$ [126]. Then, $\hat{\beta}_a^n(\omega)$ converges to an isotropic 2D Gaussian

$$\lim_{n\to\infty}\hat{\beta}_a^n(\omega) = \frac{1}{2\pi\sigma^2}e^{-\frac{\omega^T\omega}{\sigma^2}} \tag{A.25}$$

with $\sigma = \sqrt{\frac{n+1}{12}}a$. So, we have

$$\lim_{n\to\infty}(j\mathbf{r}_\theta^T\omega)^2\hat{\beta}_a^n(\mathbf{R}_\theta\omega) = \lim_{m\to\infty}(j\mathbf{r}_\theta^T\omega)^2\hat{\beta}_a^m(\omega). \tag{A.26}$$

Developing the right hand side of eq. (A.26) and applying the corresponding definition of the derivative in the Fourier domain, we get to desired convergence result. □

Jacob *et al.* [136] proposed a modified Gaussian steerable template especially sensitive to elongated image structures. Specifically, they increased the flatness of the template in the longitudinal direction near the origin. In particular, the template was modified as to satisfy the flatness criterion [131]

$$\lim_{\mathbf{x}\to 0}(\mathbf{r}_\theta'^T\cdot\nabla)^2 h_1(\mathbf{R}_\theta\mathbf{x}) = 0. \tag{A.27}$$

For the template $h_1(\mathbf{x}) = g_{\sigma,(2,0)}(\mathbf{x}) + \alpha g_{\sigma,(0,2)}(\mathbf{x})$, the left-hand side equals $(1+3\alpha)\|\mathbf{r}\|^4/\sigma^4$ and consequently, the α-factor is $\alpha = -1/3$. Luckily, the optimal template $h_1(\mathbf{x})$ is invariant to the angle θ, and thus the estimation (A.23) can still be solved using an eigenvalue decomposition of the Hessian matrix H_{f*h_1}. In particular, the optimal response and the angle are given by $r^* = \lambda_{\max}$ and $\theta^* = \arctan(v_2/v_1)$ [137]. Here, $\lambda_{\max}$ is the maximum eigenvalue of H_{f*h_1} and $\mathbf{v} = (v_1, v_2)$ is the corresponding eigenvector.

In this work, we have imposed the flatness criterion (A.27) to our B-spline template (i.e., $h_2(\mathbf{x}) = \beta_{a,(2,0)}^n(\mathbf{x}) + \alpha\beta_{a,(0,2)}^n(\mathbf{x})$) and solved for the factor α. We have

$$\alpha = -\frac{\mathcal{A}}{\mathcal{B}}\left(\frac{\cos^4\theta + 2\cos^2\theta\sin^2\theta(\mathcal{B}/\mathcal{A}-2) + \sin^4\theta}{\cos^4\theta + 2\cos^2\theta\sin^2\theta(3\mathcal{A}/\mathcal{B}) + \sin^4\theta}\right) \tag{A.28}$$

with $\mathcal{A} = 2(\beta^{n-2}(1) - \beta^{n-2}(0))^2$ and $\mathcal{B} = \beta^n(0)(\beta^{n-4}(2) - 4\beta^{n-4}(1) + 3\beta^{n-4}(0))$.

To derive this equation, we compute the α-factor for the template $h_2(\mathbf{x}) = \beta^n_{a,(2,0)}(\mathbf{x}) + \alpha\beta^n_{a,(0,2)}(\mathbf{x})$ such as the flatness criterion

$$\lim_{\mathbf{x}\to 0}(\mathbf{r}_\theta'^T \cdot \nabla)^2 h_2(\mathbf{R}_\theta\mathbf{x}) = 0. \tag{A.29}$$

is satisfied. The rotated template is given by

$$h_2(\mathbf{R}_\theta\mathbf{x}) = \cos^2\theta\vartheta_1(\mathbf{x}) + 2\cos\theta\sin\theta\vartheta_2(\mathbf{x}) + \sin^2\theta\vartheta_3(\mathbf{x}) \tag{A.30}$$

with $\vartheta_1(\mathbf{x}) = \beta^n_{a,(2,0)}(\mathbf{x}) + \alpha\beta^n_{a,(0,2)}(\mathbf{x})$; $\vartheta_2(\mathbf{x}) = (1-\alpha)\beta^n_{a,(1,1)}(\mathbf{x})$; $\vartheta_3(\mathbf{x}) = \beta^n_{a,(0,2)}(\mathbf{x}) + \alpha\beta^n_{a,(2,0)}(\mathbf{x})$.

We use that $(\mathbf{r}_\theta'^T \cdot \nabla)^2 h_2(\mathbf{R}_\theta\mathbf{x}) = \mathbf{r}_\theta'^T \cdot H_{h_2} \cdot \mathbf{r}'_\theta$ where H_{h_2} is the Hessian matrix of $h_2(\mathbf{x})$, to have $(\mathbf{r}_\theta'^T \cdot \nabla)^2 h_2(\mathbf{R}_\theta\mathbf{x}) = r_1\cos^4 g_1(\mathbf{x}) + 2r_1^3 r_2 g_2(\mathbf{x}) + r_1^2 r_2^2 g_3(\mathbf{x}) + 2r_1 r_2^3 g_4(\mathbf{x}) + r_2^4\theta g_5(\mathbf{x})$ with

$$\begin{aligned}
g_1(\mathbf{x}) &= \beta^n_{a,(2,2)}(\mathbf{x}) + \alpha\beta^n_{a,(0,4)}(\mathbf{x}) \\
g_2(\mathbf{x}) &= -\beta^n_{a,(3,1)}(\mathbf{x}) + (1-2\alpha)\beta^n_{a,(1,3)}(\mathbf{x}) \\
g_3(\mathbf{x}) &= \beta^n_{a,(4,0)}(\mathbf{x}) + (-4+6\alpha)\beta^n_{a,(2,2)}(\mathbf{x}) + \beta^n_{a,(0,4)}(\mathbf{x}) \\
g_4(\mathbf{x}) &= -\beta^n_{a,(1,3)}(\mathbf{x}) + (1-2\alpha)\beta^n_{a,(3,1)}(\mathbf{x}) \\
g_5(\mathbf{x}) &= \beta^n_{a,(2,2)}(\mathbf{x}) + \alpha\beta^n_{a,(4,0)}(\mathbf{x})
\end{aligned}$$

and $r_1 = \|\mathbf{r}\|\cos\theta$ and $r_2 = \|\mathbf{r}\|\sin\theta$.

Taking limits, we have

$$\begin{aligned}
\lim_{\mathbf{x}\to 0} g_1(\mathbf{x}) &= \lim_{\mathbf{x}\to 0} g_5(\mathbf{x}) = \mathcal{A} + \alpha\mathcal{B} \\
\lim_{\mathbf{x}\to 0} g_2(\mathbf{x}) &= \lim_{\mathbf{x}\to 0} g_4(\mathbf{x}) = 0 \\
\lim_{\mathbf{x}\to 0} g_3(\mathbf{x}) &= 2(\mathcal{B} + (-2+3\alpha)\mathcal{A})
\end{aligned} \tag{A.31}$$

with $\mathcal{A} = \lim_{\mathbf{x}\to 0}\beta^n_{a,(2,2)}(\mathbf{x})$, and $\mathcal{B} = \lim_{\mathbf{x}\to 0}\beta^n_{a,(4,0)}(\mathbf{x}) = \lim_{\mathbf{x}\to 0}\beta^n_{a,(0,4)}(\mathbf{x})$.

Substituting in eq. (A.29) and solving for α we get eq. (A.28).

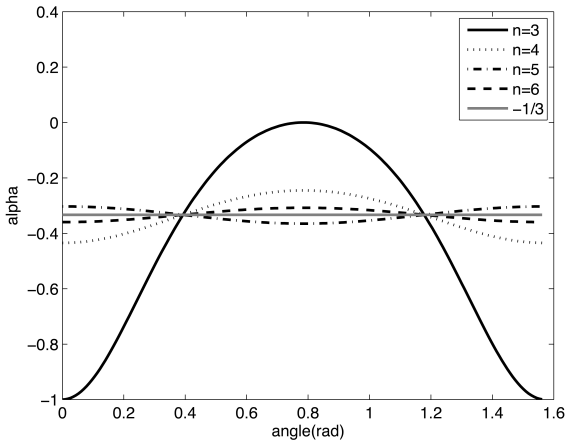


Figure A.4: *Graphical illustration of the variation of α with $\theta \in [0, \pi/2]$ for different B-spline degrees. The grey line $\alpha = -1/3$ corresponds to a rotationally invariant template around the origin.*

Figure A.4 shows how the α-factor changes as a function of the angle $\theta \in [0, \pi/2]$ for B-splines of degrees from $n = 3$ to 6. Since the α-factor is $\pi/2$ periodic, only one of the quadrants is shown. Note that for any B-spline degree and for an orientation $\theta = \pi/8$, the template obtained is rotationally invariant around the origin, i.e. $\lim_{\mathbf{x}\to 0}(\mathbf{r}_\theta^{'T} \cdot \nabla)^2 h_2(\mathbf{R}_\theta \mathbf{x}) = (1+3\alpha)\|\mathbf{r}\|^4/a^4$. In this case $\alpha = -1/3$, which is the same α-factor used in a Gaussian prototype function. Moreover, as the B-spline degree n increases we observe the convergence towards a constant α-factor of $-1/3$.

Therefore, we choose as steerable template

$$h_3(\mathbf{x}) = -\mathcal{N}(\beta^n_{a,(2,0)}(\mathbf{x}) - 1/3 \cdot \beta^n_{a,(0,2)}(\mathbf{x})). \tag{A.32}$$

The negative sign is used because the template is designed to detect bright filaments in a dark background. The normalization factor $\mathcal{N}$ is used to have a template of $a^2(n+1)$ energy, i.e., $\int |h_3(\mathbf{x})|^2 d\mathbf{x} = a^2(n+1)$. By normalizing the template energy values we will be able to compare energies calculated using different scales and/or B-splines of different degrees. In

particular, we have

$$\frac{\mathcal{N}}{a(n+1)} = \left(\frac{10}{9}\int(\beta^n_{a,(2,0)}(\mathbf{x}))^2 d\mathbf{x} - \frac{2}{3}\int \beta^n_{a,(2,0)}(\mathbf{x})\beta^n_{a,(0,2)}(\mathbf{x})d\mathbf{x}\right)^{-\frac{1}{2}}$$

where $\int(\beta^n_{a,(2,0)}(\mathbf{x}))^2 d\mathbf{x} = \frac{2}{a^2}\beta^{2n+1}(0)(3\beta^{2n-3}(0) - 4\beta^{2n-3}(1) + \beta^{2n-3}(2))$ and $\int \beta^n_{a,(2,0)}(\mathbf{x})\beta^n_{a,(0,2)}(\mathbf{x})d\mathbf{x} = \frac{4}{a^2}(\beta^{2n-1}(1) - \beta^{2n-1}(0))^2$.

Then, the rotated feature template can be written as $h_3(\mathbf{R}_\theta\mathbf{x}) = -\mathcal{N}(\cos^2\theta\vartheta_1(\mathbf{x}) + 2\cos\theta\sin\theta\vartheta_2(\mathbf{x}) + \sin^2\theta\vartheta_3(\mathbf{x}))$

$$\begin{aligned} \text{with } \vartheta_1(\mathbf{x}) &= \beta^n_{a,(2,0)}(\mathbf{x}) - \frac{1}{3}\beta^n_{a,(0,2)}(\mathbf{x}), \\ \vartheta_2(\mathbf{x}) &= \frac{4}{3}\beta^n_{a,(1,1)}(\mathbf{x}) \text{ and} \\ \vartheta_3(\mathbf{x}) &= \beta^n_{a,(0,2)}(\mathbf{x}) - \frac{1}{3}\beta^n_{a,(2,0)}(\mathbf{x}). \end{aligned}$$

Figure A.5 displays the cubic B-spline template $h_3(\mathbf{R}_\theta\mathbf{x})$ for a scale of 6 and rotation angles (a) $\theta = 0$; (b) $\theta = \pi/8$; (c) $\theta = \pi/4$.

The convolution of the input image $f(\mathbf{x})$ with the rotated feature template is $f(\mathbf{x}) * h_3(\mathbf{R}_\theta\mathbf{x}) = -\mathcal{N}(\cos^2\theta f_1(\mathbf{x}) + 2\cos\theta\sin\theta f_2(\mathbf{x}) + \sin^2\theta f_3(\mathbf{x}))$ with $f_i(\mathbf{x}) = (f * \vartheta_i)(\mathbf{x}), i = 1, 2, 3$. Therefore, the responses $f_i(\mathbf{x})$ can be computed as a linear combination of the convolution of $f(\mathbf{x})$ with the basis functions $\{\beta^n_{a,(2,0)}(\mathbf{x}), \beta^n_{a,(1,1)}(\mathbf{x}), \beta^n_{a,(0,2)}(\mathbf{x})\}$. Finally, the eigenvectors and eigenvalues used to solve for the optimal response and orientation are computed from the modified second derivative matrix $H_{f*h_3} = [f_1(\mathbf{x})\ f_2(\mathbf{x}); f_2(\mathbf{x})\ f_3(\mathbf{x})]$.

A.7 APPLICATIONS

The algorithm that we have described can be used in any task involving the spatially variant convolutions of an image with scaled B-spline kernels (or their derivatives). Here, we present two examples that illustrate the benefit of using our method: adaptive denoising and ridge detection.

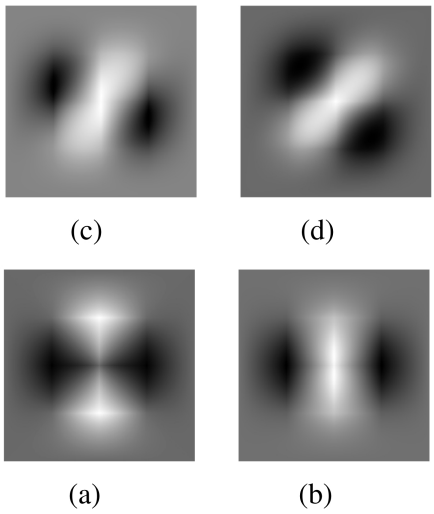


Figure A.5: *Illustration of the steerable templates: (a)* $\beta^3_{6,(2,0)}(\mathbf{x})$ *and* $h_3(\mathbf{R}_\theta \mathbf{x})$ *for different orientations: (b)* $\theta = 0$ *(c)* $\theta = \pi/8$ *(d)* $\theta = \pi/4$.

A.7.1 FEATURE ADAPTIVE DENOISING

We propose to implement an edge-preserving denoising filter as the convolution of the image with isotropic scaled B-spline kernels $f_0(b_1, b_2) = (f_i * \beta^n_a)(b_1, b_2)$ with $\beta^n_a(x_1, x_2) = \beta^n_a(x_1)\beta^n_a(x_2)$.

A.7.1.1 Scale-Vector Computation

The local scale value used to define the size of the B-spline kernel to be convolved at each image position is calculated using Saha and Udupa's method [138]. Namely, to calculate the scale at each pixel, we recursively measure the homogeneity of a hyper-ball-shaped growing region-of-interest centered at the pixel. The evaluation iterates until a given inhomogeneity value is reached. This value can be adjusted depending on the expected noise level and the degree of detail to be preserved. In presence of diffuse borders, caused for instance by motion artifacts, non-zero scale values might be assigned to edge pixels, leading to unwanted diffusion of the edges. To

avoid this problem, we added the algorithm an extra step that detects diffused borders and forces the scale value to be null on them.

A.7.1.2 Experimental Results

A.7.1.2.1 Quantitative Results on Synthetic Images

Our filter was first evaluated quantitatively, by comparing it to other commonly used denoising filters. For the comparison, we used the Shepp-Logan phantom with two different kinds of noise and artifacts. First, we modified the intensities of the phantom to force them to follow a Rician distribution, as in Magnetic Resonance Imaging (MRI) images [139]. Alternatively, we reconstructed the phantom image from a limited number of projections (128, 256 and 512) using a filtered back-projection algorithm with a Ram-Lak filter [140], to simulate the acquisition process of X-ray computed tomography (micro-CT) images. Gaussian noise was added to the images, with a standard deviation of 15, which was estimated from homogeneous regions of real micro-CT images. The performance of the filter was evaluated using two parameters: the Residual Noise (RN), which is the standard deviation of pixel intensities in homogeneous areas of the image [138], and the Average Edge Width (AEW), calculated as the ratio between the total number of pixels that belong to an edge and the number of edges in the image [141]. We compared our filter to a fixed scale smoothing filter and to an anisotropic diffusion filter [142]. The scale values used in the B-spline adaptive smoothing filter ranged from 0 to 15. The scale value used for the fixed scale smoothing filter was set to 10, which was approximately the mean scale value calculated over the entire phantom. In both fixed and variable scale filtering- cases, we used linear B-splines, i.e., $m = n = 1$. The parameters used for the anisotropic diffusion were typical: 0.1 time step, 10 iterations and 3.0 conductance.

Table A.3 shows the average RN values in 23 manually selected homogeneous regions of the phantom, for three different levels of Rician noise (Gaussian σ of 500, 1000 and 1500). The mean AEW values for the three filters compared are shown in Table A.4. Tables A.5 and A.6 show the results on the Shepp-Logan phantom with micro-CT-like noise and artifacts. In all cases our adaptive B-spline smoothing outperforms fixed scale smoothing and anisotropic diffusion, since it provides both higher homogeneity and better-defined edges. When compared to anisotropic diffusion,

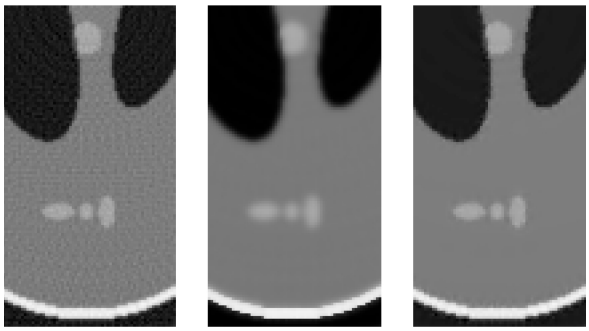


Figure A.6: *(a) Detail of the Shepp-Logan phantom reconstructed from 128 projections. (b) Shepp-Logan phantom after anisotropic diffusion filtering. (c) Shepp-Logan phantom after adaptive B-spline smoothing.*

the average improvement in homogeneity is 54% for the MRI phantom and 43% for the micro-CT phantom, with a reduction in AEW of 0.73 and 1.08 pixels respectively. The difference in homogeneity is statistically significant ($p < 0.001$), as shown using a Wilcoxon matched-pairs signed-rank test. The effect can be qualitatively observed in Figure A.6, which shows the results of applying both filters to the phantom reconstructed from 128 projections.

A.7.1.2.2 Qualitative Results on Real Images

We then tested our filter on real biomedical images. Figure A.7 shows an MRI image of a human head. We found that, for a comparable level of smoothing in homogeneous areas, our adaptive B-spline filtering blurs edges considerably less than the other filters. The same effect can be observed in Figure A.8, which shows a micro-CT image of a mouse lung with multiple small details, mainly tumors and blood vessels.

Table A.3: *Mean RN values for different levels of Rician noise and denoising filter.*

	No Filter	Fixed-Scale	Anisotropic	Adaptive
Level 1	451.57	431.00	194.92	**67.92**
	78.24	472.85	201.46	**25.57**
Level 2	920.10	445.40	253.09	**124.90**
	153.06	441.85	170.97	**42.15**
Level 3	1346.66	477.19	343.47	**181.05**
	240.38	416.97	152.42	**54.32**

Table A.4: *Mean AEW values averaged over the three different levels of Rician noise, for all three denoising filters*

	No Filter	Fixed-Scale	Anisotropic	Adaptive
AEW	7.54	11.33	8.34	**7.61**

Table A.5: *Mean RN values for different number of projections and denoising filter.*

	No Filter	Fixed-Scale	Anisotropic	Adaptive
128Pr.	1106.66	398.98	220.88	**134.51**
	61.45	422.15	210.63	**120.08**
256Pr.	648.04	423.00	221.95	**124.53**
	60.82	449.15	228.95	**110.72**
512Pr.	499.26	428.84	222.16	**121.96**
	67.93	456.53	233.73	**109.48**

A.7.2 RIDGE DETECTION

As a second example of application we used of our B-spline steerable template (A.32) as a building block of a ridge detector, in a spatially variant setting.

Table A.6: *Mean AEW values averaged over the three different numbers of projections for all three denoising filters.*

	No Filter	Fixed-Scale	Anisotropic	Adaptive
AEW	8.28	9.77	9.50	**8.42**

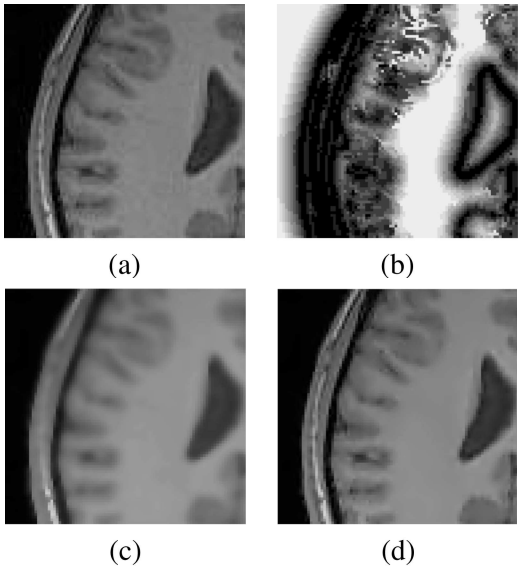


Figure A.7: *(a) Region of MRI image of human head. (b) Scale image that corresponds to (a). (c) The same region in (a) after anisotropic diffusion filtering. (d) The same region in (a) after adaptive B-spline smoothing.*

A.7.2.1 Quantitative Results on Synthetic Images

A.7.2.1.1 Synthetic image

We created a synthetic image with a spatially variant sinusoidal function. The ridges are almost vertical in the image center, and become increasingly inclined towards the sides of the image. The ridges are wider towards the bottom of the image than they are at the top. We added Gaus-

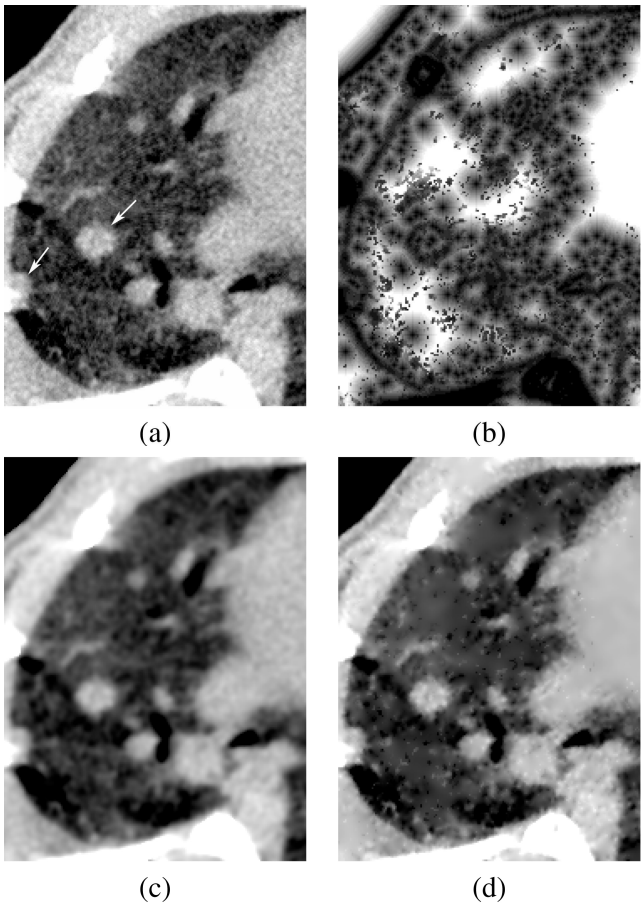


Figure A.8: *(a) Left region of a micro-CT image of a mice lung with a moderate tumor load (white arrows indicate tumors). (b) Scale-scene that corresponds to (a). (c) The same region in (a) after anisotropic diffusion filtering. (d) The same region in (a) after adaptive B-spline smoothing.*

sian noise with different standard deviations ($\sigma = 0, 0.25, 0.5$ and 0.75), to create noisy versions of the synthetic image. Figure A.9 shows the synthetic image with Gaussian noise of 0.75 standard deviation.

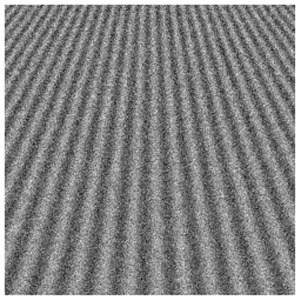


Figure A.9: *Noisy synthetic image ($\sigma = 0.75$).*

A.7.2.1.2 Improvement in sensitivity using $\alpha = -1/3$

We first compared the performance of the standard B-spline detector to that of our optimal detector. So, we performed a brute force ridge detection sequence using either a geometric or arithmetic progression of the scale steps (GPSS or APSS). The GPSS sequence was: 2, 4 and 8. In the APSS, the scale varied from 2 to 9 in 1, 0.25 or 0.1 steps.

The brute force detection method proceeded as follows:

- We first computed the basis images $\{(f*\beta^3_{a,(2,0)})(\mathbf{x}), (f*\beta^3_{a,(1,1)})(\mathbf{x}), (f*\beta^3_{a,(0,2)})(\mathbf{x})\}$ for each scale and template type.
- Then, we calculated the functions $\{f_i(\mathbf{x})\}_{i=1}^3$, as linear combinations of the basis images.
- Next, for each scale value of the series, the response of the detector and ridge orientation was calculated using an eigenvalue decomposition of the Hessian (or the modified Hessian) matrix.
- The scale that produced the maximum response, along with the corresponding response value and ridge orientation were stored for each pixel of the image, forming a maximum response map.
- Finally, a binary mask containing an estimate of the bright ridges was calculated from the maximum response map using an Isodata thresh-

olding algorithm [143]. Most of the spurious edges were eliminated by discarding objects smaller than 20 pixels.

To compare the sensitivity of the detectors we used i. the detector response (DR), ii. the ridge orientation error (ROE), iii. the ridge width error (RWE). The detector errors were calculated as the sum of squared differences (SSD) between the detected ridges and a ground truth, which was analytically pre-computed from the function used to create the synthetic image.

Specifically, ROE was measured as the difference between the ground truth angle and the estimated angle, i.e., $\theta_e = \theta_{gt} - \theta$. We took into consideration the fact that both θ and $\theta + \pi$ correspond to the same orientation. Thus, ROE was distributed between 0 and $\pi/2$. We compute the SSD of the horizontal and vertical components of the unitary vector for each orientation error [144]

$$\begin{aligned} \mathrm{SSD}_1 &= \frac{1}{L_1 L_2} \left(\sum_i \sum_j (\cos \theta_e(i,j))^2 \right)^{\frac{1}{2}}, \\ \mathrm{SSD}_2 &= \frac{1}{L_1 L_2} \left(\sum_i \sum_j (\sin \theta_e(i,j))^2 \right)^{\frac{1}{2}}. \end{aligned}$$

The ridge width is approximately proportional to the detector scale [145]. In particular, we empirically found the weight factor to be 1.3209. The SSD of the ridge width was thus computed as

$$\mathrm{SSD}_w = \frac{1}{L_1 L_2} \left(\sum_i \sum_j w_e(i,j)^2 \right)^{\frac{1}{2}}, \tag{A.33}$$

with the RWE w_e computed as the difference between the true width and the width estimation, i.e., $w_e(i,j) = w_{gt}(i,j) - w(i,j)$

To test the significance of the differences found, we applied a Wilcoxon rank test of differences of means. For the error measurements, we interpreted the mean square error as a mean of the squared differences, i.e., $\mathrm{MSE} = L_1 \cdot L_2 \cdot \mathrm{SSD}^2$.

Table A.7: *Mean (first entry) and standard deviation (second entry) of DR, ROE and RWE through three different levels of Gaussian noise for the brute force ridge detector method, the standard and the optimal template, and GPSS.*

α	0	$-1/3$
Scale vector	[2 4 8]	[2 4 8]
Response	1.6101 ± 0.3564	1.6872 ± 0.3033
Orientation Error	0.4981 0.0430	0.4987 0.0358
Width Error	0.6185	0.5217

Tables A.7 and A.8 compare the sensitivities of the optimal and standard templates, for both GPSS and APSS. In both cases the difference in ROE was not significant, since both detectors were able to accurately detect the orientation of the ridges. Therefore, we focus the rest of this section in the evaluation of DR and RWE.

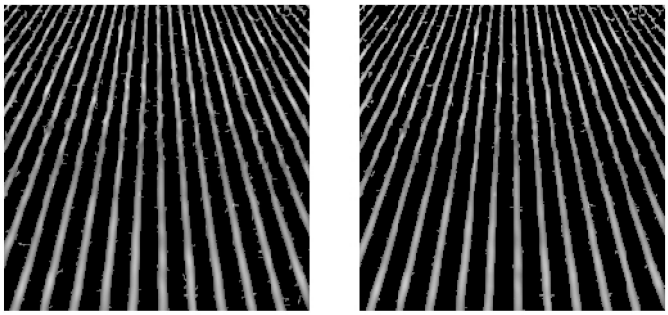
DR was significantly ($p < 0.001$) higher when we used the optimal detector compared to the standard one, both for GPSS (see Table A.7) and APSS with all tested scale steps (see Table A.8). Furthermore, RWE was significantly ($p < 0.001$) lower for the optimal detector than for the standard one, both for GPSS and APSS with scale step equal to one. Therefore, it seemed reasonable to work with the feature template (A.32) with $n = 3$ and $\alpha = -1/3$ in the rest of the experiments.

The response maps obtained using a GPSS and either the standard or optimal template are shown in Figure A.10 (a) and (b), respectively. The input image used was the synthetic noisy image with $\sigma = 0.75$. The ridges look noisier when detected by the standard template. This effect is more significant at wider filament regions.

To the best of our knowledge, this experiment is the first quantitative proof of the sensitivity improvement provided by our modified steerable template. Jacob *et al.* in [136] had predicted that a Gaussian steerable template with $\alpha = -1/3$ would be more sensitive than the standard one, but their argument was simply based on the template shape and their intuition was not quantitatively confirmed.

Table A.8: *Mean (first entry) and standard deviation (second entry) of DR, ROE and RWE through three different levels of Gaussian noise for the brute force ridge detector method, the standard and the optimal template, and three cases of APSS.*

α	0			$-1/3$		
Scale step	1.0	0.25	0.1	1.0	0.25	0.1
Response	1.7168 ± 0.3625	1.7229 ± 0.3630	1.7228 ± 0.3635	1.7953 ± 0.3037	1.7998 ± 0.3045	1.8003 ± 0.3045
Orientation Error	0.4984 0.0401	0.4984 0.0404	0.4984 0.0405	0.4989 0.0329	0.4989 0.0333	0.4989 0.0335
Width Error	0.5001	0.4880	0.4881	0.3875	0.3770	0.3762



(a) (b)

Figure A.10: *Response map for the brute force ridge detector with a GPSS for: (a) the standard template ($\alpha = 0$) (b) the optimal template ($\alpha = -1/3$). The input image was the synthetic noisy image with $\sigma = 0.75$.*

A.7.2.1.3 Spatially variant ridge detector

In our second experiment, we present a novel non-integer spatially variant ridge detector. To estimate the scale that maximizes the response at each pixel, our approach uses a local iterative optimization algorithm on a pre-computed ridge map. This is more efficient that finding the optimum output using pure brute force, specially if non-integer scale values are used. Namely, the detection is now done in two steps:

Table A.9: *Mean (first entry) and standard deviation (second entry) of DR, ROE and RWE for three different levels of Gaussian noise using the spatially variant ridge detector method, the optimal template and three cases of APSS.*

α	$-1/3$		
Scale step	1.0	0.25	0.1
Response	1.7876 ± 0.3195	1.7962 ± 0.3142	1.7976 ± 0.3128
Orientation Error	0.4989 0.0328	0.4989 0.0331	0.4989 0.0332
Width Error	0.3891	0.3727	0.3700
# Iterations	3 ± 0	6.5923 ± 1.1390	8.0989 ± 1.5724

- First, a coarse ridge detection is performed using the method described in the previous subsection, using the optimal template and GPSS. Since the progression of the scales is geometric and the scale is kept constant for the whole image, very efficient algorithms exist to compute the basis images. After the detection, a coarse scale map was computed by applying the binary mask to the scale output.

- Second, fine ridge detection is performed using our spatially variant ridge detector. For each pixel, the value stored in the coarse scale map served as initialization value of a Brent optimization algorithm [146] that finds the scale that maximizes the response of the detector. The output is now a fine scale map, the maximum response map and the orientation map.

The results are shown in Table A.9. If we compare these results to the results shown in Table A.8, we can see that our efficient two-phase algorithm provides DR, ROE and and RWE values that are not significantly different from those of the brute force algorithm. We observe also that the number of iterations needed for the optimization algorithm to converge does not increase significantly with the number of scales ($6,59$ for 36 scales, $8,09$ for 90 scales). Therefore, our algorithm becomes more efficient -compared to the classical brute force method- as the ridge width distribution and the desired resolution increases.

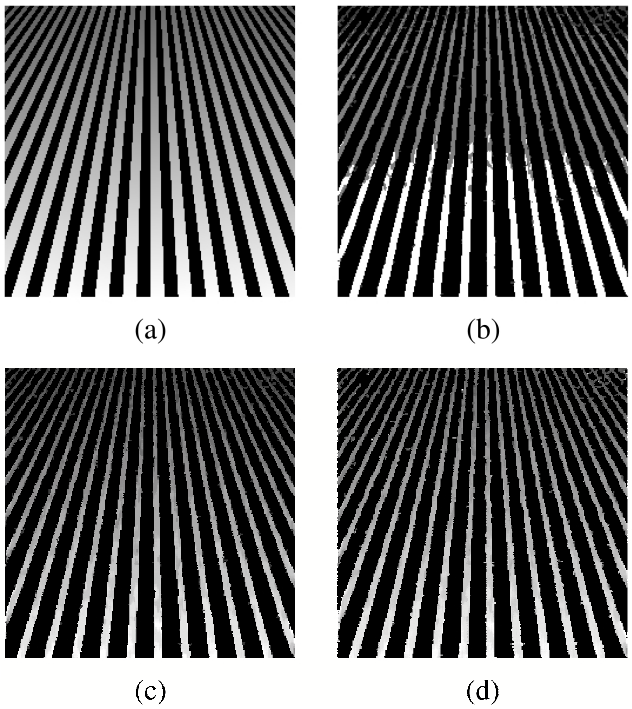


Figure A.11: *(a) Ridge width ground truth image. Width estimation for the brute force ridge detector using: (b) GPSS; (c) APSS. (d) Width estimation of the spatially variant ridge detector using an APSS with a scale step of* 0.1. *The input image was the synthetic noisy image with* $\sigma = 0.75$.

Figure A.11 illustrates the width estimates obtained using our novel spatially variant ridge detection algorithm. The input image is the noisy synthetic image with $\sigma = 0.75$. Figure A.11 (a) shows the analytically computed true ridge width. Figure A.11 (b) shows the width estimates using the brute force ridge detector with GPSS. Note the coarse width estimates with uncertainties due to the noise in the transition between widths. Figure A.11 (c) shows the width estimates using the brute force ridge detector with APSS and a scale step of 0.1. Observe how the smooth increase in the estimated ridge width from top to bottom correlates well

with the ground truth. Figure A.11 (d) shows the width estimation of our novel spatially variant ridge detector with the same scale series as in (c). Observe that both look very similar.

A.7.2.1.4 Improved Performance with the Reduction of the Scale Step

We have measured -for both the optimal and the standard templates- the reduction in RWE and the increase in the DR value as we increase the resolution of the scales. We observe in all the cases an increase in the mean of the DR and a decrease in RWE. We have found the differences in the measurements to be statistically significant ($p < 0.001$) in all cases. This result assigns an added value to carrying out ridge detection using non-integer scales, for which our method is specially suited.

A.7.2.2 Qualitative Results on a Real Image

Next, we show an example of filament detection on a real fluorescent microscopy image. Figure A.12 (a) shows fluorescent labeled actin filaments that form the cell cytoskeleton. The image consist of a cellular body surrounded by a black background. Filament detection is accomplished with our spatially variant ridge detector as described for synthetic images. The scales in geometric progression used for the estimation of the coarse scale map were [2 4 8]. The scales in arithmetic progression used for the refinement were from 2 to 8 in a step of 0.1. The resulting orientation and maximum response maps are shown in Figure A.12 (b) and (c), respectively. The scale that provided the maximum response is given in Figure A.12 (d).

A.8 CONCLUSION

In this chapter, we have presented an efficient algorithm to compute spatially, scale-variant convolutions between a ND spline signal and B-splines of any order and arbitrary size. One key aspect of our algorithm is a computational load independent of the scale, and depending only on the order of the B-splines used for the interpolation and the convolution. Our algorithm is specially designed to work with arbitrary scales but is also appropriated when dealing with integer scales. This is because an efficient moving average implementation do not hold for ND spatially variant B-spline filtering. The main drawback of the method is that is prone to round off errors and overflows as running sums are computed over entire

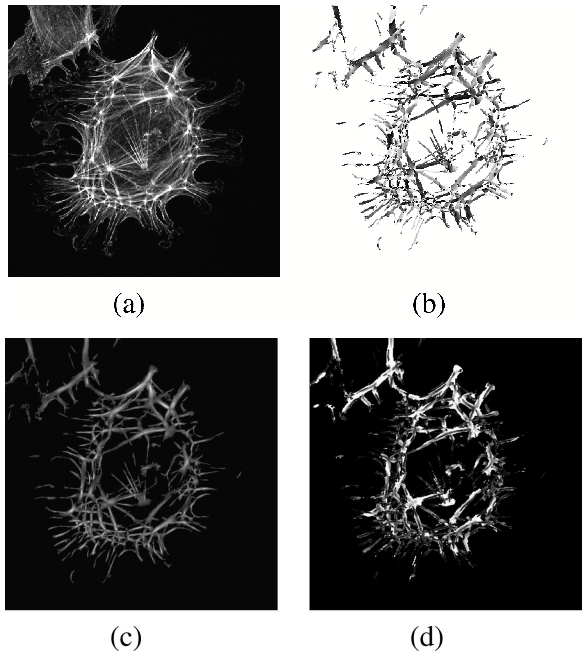


Figure A.12: *(a) Cytoskeleton image. (b) Orientation map. (c) Maximum response map. (d) Scale giving the maximum response (c) at orientation (b).*

images. In the chapter, we have discussed the ways to handle and minimize these numerical errors.

Our algorithm, particularized to approximate derivatives of isotropic Gaussians, has been used as a framework to construct steerable filters. In particular, we have modified the B-spline standard template to make it more sensitive to elongated structures.

Finally, we have included, as a proof-of-concept, two examples of application: an edge preserving smoothing filter and a ridge detector. We have shown that our spatially variant B-spline filtering yields satisfactory results at denoising biomedical images of different modalities. In our experiments, our adaptive B-spline smoothing provides a better trade-off between removal of noise and edge degradation than anisotropic diffusion.

Then we have used our filter to perform ridge detection. We started by quantifying the improvement in sensitivity obtained by our optimal template over the standard B-spline one. Next, we have presented a novel, truly spatially variant ridge detector. The algorithm performs a local -per pixel- optimization on a coarse pre-computed ridge map, to find the non-integer scale that maximizes the response. We have shown that the performance of this efficient ridge detector is not significantly different from the classical brute force one. Finally, we have evaluated the accuracy of the approach using scale series of different resolutions. We obtained a statistically significant improvement when using high-resolution scales. This result demonstrates the added value of carrying out non-integer detection, for which our method is particularly well suited.

Bibliography

[1] W. H. O., "The world health report 2004 - changing history," World Health Organization, Tech. Rep., 2004.

[2] A. T. S. / E. R. S. Task Force, "Standards for the Diagnosis and Management of Patients with COPD," American Thoracic Society / European Respiratory Society, Tech. Rep., 2004. [Online]. Available: http://www.thoracic.org/go/copd

[3] C. R. UK, "Commonly diagnosed cancers worldwide," Cancer Research UK, Tech. Rep., 2005.

[4] P. J. Barnes, "New treatments for COPD," *Nature Reviews Drug Discovery*, vol. 1, no. 6, pp. 437–446, June 2002.

[5] R. Cersosimo, "Lung cancer: a review," *American Journal of Health-System Pharmacy*, vol. 59, no. 7, pp. 611–642, 2002.

[6] M. Karol, "Animal models of occupational asthma," *European Respiratory Journal*, vol. 7, no. 3, pp. 555–568, 1994.

[7] S. D. Shapiro, "Animal Models for COPD," *Chest*, vol. 117, no. 5, pp. 223S–227S, 2000.

[8] W. N. Rom, J. G. Hay, T. C. Lee, Y. Jiang, and K.-M. Tchou-Wong, "Molecular and Genetic Aspects of Lung Cancer," *American Journal of Respiratory and Critical Care Medicine*, vol. 161, no. 4, pp. 1355–1367, 2000.

[9] W. Russell and R. Burch, *The Principles of Humane Experimental Technique*. Hyperion Books, 1992.

[10] K. A. Johnson, C. Badea, L. Hedlund, and G. A. Johnson, "Imaging Techniques for Small Animal Imaging Models of Pulmonary Disease: Micro-CT," *Toxicologic Pathology*, vol. 35, no. 1, pp. 59–64, 2007.

[11] A. Postnov, K. Meurrens, H. Weiler, D. Van Dick, H. Xu, P. Terpstra, and N. De Clerck, "In vivo assessment of emphysema in mice by high resolution x-ray microtomography," *Journal of Microscopy*, vol. 220, no. 1, pp. 70–75, 2005.

[12] N. M. De Clerck, K. Meurrens, H. Weiler, D. Van Dyck, G. Vanhoutte, P. Terpstra, and A. A. Postnov, "High-Resolution X-ray Microtomography for the Detection of Lung Tumors in Living Mice," *Neoplasia*, vol. 6, no. 4, pp. 374–9, 2004.

[13] K. Ask, R. Labiris, L. Farkas, A. Moeller, A. Froese, T. Farncombe, G. McClelland, M. Inman, J. Gauldie, and M. Kolb, "Comparison between conventional and "clinical" assessment of experimental lung fibrosis," *Journal of Translational Medicine*, vol. 6, no. 1, p. 16, 2008.

[14] B. N. Jobse, J. R. Johnson, T. H. Farncombe, R. Labiris, T. D. Walker, S. Goncharova, and M. Jordana, "Evaluation of allergic lung inflammation by computed tomography in a rat model in vivo," *European Respiratory Journal*, vol. 33, no. 6, pp. 1437–1447, 2009.

[15] C. Badea, L. W. Hedlund, and G. A. Johnson, "Micro-CT with respiratory and cardiac gating," *Medical Physics*, vol. 31, no. 12, pp. 3324–3329, 2004.

[16] N. L. Ford, H. N. Nikolov, C. J. D. Norley, M. M. Thornton, P. J. Foster, M. Drangova, and D. W. Holdsworth, "Prospective respiratory-gated micro-CT of free breathing rodents," *Medical Physics*, vol. 32, no. 9, pp. 2888–2898, 2005.

[17] E. Namati, D. Chon, J. Thiesse, E. A. Hoffman, J. de Ryk, A. Ross, and G. McLennan, "In vivo micro-CT lung imaging via a computer-controlled intermittent iso-pressure breath hold (IIBH) technique," *Physics in Medicine and Biology*, vol. 51, no. 23, pp. 6061–6075, 2006.

[18] D. Cavanaugh, E. L. Travis, R. E. Price, G. Gladish, R. A. White, M. Wang, and D. D. Cody, "Quantification of bleomycin-induced murine lung damage in vivo with micro-computed tomography," *Academic Radiology*, vol. 13, no. 12, pp. 1505 – 1512, 2006.

[19] A. M. Houghton, M. Mouded, and S. D. Shapiro, "Common origins of lung cancer and COPD," *Nature Medicine*, vol. 14, no. 10, pp. 1023–1024, October 2008.

[20] J. H. T. Bates and C. G. Irvin, "Measuring lung function in mice: the phenotyping uncertainty principle," *Journal of Applied Physiology*, vol. 94, no. 4, pp. 1297–1306, 2003.

[21] H. F. Lakatos, H. A. Burgess, T. H. Thatcher, M. R. Redonnet, E. Hernady, J. P. Williams, and P. J. Sime, "Oropharyngeal aspiration of a silica suspension produces a superior model of silicosis in the mouse when compared to intratracheal instillation," *Experimental Lung Research*, vol. 32, no. 5, pp. 181–199, 2006.

[22] A. R. Froese, K. Ask, R. Labiris, T. Farncombe, D. Warburton, M. D. Inman, J. Gauldie, and M. Kolb, "Three-dimensional computed tomography imaging in an animal model of emphysema," *European Respiratory Journal*, vol. 30, no. 6, pp. 1082–1089, 2007.

[23] W. Kalender, *Computed Tomography: Fundamentals, System Technology, Image Quality, Applications*. John Wiley and Sons, 2005.

[24] M. Paulus, S. Gleason, S. Kennel, P. Hunsicker, and D. Johnson, "High resolution X-ray computed tomography: An emerging tool for small animal cancer research," *Neoplasia*, vol. 1-2, pp. 62–70, January 2000.

[25] J. G. Ravenel, E. M. Scalzetti, W. Huda, and W. Garrisi, "Radiation Exposure and Image Quality in Chest CT Examinations," *American Journal of Roentgenology*, vol. 177, no. 2, pp. 279–284, 2001.

[26] N. L. Ford, A. R. Wheatley, D. W. Holdsworth, and M. Drangova, "Optimization of a retrospective technique for respiratory-gated high speed micro-ct of free-breathing rodents," *Physics in Medicine and Biology*, vol. 52, no. 19, pp. 5749–5769, 2007.

[27] D. Dreyfuss and G. Saumon, "Ventilator-induced Lung Injury . Lessons from Experimental Studies," *American Journal of Respiratory and Critical Care Medicine*, vol. 157, no. 1, pp. 294–323, 1998.

[28] G. B. Allen, B. T. Suratt, L. Rinaldi, J. M. Petty, and J. H. T. Bates, "Choosing the frequency of deep inflation in mice: balancing recruitment against ventilator-induced lung injury," *American Journal of Physiology - Lung Cellular and Molecular Physiology*, vol. 291, no. 4, pp. 710–717, 2006.

[29] C. G. Irvin and J. H. T. Bates, "Measuring the lung function in mice: The challenge of size," *Respiratory Research*, vol. 4, no. 1, p. 4, 2003.

[30] W. Mitzner, C. Tankersley, L. K. A. Lundblad, A. Adler, C. G. Irvin, and J. H. T. Bates, "Interpreting Penh in mice," *Journal of Applied Physiology*, vol. 94, no. 2, pp. 828–832, 2003.

[31] Z. Hantos, B. Daróczy, B. Suki, S. Nagy, and J. J. Fredberg, "Input impedance and peripheral inhomogeneity of dog lungs," *Journal of Applied Physiology*, vol. 72, no. 1, pp. 168–178, 1992.

[32] N. L. Ford, E. L. Martin, J. F. Lewis, R. A. W. Veldhuizen, M. Drangova, and D. W. Holdsworth, "In vivo characterization of lung morphology and function in anesthesized free-breathing mice using micro-computed tomography," *Journal of Applied Physiology*, vol. 102, no. 5, pp. 2046–2055, 2007.

[33] S. Hu, E. Hoffman, and J. Reinhardt, "Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 6, pp. 490–498, 2001.

[34] I. Sluimer, M. Prokop, and B. van Ginneken, "Toward automated segmentation of the pathological lung in CT," *IEEE Transactions on Medical Imaging*, vol. 24, no. 8, pp. 1025– 1038, August 2005.

[35] J. Lee, J. B. Seo, N. Kim, S. O. Park, H. Lee, Y. G. Shin, and S.-H. Kim, "Novel level-set based segmentation method of the lung at HRCT images of diffuse interstitial lung disease (DILD)," in *SPIE Medical Imaging 2009: Image Processing*, vol. 7259, no. 1, 2009.

[36] C. Jauquicoa, "Cuantificación automática de enfisema en imágenes de tomografía axial computerizada," Master's thesis, Tecnun, 2008.

[37] A. Zijdenbos, B. Dawant, R. Margolin, and A. Palmer, "Morphometric analysis of white matter lesions in MR images: Method and validation," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 716–724, December 1994.

[38] T. Rohlfing, R. Brandt, R. Menzel, D. B. Russakoff, and C. R. Maurer, Jr., "Quo vadis, atlas-based segmentation?" in *The Handbook of Medical Image Analysis – Volume III: Registration Models*. Kluwer Academic / Plenum Publishers, 2005, ch. 11, pp. 435–486.

[39] S. Klein, M. Staring, and J. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2879–2890, December 2007.

[40] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *International Journal of Computer Vision*, vol. 22, pp. 61–79, 1995.

[41] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, *ITK Software Guide*. Kitware, 2005.

[42] A. Chaturvedi and Z. Lee, "Three-dimensional segmentation and skeletonization to build an airway tree data structure for small animals," *Physics in Medicine and Biology*, vol. 50, no. 7, pp. 1405–1419, 2005.

[43] L. Shi, J. Thiesse, G. McLennan, E. A. Hoffman, and J. M. Reinhardt, "Three-dimensional murine airway segmentation in micro-CT images," in *SPIE Medical Imaging 2007: Physiology, Function, and Structure from Medical Images*, vol. 6511, 2007, p. 651105.

[44] T. Schlathoelter, C. Lorenz, I. C. Carlsen, S. Renisch, and T. Deschamps, "Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy," in *SPIE Medical Imaging 2002: Image Processing*, vol. 4684, 2002, pp. 103–113.

[45] T. Bülow, C. Lorenz, and S. Renisch, "A general framework for tree segmentation and reconstruction from medical volume data," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2004*, vol. 3216, 2004, pp. 533–540.

[46] B. van Ginneken, W. Baggerman, and E. M. van Rikxoort, "Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2008*, 2008, pp. 219–226.

[47] A. P. Kiraly, W. E. Higgins, G. McLennan, E. A. Hoffman, and J. M. Reinhardt, "Three-dimensional human airway segmentation methods for clinical virtual bronchoscopy," *Academic Radiology*, vol. 9, no. 10, pp. 1153 – 1168, 2002.

[48] D. Aykac, E. Hoffman, G. McLennan, and J. Reinhardt, "Segmentation and analysis of the human airway tree from three-dimensional

X-ray CT images," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 940–950, 2003.

[49] C. I. Fetita, F. Preteux, C. Beigleman-Aubry, and P. Grenier, "Pulmonary airways: 3-D reconstruction from multislice CT and clinical investigation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 11, pp. 1353–1364, 2004.

[50] J. Tschirren, E. Hoffman, G. McLennan, and M. Sonka, "Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans," *IEEE Transactions on Medical Imaging*, vol. 24, no. 12, pp. 1529–1539, December 2005.

[51] M. W. Graham, J. D. Gibbs, and W. E. Higgins, "Robust system for human airway-tree segmentation," in *SPIE Medical Imaging 2008: Image Processing*, vol. 6914, 2008, pp. 1 –18.

[52] N. L. Ford, M. M. Thornton, and D. W. Holdsworth, "Fundamental image quality limits for microcomputed tomography in small animals," *Medical Physics*, vol. 30, no. 11, pp. 2869–2877, 2003.

[53] R. A. Collins, M. Ikegami, T. R. Korfhagen, J. A. Whitsett, and P. D. Sly, "In Vivo Measurements of Changes in Respiratory Mechanics with Age in Mice Deficient in Surfactant Protein D," *Pediatric Research*, vol. 53, no. 3, pp. 463–467, 2003.

[54] S. Tomioka, J. H. T. Bates, and C. G. Irvin, "Airway and tissue mechanics in a murine model of asthma: alverolar capsule vs. forced oscillations," *Journal of Applied Physiology*, vol. 93, no. 1, pp. 263–270, 2002.

[55] L. K. A. Lundblad, J. Thompson-Figueroa, G. B. Allen, L. Rinaldi, R. J. Norton, C. G. Irvin, and J. H. Bates, "Airways Hyperresponsiveness in Allergically Inflamed Mice: The Role of Airway Closure," *American Journal of Respiratory and Critical Care Medicine*, vol. 175, no. 8, pp. 768–774, 2007.

[56] L. Vincent, "Morphological grayscale reconstruction in image analysis: applications and efficient algorithms," *IEEE Transactions on Image Processing*, vol. 2, no. 2, pp. 176–201, 1993.

[57] T. Deschamps and L. D. Cohen, "Fast extraction of minimal paths in 3D images and applications to virtual endoscopy," *Medical Image Analysis*, vol. 5, pp. 281–299, 2001.

[58] E. Bribiesca, "An easy measure of compactness for 2D and 3D shapes," *Pattern Recognition*, vol. 41, no. 2, pp. 543–554, 2008.

[59] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, no. 6, pp. 80–83, December 1945.

[60] H. Park, P. Bland, and C. Meyer, "Construction of an abdominal probabilistic atlas and its application in segmentation," *IEEE Transactions on Medical Imaging*, vol. 22, no. 4, pp. 483–492, April 2003.

[61] B. Dawant, S. Hartmann, J. Thirion, F. Maes, D. Vandermeulen, and P. Demaerel, "Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations. I. Methodology and validation on normal subjects," *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 909–916, Oct 1999.

[62] M. Lorenzo-Valdes, G. I. Sanchez-Ortiz, A. G. Elkington, R. H. Mohiaddin, and D. Rueckert, "Segmentation of 4D cardiac MR images using a probabilistic atlas and the EM algorithm," *Medical Image Analysis*, vol. 8, no. 3, pp. 255–265, September 2004.

[63] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer, Jr., "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, pp. 1428–1442, 2004.

[64] J. Sethian, *Level Set Methods and Fast Marching Methods.* Cambridge University Press, 1999.

[65] S. Beucher and F. Meyer, *The morphological approach to segmentation: The watershed transformation. In: Mathematical Morphology in Image Processing.*, E. Dougherty, Ed. Marcel Dekker,New York, 1992.

[66] A. Klein, B. Mensh, S. Ghosh, J. Tourville, and J. Hirsch, "Mindboggle: Automated brain labeling with multiple atlases," *BMC Medical Imaging*, vol. 5, no. 1, October 2005.

[67] C. Svarer, K. Madsen, S. Hasselbach, L. Pinborg, S. Haugbol, V. Frojaer, S. Holm, O. B. Paulson, and G. Knudsen, "MR-based automatic delineation of interest in human brain PET images using probability maps," *NeuroImage*, vol. 24, no. 4, pp. 969–979, February 2005.

[68] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, October 2006.

[69] T. Rohlfing, R. Brandt, C. R. Maurer, Jr., and R. Menzel, "Bee brains, B-splines and computational democracy: Generating an average shape atlas," in *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, L. Staib, Ed. Kauai, HI: IEEE Computer Society, Los Alamitos, CA, 2001, pp. 187–194.

[70] P. Kochunov, J. L. Lancaster, P. Thomson, R. Woods, J. Maziotta, J. Hardies, and P. Fox, "Regional Spatial Normalization: Toward an optimal target," *Journal for Computer Assisted Tomography*, vol. 25, no. 5, pp. 805–816, 2001.

[71] T. Rohlfing and C. R. Maurer, Jr., "Multi-classifier framework for atlas-based image segmentation," *Pattern Recognition Letters,*, vol. 26, no. 13, pp. 2070–2079, October 2005.

[72] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[73] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. Wiley, 2004.

[74] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 7, pp. 903–921, July 2004.

[75] T. Rohlfing, D. Russakoff, and C. R. Maurer, Jr., "Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 983–994, August 2004.

[76] T. Rohlfing and C. R. Maurer, Jr., "Shape-based averaging," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 153–161, January 2007.

[77] S. Klein, U. van der Heide, I. Lips, M. van Vulpen, M. Staring, and J. Pluim, "Automatic Segmentation of the Prostate in 3D MR Images by Atlas Matching using Localized Mutual Information," *Medical Physics*, vol. 35, no. 4, pp. 1407–1417, April 2008.

[78] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, "Nonrigid multimodality image registration," in *Proceedings SPIE Medical Imaging 2001: Image Processing*, M. Sonka and K. M. Hanson, Eds., vol. 4322, 2001, pp. 1609–1620.

[79] M. Wu, C. Rosano, P. Lopez-Garcia, C. S. Carter, and H. J. Aizenstein, "Optimum template selection for atlas-based segmentation," *NeuroImage*, vol. 34, no. 4, pp. 1612–1618, February 2007.

[80] X. Artaechevarria, A. Munoz-Barrutia, and C. O. de Solórzano, "Efficient classifier generation and weighted voting for atlas-based segmentation: Two small steps faster and closer to the Combination Oracle," in *SPIE Medical Imaging: Image Processing*, 2008, p. 69141W.

[81] W. R. Crum, L. D. Griffin, D. G. Hill, and D. J. Hawkes, "Zen and the art of medical image registration: Correspondence, homology, and quality," *NeuroImage*, vol. 20, no. 3, pp. 1425–1437, November 2003.

[82] A. Roche, G. Malandain, and N. Ayache, "Unifying Maximum Likelihood approaches in medical image registration," in *in Medical Image Registration. International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging*, 2000, pp. 71–80.

[83] R. L. Gregg and R. D. Nowak, "Noise removal methods for high resolution MRI," in *IEEE 1997 Nuclear Science Symposium*, vol. 2, 1997, pp. 1117–1121.

[84] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hil, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, August 1999.

[85] S. Klein, M. Staring, K. Murphy, M. Viergever, and J. Pluim, "elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, 2009.

[86] G. Gerig, M. Jomier, and M. Chakos, "VALMET: A new validation tool for assesing and improving 3D object segmentation," in *Lecture*

Notes in Computer Science, vol. 2208. Springer-Verlag, 2001, pp. 516–523.

[87] "Internet Brain Segmentation Repository." [Online]. Available: http://www.cma.mgh.harvard.edu/ibsr/

[88] E. van Rikxoort, Y. Arzhaeva, and B. van Ginneken, "A multi-atlas approach to automatic segmentation of the caudate nucleus in MR brain images," in *3D Segmentation In The Clinic: A Grand Challenge*, T. Heimann, M. Styner, and B. van Ginneken, Eds., 2007, pp. 29–36.

[89] R. Hu and R. I. Damper, "A No Panacea Theorem for classifier combination," *Pattern Recognition*, vol. 41, no. 8, pp. 2665–2673, 8 2008.

[90] N. Suzuki, K. Ohta, T. Horiuchi, H. Takizawa, T. Ueda, M. Kuwabara, J. Shiga, and K. Ito, "T lymphocytes and silica-induced pulmonary inflammation and fibrosis in mice." *Thorax*, vol. 51, no. 10, pp. 1036–1042, 1996.

[91] H. J. Lee, J. M. Goo, N. R. Kim, M. A. Kim, D. H. Chung, K.-R. Son, H.-C. Kim, C. H. Lee, C. M. Park, E. J. Chun, and J.-G. Im, "Semiquantitative measurement of murine bleomycin-induced lung fibrosis in in vivo and postmortem conditions using microcomputed tomography: Correlation with pathologic scores-initial results," *Investigative Radiology*, vol. 43, no. 6, pp. 453–460, 2008.

[92] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861 – 874, 2006.

[93] J. W. Eaton, *GNU Octave Manual*. Network Theory Limited, 2002.

[94] U. Saffiotti, A. Williams, L. Daniel, M. Kaighn, Y. Mao, and X. Shi, *Silica and silica-induced lung diseases*. CRC Press, 1996, ch. Carcinogenesis by crystalline silica: Animal, cellular, and molecular studies.

[95] H. Gietema, A. Schilham, B. van Ginneken, R. van Klaveren, J. Lammers, and M. Prokop, "Monitoring of smoking-induced emphysema with multidetector-row computed tomography in a lung cancer screening setting: What is the minimum increase in emphysema scores required to distinguish real increase in extent of emphysema from interscan variability," *Radiology*, vol. 244, no. 3, pp. 890–897, 2007.

[96] S. Ito, E. P. Ingenito, K. K. Brewer, L. D. Black, H. Parameswaran, K. R. Lutchen, and B. Suki, "Mechanics, nonlinearity, and failure strength of lung tissue in a mouse model of emphysema: possible role of collagen remodeling," *Journal of Applied Physiology*, vol. 98, no. 2, pp. 503–511, 2005.

[97] K. R. Lutchen, Z. Hantos, F. Petak, A. Adamicza, and B. Suki, "Airway inhomogeneities contribute to apparent lung tissue mechanics during constriction," *Journal of Applied Physiology*, vol. 80, no. 5, pp. 1841–1849, 1996.

[98] WHO, "Who world health statistics," World Health Organization, Tech. Rep., 2008.

[99] "The definition of emphysema. report of a national heart, lung, and blood institute, division of lung diseases workshop," *Am Rev Respir Dis*, vol. 132, pp. 182–185, 1985.

[100] B. Celli, W. MacNee, and ATS/ERS Task Force, "Standards for the diagnosis and treatment of patients with copd: a summary of the ats/ers position paper," *Eur Respir J*, vol. 23, pp. 932–946, 2004.

[101] R. Hautamaki, D. Kobayashi, R. Senior, and S. Shapiro, "Requirement for macrophage elastase for cigarette smoke-induced emphysema in mice," *Science*, vol. 277, no. 5334, p. 2002, 1997.

[102] G. Finlay, M. O'Donnell, C. O'Connor, J. Hayes, and M. FitzGerald, "Elastin and collagen remodeling in emphysema. A scanning electron microscopy study." *The American Journal of Pathology*, vol. 149, no. 4, p. 1405, 1996.

[103] E. Lucey, R. Goldstein, P. Stone, and G. Snider, "Remodeling of alveolar walls after elastase treatment of hamsters. Results of elastin and collagen mRNA in situ hybridization," *American Journal of Respiratory and Critical Care Medicine*, vol. 158, no. 2, p. 555, 1998.

[104] B. Suki, K. Lutchen, and E. Ingenito, "On the progressive nature of emphysema: roles of proteases, inflammation, and mechanical forces," *American Journal of Respiratory and Critical Care Medicine*, vol. 168, no. 5, p. 516, 2003.

[105] A. Guerassimov, Y. Hoshino, Y. Takubo, A. Turcotte, M. Yamamoto, H. Ghezzo, A. Triantafillopoulos, K. Whittaker, J. Hoidal, and M. Cosio, "The development of emphysema in cigarette smoke-exposed mice

is strain dependent," *American Journal of Respiratory and Critical Care Medicine*, vol. 170, no. 9, p. 974, 2004.

[106] H. Parameswaran, A. Majumdar, S. Ito, A. Alencar, and B. Suki, "Quantitative characterization of airspace enlargement in emphysema," *Journal of Applied Physiology*, vol. 100, no. 1, p. 186, 2006.

[107] K. Chung, "Cytokines in chronic obstructive pulmonary disease," *European Respiratory Journal*, vol. 18, no. Supplement 34, p. 50S, 2001.

[108] R. Valentine, R. B. Rucker, C. E. Chrisp, and G. L. Fisher, "Morphological and biochemical features of elastase-induced emphysema in strain A/J mice," *Toxicology and Applied Pharmacology*, vol. 68, no. 3, pp. 451 – 461, 1983.

[109] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2009, ISBN 3-900051-07-0. [Online]. Available: http://www.R-project.org

[110] M. Kawakami, Y. Matsuo, K. Yoshiura, T. Nagase, and N. Yamashita, "Sequential and Quantitative Analysis of a Murine Model of Elastase-Induced Emphysema," *Biological & Pharmaceutical Bulletin*, vol. 31, no. 7, pp. 1434–1438, 2008.

[111] N. Ford, E. Martin, J. Lewis, R. Veldhuizen, D. Holdsworth, and M. Drangova, "Quantifying lung morphology with respiratory-gated micro-CT in a murine model of emphysema," *Physics in Medicine and Biology*, vol. 54, pp. 2121–2130, 2009.

[112] C. Onclinx, V. De Maertelaer, P. Gustin, and P. Gevenois, "Elastase-induced pulmonary emphysema in rats: comparison of computed density and microscopic morphometry," *Radiology*, vol. 241, no. 3, p. 763, 2006.

[113] G. Snider, E. Lucey, and P. Stone, "State of the art: Animal models of emphysema," *Am Rev Resp Dis*, vol. 133, pp. 149–169, 1986.

[114] R. Foronjy, B. Mercer, M. Maxfield, C. Powell, J. D'Armiento, and Y. Okada, "Structural emphysema does not correlate with lung compliance: lessons from the mouse smoking model," *Experimental Lung Research*, vol. 31, no. 6, pp. 547–562, 2005.

[115] E. Lucey, J. Keane, P. Kuang, G. Snider, and R. Goldstein, "Severity of Elastase-Induced Emphysema Is Decreased in Tumor Necrosis Factor-&agr; and Interleukin-1&bgr; Receptor-Deficient Mice," *Laboratory Investigation*, vol. 82, no. 1, pp. 79–85, 2002.

[116] J. P. de Torres, G. Bastarrika, J. P. Wisnivesky, A. B. Alcaide, A. Campo, L. M. Seijo, J. C. Pueyo, A. Villanueva, M. D. Lozano, U. Montes, L. Montuenga, and J. J. Zulueta, "Assessing the Relationship Between Lung Cancer Risk and Emphysema Detected on Low-Dose CT of the Chest," *Chest*, vol. 132, no. 6, pp. 1932–1938, 2007.

[117] D. O. Wilson, J. L. Weissfeld, A. Balkan, J. G. Schragin, C. R. Fuhrman, S. N. Fisher, J. Wilson, J. K. Leader, J. M. Siegfried, S. D. Shapiro, and F. C. Sciurba, "Association of Radiographic Emphysema and Airflow Obstruction with Lung Cancer," *American Journal of Respiratory and Critical Care Medicine*, vol. 178, no. 7, pp. 738–744, 2008.

[118] L. M. Coussens and Z. Werb, "Inflammation and cancer," *Nature*, vol. 420, no. 6917, pp. 860–867, December 2002.

[119] Y. Horio, A. Chen, P. Rice, J. A. Roth, A. M. Malkinson, and D. S. Schrump, "Ki-ras and p53 mutations are early and late events, respectively, in urethane-induced pulmonary carcinogenesis in A/J mice," *Molecular Carcinogenesis*, vol. 17, no. 4, pp. 217–223, 1996.

[120] D. D. Cody, C. L. Nelson, W. M. Bradley, M. Wislez, D. Juroske, R. E. Price, X. Zhou, B. N. Bekele, and J. M. Kurie, "Murine lung tumor measurement using respiratory-gated micro-computed tomography," *Investigative Radiology*, vol. 40, no. 5, pp. 263–269, 2005.

[121] D. A. McNeill, C. E. Chrisp, and G. L. Fisher, "Pulmonary adenomas in A/J mice treated with silica," *Drug and Chemical Toxicology*, vol. 13, no. 1, pp. 87–92, 1990.

[122] M. Unser, "Splines, a perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, November 1999.

[123] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, May 2004.

[124] K. G. Derpanis, E. T. H. Leung, and M. Sizintsev, "Fast scale-space feature representations by generalized integral images," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 4, September 16-19 2007, pp. 521–524.

[125] Z. Weiping and S. Huazhong, "Detection of cerebral vessels in MRA using 3D steerable filters," in *Proceedings of the 27th Annual Interantional Conference of the Engineering in Medicine and Biology Society*, September 1-4 2005, pp. 3249–3253.

[126] M. Unser, A. Aldroubi, and M. Eden, "On the asymptotic convergence of B-spline Wavelets to Gabor functions," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 864–872, March 1992.

[127] P. S. Heckbert, "Filtering by repeated integration," in *Proceedings of the International Conference on Computer Graphics and Interactive Techniques*, vol. 20 (4), August 1986, pp. 315–321.

[128] A. Muñoz-Barrutia, R. Ertlé, and M. Unser, "Continuous Wavelet Transform with arbitrary scales and $O(N)$ complexity," *Signal Processing*, vol. 82, no. 5, pp. 749–757, May 2002.

[129] K. N. Chaudhury, A. Muñoz-Barrutia, and M. Unser, "Fast adaptive elliptical filtering using box splines," in *Proceedings of the 2008 IEEE International Conference on Image Processing*, October 12-15 2008, pp. 785–788.

[130] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, September 1991.

[131] 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. 58A, no. 2, pp. 167–176, April 2004.

[132] A. Muñoz-Barrutia, T. Blu, and M. Unser, "Least-Squares image resizing using finite differences," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1365–1378, September 2001.

[133] M. Unser, A. Aldroubi, and S. J. Schiff, "Fast implementation of the Continuous Wavelet Transform with integer scales," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3519–3523, December 1994.

[134] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical recipes in C (2nd ed.): The art of scientific computing*, New York, NY, USA, 1992.

[135] D. Goldberg, "What every computer scientist should know about floating-point arithmetic," *ACM Computing Surveys*, 1991.

[136] M. Jacob and M. Unser, "Design of steerable filters for feature detection using Canny-like criteria," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1007–1019, August 2004.

[137] E. Eberly, R. Gardner, B. Morse, S. Pizer, and C. Scharlach, "Ridges for image analysis," *Journal of Mathematical Imaging and Vision*, vol. 4, no. 4, pp. 353–373, December 1994.

[138] P. Saha and J. Udupa, "Scale-based diffusive image filtering preserving boundary sharpness and fine structures," *IEEE Transactions on Medical Imaging*, vol. 20, no. 11, pp. 1140–1155, November 2001.

[139] H. Gubdjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magnetic Resonance in Medicine*, vol. 34, no. 6, pp. 910–914, December 1995.

[140] H. P. Hiriyannaiah, "X-ray computed tomography for medical imaging," *IEEE Signal Processing Magazine*, vol. 14, no. 2, pp. 42–59, March 1997.

[141] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proceedings of the International Conference on Image Processing*, vol. 3, September 2002, pp. 57–60.

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

[143] S. Ridler and S. Calvard, "Picture thresholding using an iterative selection method," *IEEE Transactions Systems, Man and Cybernetics*, vol. 8, no. 8, pp. 630–632, August 1978.

[144] N. I. Fisher, *Statistical Analysis of Circular Data*, C. U. Press, Ed., 1993.

[145] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, no. 2, 1998.

[146] R. P. Brent, *Computer Methods for Mathematical Computations*, Prentice-Hall, Ed., 1973.

Publications

PAPERS IN INTERNATIONAL JOURNALS

- X. Artaechevarria, D. Blanco, Daniel Pérez-Martín, Gabriel de Biurrun, Luis M. Montuenga, Juan Pablo de Torres, Javier J. Zulueta, Gorka Bastarrika, Arrate Muñoz-Barrutia, Carlos Ortiz-de-Solorzano, "Longitudinal study of a mouse model of chronic pulmonary inflammation using breath hold gated micro-CT", *submitted*.

- X. Artaechevarria, D. Blanco, Gabriel de Biurrun, Mario Ceresa, Daniel Pérez-Martín, Gorka Bastarrika, Juan Pablo de Torres, Javier J. Zulueta, Luis M. Montuenga, Carlos Ortiz-de-Solorzano, Arrate Muñoz-Barrutia, "A multiplatform approach for emphysema assessment in mice applied to the elastase model", *submitted*.

- A. Muñoz-Barrutia, X. Artaechevarria and C. Ortiz-de-Solorzano, "Spatially Variant Convolution with Scaled B-Splines", **IEEE Transactions on Image Processing**, vol 19(1), pp. 11-24, January 2010.

- X. Artaechevarria, D. Pérez-Martín, M. Ceresa, G. de Biurrun, D. Blanco, L. M. Montuenga, B. van Ginneken, C. Ortiz-de-Solorzano and A. Muñoz-Barrutia, "Airway Segmentation and Analysis for the Study of Mouse Models of Lung Disease using Micro-CT", **Physics in Medicine and Biology**, vol. 54(22), pp. 7009-7024, November 2009.

- X. Artaechevarria, A. Muñoz-Barrutia and C. Ortiz-de-Solorzano, "Combination Strategies in Multi-Atlas Image Segmentation: Application to Brain MR Data", **IEEE Transactions on Medical Imaging**, vol. 28(8), pp. 1266-77, August 2009.

PAPERS IN CONFERENCE PROCEEDINGS

- M. Ceresa, X. Artaechevarria, A. Muñoz-Barrutia and C. Ortiz-de-Solorzano, "Early emphysema detection in a large serie of histological images", *Focus on Microscopy 2010*, accepted for presentation.

- M. Ceresa, X. Artaechevarria, A. Muñoz-Barrutia and C. Ortiz-de-Solorzano, "Automatic Leakage Detection and Recovery for Airway Tree Extraction in Chest CT images", *IEEE International Symposium on Biomedical Imaging 2010*, accepted for presentation.

- X. Artaechevarria, D. Pérez-Martín, J. M. Reinhardt, A. Muñoz-Barrutia, and C. Ortiz-de-Solorzano, "Automated Quantitative Analysis of a Mouse Model of Chronic Pulmonary Inflammation using Micro X-ray Computed Tomography", In M. Brown, M. de Bruijne, B. van Ginneken, A. Kiraly, J.-M. Kuhnigk, C. Lorenz, J. R. McClelland, K. Mori, A. Reeves, and J. M. Reinhardt, eds., *Second International Workshop on Pulmonary Image Analysis*, pp. 115-124, London, UK, 2009.

- X. Artaechevarria, A. Muñoz-Barrutia, B. van Ginneken and C. Ortiz-de-Solorzano, "Fast Murine Airway Segmentation and Reconstruction in Micro-CT images",*SPIE Medical Imaging 2009*, vol. 7262, pp. 72620B, Orlando (FL), USA, 2009.

- X. Artaechevarria, A. Muñoz-Barrutia and C. Ortiz-de-Solorzano, "Efficient Classifier Generation and Weighted Voting for Atlas-based Segmentation: Two Small Steps Faster and Closer to the Combination Oracle", *SPIE Medical Imaging 2008*, vol 6914, pp. 69141W, San Diego (CA), USA, 2008, **Cum Laude Best Poster Award**.

- X. Artaechevarria, A. Muñoz-Barrutia and C. Ortiz-de-Solorzano, "Restoration of Biomedical Images using Locally Adaptive B-Spline Smoothing", *IEEE International Conference on Image Processing 2007*, vol. 2, pp. 425-428, San Antonio (TX), USA, 2007.

- X. Artaechevarria, A. Muñoz-Barrutia and C. Ortiz-de-Solorzano, "Restoration of Micro-CT Images using Locally Adaptive B-Spline Smoothing", *IEEE International Symposium on Biomedical Imaging 2007*, pp. 800-803, Arlington (VA), USA, 2007.

Curriculum Vitae

Xabier Artaechevarria studied Telecommunication Engineering at Tecnun, the School of Engineering of the University of Navarra in San Sebastian, Spain. He received the MSc degree with a thesis on frequency synchronization for mobile communications, developed at Siemens Communications, Munich, Germany. His academic record was awarded with a special mention at the 2006 National End of Degree Awards.

This PhD thesis was carried out at the Cancer Imaging Laboratory of the Center for Applied Medical Research (CIMA) of the University of Navarra in Pamplona, Spain, under the supervision of Dr. Muñoz-Barrutia and Dr. Ortiz-de-Solorzano. Two research stays at the Image Sciences Institute (Dr. van Ginneken's group), Utrecht, The Netherlands, and the Biomedical Imaging Institute of the University of Iowa (Dr. Reinhardt's group), USA, contributed to improve the research work.