Universitat de Girona

## NEW MEDICAL REGISTRATION AND SEGMENTATION TECHNIQUES USING INFORMATION-THEORETIC TOOLS

## Antoni BARDERA I REIG

ISBN: 978-84-692-1442-8 Dipòsit legal: GI-120-2009



## UNIVERSITAT DE GIRONA

PHD THESIS

# New Medical Registration and Segmentation Techniques using Information-theoretic Tools

Author: Anton BARDERA I REIG

*Advisors:* Dra. Imma BOADA I OLIVERAS Dr. Miquel FEIXAS I FEIXAS

July 18, 2008

# Contents

1	Intr	oductio	n		1
	1.1	Motiva	ation		1
	1.2	Object	ive		3
	1.3	Thesis	Outline		4
	1.4	Main c	contributions		7
2	Bacl	kground	d		9
	2.1	Medic	al Imaging		9
		2.1.1	Medical Image Pipeline		9
			2.1.1.1 Data acquisition		10
			2.1.1.2 Data processing		12
			2.1.1.3 Information retrieval		13
		2.1.2	Medical Image Modalities		17
			2.1.2.1 Computed Tomography (CT) .		17
			2.1.2.2 Magnetic Resonance (MR)		18
			2.1.2.3 Nuclear tomographies (PET, SPE	ЕСТ)	19
	2.2	Inform	nation Theory		20
		2.2.1	Entropy		20
		2.2.2	Mutual Information		23
		2.2.3	Entropy Rate of a Markov Chain		24
		2.2.4	Important Inequalities		25
			2.2.4.1 Jensen's inequality		25
			2.2.4.2 The log-sum inequality		26
			2.2.4.3 Data processing inequality		27
		2.2.5	Entropy and Coding		27
	2.3	Image	processing		28
		2.3.1	Image Registration		28
			2.3.1.1 The image registration pipeline		29
		2.3.2	Similarity metrics		31
			2.3.2.1 Feature-based registration		31
			2.3.2.2 Voxel-based similarity measures		32
		2.3.3	Challenges in image registration		37
		2.3.4	Image Segmentation		38

		2.3.4.1 Global segmentation	39
		2.3.4.2 Local segmentation	41
		2.3.4.3 Challenges in image segmentation 4	45
3	Ima	ge Registration based on Global Lines 4	47
	3.1	Introduction	48
	3.2	Background and Related Work	49
		3.2.1 Information-theoretic measures	49
		3.2.2 Histogram estimation	50
		3.2.2.1 Parzen window estimation	50
		3.2.2.2 Binning estimation	51
		3.2.3 MI-based image registration with spatial information 5	52
	3.3	Uniformly Distributed Global Lines for Image Registration 5	53
	3.4	Histogram Estimation using Global Lines	55
	3.5	Generalization of Mutual Information-based Measures 5	58
		3.5.1 High-dimensional normalized mutual information 5	58
		3.5.2 Implementation using global lines	50
	3.6	Results and Discussion	51
	3.7	Conclusions	53
4	Ima	ge Registration by Compression 6	55
	4.1	Introduction	55
	4.2	The Similarity Metric	56
	4.3	Registration based on the Similarity Metric	57
	4.4	Kolmogorov Approach	59
		4.4.1 Image compressor-based registration	59
		4.4.2 Text compressor-based registration	71
	4.5	Shannon Approach	72
	4.6	Results and Discussion	73
		4.6.1 Image compression-based method	73
		4.6.2 Text compression-based method	77
		4.6.3 Entropy-based method	77
	4.7	Conclusions	79
5	Mul	tiresolution Image Registration Based on Tree Data Structures	31
	5.1	Introduction	31
	5.2	Previous Work	33
		5.2.1 Hierarchical data structures	33
		5.2.2 MI-based partitioning algorithm	34
	5.3	Multiresolution Image Registration	35
	5.4	Pixel-based Image Registration	36
	5.5	Node-based Registration	39
		5.5.1 Metric computation	39
		5.5.2 Overlapping between nodes	<del>)</del> 1

	5.6	Results and Discussion	93
		5.6.1 MR-CT tests	93
		5.6.2 MR-PET tests	96
	5.7	Conclusions	97
6	Ima	as Segmentation using Information Pottlaneak Mathed	01
U	1111a	Introduction	01
	0.1	Introduction Dettlengels Method	01
	0.2 6.2	Split and Marga Algorithm	03
	0.5	Spint-and-Merge Argonunin $\dots$ 1	04
		0.3.1 Spitting	00
	<i>C</i> <b>A</b>	6.3.2 Merging	10
	6.4	Histogram Clustering Algorithm	10
	6.5	Registration-based Segmentation	12
		6.5.1 One-sided clustering algorithm	13
		6.5.2 Co-clustering algorithm	14
		6.5.3 Results and discussion	15
	6.6	Conclusions	18
7	Ima	ge Segmentation using Excess Entropy 1	21
	7.1	Introduction	21
	7.2	Background and Related Work	22
		7.2.1 Information-Theoretic Measures	22
		7.2.2 Image Segmentation based on Thresholding	23
	73	Snatial structure of an image	24
	74	Thresholding based on Excess Entropy	26
	75	Results and Discussion	20
	7.6	Conclusions 1	34
	7.0		54
8	Con	clusions and Future Work 1	37
	8.1	Conclusions	37
	8.2	Future Work	38
	8.3	Publications	39
	Stor	Studios o tool to Assist Asuto Studios Diagnosis	15
A		Introduction	45
	A.1	Mathad	43
	A.Z		40
		A.2.1 PWI-DWI registration	4/
		A.2.2 Segmentation	49
		A.2.2.1 DWI segmentation	49
		A.2.2.2 PWI segmentation	50
		A.2.3 Volume assessment	51
	A.3	Visualization	52
	A.4	Evaluation	53
	A.5	Conclusions	54

A Ne	w Method for Hematoma and Edema Quantification using CT	155
<b>B</b> .1	Introduction	155
B.2	Related Work	156
B.3	Segmentation Framework	157
	B.3.1 Hematoma segmentation	157
	B.3.2 Edema segmentation	159
B.4	Results and Discussion	163
B.5	Conclusions	166
	A Ne B.1 B.2 B.3 B.4 B.5	A New Method for Hematoma and Edema Quantification using CT         B.1       Introduction         B.2       Related Work         B.3       Segmentation Framework         B.3.1       Hematoma segmentation         B.3.2       Edema segmentation         B.4       Results and Discussion         B.5       Conclusions

# **List of Figures**

2.1	Visualization pipeline	10
2.2	Grid structures	11
2.3	Rendering techniques	16
2.4	CT images	18
2.5	MR images	19
2.6	NT images	20
2.7	Grouping property of the entropy	21
2.8	Binary entropy	22
2.9	Venn diagram of a discrete channel	24
2.10	The main components of the registration framework	29
2.11	Joint histogram plots	35
3.1	Weights for 2D linear interpolation	52
3.2	Probability density function of the global line directions	54
3.3	Global lines in a bounding box	54
3.4	PWI-DWI and CT-MR test image pairs	55
3.5	Probability density function estimation of the histogram image	56
3.6	NMI plots using the global line method	57
3.7	Computational cost of the global line method	58
3.8	Interpolation artifacts	59
3.9	High dimensional data generation from global lines	61
3.10	CT-MR registration results for different $L$ values $\ldots \ldots \ldots$	64
4.1	Image compressor method scheme	68
4.2	Text compressor method scheme	68
4.3	Entropy rate method scheme	69
4.4	Lena image and its R, G and B channels	70
4.5	Different scan paths on the overlap area between two images	72
4.6	Compression and distance results of the Lena image	74
4.7	Synthetic MR T1 and T2 and real MR-CT pairs of test images	75
4.8	Fused test images	75
4.9	Compression and distance results of the synthetic MR images	76
4.10	Compression and distance results of MR and CT images	78

4.11	Compression and distance results with the text compressor	78
4.12	Shannon-based measure results	79
5.1	Image simplifications	84
5.2	Main steps to obtain the simplified images	86
5.3	Test images from the Vanderbilt database	86
5.4	MR an CT simplified images using quadtree and BSP partitions .	87
5.5	Registration results for the simplifications of the MR-CT pair	88
5.6	Overlapping area between the floating and the reference images .	89
5.7	Algorithm of the similarity measure computation	90
5.8	Representations of the initial reference and floating images	91
5.9	Reference and floating node intersections	92
5.10	Region overlapping	93
5.11	Degenerate cases	94
5.12	Steps of the polygon area computation	95
5.13	MR-CT registration results considering translation	96
5.14	MR-CT registration results considering rotation	97
5.15	MR-PET registration results considering translation	98
5.16	MR-PET registration results considering rotation	99
(1	T. Commenting of the second state of the transfer the transfer the second state of the transfer the second state of the second	104
0.1	Information channel between regions and intensity bins	104
6.2	Pertition of the Long income	105
0.3	Partition of the Lena image	107
6.4	Partition of the Hematoma image	108
6.5	Results of the split-and-merge algorithm for the test images	110
6.6	Results of the split-and-merge algorithm for different images	111
6.7	Results of the histogram clustering algorithm for the test images .	112
6.8	One-sided clustering algorithm	113
6.9	Information channel between the intensities of images A and B	114
6.10	Co-clustering algorithm	115
6.11	Results for a synthetic MR image with 3% of noise	117
6.12	Results of the one-sided algorithm for a MRI with 7% of noise	118
6.13	Results for the MR-CT test images	119
6.14	$MIR$ and NMI vs the number of clusters $\ldots \ldots \ldots \ldots$	120
71	Two representations of the excess entropy measure	124
7.2	Global line generation and L-blocks	124
73	Synthetic images and their excess entrony values	124
7.5 7.4	Preliminary experiment	120
7. <del>-</del>	Excess entrony values for different threshold levels	127
7.5 7.6	Block diagram of the excess entrony based thresholding method	120
ט.י ד ד	Histogram of the T1 image with $20\%$ of poiso	120
1.1 7 0	T1 and T2 toot images and the corresponding accounted images	129
7.ð	Comparison of the comparted images with the around to the	130
7.9	Comparison of the segmented images with the ground truth	131

7.10	Mammographic images with their segmentations	135
7.11	MR and CT images with their corresponding segmentations	135
A.1	Block diagram of the new method for acute stroke assessment	147
A.2	Registration evaluation plots	148
A.3	The original DWI image before and after segmentation	150
A.4	The original PWI image before and after segmentation	151
A.5	Captures of the DWI and PWI segmentation processes	152
A.6	Visualizations obtained from the DWI-PWI registered model	153
B.1	Block diagram of the segmentation process	157
B.2	CT scans with a hematoma and the surrounding edema	158
B.3	Original hematoma image and connected threshold segmentation .	159
B.4	CT image and auxiliary images for edema segmentation	161
B.5	The proposed propagation term	162
B.6	Contours for different k values	162
<b>B</b> .7	Hematoma plus edema volume comparison	164
<b>B.</b> 8	${\it CR}$ ratio and ${\it MR}$ ratio versus the edema plus hematoma volume .	166

# **List of Tables**

2.1 2.2	The main image registration similarity metrics	37 45
3.1 3.2	The mean of the error for NMI and $NMI(X^2, Y^2)$ The mean, median and maximum of the distances between VOIs .	62 62
5.1 5.2	Percentage of regions of the simplified MR and CT images Joint histogram computation time	87 97
7.1	Distribution of brain tissues in the different clusters	133
A.1	Evaluation of the segmentation results	154
B.1 B.2	Interoperator variance	164 165

## **Chapter 1**

## Introduction

## 1.1 Motivation

Over the last decades, technological advances have led to great transformations in medical applications. Current capture devices are able to scan any part of the body with millimetric precision, providing accurate information of the patient's anatomy. Not surprisingly, medical image processing has became a priority task at the present time in any medical scenario. However, despite the advantages introduced by these new technologies, there are still many aspects that need further development. A lot of research is being done in different aspects such as the automation of medical image processing in order to reduce the manual intervention, the development of acceleration techniques to reduce the processing time required to deal with the large amount of data obtained from the devices, the improvement of image quality in order to facilitate a more accurate interpretation of the data and assist practitioners in diagnosis.

Amongst the different research topics related with medical image processing, our research is focused in in two of them, registration and segmentation:

• The **registration** process consists in aligning, in a common model, two or more datasets (or images) containing different information of a patient. Since the information from the two images is usually complementary, its integration provides valuable information for diagnosis. For instance, the registration of magnetic resonance (MR) images and positron emission to-mography (PET) images permits the experts to analyze the anatomy and functionality of internal structures simultaneously. Applications of medical image registration include diagnosis of cancer, cardiac studies, different neurological disorders including Alzheimer's disease or schizophrenia, etc. Registration is also utilized to prepare and simulate complex surgical procedures, to verify a treatment by comparing the pre- and post-intervention images or to analyze growth monitoring.

Image registration is based on the search of a transformation that maps each point of one image to the corresponding anatomical point of another image by optimizing a certain similarity measure. The four principal elements involved in the registration process are: the similarity metric, the optimizer, the transformation, and the interpolator [113].

The new image modalities, which give to the specialists complementary information of the patient's disease, have led the registration process to an emerging focus of research. Current investigations are focused on the development of new methods that improve the results in terms of robustness, accuracy, and computational time.

• The **segmentation** process can be described as the identification of structures (for instance, representing organs or pathologies) in an image. It consists in subdividing an image into its constituent parts, a significant step towards image understanding [55]. The demand for automatic or semiautomatic image interpretation tools designed to manage the large amount of information made available by current imaging technologies has led segmentation to be a main focus of research in medical image processing.

Image segmentation plays an important role in medical imaging applications such as the delineation of anatomical structures, the identification of lesions or tumors, and the monitoring of a lesion volume. Therefore, the accuracy of these methods is a critical issue, since the diagnosis or treatment decisions are taken from their results.

Segmenting anatomical structures on medical images is a challenging problem due to the complexity and variability of human anatomy. Other difficulties appear in practice, especially when dealing with inherently noisy images or with low spatial resolution. Even though most of the structures are easily detected by an expert, they cannot always be detected automatically. Therefore, in most cases user intervention is required to separate adjacent structures of similar image intensities.

Most of the contributions of this thesis are based on Information Theory (IT). This theory deals with the transmission, storage and processing of information [138] and it is used in fields such as physics, computer science, mathematics, statistics, economics, biology, linguistics, neurology, learning, computer graphics, etc. [39, 48]. In particular, IT provides powerful tools for medical image registration [91, 156, 144] and segmentation [118, 74, 18, 61]. In this thesis, the two most basic information-theoretic measures (*Shannon entropy* and *mutual information*) are used to propose new registration and segmentation techniques. Entropy expresses the information content or the uncertainty of a random variable, and is used as a measure of diversity of a probability distribution. Mutual information expresses the information transfer in a communication channel, and is used as a measure of dependence or correlation between two random variables. Both measures can be utilized to capture different aspects of the image registration and segmentation processes.

Despite the large number of algorithms dealing with image registration and segmentation, a lot of research can be done to improve the performance, accuracy, and robustness of these techniques. Although the methods proposed throughout the thesis can be applied to any kind of 2D or 3D images, we focus our interest in medical images due to, in part, their great impact in the improvement of the quality of life.

## 1.2 Objective

The main goal of this thesis is the development of new registration and segmentation techniques that support and enhance computer aided diagnosis tools. To reach this objective we aim to:

1. Improve the registration process in terms of accuracy, robustness and performance.

One of the key aspects, and maybe the most important, of the registration process is the performance of the metric that quantifies the similarity between two or more images. Currently, the most used similarity metrics for multimodal image registration are both mutual information (MI) [91, 156] and normalized mutual information (NMI) [144]. However, MI-based methods have a number of well-known drawbacks, such as grid [154] and interpolation effects [91, 151]. Another limitation of these methods is that the computation of the joint histogram is habitually calculated from the correspondence between individual voxels in the two images, without taking into account any spatial information. To overcome these limitations, our main objective is to explore new similarity metrics and to deal with some aspects closely related to their computation.

2. Introduce new information-theoretic measures and algorithms for image segmentation.

Some information-theoretic based segmentation algorithms have been proposed in the literature [118, 74, 18, 61]. In this thesis, we will study new techniques, such as the information bottleneck algorithm and the excess entropy measure, proposed in the IT field, analyzing its suitability for 3D image processing. Basically, our objective is twofold. On the one hand, to explore new information channels involved in the information bottleneck method, which can incorporate spatial or multimodal information. On the other hand, to study the use the structural information contained in an image as a criterion to automatically define the threshold levels.

## 3. Evaluate the obtained results in a clinical environment.

An important step towards the development of new computer aided diagnosis tools is the validation of proposed techniques in a real environment. Therefore, an essential part of our work is devoted to the evaluation of the techniques in medical practice. Such an evaluation is carried out at the Josep Trueta's Hospital of Girona in collaboration with a team of clinical experts who determine the pathologies to study.

## 1.3 Thesis Outline

This dissertation is organized into eight chapters. The first two chapters are introductory and deal with previous work. The next three chapters are focused on different aspects of the registration process and the following two chapters tackle the segmentation process. Finally, a concluding chapter is presented. In more detail:

• Chapter 2: Background

In this chapter, an introduction of the topics needed to develop this thesis is outlined. First, since the application of the proposed method will be focused on medical applications, an introduction to the characteristics and the different types of medical images is presented. Second, a brief overview of the main concepts of Information Theory is given. Finally, an state of the art of the main image processing techniques is presented, giving special attention to the registration and segmentation processes.

• Chapter 3: Image Registration based on Global Lines and High-dimensional Normalized Mutual Information

In this chapter, a new stochastic sampling approach to compute a 3D image histogram using uniformly distributed random lines is proposed. This sampling strategy is introduced to compute the similarity between two images, optimizing the computational time and reducing some undesired artifacts. Moreover, the information channel used for image registration can be generalized using this sampling strategy, considering blocks of voxels instead of individual voxels. That is, while the standard methods only consider the channel between individual voxels, this generalization considers blocks of voxels, taking into account the spatial distribution of the samples. The ambiguity and high-dimensionality that appears when dealing with the voxel neighbourhood is solved using uniformly distributed random lines and reducing the number of bins of the images.

### • Chapter 4: Image Registration by Compression

In this chapter, a new framework for image registration based on compression is presented. The basic idea underlying our approach is the conjecture that two images are correctly registered when we can maximally compress one image given the information in the other. The contribution of this chapter is twofold. First, we show that the image registration process can be dealt with from the perspective of a compression problem. Second, we demonstrate that the similarity metric, introduced by Li et al., performs well in image registration. Two different versions of the similarity metric have been used: the Kolmogorov version, computed using standard real-world compressors, and the Shannon version, calculated from an estimation of the entropy rate of the images.

## • Chapter 5: Multiresolution Image Registration Based on Tree Data Structures

This chapter presents a new approach for image registration based on the partitioning of the source images in binary-space (BSP) and quadtree structures. These partitioned images have been obtained with a maximum mutual information gain algorithm. Two different approaches to compute the similarity metric are proposed: one uses simplified images and the other deals directly with the tree data structures. Multimodal registration experiments with regularly downsampled images, and BSP and quadtree partitioned images show an outstanding accuracy and robustness by using BSP images, since the grid effects are drastically reduced. The obtained results indicate that BSP partitioning can provide a suitable framework for multiresolution registration.

#### • Chapter 6: Image Segmentation using Information Bottleneck Method

In this chapter, several new image segmentation algorithms based on a hard version of the information bottleneck method are presented. With this method, a compact representation of a variable can be obtained from another variable and the definition of an information channel. First, the definition of an information channel between the image histogram and the image regions leads to two different image segmentation algorithms: split-and-merge and histogram clustering. In addition, the definition of an of another information channel between two multimodal images leads to a histogram bin clustering method. With this last approach, two different algorithms are proposed: a one-sided clustering algorithm, where only one image is segmented, and a co-clustering algorithm, where both images are simultaneously segmented. Several experiments on 2D and 3D images show the behavibour of the proposed algorithms.

## • Chapter 7: Segmentation using Excess entropy

In this chapter, a novel information-theoretic approach for thresholding-based segmentation that uses the excess entropy to measure the structural information of an image and to locate the optimal thresholds is presented. This approach is based on the conjecture that the optimal thresholding corresponds to the segmentation with maximum structure, i.e., maximum excess entropy. The contributions of this chapter are severalfold. First, the excess entropy as a measure of the spatial structure of an image is introduced. Second, an adaptive thresholding method based on the maximization of excess entropy

is introduced. Third, the use of uniformly distributed random lines to overcome the main drawbacks of the excess entropy computation is proposed. Finally, experimental results show the good performance of our segmentation method.

#### • Chapter 8: Conclusions

The conclusions of this thesis are presented, as well as some indications about our current and future research. In addition, the publications related with this thesis are summarized.

Moreover, two appendix chapters describe the two modules developed in the scope of this thesis to assist radiologist in diagnosis of two different pathologies. They have not been included as standard chapters since they deal with more applied problems, while the research contributions of them are not so important. In more detail:

#### • Appendix A: StarStroke: a tool to Assist Acute Stroke Diagnosis

In this chapter, a framework developed to assist and support treatment and diagnosis of acute stroke patients is presented. The current clinical study is based on manual image editing and manual segmentation of diffusion-weighted imaging (DWI) and perfusion weighted imaging (PWI) by an integrated framework with registration, segmentation, manipulation and visualization strategies specifically designed for the DWI/PWI processing. A robust and accurate technique to register DWI/PWI data which uses an acceleration strategy based on uniformly distributed random lines that reduces considerably the processing time is described. Then, a segmentation method based on PWI-DWI fused data to detect stroke and hypoperfusion areas that reduces the manual intervention of radiologist is proposed. With our approach, the clinical study of an acute stroke patient is reduced from 15/20 minutes to 5 minutes, including the manual editing.

## • Appendix B: A New Method for Hematoma and Edema Quantification using CT

In this chapter, an automated method for brain hematoma and edema segmentation, and volume measurement using computed tomography imaging is presented. The method combines a region growing approach to segment the hematoma and a level set segmentation technique to segment the edema. The main novelty of this method is the strategy applied to define the propagation function required by the level set approach. To evaluate the method, 18 patients with brain hematoma and edema of different size, shape and location were selected. The obtained results demonstrate that the proposed approach provides objective and reproducible segmentations that are similar to the results obtained manually. Moreover, the processing time is reduced to 4 minutes, while the manual segmentation is about 10 minutes.

## **1.4 Main contributions**

In this section, the principal contributions of this thesis are described. We also indicate the papers related to each contribution.

The main contributions in medical image registration are:

- The introduction of a new sampling strategy based on uniformly distributed random lines. This sampling technique has been used to decrease the computational cost of image registration and to capture the spatial relations between individual samples. Based on [9].
- The incorporation of spatial information in the similarity metric. This has been done by generalizing the concept of mutual information considering voxel blocks instead of individual voxels, as the standard methods do. Based on [7].
- The proposal of a new similarity metric paradigm based on the idea that registration can be seen as a compression problem: the registration position is reached when then compression of one image is maximum when the other image is known. Experimental results show the sensitivity of this new paradigm. Based on [10].
- The study of the performance of different image simplification schemes in order to define a general multi-resolution framework. This preliminary study demonstrates that the binary space partition of a 3D image is an appropriate option for this general multiresolution scheme. Based on [11].

The main contributions in the medical image segmentation context are:

- A new general framework for image segmentation based on a hard version of the information bottleneck method. Three different algorithms are presented: split-and-merge, histogram clustering, and multimodal clustering. Based on [122, 8].
- The introduction of a new measure of image structure and its application to image segmentation. The maximization of this measure is the basis of a new thresholding method. Based on [12, 150].
- The development of a new application to assist the clinicians in acute stroke diagnosis. This application incorporates both registration and segmentation methods. Some of the methods presented in this thesis have been implemented within this software. Based on [108, 5].

## **Chapter 2**

## Background

In the last decades, image processing has acquired great importance due to the large development of imaging technology and computational capabilities. Medical imaging technologies have also developed similarly, enabling us to obtain images with more resolution and new image modalities that give complementary information to the specialists. These technological advances have meant that a great number of researchers focus their work on developing new methods in order to automatically process this large amount of data to make the clinician's work easier. In this context, information theory has been applied to develop a large variety of image processing methods. In this chapter, the basic concepts of this dissertation are introduced. First, a quick overview of medical image features is carried out. Then, the main information theory concepts are reviewed and, finally, the main information processing methods are presented, giving special attention to image registration and segmentation.

## 2.1 Medical Imaging

The medical imaging pipeline defines the different steps that are required to obtain medical images. This ranges from patient scanning to the representation of the medical images on the screen. In this section, all the steps of this pipeline are described. The most important medical image modalities and the basic nomenclature related to our research are also presented.

## 2.1.1 Medical Image Pipeline

Over the last decades, medical imaging has become standard for routine patient diagnosis care. The capabilities of medical imaging devices to scan functional and anatomical information of the internal structures of the patient and to represent it by images have made these images crucial in medical diagnosis and treatment.



Figure 2.1: The medical imaging pipeline describes the process from data acquisition to picture generation.

Currently, different medical imaging modalities can be obtained from medical scanning devices. However, despite the differences between these images, the process required to obtain them can be described by a common pipeline. In this section, the three steps of this pipeline: patient scanning (or image acquisition), data processing and information retrieval are described. In Figure 2.1, these three steps of the medical imaging pipeline are represented.

#### 2.1.1.1 Data acquisition

The medical imaging pipeline starts with a scanning process that captures the information of the patient using a specialized device. There are many different acquisition devices which are being constantly improved. Amongst them are computed tomography (CT), magnetic resonance (MR), position emission tomography (PET), and single photon emission computed tomography (SPECT). The information provided by each device is based on its own physical principles. Therefore, each type of data is a measure of a different property, either anatomical or functional, and it is provided at a specific spatial intensity resolution and scale with a given range of capture error (see Section 2.2)

Commonly, the acquired data are represented as a set S of samples  $(x, y, z, \mathbf{v})$ , where the value **v** corresponds to some property of the patient at a certain 3D lo-



Figure 2.2: Different grid structures.

cation (x, y, z). This property can be represented by a scalar, a vector or a tensor. In some cases, the measured property **v** changes along the time and, in this case, each sample is represented as  $(x, y, z, t, \mathbf{v})$ . Since a uniform standard format to represent them does not exist, the datasets are characterized in terms of their geometry and topology (see Figure 2.2). Following this characterization, datasets can be classified as:

- *Rectilinear*, when data samples are distributed on regular grids which partition the physical domain into identical rectangular hexahedral cells.
- *Curvilinear*, when data samples lie on regular grids defined in *computational space*, (i.e. the rectilinear grid defining the logical organization), and a warping function is defined to give a sample position in the physical space.
- *Unstructured or irregular*, when data points have no regular structure. Samples are given as a list of physical space location plus the measured field values. The cell connectivity is specified explicitly and the cells can be of an arbitrary shape, such as tetrahedra, hexahedra, and prisms.

Due to the regular distribution of the sampled data obtained from medical devices, the most widely used representation scheme in medical applications is the spatial enumeration model or *voxel model* [76], which is a rectilinear representation. The *voxel model* is based on the decomposition of the 3D space into a regular set of identical cubical cells, known as *voxels*, whose edges are parallel to the coordinate axis. This spatial structure allows for voxels to be directly represented by a 3D point (i, j, k), where  $1 \le i, j, k \le n$ , being *n* the number of voxels per axis from which the whole geometrical and topological information may be retrieved. Each voxel is associated with a set of values. According to the location of these values, two different approaches can be considered:

• The *cell approach*, which represents samples at the eight corner vertices of each voxel. The behavibour inside the voxel is estimated through an interpolation of the vertices values. There are many possible interpolation functions. One of the most common is the piecewise function known as *first order interpolation*, or *trilinear interpolation*.

• The *voxel approach* assigns one sample value at the center of the voxel. The inside of the voxel is considered homogeneous.

The majority of acquisition devices (CT, MRI, PET, etc.) scan data by following a planar distribution and so output sequences of slices (images) with a regular distribution are returned. Therefore, the definition of the voxel model from these data is not difficult, since it is only required to correctly set the slices on a 3D regular grid.

#### 2.1.1.2 Data processing

Acquisition devices convey increasingly vast and detailed information for clinical or research applications. The examination and interpretation of these data require some data processing techniques to communicate information more effectively and enhance its interpretation. To assist visual interpretation of medical images, the second step of the medical imaging pipeline, the data processing, includes different techniques which are applied according to the features of the original datasets and user requirements. Next, some of the most representative techniques are outlined:

• Resampling and restructuration

Each manufacturer and each acquisition modality has an in-house image structure and particular noise characteristics, as well as artifacts introduced by the particular capturing mechanisms. It is necessary, therefore, to resample and restructure data in order to define the representation scheme. Such an operation is specific to both particular datasets and the hardware used to generate the information.

• Filtering

Usually, images are affected by some undesirable effects as noise or artifacts, commonly produced in the acquisition process. Filters are applied in the images to minimize these effects. Amongst these filters, we highlight the gaussian filters, which have a blurring effect, and the anisotropic diffusion filters, which preserve the edges of the image.

• Registration

The goal of image registration is to determine the spatial transformation that will bring the images to be registered into correspondence. Such an operation is of special interest in clinical practice since it allows us to combine in a single model the information acquired with different devices, at different times or from different patients. Image registration allows us, for instance, to complement different data in order to complete the anatomical representation or combine functional information with anatomical representation. Such a capability is of special interest if we are interested in performing studies between different patients and carrying out studies along the time or intrapatient studies. • Segmentation and classification

The values in a volume dataset are used to represent different properties of matter. Segmentation is defined as the division of an image into coherent regions using some local image characteristic criteria. In medical imaging, segmentation is used for analyzing anatomical structure and tissue types, spatial distribution of function and activity, and pathological regions. Since segmentation requires classification of pixels, it is often treated as a pattern recognition problem and addressed with related techniques.

### 2.1.1.3 Information retrieval

In the last step of the pipeline, the information required by the doctor has to be projected onto the screen. Such information depends on user requirements and it can be an image, a set of parameters, the visualization of a region of interest, flows of a determined fluid, plots of selected features, ratios, etc. Although the user may only be interested in a set of numerical parameters, in most cases, information retrieval is by means of an image, requiring visualization techniques to obtain it.

At a broad level, the visualization process requires two different steps to be carried out: the mapping and the rendering. The mapping process transforms abstract data into geometric primitives like points, surfaces or voxels and their associated graphic attributes like color, transparency, etc. Mapping is the main core of this visualization process as it decides which geometric primitive should be generated and what should be assigned to its attributes.

In the context of medical imaging, the three main approaches in which mapping strategies can be grouped are:

• Slicing

Slicing reduces the 3D visualization into a 2D visualization problem. This technique considers volume samples distributed on a plane (or slice) that intersect the volume model. Although, the orientation of the plane can be any one in most cases, the one parallel to the coordinate planes is selected. To represent the data contained in the plane, different strategies can be used, color maps, contour extraction, etc. This method is used to obtain rapid exploration of volume data.

• Surface fitting

Surface fitting techniques generate the polygonal approximation of an isosurface, i.e. the surface that approximates the volume dataset points with a given property, known as the *isovalue*. This is a good approach for objects with sharply determined borders, like bones in CT, but it is inadequate for amorphous objects which are difficult to represent by thin surfaces. The most popular approach in this group is the marching cubes algorithm, proposed by Lorensen et al. [90].

#### • Direct volume rendering

When volumetric data are modeled using a surface rendering technique, a dimension of information is lost. In response to this, volume rendering techniques were developed to capture the entire 3D data in a 2D image. Volume rendering aims to visualize all the data and not only a subset. A direct volume renderer requires every sample of the model to be mapped to both opacity and color values. This is done with a transfer function which can be a simple ramp, a piecewise linear function or an arbitrary table. Once this assignment is carried out, a rendering technique determines how colors have to be composed to obtain the final image.

The rendering process generates the image by using the geometric primitives from the mapping process. Following the classification of mapping strategies, the rendering methods can be grouped in three approaches:

• Slice rendering approach

In clinical practice, this is the most popular approach, since it allows for a rapid exploration of the model. The data contained in a plane traversing the model is represented as a grey scale or color map. Although only 2D views are possible, this approach is used extensively and it is easy to obtain and interpret. An extension of this approach is the multiplanar reconstruction, which combines slices located in three different orientations, is used.

Surface rendering approach

In general, surfaces are represented as triangular meshes. In the simplest surface rendering form, the triangles are passed onto the renderer without specific normal vectors at the vertices. Then the renderer will apply the shader (e.g. flat, Gouraud or Phong shading). It is also possible to assign the normal vector at the triangle vertices. The normal is equal to the gradient vector of the isosurface. This gradient vector can be easily calculated by central differences at each vertex and then linear interpolation gives the gradient.

### • Volume rendering approach

The volume rendering approach displays data directly as a transparent cloudy object, without any intermediate conversion, assigning attributes like color, opacity or gradient vector to each volume element [82, 158, 159]. Volume rendering techniques require an illumination model to determine how the data volume generates, reflects, scatters, or occludes light. An analysis of the main proposed optical models is presented by Max [100].

Volume rendering techniques can be classified into three main groups: objectorder, image-order or domain-order. Object-order volume rendering techniques use a forward mapping scheme where the volume data are mapped onto the image plane. In image-order algorithms, a backward mapping scheme is used. Rays are cast from each pixel to the image plane through the volume data to determine the final pixel value. In a domain-based technique, spatial volume data are first transformed into an alternative domain, such as compression, frequency, and wavelet and then a projection is generated directly from that domain.

Some of the most representative volume rendering techniques are [102]:

- Volume Ray Casting. This technique can be derived directly from the rendering equation [72]. In this technique, a ray is generated for each desired image pixel. Using a simple camera model, the ray starts at the center of the projection of the camera and passes through the image pixel on the imaginary image plane floating in-between the camera and the volume to be rendered. The ray is clipped by the boundaries of the volume in order to save time. Then the ray is sampled at regular or adaptive intervals throughout the volume. The data is interpolated at each sample point, the transfer function applied to form an RGBA sample, the sample is composited onto the accumulated RGBA of the ray, and the process repeated until the ray exits the volume. The RGBA color is converted to an RGB color and deposited in the corresponding image pixel. The process is repeated for every pixel on the screen to form the completed image. It provides very high quality results and is usually considered to give the best image quality. Volume ray casting is classified as image-order algorithm, as the computation emanates from the output image, not the input volume data.
- Splatting. Splatting was proposed by Westover [159], and it works by representing the volume as an array of overlapping basis functions, commonly Gaussian kernels with amplitudes scaled by the voxel values. An image is then generated by projecting these basis functions to the screen. The screen projection of these radially symmetric basis functions can be efficiently achieved by the rasterization of a precomputed footprint lookup table. Here, each footprint table entry stores the analytically integrated kernel function along a traversing ray. A major advantage of splatting is that only voxels relevant to the image must be projected and rasterized. This can tremendously reduce the volume data that needs to be both processed and stored [103]. However, depending on the zooming factor, each splat can cover up to hundreds of pixels which need to be processed.
- Shear Warp. In this technique, the viewing transformation is transformed such that the nearest face of the volume becomes axis aligned with an off-screen image buffer with a fixed scale of voxels to pixels. The volume is then renderered into this buffer using the far more favourable memory alignment and fixed scaling and blending factors.



Figure 2.3: Different examples of image rendering techniques: (a) slicing, (b) surface rendering and (c) volume rendering.

Once all the slices of the volume have been rendered, the buffer is then warped into the desired orientation and scale in the displayed image [78]. This technique is relatively fast in software at the cost of less accurate sampling and potentially worse image quality compared to ray casting. There is memory overhead for storing multiple copies of the volume, enabling us to have near axis aligned volumes. This overhead can be mitigated using run length encoding.

- Texture Mapping. The use of 3D texture mapping was popularized by Cabral [23] for non-shaded volume rendering. The volume is loaded into texture memory and the hardware rasterizes polygonal slices parallel to the viewplane. The slices are then blended back to front. Commodity PC graphics cards are fast at texturing and can efficiently render slices of a 3D volume, with realtime interaction capabilities. These slices can either be aligned with the volume and rendered at an angle to the viewer, or aligned with the viewing plane and sampled from unaligned slices through the volume. Volume aligned texturing produces images of reasonable quality, though there is often a noticeable transition when the volume is rotated.

Some examples of the three basic rendering approaches are represented in Figure 2.3. It is important to remark that, when the visualization involves more than one feature per voxel, as occurs after a registration process, the visualization process varies considerably. Basically, two main strategies to solve this problem have been proposed. The first consists in fusing all the models into only one, combining the intensity values of all the images, and then the classical approaches can be applied. The second carries out the mapping step taking into account all the features and, at each step of the rendering process, it has to determine which information has to be rendered.

## 2.1.2 Medical Image Modalities

Medical imaging modalities are based on various physical phenomena, such as x-ray attenuation, magnetized hydrogen nuclei, and sound reflections. Each phenomenon evaluates a certain characteristic (water density, blood diffusion, etc.), that indirectly allows for the inspection of the internal body structures. The choice of the image modality depends on the part of the body to be explored, the patient disease, its cost, and some patient incompatibilities.

Imaging modalities can be divided into two categories:

• Structural (anatomical and morphological)

The main purpose of these modalities is to capture characteristics such as size, shape, texture, position, color, and composition of anatomical structures. Examples of this imaging modality are: computed tomography, magnetic resonance, X-ray, ultrasound, etc.

• Functional (physiological and biochemical)

These kinds of images measure some characteristics such as flow perfusion, metabolism, and chemistry. Usually, these images are used to study which parts of the body are activated by a certain stimulus. Examples of these imaging modalities are: PET, SPECT, fMR, etc.

A more detailed description of the most representative imaging modalities is now given.

#### 2.1.2.1 Computed Tomography (CT)

This modality consists in irradiating an area of the body with X-rays emissions from different angles. These emissions are back captured after crossing the body and the quantity of radiation absorbed is measured [15]. The quantity of emission is proportional to the type of crossed tissue.

The resolution of the final image depends on some parameters of the acquisition machine. The output is a set of 2D images, usually from 30 to 50 corresponding to axial, coronal or sagittal slices of the body. The resolution of images goes from  $256 \times 256$  to  $2048 \times 2048$ . The values are sampled at a distance ranging from 0.5 to 2 mm and the distance between planes goes from 1 to 10 mm.

The quantification of X-rays emission in each sample provides an integer number that represents the type of tissue corresponding to the real object at that point. The CT units of measure used are Housenfeld (Hu). The origin of the range values are four basic densities: air (-1000 Hu), fat (-300 to -100 Hu), water (0 Hu) and bone (around 200 Hu). The rest are calculated from them (muscle 10 to 70 Hu,...).

CT is used in applications in which the high intensity contrast of different structures is required. This type of image is used for all the body, like the brain, the colon, or the extremities. Thanks to the high contrast and resolution of the resulting image, different structures can be easily segmented, giving high quality



Figure 2.4: Different examples of CT images: (a) head, (b) abdominal and (c) thorax images.

3D reconstructions. One of the main applications is virtual colonoscopy, giving doctors the opportunity to navigate within the colon without have to introduce a catheter inside the patient. In Figure 2.4, some CT images from different parts of the body are shown.

#### 2.1.2.2 Magnetic Resonance (MR)

MR is based on the behaviour of atoms with an even number of protons and neutrons in a magnetic field [15], particularly hydrogen and oxygen atoms. Hydrogen is the most abundant element and with a higher concentration variation in the nervous system tissues. Traditional MR images reflect the density of hydrogen, generally in the human body in the form of water or fat. To be more exact, MR signal intensity reflects the density of mobile hydrogen nuclei modified by the chemical environment. Depending on the physical measure of this phenomenon, different image modalities, as T1, T2, or FLAIR, are obtained.

Modifications of the magnetic field lead to modalities different from the traditional magnetic resonance imaging:

- MRA (Magnetic Resonance Angiography) captures the presence of flows,
- fMRI (Functional Magnetic Resonance Imaging) measures functional aspects,
- DWI (Diffusion-Weighted Imaging) reflects microscopic random motion of water molecules,
- DTI (Diffusion Tensor Imaging), similarly than DWI, reflects microscopic random motion of water molecules assigning at each voxel a tensor instead of a single value, and
- PWI (Perfusion-Weighted Imaging) measures the passage of MR contrast through brain tissue.



Figure 2.5: Different MR image modalities: (a) MR-T1, (b) MR-T2, (c) FLAIR, (d) MR Angiography, (e) functional MRI, (f) diffusion weighted imaging, (g) diffusion tensor imaging, and (h) perfusion weighted imaging.

The images obtained by MR usually have a lower resolution than CT scans, and, for example, typical voxel dimensions are  $1 \times 1 \times 3 mm$ . The data associated to each point of the image represent tissue density, but they have arbitrary units. MR images are usually applied in the analysis of organs and soft tissues, since they differentiate clearly the tissues with slightly different densities. On the contrary, in these modalities, the contrast of large body structures, like bones, is not as high as CT images. Figure 2.5 illustrates different MR image modalities.

## 2.1.2.3 Nuclear tomographies (PET, SPECT)

This image modality is based on the measure of the distribution of a radioactive source in the body. A radioactive substance is first injected to the patient who is introduced into a device that detects the radiation emitted by the region of the body under study. It measures the density of the property during physiological activities and, therefore, it is used for the acquisition of functional data.

The two main nuclear tomography techniques are: SPECT (Spectral Positron Emission Computed Tomography) and PET (Positron Emission Tomography). The property measured is the metabolic activity. They give a stack of tridimensional images with a resolution of  $256 \times 256$  or  $512 \times 512$ . Each pixel corresponds to a real width of 2 to 10 mm. and the space between slices is about 3 mm. The units of measure are Curies (Ci). In Figure 2.6, different images of these modalities are plotted.



Figure 2.6: Different Functional image modalities: (a-b) PET and (c-d) SPECT.

## 2.2 Information Theory

In 1948, Claude Shannon published "A mathematical theory of communication" [138] which marks the beginning of information theory. In this paper, he defined measures such as entropy and mutual information<sup>1</sup>, and introduced the fundamental laws of data compression and transmission.

In this section, we present some basic concepts of information theory. A very good reference is the text by Cover and Thomas [39]. Other main references used in this thesis are Blahut [16], Lubbe [155] and Yeung [162].

### 2.2.1 Entropy

After representing a discrete information source as a Markov process, Shannon asks himself: "Can we define a quantity which will measure, in some sense, how much *information* is "produced" by such a process, or better, at what rate information is produced?" [138].

His answer is: "Suppose we have a *set of possible* events whose *probabilities* of occurrence are  $p_1, p_2, \ldots, p_n$ . These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?

If there is such a measure, say  $H(p_1, p_2, ..., p_n)$ , it is reasonable to require of it the following properties:

- 1. *H* would be continuous in the  $p_i$ .
- 2. If all the  $p_i$  are equal,  $p_i = \frac{1}{n}$ , then *H* should be a monotonic increasing function of *n*. With equally likely events there is more choice, or uncertainty, when there are more possible events.
- 3. If a choice is broken down into two successive choices, the original H should be the weighted sum of the individual values of H. The meaning of this is illustrated in Figure 2.7.

<sup>&</sup>lt;sup>1</sup>In Shannon's paper, the mutual information is called rate of transmission.



Figure 2.7: Grouping property of the entropy.

On the left, we have three possibilities  $p_1 = \frac{1}{2}$ ,  $p_2 = \frac{1}{3}$ ,  $p_3 = \frac{1}{6}$ . On the right, we first choose between two possibilities each with probability  $\frac{1}{2}$ , and if the second occurs, we make another choice with probabilities  $\frac{2}{3}$ ,  $\frac{1}{3}$ . The final results have the same probabilities as before. We require, in this special case, that  $H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}H(\frac{2}{3}, \frac{1}{3})$ . The coefficient  $\frac{1}{2}$  is because this second choice only occurs half the time."

After these requirements, he introduces the following *theorem*: "The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{x \in \mathcal{X}} p(x) \log p(x), \qquad (2.1)$$

where K is a positive constant". When K = 1 and the logarithm is  $\log_2$ , information is measured in bits.

Shannon calls this quantity *entropy*, as "the form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics where  $p_i$  is the probability of a system being in cell i of its phase space". There are other axiomatic formulations which involve the same definition of entropy [39].

The Shannon entropy is the classical measure of *information*, where information is simply *the outcome of a selection among a finite number of possibilities*. Entropy also measures *uncertainty* or *ignorance*.

Thus, the Shannon entropy H(X) of a discrete random variable X with values in the set  $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$  is defined as

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x), \qquad (2.2)$$

where p(x) = Pr[X = x], the logarithms are taken in base 2 (entropy is expressed in bits), and we use the convention that  $0 \log 0 = 0$ , which is justified by continuity. We can use interchangeably the notation H(X) or H(p) for the entropy, where pis the probability distribution  $\{p_1, p_2, \ldots, p_n\}$ . As  $-\log p(x)$  represents the *information* associated with the result x, the entropy gives us the *average information* 



Figure 2.8: Binary entropy.

or *uncertainty* of a random variable. Information and uncertainty are opposite. Uncertainty is considered before the event, information after. So, information reduces uncertainty. Note that the entropy depends only on the probabilities.

Some other relevant properties [138] of the entropy are

- 1.  $0 \le H(X) \le \log n$ 
  - H(X) = 0 if and only if all the probabilities except one are zero, this one having the unit value, i.e., when we are certain of the outcome.
  - $H(X) = \log n$  when all the probabilities are equal. This is the most uncertain situation.
- 2. If we equalize the probabilities, entropy increases.

When n = 2, the *binary* entropy (Figure 2.8) is given by

$$H(X) = -p\log p - (1-p)\log(1-p),$$
(2.3)

where the variable X is defined by

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

If we consider another random variable Y with probability distribution p(y) corresponding to values in the set  $\mathcal{Y} = \{y_1, y_2, \dots, y_m\}$ , the *joint entropy* of X and Y is defined as

$$H(X,Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log p(x,y), \qquad (2.4)$$

where p(x, y) = Pr[X = x, Y = y] is the joint probability.

Also, the conditional entropy is defined as

$$H(X|Y) = -\sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x, y) \log p(x|y), \qquad (2.5)$$

where p(x|y) = Pr[X = x|Y = y] is the conditional probability.

The Bayes theorem expresses the relation between the different probabilities:

$$p(x,y) = p(x)p(y|x) = p(y)p(x|y).$$
 (2.6)

If X and Y are *independent*, then p(x, y) = p(x)p(y).

The conditional entropy can be thought of in terms of a *channel* whose input is the random variable X and whose output is the random variable Y. H(X|Y)corresponds to the uncertainty in the channel input from the receiver's point of view, and vice versa for H(Y|X). Note that in general  $H(X|Y) \neq H(Y|X)$ .

The following properties are also met:

- 1.  $H(X, Y) \le H(X) + H(Y)$
- 2. H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)
- 3.  $H(X) \ge H(X|Y) \ge 0$

### 2.2.2 Mutual Information

The *mutual information* between two random variables X and Y is defined as

$$I(X,Y) = H(X) - H(X|Y)$$
  
=  $H(Y) - H(Y|X)$   
=  $-\sum_{x \in \mathcal{X}} p(x) \log p(x) + \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x,y) \log p(x|y)$   
=  $\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}.$  (2.7)

Mutual information represents the amount of information that one random variable, the output of the channel, gives (or contains) about a second random variable, the input of the channel, and vice versa, i.e., how much the knowledge of X decreases the uncertainty of Y and vice versa. Therefore, I(X, Y) is a measure of the shared information between X and Y.

Mutual information I(X, Y) has the following properties:

- 1.  $I(X, Y) \ge 0$  with equality if, and only if, X and Y are independent.
- 2. I(X, Y) = I(Y, X)
- 3. I(X,Y) = H(X) + H(Y) H(X,Y)


Figure 2.9: Venn diagram of a discrete channel.

#### 4. $I(X, Y) \leq H(X)$

The relationship between all the above measures can be expressed by the Venn diagram, as shown in Figure 2.9.

The *relative entropy* or *Kullback-Leibler distance* between two probability distributions p(x) and q(x), that are defined over the set  $\mathcal{X}$ , is defined as

$$D_{KL}(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)},$$
(2.8)

where, from continuity, we use the convention that  $0 \log 0 = 0$ ,  $a \log \frac{a}{0} = \infty$  if a > 0, and  $0 \log \frac{0}{0} = 0$ .

The relative entropy is "a measure of the inefficiency of assuming that the distribution is q when the true distribution is p" [39].

The relative entropy satisfies the *information inequality*  $D_{KL}(p||q) \ge 0$ , with equality only if p = q. The relative entropy is also called *discrimination* and it is not strictly a distance, since it is not symmetric and does not satisfy the triangle inequality. Moreover, we have to emphasize that the mutual information can be expressed as

$$I(X,Y) = D_{KL}(\{p(x,y)\} \| \{p(x)p(y)\}).$$
(2.9)

#### 2.2.3 Entropy Rate of a Markov Chain

The joint entropy of a collection of n random variables is given by

 $H(X_1, \dots, X_n) = H(X_1) + H(X_2|X_1) + \dots + H(X_n|X_{n-1}, \dots, X_1).$  (2.10)

The entropy rate or entropy density of a stochastic process  $\{X_i\}$  is defined by

$$h = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n) = \lim_{n \to \infty} H(X_n | X_{n-1}, \dots, X_1),$$
(2.11)

representing the *average information content* per output symbol<sup>2</sup> [39]. It is the "uncertainty associated with a given symbol if all the preceding symbols are known"

<sup>&</sup>lt;sup>2</sup>At least, h exists for all stationary stochastic processes.

and can be viewed as "the intrinsic *unpredictability*" or "the irreducible *random-ness*" associated with the chain [51].

In particular, a Markov chain can be considered as a chain of random variables complying with

$$H(X_n|X_1, X_2, \dots, X_{n-1}) = H(X_n|X_{n-1}).$$
(2.12)

An important result is the following theorem: For a stationary Markov chain, with stationary distribution  $w_i$ , the entropy rate or information content is given by

$$h = \lim_{n \to \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n)$$
  
= 
$$\lim_{n \to \infty} H(X_n | X_{n-1})$$
  
= 
$$H(X_2 | X_1) = -\sum_{i=1}^n w_i \sum_{j=1}^n P_{ij} \log P_{ij},$$
 (2.13)

where  $w_i$  is the equilibrium distribution and  $P_{ij}$  is the transition probability from state *i* to state *j*.

In addition, the *excess entropy* or *effective measure complexity* [41, 56, 139, 148] of an infinite chain is defined by

$$E = \lim_{n \to \infty} (H(X_1, X_2, \dots, X_n) - nh),$$
 (2.14)

where h is the entropy rate of the chain and n is the length of this chain. The excess entropy can be interpreted as the mutual information between two semiinfinite halves of the chain. "Another way of viewing this, is that excess entropy is the *cost of amnesia* – the excess entropy measures how much more random the system would become if we suddenly forgot all information about the left half of the string" [50].

#### 2.2.4 Important Inequalities

Some of the above properties can be deduced from the inequalities presented in this section [39]. In addition, these will also play an important role in obtaining fundamental results in this thesis.

#### 2.2.4.1 Jensen's inequality

A function f(x) is *convex* over an interval (a, b) (the graph of the function lies below any chord) if for every  $x_1, x_2 \in (a, b)$  and  $0 \le \lambda \le 1$ ,

$$f(\lambda x_1 + (1 - \lambda)x_2) \le \lambda f(x_1) + (1 - \lambda)f(x_2).$$
(2.15)

A function is strictly convex if equality holds only if  $\lambda = 0$  or  $\lambda = 1$ . A function f(x) is *concave* (the graph of the function lies above any chord) if -f(x) is convex.

For instance,  $x \log x$  for  $x \ge 0$  is a strictly convex function, and  $\log x$  for  $x \ge 0$  is a strictly concave function [39].

Jensen's inequality: If f is convex on the range of a random variable X, then

$$f(E[X]) \le E[f(X)], \tag{2.16}$$

where E denotes expectation. Moreover, if f(x) is strictly convex, the equality implies that X = E[X] with probability 1, i.e., X is a deterministic random variable with  $Pr[X = x_0] = 1$  for some  $x_0$ .

One of the most important consequences of Jensen's inequality is the information inequality  $D_{KL}(p||q) \ge 0$ . Other previous properties can also be derived from this inequality.

Observe that if  $f(x) = x^2$  (convex function), then  $E[X^2] - (E[X])^2 \ge 0$ . So, the variance is invariably positive.

If f is substituted by the Shannon entropy, which is a concave function, we obtain the *Jensen-Shannon inequality* [19]:

$$JS(\pi_1, \pi_2, \dots, \pi_n; p_1, p_2, \dots, p_n) \equiv H\left(\sum_{i=1}^n \pi_i p_i\right) - \sum_{i=1}^n \pi_i H(p_i) \ge 0, \quad (2.17)$$

where  $JS(\pi_1, \pi_2, \ldots, \pi_n; p_1, p_2, \ldots, p_n)$  is the Jensen-Shannon divergence of probability distributions  $p_1, p_2, \ldots, p_n$  with prior probabilities or weights  $\pi_1, \pi_2, \ldots, \pi_n$ , fulfilling  $\sum_{i=1}^n \pi_i = 1$ . The JS-divergence measures how 'far' are the probabilities  $p_i$  from their likely joint source  $\sum_{i=1}^n \pi_i p_i$  and equals zero if and only if all the  $p_i$  are equal. It is important to note that the JS-divergence is identical to I(X, Y)when  $\pi_i = p(x_i)$  and  $p_i = p(Y|x_i)$  for each  $x_i \in \mathcal{X}$ , where  $p(X) = \{p(x_i)\}$  is the input distribution,  $p(Y|x_i) = \{p(y_1|x_i), p(y_2|x_i), \ldots, p(y_m|x_i)\}, n = |\mathcal{X}|$ , and  $m = |\mathcal{Y}|$  [19, 141].

#### 2.2.4.2 The log-sum inequality

Log-sum inequality: If  $a_1, a_2, \ldots, a_n$  and  $b_1, b_2, \ldots, b_n$  are non-negative numbers, then

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \ge (\sum_{i=1}^{n} a_i) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i},$$
(2.18)

with equality if and only if  $\frac{a_i}{b_i} = constant$ .

Note that the conditions in this inequality are much weaker than for Jensen's inequality.

From this inequality, certain results can be derived:

- 1.  $D_{KL}(p||q)$  is convex in the pair (p,q).
- 2. H(X) is a concave function of p.
- 3. If X and Y have the joint pdf p(x,y) = p(x)p(y|x), then I(X,Y) is a concave function of p(x) for fixed p(y|x) and a convex function of p(y|x) for fixed p(x).

#### 2.2.4.3 Data processing inequality

Data processing inequality: If  $X \to Y \to Z$  is a Markov chain, then

$$I(X,Y) \ge I(X,Z). \tag{2.19}$$

This result demonstrates that no processing of Y, deterministic or random, can increase the information that Y contains about X.

#### 2.2.5 Entropy and Coding

Other ways of interpreting the Shannon entropy are possible:

• As we have seen in section 2.2.1,  $-\log p_i$  represents the *information* associated with the result  $x_i$ . But  $-\log p_i$  can also be interpreted as the *surprise* associated with the outcome  $x_i$ . If  $p_i$  is small, the surprise is large; if  $p_i$  is large, the surprise is small. Thus, the entropy

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i$$

is the expectation value of the surprise [50].

• Entropy is also related to the difficulty in guessing the outcome of a random variable. Thus, it can be seen [39, 50] that

$$H(X) \le \overline{questions} < H(X) + 1, \tag{2.20}$$

where  $\overline{questions}$  is the average minimum number of binary questions to determine X. This idea agrees with the interpretation of entropy as a measure of uncertainty and also with the next interpretation.

- A fundamental result of information theory is the Shannon source coding theorem, which deals with the encoding of an object in order to store or transmit it efficiently [39, 50]. "Data compression can be achieved by assigning short descriptions to the most frequent outcomes of the data source and necessarily longer descriptions to the less frequent outcomes" [39]. For instance, Huffman instantaneous coding <sup>3</sup> is optimal and fulfils the following theorems:
  - Similarly to (2.20), we have

$$H(X) \le \ell < H(X) + 1,$$
 (2.21)

where  $\overline{\ell}$  is the expected length of the optimal binary code for X.

<sup>&</sup>lt;sup>3</sup>A code is called a prefix or instantaneous code if no codeword is a prefix of any other codeword.

- If we encode *n* identically distributed random variables *X* with a binary code, the Shannon source coding theorem can be enunciated in the following way:

$$H(X) \le \overline{\ell_n} < H(X) + \frac{1}{n}, \tag{2.22}$$

where  $\overline{\ell_n}$  is the expected codeword length per unit symbol. Thus, by using large block lengths, we can achieve an expected codelength per symbol arbitrarily close to the entropy [39].

- For a stationary stochastic process, we have

$$\frac{H(X_1, X_2, \dots, X_n)}{n} \le \overline{\ell_n} < \frac{H(X_1, X_2, \dots, X_n)}{n} + 1 \quad (2.23)$$

and thus, by definition of entropy rate h (2.11),

$$\lim_{n \to \infty} \overline{\ell_n} \to h. \tag{2.24}$$

Thus, the entropy rate is the expected number of bits per symbol required to describe the stochastic process.

We can conclude that the entropy of a random variable is a measure of the amount of information required on average to describe it.

# 2.3 Image processing

Advances in medical imaging have lead to the development of more sophisticated image processing techniques in order to satisfy new user requirements. In this section, two of the most representative image processing techniques: registration and segmentation are reviewed.

#### 2.3.1 Image Registration

Registration is a fundamental task in image processing used to match two or more images or volumes obtained at different times, from different devices or from different viewpoints. Basically, it consists in finding the geometrical transformation that enables us to align volumes into a unique coordinate space. In medical applications it is of special interest since it allows us to integrate complementary information in a single model. The integration of information from different imaging modalities is difficult and, in most cases, dependent on the data we have to deal with. In the scope of this thesis we will focus on *multimodal rigid registration* techniques, multimodal since images are from different image modalities and rigid because only transformations that consider translations and rotations are allowed.

In this section, the main components of the image registration pipeline are presented. A classification of the most representative registration methods that have been proposed is also given. To end the section, the main challenges in the registration field are described.



Figure 2.10: The main components of the registration framework are the two input images, a transformation, a metric, an interpolator, and an optimizer.

#### 2.3.1.1 The image registration pipeline

The image registration pipeline starts with the selection of the two images to be registered. One of the two images is defined as the fixed image and the other one as the moving image. Given these images, registration is treated as an optimization problem with the goal of finding the spatial mapping that will bring the moving image into alignment with the fixed one. This process can be described as a process composed of four basic elements [79, 113]: the transformation, the interpolator, the metric and the optimizer (see Figure 2.10). The *transformation* component represents the spatial mapping of points from the fixed image space to points in the moving image space. The *interpolator* is used to evaluate moving image intensity at non-grid positions. The *metric* component provides a measure of how well the fixed image is matched by the transformed moving image. This measure forms the quantitative criterion to be optimized by the *optimizer* over the search space defined by the parameters of the transformation. Each of these components is now described in more detail .

1. **Spatial transformation.** The registration process consists in reading the input data models, defining the reference space (i.e. its resolution, positioning and orientation of the objects) for each of these models, and establishing the *correspondence* between them (i.e. how to transform the coordinates from one data model to the coordinates of the other data model).

The spatial transformation defines the spatial relationship between both images. Basically, two groups of transformations can be considered:

• *Rigid or affine transformations*. These transformations are defined as geometrical transformations that preserve all distances. They also preserve the straightness of lines (and the planarity of surfaces) and all nonzero angles between straight lines. These kinds of transformations are mainly used in intrapatient registration, i.e., the registration of images of the same patient.

• *Nonrigid or elastic transformations*. Using these kinds of transformations, the straightness of the lines are not ensured. These transformations are very useful for interpatient registration, i.e., the registration between different subjects. The fusion of intrapatient volumes allows the researchers to obtain the atlas of a given population, which permits us to establish the general templates of diseases like schizophrenia or Alzheimer's.

In this thesis, rigid image registration is our reference point.

2. **Interpolation.** The interpolation strategy determines the intensity value of a point at a non-grid position. When a general transformation is applied to an image, the transformed points cannot coincide with the regular grid. So, an interpolation scheme is needed to estimate the values at these positions.

Several interpolation schemes have been introduced [80]. The most common are:

- *Nearest neighbour interpolation:* the intensity of each point is given by the one of the nearest grid-point.
- *Trilinear interpolation:* the intensity of a point is obtained from the linear-weighted combination of the intensities of its neighbours.
- *Splines:* the intensity of a point is obtained from the spline-weighted combination of a grid-point kernel [153].
- *Partial volume interpolation:* the weights of the linear interpolation are used to update the histogram, without introducing new intensity values [35].
- 3. **Metric.** The metric evaluates the similarity (or disparity) between the two images to be registered. Several image similarity measures have been proposed. They can be classified depending on the used features which are:
  - *Geometrical features*. A segmentation process detects some features and, then, they are aligned. These methods obtain subvoxel accuracy and they do not have high computational cost. Nevertheless, there is a great dependence on the initial segmentation results.
  - *Correlation measures*. The intensity values of each image are analyzed and the alignment is achieved when a certain correlation measure is maximized. Usually, a priori information is used in these metrics.
  - *Intensity occurrence*. These measures depend on the probability of each intensity value and are based on information theory [138].

Despite this variety of measures, this last group has become the most popular. Due to the importance of the similarity measure in our research, a classification of registration techniques according to this parameter will be given in Section 2.3.2.

4. **Optimization.** The optimizer finds the maximum (or minimum) value of the metric varying the spatial transformation. A lot of numerical methods have been developed in order to obtain the global extreme of a non analytical function. The following methods are the most used in medical image registration: Powell's method, gradient descent, one-plus-one evolutionary, and the simplex method. The choice of method will depend on the implementation criteria and the measure features (smoothness, robustness, etc.).

#### 2.3.2 Similarity metrics

The registration metric characterizes the similarity (or disparity) of both images for a given transformation. It is considered that the two models are registered when this similarity (or disparity) function is maximum (or minimum).

The registration methods that have been proposed can be classified into two main groups according to the information considered to compute the measure: (i) feature-based registration, which uses previously segmented objects from the images to achieve the alignment and (ii) voxel-based methods, which use the whole data. A more detailed description of both groups is given below.

#### 2.3.2.1 Feature-based registration

Measures based on geometric features minimize spatial disparity between selected features from the images (e.g. distance between corresponding points). The main difference between the methods of these group is the feature selected for the registration, which can be points, surface, intrinsic features such as landmarks, or extrinsic measures such as implanted markers. According to the features, two main categories of algorithms can be considered:

Point-based registration algorithms

The basis of these algorithms is the selection of a set of points in each of the images and then the minimum euclidian distance between them gives the best alignment. The set of selected points can be:

- Anatomical landmarks. Usually, these anatomical points are set manually, requiring expert support and are quite time-consuming [64].
- Skin affixed markers. This method is automatic and non-invasive, but during the exploration the markers can experience small movements, affecting the final results.
- Bone implanted markers. This method avoids the previously mentioned problem of the movement of the markers, but it is very invasive for the patient [99].

Since, in general, the point sets of each image do not exactly coincide, an iterative algorithm is performed until the distance between these sets of points is minimal [57]. These methods are used extensively in the medical scenario due to their simplicity.

• Segmented-based registration algorithms

Segmentation-based registration algorithms are based on the alignment of segmented structures, which in most cases correspond to object surfaces of the images. The segmentation process takes an image and separates its elements into connected regions, which present the same desired property or characteristic. Below, some of the most representative methods of this group used in medical applications are described.

The most popular approach is the *head-hat* algorithm [110]. This method segments the skin surfaces of both images and tries to fit one with the other one, similar to a hat on a head. Another approach is presented in [87], where a crest or valley-like features are assumed in the image, as with MR or CT images.

The segmentation-based algorithms are generally accurate and fast if a good choice of features is performed. The main drawback of this approach is that the registration accuracy is limited to the accuracy of the segmentation step, which, in the case of functional data such as PET, is almost impossible to be carried out due to its noisy nature. In theory, segmentation-based registration is applicable to images of many areas of the body, but in practice the application areas have been limited to neuroimaging and orthopedic imaging. Moreover, feature-based registration requires specialized segmentation and feature extraction for each application. In addition, as not all voxels are used, the methodology is not immune to noise and is sensitive to outliers. The main advantages of the segmentation-based methods are that these can yield subvoxel accuracy, giving more accurate results than the intensity-based approach. They are faster than the intensity-based registration as they use a lower number of features and the optimization procedure needs less iterations.

#### 2.3.2.2 Voxel-based similarity measures

The alternative to the feature-based approach is the intensity-based registration. This approach assumes some relation between the optical densities of voxels and operates directly on the image grey values without prior data reduction by the user nor segmentation. The registration is implicitly performed by the definition of a function which evaluates the quality of alignment and thereby controls the optimization procedure. The information used for the alignment is not restricted to any specific feature and therefore this approach is more flexible than the feature-based one.

There are two different methodologies distinguishing the methods in this group:

• Intensity-based methods

These methods base the alignment on the evaluation of the intensity values considering the images aligned when the differences between grey values are minimal. This restriction is ideal in cases where two images are identical except for noise. Thus, most of the techniques of this group has been developed for monomodal registration. A known relation between the intensity values in the datasets to be registered is assumed, which is not valid for a general multimodal registration. An important aspect to be considered is that the proposed functions are only computed on the overlap area between both image, which varies for different transformations.

Some of the functions that have been proposed to describe the relation between grey values are [144]:

 The sum of absolute value differences. This is the simplest and most direct measure of similarity of two image values. This measure is defined as

$$S(A,B) = \sum_{x \in \mathsf{A} \bigcap \mathsf{B}} |I_{\mathsf{A}}(x) - I_{\mathsf{B}}(x)|, \qquad (2.25)$$

where  $I_A(x)$  and  $I_B(x)$  represent the intensity at a point x of the image A and B, respectively. When this measure is applied we assume that the image values are calibrated to the same scale.

Correlation. In the alignment of two images of the same modality, registration results in a strong linear relationship between corresponding values in the two images. A measure of similarity would be the correlation, which determines the fit of a line to the distribution of corresponding values. Correlation is expressed as

$$C(A,B) = \sum_{x \in \mathsf{A} \cap \mathsf{B}} I_{\mathsf{A}}(x) \times I_{\mathsf{B}}(x).$$
(2.26)

The main limitations of this measure are:

- \* Its dependence on the number of points over which it is evaluated. This tends to favour transformations yielding large overlap. The *normalized cross-correlation* solves this problem simply by dividing correlation by the number of points.
- \* Its dependence on the intensity values, which tends to favour high intensity values. As a solution to this second limitation a better

measure of alignment was proposed: the *correlation coefficient*. The correlation coefficient is a measure of the residual errors from the fitting of a line to the data by minimization of the least squares.

 Variance of intensity ratio. The correlation measure was widely used until the mid-1980's for 2D registration. In the 90's several researchers started looking for alternatives. One of these alternatives proposed by Woods et al. [161] developed a 3D registration algorithm based on local intensity variances.

Woods et al. introduced a registration measure for multimodal images based on the assumption that regions of similar tissue (similar grey values) in one image would correspond to regions in the other image that also consist of similar grey values. Ideally, the ratio of the grey values for all corresponding points in a certain region in both images varies little. Consequently, the average variance of this ratio for all regions is minimized to achieve registration.

#### • Methods based on the occurrences of intensity values

The basic idea behind these methods is that two values are related or similar if there are many other examples of those values occurring together in the overlapping image volume. These measures are a class of more generic statistical measures which only look at the occurrence of image values and not at the values themselves.

Most of these techniques are based on the *feature space* or *joint histogram*. The joint histogram is a two-dimensional plot of the corresponding grey values in the images showing the combinations of grey values in each of the two images for all corresponding points. The joint histogram is constructed by counting the number of times a combination of grey values occurs. For each pair of corresponding points (x, y), where x is a point in the first image and y a point in the second image, the entry  $(I_A(x), I_B(y))$  in the joint histogram is increased.

The joint histogram depends on the alignment of the images. When the images are correctly registered, corresponding anatomical structures overlap and the joint histogram will show certain clusters for the grey values of those structures. Conversely, when the images are misaligned, structures in one image will overlap with structures in the other image that are not their anatomical counterparts. In Figure 2.11, the joint histogram plots of an MR-PET image pair in the registration position (a) and with a lateral translation of 2 mm (b), 4 mm (c), and 10 mm (d) are shown. Note how the intensity of the clusters for corresponding anatomical structures will decrease and new combinations of grey values emerge. This will be manifested in the joint histogram by a dispersion of the clustering. This property is exploited by defining measures of clustering or dispersion which have to be maximized



Figure 2.11: Joint histogram plots of a MR (horizontal scale) and PET (vertical scale) test image pair in the registration position (a) and with a lateral translation of 2 mm (b), 4 mm (c), and 10 mm (d). Images obtained from [145].

and minimized respectively. Most of these measures are based on *information theory*. For a detailed description of this theory see the previous section. In the information theory context, the registration of two images is represented by an information channel  $X \to Y$ , where the random variables Xand Y represent the images. Their marginal probability distributions, p(x)and p(y), and the joint probability distribution, p(x, y), are obtained by simple normalization of the marginal and joint intensity histograms of the overlapping areas of both images [93].

Some of the measures based on the occurrences of intensity values are

- Moments of the joint probability distribution. The joint probability tells us the proportion of times one or more variables hold some specific values. The computation of the moments of the joint probability distribution was first proposed by Hill [62] from visual examination of the effects of misregistration on the feature space. Empirically, as the images approach the registration position, the values of the peaks in the joint probability distribution increase in height and the values on the regions of the probability distribution which contain lower counts decrease in height. Therefore, the registration process has to re-arrange the voxels so that they occur with their most probable corresponding value in the other image. One possible approach to quantify this shift from lower probabilities in the joint probability distribution to a smaller number of higher probabilities is to measure *skewness* (or the third moment) in the distribution of probabilities in the joint histogram. The skewness characterizes the degree of asymmetry of a distribution around its mean. It is a pure number that characterizes only the shape of the distribution [117].

- Joint entropy. In the joint histogram of two images, grey values disperse with misregistration and the joint entropy is a measure of this dispersion. By finding the transformation that minimizes their joint entropy, images should be registered [34, 146]. The main drawback of this method is its high sensitivity to the overlap area.
- Mutual information (MI). Another measure is mutual information which is less sensitive to the overlap area. The more dependent the datasets are, the higher the MI between them. Registration is assumed to correspond to the maximum mutual information: the images have to be aligned in such a manner that the amount of information they contain about each other is maximal [92, 157].

In the image registration context, Studholme [144] proposed a normalized measure of mutual information defined by

$$NMI(X,Y) = \frac{H(X) + H(Y)}{H(X,Y)} = 1 + \frac{I(X,Y)}{H(X,Y)},$$
 (2.27)

which is more robust than MI, due to its greater independence of the overlap area.

To conclude this section, the most relevant properties of the intensity-based registration approach are summarized. The main feature of intensity-based registration is its generality; it can be applied to any dataset with no previous pre-processing nor segmentation. Moreover, as all the voxels are considered on the alignment process, the method is quite immune to noise and is insensitive to outliers. As voxels are used as the features in registration, it is very difficult to obtain a sub-voxel registration accuracy. Maes et al. [91] attempts to reach sub-voxel accuracy by interpolating densities between neighbouring voxels. However, careless interpolation introduces artifacts through fictitious density values.

The convergence of intensity-based registration is in general very slow. Several strategies have been proposed to speed up the process, first registering at lower resolutions and then increasing the resolution. Due to the considerable computational cost required by these methodologies, multi-resolution and multi-scale approaches are incorporated to the process in order to speed up the convergence of this method.

	Point-based	Landmarks [64]
Feature-based	registration	Skin implants [99]
registration	Segmentation-based	Head-hat method [110]
	registration	Crest or valley features [87]
		Absolute differences [144]
	Intensity-based	Correlation [144]
	registration	Normalized cross-correlation
		Correlation coefficient
Voxel-based		Variance of intensity ratio [161]
registration		Moments of joint probability
	Occurrence-based	Joint entropy [93]
	registration	Mutual Information [92, 157]
		Normalized
		Mutual Information [144]

Table 2.1: The main image registration similarity metrics.

In Table 2.1 a summary of the explained methods is presented. This is only a small part of the several proposed approaches. More extensive surveys are presented in [44, 63, 116].

#### 2.3.3 Challenges in image registration

Advances in clinical imaging have lead to the development of new imaging registration techniques in order to satisfy requirements demanded by clinical experts. This fact has lead registration to be an active focus of research.

In this section, the main problems currently being addressed by image registration researchers are briefly summarized.

#### **Robustness and accuracy**

To evaluate the behaviour of a registration method robustness and accuracy are the main parameters to be considered. The first parameter, robustness, refers to how the method behaves with respect to different initial states, i.e. different initial positions of the images, image noise, modality of the images, etc. The second parameter, accuracy, refers to how the final method solution is closer to the ideal solution. In the multimodal rigid registration framework, certain methods achieve a reasonable robustness and accuracy, but a lot of research is being currently done in this area. Constantly, new measures and new interpolation schemes appear trying to improve the robustness and the accuracy of the standard measures.

#### Artifacts

In the registration process, the interpolator algorithm plays an important role, since usually the transformation brings the point to be evaluated into a non-grid position. This importance is greater when the grid size coincides in both images, since the interpolator pattern is repeated for each point. When the mutual information or its derivations, which are the most common measures used in multimodal image registration, are computed, their value is affected by both the interpolation scheme and the selected sampling strategy, limiting the accuracy of the registration. The fluctuations of the measure are called artifacts and are well studied by Tsao [151].

#### Speed-up

One of main user requirements when using registration techniques is speed. Users desire results as fast as possible. The large amount of data acquired by current capture devices makes its processing difficult in terms of time. Therefore, the definition of strategies able to accelerate the registration process is fundamental. Several multiresolution frameworks have been proposed achieving better robustness and speeding up the process.

#### 2.3.4 Image Segmentation

Image segmentation is the process of labeling each voxel in a image dataset according to certain parameter or features. In the case of medical images, the segmentation considers the tissue type or the anatomical structure contained in the voxel. Since a segmented image provides richer information than the original one, it is an essential tool in medical environments. It is used to improve visualization of medical imagery, to perform quantitative measurements of image structures, to build anatomical atlases, to research shapes of anatomical structures, or to track anatomical changes over time. Segmentation is considered a very difficult task and a lot of research is being done to develop automatic segmentation techniques. The main aspects that make this process so difficult are the imaging process itself and the anatomy that is represented in the images. The imaging process is chosen so that its interactions with the tissues of interest will provide clinically relevant information about the tissue in the resulting output image. But this does not mean that the anatomical feature of interest will be particularly separable from its surroundings. The second aspect that makes segmentation difficult is the complexity and variability of the anatomy that is being imaged. It may not be possible to locate or delineate certain structures without detailed anatomical knowledge. For this reason, in most of medical applications, segmentation is carried out manually by an expert radiologist. This is a very time consuming task and also sensitive to subjective errors.

To overcome these limitations, a lot of research is being done to develop new segmentation methods as accurate as manual editions but reducing user interaction as much as possible. Unfortunately, the automatic process is not easy since the regions to be segmented vary with the anatomy being imaged. Consequently, most proposed methods are anatomy specific assuming usually some *a priori* information that must either be built into the system or provided by a human operator. In this section, we briefly describe some of the main proposed segmentation algorithms. In the image processing literature, we can find a lot of segmentation methods and also very diverse ways of classifying them [55, 112, 53]. Nevertheless, we present in this section a classification of the automatic segmentation processes into two groups: the *global segmentation* methods, where all image voxels (or pixels) are collected in some clusters, and the *local segmentation* methods, where only a region is taken into account classifying the voxels (or pixels) inside or outside of this region. To end the section, main challenges in image segmentation are described.

#### 2.3.4.1 Global segmentation

These methods are also referred to as classification methods, since each point is classified into a cluster, usually depending on its intensity value and the intensity of its neighbours, and not on its position in the image. Global methods are very useful for tissue segmentation and morphological studies since a whole image segmentation is required and a manual process is extremely time-consuming. In most cases, some information of the image modality, such as default tissue intensity, is needed to start the segmentation process.

The main global segmentation methods can be classified in these groups:

• Thresholding

This segmentation scheme relies upon the selection of a range of intensity levels, called *threshold values*, for each material class. These intensity ranges are exclusive to a single class, and span the dynamic range of the image. Subsequently a feature is classified by selecting the class in which the value of the feature falls within the range of feature values of the class. The determination of more than one threshold value is a process called multithresholding.

The selection of the threshold generally depends on the visual identification of a peak in the histogram corresponding to a material class, and the selection of a range of intensities around the peak to include only the material class. A possible criterion is to assign the histogram minima as the threshold values. More refined criteria are summarized in [129, 137].

• Segmentation by image enhancement

In image processing terminology, an operation for image enhancement improves the quality of the image in a particular manner, either subjectively or objectively. This segmentation model assumes that a material class ideally has a single intensity, and that noise and scanning artifacts corrupt this level to produce the distribution of intensities observed for a material class. Thus, by the application of image enhancement techniques for reducing noise and smoothing the image, the enhanced image approximates the ideal image (the segmented one).

A class of filters known as nonlinear anisotropic diffusion filters can be useful for image enhancement [111]. This class of filters performs smoothing while preserving edges.

The two main drawbacks of this segmentation approach are the following. First, the structures that do not have strong edge on all sides are smoothed, leading to large classification errors when subsequent labelling is applied. Second, large numbers of iterations of the filters are required before regions of similar tissue have nearly the same intensity

#### • Segmentation by unsupervised clustering

Clustering methods are algorithms that operate on an input dataset, grouping data into clusters based on the similarity of the data in these clusters. Clustering algorithms are unsupervised classifiers, assigning states from scratch. They are also useful for data exploration, allowing a user to discover patterns of similarities in a dataset.

A well-known clustering algorithm is the k-means [58]. The k-means algorithm accepts as input the number of clusters to organize data within, initial location of cluster centers, and a dataset to cluster. The number of clusters in which the algorithm fits the data is specified to the algorithm, and represents a parameter the user desires to experiment with, or, also, the expected or desired number of classes to discern from the data. There are no conditions upon which data is excluded or included in consideration to fit into a class; all data provided as input are classified. A given sample or feature measurement is assigned exclusively to one class (fuzzy k-means clustering assigns a degree of membership to each data item for each class).

The algorithm is an iterative algorithm, assigning a class at each iteration to each data element. The algorithm iteration ceases when there are no changes in the classification solution. Each iteration consists in classifying the dataset by comparison of the dataset to the current cluster centers. A data item is assigned to the same class as a cluster center if the Euclidian distance between the data item and the cluster center is the least distance between the data item and all the cluster centers. Following class assignment, cluster centers are updated by computation of the centroid of the dataset classified as the same class.

Another popular approach is the Markov random fields [86], which is not a segmentation method by itself, but it is often incorporated into clustering segmentation algorithms such as the k-means algorithm under a Bayesian prior model. This method assumes that most voxels belong to the same structure as its neighbours. The segmentation is then obtained by maximizing the *a posteriori* probability of the segmentation given the image data using iterative methods such as iterated conditional modes or simulated annealing. The main problem of this method is the high sensitivity to the parameters. Nevertheless, they are widely used in medical image classification.

#### 2.3.4.2 Local segmentation

The main objective of the local segmentation methods is to classify each voxel into inside or outside the region of interest. These methods focus their attention on the spatial distribution of voxels and are very specific to the problem to be solved. Most of them assume some region features as shape, intensity or texture. These methods are widely used in medical image segmentation due to their accurate results, and are used extensively to segment diseases, as tumors or strokes, or structures, as the corpus callosum in the brain or the left ventricle in the heart.

Two main approaches exist: in the first one, all voxels of an initial region are taken into account to iteratively modify this region until it converges to the desired one, and in the second one, only the boundary conditions of the region are considered, deforming the region with some particular criteria.

• Region-based segmentation

The most popular and simple algorithm of the region-based segmentation is the *region growing*. The region growing method extracts the region of interest assuming a predefined connection criterion [165]. This criterion can be based on intensity information and/or edges in the image. In its simplest form, region growing requires a seed point or region that is manually selected by an operator, and extracts all pixels connected to the initial seed with the same intensity value. Then, some different criteria have appeared based on an intensity range, gradient threshold, contrast or texture [53]. In medical applications, region growing is not often used alone but within a set of image processing operations, particularly for the delineation of small, simple structures such as tumors and lesions. Its main disadvantages are that it requires manual interaction to obtain the initial region and there is no subvoxel accuracy. Moreover, it can only segment connected regions.

Another popular approach is the *split-and-merge* method [69], that tackles the problem in the opposite way to the region growing algorithm. This approach starts from the entire image and subdivides it into smaller regions where the image is not homogeneous. To divide the image, different strategies can be adopted: octrees (where each region is subdivided into eight equal regions), BSP (where an optimal partition plane is selected to divide the region), etc. After this first splitting step, the neighbour regions that are homogeneous in a certain criterion are merged.

A similar approach is the *watershed* method. This method treats the image as a topographic surface, being the intensity of the height of each point. First, the gradient image is computed and a high number of seed points are placed, usually automatically, in the "valley" points. Then, these seed points are "flooded" until the whole image is completed. Each "flooded" region is the result of the segmentation. However, in practice, this process produces an important over-segmentation due to noise or local irregularities in the gradient image. To avoid this over-segmentation, a merging step is added at the end of the method.

Despite the fact that the use of the split-and-merge methods in realistic and biological image segmentation is very common, they are not much used in medical applications.

• Edge-based segmentation

In edge-based segmentation approaches, the image to be segmented is described in terms of the boundaries between different regions. In this approach, *deformable models* have been extensively studied and widely used in medical image segmentation. For instance, in Chapter B of the appendix, this kind of segmentation is used in order to determine the edema region surrounding a hematoma lesion in a CT image. For that reason, edge-based segmentation methods are introduced here in more detail than the previous methods.

Deformable models are curves or surfaces defined within an image domain that can move under the influence of *internal forces*, which are defined within the curve or surface itself, and *external forces*, which are computed from image data [142]. The internal forces are designed to keep the model smooth during the deformation, while the external forces are defined to move the model towards an object boundary or other desired features within the image. There are two main types of deformable models: parametric deformable models and geometric deformable models. The first ones represent the boundary in its parametric form during the deformation, not requiring high computational cost. On the other hand, the change of the curve topology is not possible in most cases. The geometric models represent curves and surfaces as a level set of a higher-dimensional scalar function and their parametrizations are computed only after the curve computation. Thus, a priori topology is not assumed.

- Parametric deformable models. In the parametric deformable models methods, two different formulations exist: an energy minimizing formulation and a dynamic force formulation. For both cases, the deformable parametric contour can be defined as  $\mathcal{X}(s) = (X(s), Y(s))$ ,  $s \in [0, 1]$ .

In the first formulation, the curve  $\mathcal{X}(s)$  moves through the spatial domain trying to minimize the following energy functional:

$$E(\mathcal{X}) = S(\mathcal{X}) + P(\mathcal{X}), \qquad (2.28)$$

where  $S(\mathcal{X})$  is the internal energy functional and  $P(\mathcal{X})$  is the potencial energy functional. The first term is defined by

$$S(\mathcal{X}) = \frac{1}{2} \int_0^1 \alpha(s) \left| \frac{\partial \mathcal{X}}{\partial s} \right|^2 + \beta(s) \left| \frac{\partial^2 \mathcal{X}}{\partial s^2} \right|^2 ds, \qquad (2.29)$$

and controls the contour tension and rigidity, i.e, the coherence of the curve. The second term is computed integrating a potential energy function  $\mathcal{P}(x, y)$  along the contour:

$$P(\mathcal{X}) = \int_0^1 \mathcal{P}(\mathcal{X}(s)).$$
 (2.30)

This potential energy function is derived from the image data and takes smaller values at object boundaries as well as other features of interest. Thus, this second term pulls the contour to the desired object boundaries.

The dynamic force formulation arises from the necessity to use more general external forces. According to Newton's second law, the dynamics of a contour  $\mathcal{X}(s,t)$  must satisfy

$$\mu \frac{\partial^2 \mathcal{X}}{\partial t^2} = \mathcal{F}_{damp}(\mathcal{X}) + \mathcal{F}_{int}(\mathcal{X}) + \mathcal{F}_{ext}(\mathcal{X}), \qquad (2.31)$$

where  $\mu$  is a coefficient that has a mass unit and  $F_{damp}$  is a damping (or viscous) force defined as  $-\gamma \partial \mathcal{X} / \partial t$ , with  $\gamma$  being the damping coefficient. In the image segmentation context, the  $\mu$  parameter is often set to zero, becoming the previous equation

$$\gamma \frac{\partial \mathcal{X}}{\partial t} = \mathbf{F}_{int}(\mathcal{X}) + \mathbf{F}_{ext}(\mathcal{X}).$$
(2.32)

The internal forces control the coherence of the curve, similar to the previous formulation, while external forces can be either potential forces or non-potential forces. These external forces are often expressed as the superposition of several different forces, each one depending on the feature that we want to be highlighted. A wide study of these forces is presented in [142].

Although these two formulations yield similar results, the first formulation has the advantage that its solution satisfies a minimum principle, which is mathematically more satisfactory, whereas the second formulation has the flexibility of allowing for the use of more general types of external forces.

One of the first approaches, and probably the most popular, is the *snakes* algorithm [75]. Snakes are planar deformable contours that are useful in several image analysis tasks. They are often used to approximate the locations and shapes of object boundaries in images based on

the reasonable assumption that boundaries are piecewise continuous or smooth. For more information, an interesting survey of the parametric deformable model methods can be found in [101].

- Geometric deformable models

In the geometric deformable models, introduced in the image segmentation context in [25], curves and surfaces are evolved using only geometric measures, dissociating the curve evolution from its parametrization. Then, topology changes can be handled automatically. During the last years, several investigations have been done in this direction, with the *level sets* method [106] as one of the most popular approaches. The main difference of this method compared to the parametric ones is that, instead of wrapping the curve, it deforms, at each iteration, a certain higher-dimensional function previously defined, depending on the a priori information of the model. Being mathematically more accurate, we can define a level set function  $\phi(x, y, t)$  with the contour  $\mathcal{X}(s, t)$  as its zero level set. Then,

$$\phi(\mathcal{X}(s,t),t) = 0 \tag{2.33}$$

and, if this equation is derived with respect to t, we obtain

$$\frac{\partial \phi}{\partial t} + \nabla \phi \cdot \frac{\partial \mathcal{X}}{\partial t} = 0, \qquad (2.34)$$

where  $\nabla \phi$  denotes the gradient of  $\phi$ . Assuming that  $\phi$  is negative inside the zero level and positive outside, the equation can be rewritten as

$$\frac{\partial \phi}{\partial t} = -\nabla \phi \cdot \frac{\partial \mathcal{X}}{\partial t} = |\nabla \phi| \frac{\partial \mathcal{X}}{\partial t} \cdot \mathcal{N}(\mathcal{X}), \qquad (2.35)$$

where  $N(\mathcal{X})$  is the inward unit normal to the level set curve. Thus, the evolution of the curve can be written as

$$\frac{\partial \phi}{\partial t} = \mathcal{V}(\kappa) |\nabla \phi|, \qquad (2.36)$$

where  $V(\kappa)$  is the speed function and  $\kappa$  the level set curvature defined as  $\kappa = \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|}$  [142].

Several speed functions have been proposed depending on the features which rely on the segmentation process (see [136] for a detailed discussion on this topic). An interesting approach is proposed in [61, 60], where the speed function depends on some information theoretic concepts as the conditional entropy and mutual information.

A summary of the segmentation methods is shown in Table 2.2.

	Thresho	olding [129, 137]
Global segmentation	Enhanced-bas	sed segmentation [111]
	Unsupervise	ed clustering [58, 86]
		Region Growing [165]
	Region-based	Split-and-merge [69]
	segmentation	Watersheds [123]
Local segmentation		Parametric deformable
	Edge-based	models [75, 101, 142]
	segmentation	Geometric deformable
		models [25, 106, 142]

Table 2.2: The main image segmentation methods.

#### 2.3.4.3 Challenges in image segmentation

Image segmentation is an open problem that probably will never be solved at all. Despite the big evolution of the segmentation algorithms in the last decades, there is no universal segmentation method and most of the existing algorithms solve only few particular problems. This is due to the difficulty to include a priori information in a segmentation algorithm, although this problem is in most cases obvious for human perception. Current algorithms usually require human interaction and are very sensitive to the initial parameters. Moreover, they are not able to segment any kind of images, and are very sensitive to image "outliers", as noise or blurring. The image resolution is usually another parameter to be taken into account in order to obtain a correct segmentation, due to the partial volume effect, i.e., one pixel (or voxel) represents two or more regions. Therefore, the main image segmentation challenge is probably to develop a general method able to segment any kind of image, robust to the image imperfections as noise, artifacts, blurring, etc. and completely automatic, i.e., without any human interaction during the segmentation process. Even though we are conscious that this objective is difficult to reach, our purpose is to contribute to the development of better algorithms than enhance the existing ones, reducing human interaction as far as possible, and making them more robust to image noise.

# **Chapter 3**

# Image Registration based on Global Lines and High-dimensional Normalized Mutual Information

One of the key aspects in 3D-image registration is the computation of the similarity metric. Some of the most commonly used measures, like mutual information or normalized mutual information, require the computation of the joint intensity histogram between the two images. In this chapter we propose a new approach to compute this histogram using uniformly distributed random lines to sample stochastically the overlapping volume between two 3D-images. This method provides us with an accurate, robust and fast mutual information-based registration. The interpolation effects are drastically reduced, due to the stochastic nature of the line generation, and the alignment process is also accelerated. Despite the fact that the mutual informationbased measures have been successfully used as an effective similarity measure for multimodal image registration, a drawback is that the joint histogram is only calculated from the correspondence between individual voxels in the two images. To tackle this problem, the mutual information-based similarity measures referring to the correspondence between voxel blocks in multimodal rigid registration is generalized. With this approach the spatial information is taken into account in the similarity metric computation. The ambiguity and highdimensionality that appears when dealing with the voxel neighbourhood is solved using uniformly distributed random lines and reducing the number of bins of the images. Experimental results show a significant improvement with respect to the standard normalized mutual information.

### 3.1 Introduction

Recently, several similarity measures have been proposed to solve the multimodal image registration problem. Among them, mutual information (MI) [91, 156] and normalized mutual information (NMI) [144] are the most commonly used since they produce satisfactory results in terms of accuracy, robustness and reliability. However, MI-based methods have a number of well-known drawbacks, such as grid [154] and interpolation effects [91, 151]. In particular, the way of estimating the probability distributions, the choice of the interpolator and the sampling strategy have a great influence on the accuracy and robustness of the registration results. Another limitation of these methods is that the computation of the joint histogram is calculated from the correspondence between individual voxels in the two images. In this sense, the spatial distribution of the samples is not taken into account. For instance, because of this the swapping of voxels in the overlap area of the two images does not change the MI value.

Therefore, one of the key aspects of the similarity calculation process is the computation of the joint histogram. As proposed in [91], this step is usually done by taking all the points of the reference image and the corresponding values of the transformed floating image (see Section 2.3.1). In general, an interpolation scheme is needed to estimate these transformed values at non-grid positions. This interpolation provokes undesirable artifacts when the voxel grids have coinciding periodicities [115, 151], reducing the robustness and accuracy of the MI-based methods. Moreover, to accelerate the matching process, different multiresolution and multisampling schemes have been proposed. In particular, downsampling techniques are used to speed up the registration process [116].

As mentioned before, the spatial information is not taken into account in the standard methods. In recent years, different approaches have considered a region-based correlation to compute image similarity [126, 114, 128, 127, 66]. These approaches deal with gradient magnitude or high-dimensional vectors in order to incorporate spatial information on the similarity measure.

In this chapter, a new approach to compute MI-based similarity measures by using uniformly distributed random lines is introduced. This method accelerates the alignment and almost suppresses the interpolation artifacts due to the stochastic nature of the process. Several experiments show clearly the suitability of our approach to speed up the registration process and to improve its accuracy and robustness. An MI-based framework that uses structural information in an image is also proposed. A high dimensional information channel which incorporates the spatial information by considering the voxel neighbourhood probabilities is defined. Then, the standard similarity measures, MI and NMI, are extended to consider the correspondence between regions of voxels in multimodal rigid registration. The problems that appear when dealing with the voxel neighbourhood are tackled using uniformly distributed random lines and reducing the number of bins of the histograms. Experimental results over real medical data analyze the behaviour of our approach when neighbour intensity values are considered, showing better accuracy and robustness.

This chapter is organized as follows. In Section 3.2, background and related work is surveyed. In Section 3.3, the use of uniformly distributed random lines for image registration is presented. In Section 3.4, a method to compute the intensity histogram based on these random lines is proposed and the registration results obtained using this method over different multimodal images are analyzed. In Section 3.5, the high-dimensional normalized mutual information measure for image registration which considers the spatial information of both images is presented. Experimental results are given in Section 3.6. Finally, our conclusions are presented in Section 3.7.

# 3.2 Background and Related Work

In this section some basic topics related with this chapter are examined. First, some high-dimensional information-theoretic measures are reviewed [39, 49]. Then, different methods for the histogram estimation as well as some recent approaches which incorporate spatial information in the information-theoretic similarity measures for image registration are presented.

#### 3.2.1 Information-theoretic measures

The definition of block entropy and entropy rate is now reviewed. The notation used here is inspired by the work of Feldman and Crutchfield [49]. Given a chain  $\dots X_{-2}X_{-1}X_0X_1X_2\dots$  of random variables  $X_i$  taking values in  $\mathcal{X}$ , a block of L consecutive random variables is denoted by  $X^L = X_1 \dots X_L$ . The probability that the particular L-block  $x^L$  occurs is denoted by  $p(x^L)$ . The Shannon entropy of length-L sequences or L-block entropy is defined by

$$H(X^L) = -\sum_{x^L \in \mathcal{X}^L} p(x^L) \log p(x^L), \qquad (3.1)$$

where the sum runs over all possible L-blocks. The entropy rate is defined by

$$h = \lim_{L \to \infty} \frac{H(X^L)}{L},\tag{3.2}$$

and it can be written as

$$h = \lim_{L \to \infty} h(L),$$

where

$$h(L) = H(X^{L}) - H(X^{L-1}) = H(X_{L}|X_{L-1}...X_{1})$$

is the entropy of a symbol conditioned on a block of L-1 adjacent symbols. The entropy rate of a sequence measures the average amount of information per symbol x and the optimal achievement for any possible compression algorithm [39].

#### 3.2.2 Histogram estimation

MI is a basic concept in information theory, which is usually used to measure the statistical dependence between two random variables A and B, and is defined by

$$I(X,Y) = \sum \sum p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

where  $p_{XY}(x, y)$  is the joint probability density function (pdf) and  $p_X(x)$  and  $p_Y(y)$  are the marginal pdf's. In the context of image registration, the random variables X and Y correspond to the reference and floating images, respectively, and registration is achieved by maximizing the MI. The normalization of MI, defined by

$$NMI(X,Y) = \frac{H(X) + H(Y)}{H(X,Y)} = 1 + \frac{I(X,Y)}{H(X,Y)},$$

is more robust than MI due to its greater independence of the overlap area [144].

All these methods are based on the same four steps: the superposition of the reference and floating images, the computation of the pdf's, the computation of the similarity metric, and the optimization of the metric by reorienting the images. In this section we focus on the technique to estimate the marginal and the joint pdf's. To compute them, two different methods are usually applied: the Parzen window method [107, 156] and the binning method [91].

#### 3.2.2.1 Parzen window estimation

Given an instance of the random sample, Parzen-windowing [107] estimates the pdf from which the sample was derived. It essentially superposes kernel functions placed at each sample value. The general form of the density is

$$P^*(x,a) \equiv \frac{1}{N_a} \sum_{x_a \in a} R(x - x_a),$$

where a is a sample and R is a window function or kernel such that

$$\int_{x \in X} R(x) dx = 1,$$

where X is the domain of function R.

This function R is often called the smoothing or window function. In this way, each observation contributes to the pdf estimate. Unlike parametric estimations, Parzen estimation does not suppose any a priori distribution and only requires the density to be smooth [156]. Intuitively, the Parzen density estimator can be seen as a computation of a windowed average of the sample. The most common window functions are unimodal, symmetric about the origin, and fall off quickly to zero. Among these functions, the Gaussian window is a popular kernel for Parzen-window density estimation, being infinitely differentiable and thereby lending the

same property to the Parzen-window pdf estimate. The quality of the approximation is dependent both on the functional form of R and its width. Different window functions can lead to very different density estimates.

In terms of memory, the computation of the Parzen estimation is cheap, since one needs only to memorize the sample. On the contrary, the evaluation of  $P^*(x, a)$ is more expensive, requiring a time proportional to the sample size. Moreover, in the joint density function estimation, the computational cost increases with the square of the sample size. Fortunately, the efficiency can be improved if a limited window width is employed [98].

This estimation method has other advantages too, such as the computation of the entropy and the mutual information directly from the image samples. It is also computationally simple to compute the derivatives of these measures, which can help the optimizer to find the best transformation in the registration process [156].

#### 3.2.2.2 Binning estimation

Another way to compute the joint intensity histogram is by binning the intensity pairs (A, T(B)) of the overlapping parts of the reference image A and the transformed image T(B). Since generally the grid points of T(B) do not coincide with the grid points of image A, the application of this method requires the selection of an interpolator.

The most used interpolation schemes are as follows (see Section 2.3.1):

- *trilinear (TL) interpolation:* the intensity of a point is obtained from the weighted combination of the intensities of its neighbours. These weights for the 2D case are plotted in Figure 3.1.
- *nearest neighbour (NN) interpolation:* the intensity of a point is the same as the nearest grid point.
- *partial volume (PV) interpolation:* the weights of the linear interpolation (see Figure 3.1) are used to update the histogram bin of the corresponding grid intensity values. In this way, any new intensity value pair is not introduced in the joint histogram.

Although there are different interpolators, all of them introduce artifacts that will deteriorate the accuracy and reliability of the registration. These artifacts are from different natural sources depending on the interpolation scheme. For the TL interpolation, these artifacts result from the low filter effect of the interpolator. For the PV interpolation, these artifacts are due to the higher histogram dispersion at non-grid positions [115]. The NN interpolation causes artifacts due to the discontinuities that can appear for small transformation differences.

Some recent research has deal with the problem of reducing these artifacts. Tsao [151] has shown that jittered sampling is extremely beneficial to the robustness and accuracy of registration, reducing considerably the interpolation artifacts.



Figure 3.1: Interpolation weights, the areas  $w_i$ , for 2D linear interpolation. This Figure is based on [116].

Since grid effects are caused by the regular grid sampling of the images, other stochastic sampling strategies have also been proposed [154]. Also, Salvado and Wilson [130] proposed a constant variance filter to the images to be registered in order to compensate the blurring effect of the linear interpolation.

Binning techniques are more easily implemented and, probably for that reason, more commonly used. Our research is focused on these techniques.

#### **3.2.3** MI-based image registration with spatial information

Standard information theory-based measures ignore the spatial information contained in the images. Recent research has been focused on overcoming this problem. Rueckert et al. [126] propose a second-order MI to incorporate spatial information. The neighbourhood has been defined by the nearest neighbours of each pixel. With this strategy mutual information is computed considering voxel pairs, and not only single voxel values as the standard method. Therefore, this method takes into account some spatial information. The results of this method demonstrate that this increases the robustness of the registration process. Pluim et al. [114] include spatial information by combining MI with a term based on the image gradient of the images to be registered. The gradient term seeks to align locations of high gradient magnitude and similar orientations of the gradients at these locations. In this approach spatial information is incorporated using the gradient values, which measure the difference of a voxel and its neighbours. A more general framework was presented by Butz et al. [22], where the MI is computed not from the individual pixel pairs but from the choice of various feature spaces. Sabuncu and Ramadge [128] include spatial information in the MI-based approach by using spatial feature vectors obtained from the images and use a minimum spanning tree algorithm to estimate the conditional entropy in higher dimensions. Russakoff et al. [127] propose an MI extension which takes neighbourhood regions of corresponding pixels into account. They assume that the high-dimensional distribution

is approximately normally distributed. Holden et al. [66] use the derivatives of gaussian scale space to provide structural information in the form of a feature vector for each voxel. Gan and Chung [54] integrate a maximum distance-gradient-magnitude feature with intensity into a two-element attribute vector and adopt multidimensional MI as a similarity measure on the vector space.

In all of these investigations the introduction of spatial information improves the registration results notably and this encourages us to define a high dimensional channel which incorporates this kind of information.

# **3.3 Uniformly Distributed Global Lines for Image Registration**

In this section, a new method based on the use of uniformly distributed random lines to compute the joint intensity histogram for 3D-image registration, which is the most demanding step in the similarity measure computation is introduced. The overlapping volume between two 3D-images is stochastically sampled using a uniform distribution of lines in the sense of integral geometry, i.e., invariant to translations and rotations [132]. Points chosen on each line provide us with the intensities to calculate the probability distributions using a binning strategy. With this sampling strategy each point has the same probability to be considered and, hence, the resulting sample is a representative one. This global line density was first used in computer graphics to compute the illumination in a scene. The lines generated using this density were called *global lines* [133, 26].

Two alternatives to generate a global line density can be used. As a first alternative, a global line can be generated taking two random points on a sphere circumscribing the object or the scene [133]. This is only valid for a sphere, since taking pairs of points on the surface of any convex body do not result in a uniform density. Observe also that the sphere density is equivalent to taking a single point on the sphere and a uniform direction from this point, weighted according to the cosinus of the angle  $\theta$  between the radius at this point and the tangent plane, i.e., between the tangent plane and the normal to the plane. Thus, taking only a uniformly distributed direction does not result in a uniform density. As a second alternative, a global line can also be generated from the walls of a convex bounding box containing the object or the scene [26]. This can be done taking a random point on the surface of the convex bounding box and a cosinus distributed random direction. In Figure 3.2, the probability density function of the random line direction is plotted. The vertical line represents the wall where the random line is being sorted and the horizontal arrow is the normal of this wall.

For our registration intentions, this second alternative, taking the reference image as the bounding box (see Figure 7.2), since the histogram computation just take into account the voxels that belong to both images is adopted. The intensity values are captured from the lines at evenly spaced positions, taking an initial random offset different for each line. The random offset ranges from 0 to the step size. The



Figure 3.2: Probability density function of the global line directions from a wall of a convex bounding box.



Figure 3.3: Global lines are cast from the walls of the bounding box.



Figure 3.4: PWI-DWI and CT-MR test image pairs.

regular grid sampling is thus substituted in our method by sampling with random lines. Although we skip with regular steps, the use of a random offset ensures the stochasticity of the process. The cost of the histogram computation depends on both the number of lines cast and the number of points taken for each line, which is inversely proportional to the step size.

Note the fact that using this strategy the images will be considered at non grid positions, independently if they have coinciding grids. Therefore, the interpolation effects, that are due to these coinciding grids, will be drastically reduced. In other words, using this strategy the image is considered as a continuous function which is stochastically sampled in order to compute the joint histogram, instead of considering the image as a grid of points and computing the histogram just on these points. Also note that this strategy is not directly related with any interpolator and therefore the artifact suppression is not due to the interpolator scheme, but the sampling strategy.

This strategy can also be used to obtain random 1-D strings from the image data in order to study spatial correlation. The next sections will explore this feature in more detail.

## 3.4 Histogram Estimation using Global Lines

To evaluate the behaviour of our global line method, different registration experiments are carried out. The data sets used in our tests (see Figure 3.4) are a diffusion-weighted imaging (DWI) and perfusion weighted imaging (PWI), from the Josep Trueta Hospital, and a CT and an MR image, from the Vanderbilt database. The DWI and PWI sets have a resolution of  $256 \times 256 \times 20$  and  $128 \times 128 \times 12$ , respectively. The voxel size is  $0.977 \times 0.977 \times 7.0 \ mm^3$  for the DWI and  $1.797 \times 10^{-10} \ mm^3$ .



Figure 3.5: Probability density function estimation of the histogram image. Results obtained with 1.000, 10.000, 100.000, 1.000.000 and 10.000.000 sampling points using the nearest neighbour interpolator. Also the real density function is plotted.

 $1.797 \times 10.0 \ mm^3$  for the PWI. The CT image has a resolution of  $512 \times 512 \times 28$ and a voxel size of  $0.654 \times 0.654 \times 4 \ mm^3$ . The resolution of the MR image is  $256 \times 256 \times 26$  and the voxel size is  $1.25 \times 1.25 \times 4 \ mm^3$ .

In the first experiment, the convergence of the marginal histogram of an MR image computed using global lines to the histogram computed using the standard binning strategy is observed. In Figure 7.7, the pdf estimation convergence by increasing the number of sampling points is shown. As expected, by increasing the number of sampling points, the estimated distribution tends to the real one.

For our two next experiments, we have compared the NMI similarity measure [144] computed with global lines with the same measure computed with the standard binning strategy. The standard NMI measure computation has been represented in all the plots as the bold bottom curve. These two experiments aim to analyze the influence of the step size and the number of lines cast on our method, respectively. The behaviour of the NMI has been analyzed moving the floating image through the X axis from -10 mm to 10 mm around the origin. In these experiments, the PWI-DWI images have been used as the testing set.

In the second experiment, we have fixed the number of sampled points to 150000, which corresponds to 11% of voxels of the original model. Figure 3.6(a) shows the obtained results considering different step sizes of 1, 2, 4, 6 and 8 mm, from top to bottom. It is important to note that the step size determines the number of cast lines since the number of points is fixed, i.e., a small step size implies a small number of lines, and viceversa. Two facts need to be noted. First, the curves corresponding to the step sizes of 1 and 2 mm give undesired results, as the maxima of these curves do not correspond to the perfect registration. This is a consequence of supersampling, since in general a lower step size can produce several consecutive sampling points into the same voxel. For step sizes greater than



Figure 3.6: (a) NMI using the global line method with 150000 sampling points vs. the slice translation on the X-axis. The step size increases from top to bottom (1, 2, 4, 6 and 8 mm.). The standard NMI curve in bold. (b) NMI using the global line method with a step size of 5 mm. vs. the slice translation on the X-axis. The number of sampled points increases from top to bottom (50000, 75000, 100000, 125000 and 150000 points). The standard NMI curve in bold.

4 mm, the differences between the NMI curves are very low, and they are due to the stochasticity of the method. Second, a reduced number of lines cast does not ensure that the model has been probed in a sufficient number of directions. Thus, taking into account that the computational cost of the method increases with the step size, since more lines have to be generated, the optimal process is a trade-off between the number of lines and the step size.

In the next experiment, the step size has been fixed to 5 mm and a different number of lines has been considered. The obtained results are represented in Figure 3.6(b), where the plotted curves correspond, from top to bottom, to the NMI measure computed using 50, 75, 100, 125 and 150 thousand points. Observe that the NMI value decreases when the number of points taken increases, converging to the NMI value measured using the standard strategy. This behavibour results from the fact that the joint entropy increases with the number of points [154]. In Figure 3.7, the cost of the standard (dash-dotted line) and the global line (solid line) methods are compared. Note that for our method the time increases linearly with the number of points. Therefore, using global lines, the time can be outstandingly reduced. For example, an acceptable estimation is obtained with 50000 points and its correspondent processing time is almost five times lower (see Figure 3.6(b) and 3.7).

In the last experiment, interpolation artifacts are studied. As these especially occur when the voxel grids of the images have coinciding periodicities, the CT-MR pair has been properly rescaled in order to maximize these periodicities. The NMI values have been determined as a function of translation along the X-axis in the range of  $\pm$  10 voxels. Figures 3.8(a) and 3.8(b) show the curves of the NMI values obtained with the global line method using the nearest neighbour and the



Figure 3.7: Computational cost of the global line method (continuous line) vs. the value obtained with the standard NMI method (dash-dotted line). The horizontal axis represents the number of sampled points and the vertical axis the time units.

linear interpolator schemes, respectively. We have considered a different number of sampled points, from top to bottom: 75000, 100000, 150000 and 200000. In both plots, the bottom curve corresponds to the NMI value computed with the standard method. Note the interpolation artifacts, in a stairs-like mode for the NN scheme and as a set of local minima at every integer-voxel step for the linear scheme [151]. Observe in Figure 3.8(a) that the curves obtained by using the NN scheme with the global line method also present small artifacts. In this case the grid alignment causes local maxima, but not the stair-like mode as in the standard NN approach. Interestingly, the behavibour of the NN interpolator coincides with the one of the partial volume method [116], since our estimation method converges to the PV scheme when the number of points tends to  $\infty$  and there is no rotation. Therefore, these artifacts are because of the higher histogram dispersion at non-grid positions, as in the PV scheme [115]. Figure 3.8(b) illustrates the results of the global line method using the TL interpolation scheme. It can be seen that the interpolation artifacts have completely disappeared. This is due to the fact that the grid effects have been eliminated by the stochastic sampling and the linear interpolator has a blurring effect that avoids the artifacts which appeared in the NN interpolation scheme.

# 3.5 Generalization of Mutual Information-based Measures

In this section, the high-dimensional normalized mutual information and its computation using uniformly distributed random lines is defined.

#### 3.5.1 High-dimensional normalized mutual information

The registration of two medical images is commonly represented by an information channel  $X \to Y$ , where the random variables X and Y represent, respectively,



Figure 3.8: NMI values using (a) the nearest neighbour and (b) the linear interpolator schemes.

the images A and B. This approach does not take into account the spatial distribution of the samples. Here, a new approach based on taking *L*-blocks of samples from the images instead of single values is proposed. Then, the new information channel can be described as  $X^L \to Y^L$ . Both marginal probability distributions of the *L*-blocks of each image,  $p(x^L) = p(x_1, x_2, \ldots, x_L)$  and  $p(y^L) =$  $p(y_1, y_2, \ldots, y_L)$ , are computed by the normalization of the *L*-dimensional histograms of the *L*-blocks. Similar to the standard method, the joint probability distribution,  $p(x^L, y^L) = p(x_1, x_2, \ldots, x_L, y_1, y_2, \ldots, y_L)$ , is obtained by the normalization of the joint intensity histograms of the *L*-blocks in the overlapping area of both images. The conditional probability can be calculated using the Bayes theorem

$$p(y^L|x^L) = \frac{p(x^L, y^L)}{p(x^L)},$$

leading to the transition probability matrix P of the channel (conditional probability matrix):

$$P = \begin{pmatrix} p_{1,1,\dots,1|1,1,\dots,1} & \cdots & p_{m,1,\dots,1|1,1,\dots,1} & \cdots & p_{m,m,\dots,m|1,1,\dots,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{1,1,\dots,1|n,1,\dots,1} & \cdots & p_{m,1,\dots,1|n,1,\dots,1} & \cdots & p_{m,m,\dots,m|n,1,\dots,1} \\ p_{1,1,\dots,1|1,2,\dots,1} & \cdots & p_{m,1,\dots,1|1,2,\dots,1} & \cdots & p_{m,m,\dots,m|1,2,\dots,1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_{1,1,\dots,1|n,n,\dots,n} & \cdots & p_{m,1,\dots,1|n,n,\dots,n} & \cdots & p_{m,m,\dots,m|n,n,\dots,n} \end{pmatrix}, (3.3)$$

where n and m are, respectively, the number of bins of the intensity histograms of images A and B. Note that the dimensionality of this matrix is  $n^L \cdot m^L$ . This huge value, which will be discussed in the next section, has critical consequences in the design of the implementation strategy.

From these probability distributions, the joint entropy  $H(X^L, Y^L)$ , which mea-
sures the dispersion of the joint histogram, is given by

$$H(X^L, Y^L) = -\sum_{x^L \in \mathcal{X}^L, y^L \in \mathcal{Y}^L} p(x^L, y^L) \log p(x^L, y^L).$$
(3.4)

Then, the mutual information  $I(X^L, Y^L)$ , which quantifies how much information both random variables share (the *L*-block intensities, in our case), is defined as

$$I(X^{L}, Y^{L}) = H(X^{L}) + H(Y^{L}) - H(X^{L}, Y^{L}).$$
(3.5)

Similar to the standard approach, the normalization of the mutual information by the joint entropy is defined by

$$NMI(X^L, Y^L) = \frac{I(X^L, Y^L)}{H(X^L, Y^L)} = \frac{H(X^L) + H(Y^L) - H(X^L, Y^L)}{H(X^L, Y^L)}.$$
 (3.6)

To compute the marginal and joint probabilities, how is  $x^L$  selected?, i.e., how is the neighbourhood ambiguity solved? With respect to this problem, different templates have been proposed for two dimensions [49], but the generalization from 1-block approach to L-block approaches is a difficult problem.

Another aspect to be considered is the size of an *L*-block. As *L* increases,  $\frac{H(X^L)}{L}$  decreases, converging to the entropy rate *h* of an image, which expresses its maximum compressibility or, equivalently, its irreducible randomness. In fact, the entropy of an image appears more random than it actually is, and the difference is given by  $h^L - h$  [49] (see Chapter 7). Therefore, the higher the *L* value, the better the approximation of  $I(X^L, Y^L)$  to the true shared information. On the other hand, note that the number of elements of the joint histogram is given by  $N^{2L}$ , where *N* is the number of bins of an image.

From these limitations, how is  $NMI(X^L, Y^L)$  computed for image registration? Rueckert et al. [126] propose a second-order MI using neighbour voxel pairs and reducing the number of bins to 16. One drawback of this approach is that not all directions are taken into account. In the next section, this problem using uniformly distributed random lines is tackled.

### **3.5.2** Implementation using uniformly distributed random lines

In a 3D image volume, the definition of a neighbourhood is not a simple task. The simplest approach is to take as a neighbourhood the six nearest voxels. But this approach has several drawbacks. The main one is the high dimensionality of the joint histogram, which has to consider 14 different symbols (the central voxel and the six neighbours for both images). Another drawback is that only the directions parallel to the coordinate system are considered. Another possible solution is to take only pairs of voxels as proposed by Rueckert in [126], but in this case just a little neighbour information is taken into account and as in the previous approach only few directions are considered.



Figure 3.9: Diagram of the high dimensional data generation from global lines.

In order to implement the  $NMI(X^L, Y^L)$  measure, taking L-blocks on the global lines is proposed. The sequence of intensity values (L-block  $X^L$ ) needed to estimate the joint probabilities is captured at evenly spaced positions over the global line. Points chosen on each line provide us with the intensities to calculate the L-block entropies. In this manner, the 3D-neighbourhood problem is reduced to a 1D problem, where the ambiguity about how to sample the neighbourhood disappears as the order in which we take the neighbours is well defined on a global line. In Figure 3.9 a representation of the process is illustrated. With this strategy, blocks from all directions are considered and not only the ones parallels to the coordinate system. In spite of this, the dimensionality problem persists due to the fact that the computation of the joint histogram of the channel  $X^L \to Y^L$  has an excessive cost if  $L \ge 2$  (see the probability transition matrix of Equation (3.3)). Another problem is the sparsity of the resulting histograms since the number of bins of the joint histogram is  $N^{2L}$ . A possible solution to these problems is to reduce the number of bins of the image, preserving a trade-off between levels of binning and the accuracy of the entropy rate.

## 3.6 Results and Discussion

In this section, two sets of experiments are carried out. First, we analyze the behaviour of  $NMI(X^L, Y^L)$  when two neighbour intensity values are considered on global lines and, second, we show some results taking 3-blocks. The clinical images and the standard transformations used for the tests are provided as part of the Retrospective Image Registration Evaluation project [104], which is one of the main standards of validation of multimodal rigid registration results.

The first experiments use the ITK MI\_Validation application [1] with a multiresolution optimization approach using the quaternion rigid transformation gradient descent. In our implementation, three multiresolution steps, with 1500 it-

CT-MR_T1	NMI	$NMI(X^2, Y^2)$	CT-MR_T1	NMI	$NMI(X^2, Y^2)$
pat_001	2.05	0.97	pat_102	5.57	1.54
pat_002	1.98	1.28	pat_103	2.28	2.44
pat_003	1.78	1.10	pat_104	2.91	1.22
pat_004	13.16	2.21	pat_105	11.49	2.45
pat_005	1.63	0.84	pat_106	5.81	1.74
pat_006	12.62	1.44	pat_107	4.18	0.94
pat_007	0.78	1.13	pat_108	9.21	1.69
pat_101	6.57	2.10	pat_109	1.27	1.15

Table 3.1: The mean of the error in mm. for NMI and  $NMI(X^2, Y^2)$ , measured at different VOIs for each patient

Method	Mean	Median	Maximum
NMI	5.02	3.00	19.74
$NMI(X^2, Y^2)$	1.50	1.31	2.92

Table 3.2: The mean, median and maximum of the distances in mm. between VOIs from all patient results

erations each one, are used. To reduce the computational cost, all the measures are computed using global lines, with a total number of 100000 sampling points and a step size of 2 mm. The NMI and  $NMI(X^2, Y^2)$  have been computed, respectively, over a range of 256 and 16 bins per dimension. The results of CT-MR\_T1 image registrations for 16 patients are shown in Table 3.1. Each entry in the table is the mean of the distances in mm between the evaluated registration method and the gold standard measured at different volumes of interest (VOI) of the patient. Note that in 14 out of 16 tested patients the proposed approach behaves better than the NMI method, and, in all the cases, the error of our method is lower than 2.5 mm, which reveals its good performance. This good behaviour is also shown in Table 3.2, where the mean, the median and the maximum of the distances between VOIs have been computed from all patient results. These results show the necessity to incorporate spatial information in the similarity measures, and the great improvement shown in the registration results.

The second experiment deals with the improvement considering different L values. Our method is now used to register the CT-MR\_T1 pair of patient 3 of the Vanderbilt database, since this allows a good visualization of the registration results. In Figure 3.10, a preliminary evaluation of the proposed algorithm with blocks of three neighbours is presented. To overcome the high-dimensionality of the joint histogram, the number of bins has been reduced to 16 bins and 8 bins for L = 2 and L = 3, respectively. In Figure 3.10, we show the original CT with a rectangu-

lar marker and the transformed MR obtained with L = 1, L = 2, and L = 3 and with the marker in the same geometrical position as in the CT image. In the second row of Figure 3.10 a zoom of the first row images is shown. In this test, the correctness of the registration for each case is determined by the degree of coincidence between the white dot in the center of the marker of the CT image and the black dot in the MR one. Observe that, for  $NMI(X^2, Y^2)$  and  $NMI(X^3, Y^3)$ , the registration achieves more accurate results than the standard method NMI. This behaviour encourages us to explore the proposed measure with higher L values.

## 3.7 Conclusions

In this chapter, a new stochastic approach for 3D-image registration based on sampling the images with uniformly distributed random lines has been proposed. The advantages of this method can be summarized as follows. First, similarly to the jittered sampling [151], the global line sampling reduces considerably the interpolation artifacts, almost suppressing the periodicities of the voxel grids. Second, the registration accuracy is preserved with a high reduction of sampled points, accelerating the computation. Third, the randomness of the sampling lines permits us to generalize the mutual information-based measures, taking into account the dependence between image regions. Experimental results show that this approach achieves accurate and robust registration results using only two neighbour intensity values in the normalized mutual information computation.



Figure 3.10: CT-MR registration results for different L values.

## **Chapter 4**

# **Image Registration by Compression**

In this chapter, a new framework for image registration based on compression is presented. The basic idea underlying our approach is the conjecture that two images are correctly registered when we can maximally compress one image given the information in the other. The contribution of this chapter is twofold. First, we show that the image registration process can be formulated as a compression problem. Second, we demonstrate the good performance of the similarity metric, introduced by Li et al., in image registration. Two different approaches for the computation of this similarity metric are described: the Kolmogorov version, computed using standard real-world compressors, and the Shannon version, calculated from an estimation of the entropy rate of the images.

## 4.1 Introduction

As we have seen, image registration consists in aligning two images using a similarity measure which quantifies the quality of the alignment. The task of finding out the correct registration between two images is frequently based on the maximization of a similarity measure or the minimization of a given distance. Some information-theoretic measures, such as *mutual information* (MI) [91, 156] and *normalized mutual information* (NMI) [144], have become a standard reference, mainly in medical imaging, due to their accuracy and robustness.

In this chapter, the *normalized information distance* (*NID*), also called *the similarity metric*, as a new similarity measure for image registration is proposed. *NID*, introduced by Li et al. [84] for measuring similarity between sequences, is based on the non-computable notion of *Kolmogorov complexity* and it is a normalized version of the information metric [13]. In essence, the main idea is that two

objects are similar if we can significantly compress one given the information in the other. It has been successfully applied in areas such as genome phylogeny [83], language phylogeny [84] and classification of music pieces [29].

However, the application of NID is limited by its non-computability. To tackle this problem, we propose two different approaches. The first one is based on the *normalized compression distance* (*NCD*) [30], which approximates the Kolmogorov complexity using real-world compressors. In this case, the capability of the compressor to approximate the Kolmogorov complexity will determine the registration accuracy. The second approach is based on the *normalized entropy rate distance* (*NED*) [73, 45], which substitutes the Kolmogorov complexity by the entropy rate. This is a measure of the degree of compressibility of an image from a Shannon perspective. In both cases, experimental results demonstrate that *the similarity metric* performs well in image registration. We show that the Kolmogorov version is less accurate and robust than the entropy rate approach probably due to compressor imperfections.

This chapter is organized as follows. In Section 4.2, the concept of *the similar-ity metric* introduced in [84] is described. In Section 4.3, our framework for image registration is presented. In Sections 4.4 and 4.5 the two proposed approaches are detailed. Experimental results are given in Section 4.6. Finally, the conclusions are summarized in Section 4.7.

## 4.2 The Similarity Metric

The Kolmogorov complexity K(x) of a string x is the length of the shortest program to compute x on an appropriate universal computer<sup>1</sup>. Essentially, the Kolmogorov complexity of a string is the length of the ultimate compressed version of the string. The conditional complexity K(x|y) of x relative to y is defined as the length of the shortest program to compute x given y as an auxiliary input to the computation. The joint complexity K(x, y) represents the length of the shortest program for the pair (x, y) [84]. For a detailed review see [85].

In [13], the *information distance* is defined as the length of the shortest program that computes x from y and y from x. It was shown there that, up to an additive logarithmic term, the information distance is given by

$$E(x, y) = \max\{K(y|x), K(x|y)\}.$$
(4.1)

It was also shown that E(x, y) is a metric. It is interesting to note that long strings

<sup>&</sup>lt;sup>1</sup>A universal computer, or Turing machine, is a theoretical computing machine, invented by Alan Turing [152], to serve as an idealized model for mathematical calculation. A Turing machine consists of a line of cells known as a "tape" that can be moved back and forth, an active element known as the "head" that possesses a property known as "state" and that can change the property known as "color" of the active cell underneath it, and a set of instructions for how the head should modify the active cell and move the tape [160]. In our context, the shortest program to compute x on an appropriate universal computer is equivalent to considering the length of the shortest binary program to compute x in a universal programming language, such as Java [84].

that differ by a tiny part are intuitively closer than short strings that differ by the same amount. Hence, there arises the necessity to normalize the information distance. In [84], the normalized version of E(x, y), called the *normalized information distance (NID)* or *the similarity metric*, is defined by

$$NID(x,y) = \frac{\max\{K(x|y), K(y|x)\}}{\max\{K(x), K(y)\}} \\ = \frac{K(x,y) - \min\{K(x), K(y)\}}{\max\{K(x), K(y)\}}.$$
 (4.2)

Li et al. show that NID(x, y) is a metric and it takes values in [0, 1]. This metric is universal in the sense that if two strings are similar according to the particular feature described by a particular normalized admissible distance (not necessarily metric), then they are also similar in the sense of the normalized information metric [30].

The Kolmogorov complexity K is a non-computable measure in the Turing sense [85] and, therefore, for real-world applications, we will need an approximation of it. An upper bound of the non-computable complexity K is the length of compressed string x (or y), C(x) (or C(y)), generated by a compression algorithm. The better the compression algorithm, the better the approximation to K. Then, a feasible version of the normalized information distance (4.2), called the *normalized compression distance* (*NCD*), is defined [30] as

$$NCD(x,y) = \frac{C(x,y) - \min\{C(x), C(y)\}}{\max\{C(x), C(y)\}},$$
(4.3)

where C(x) (or C(y)) represents the length of compressed string x (or y) and C(x, y) the length of the compressed pair (x, y). Thus, NCD is computed from the lengths of compressed data files and therefore, NCD approximates NID by using standard real-world compressors.

## 4.3 Registration based on the Similarity Metric

In this section, a new approach for image registration based on the similarity metric is presented. This approach is based on the conjecture that two images are correctly registered when the NID between them is minimum. Due to the noncomputability of the Kolmogorov complexity, two feasible versions of NID are implemented. First, the NCD is applied as an approximation of NID. Second, a Shannon version of NID, obtained by substituting the Kolmogorov complexity in (4.2) by the entropy rate, is proposed. These two registration approaches and their respective implementations are outlined below.

• Kolmogorov version



Figure 4.1: Image compressor method scheme.



Figure 4.2: Text compressor method scheme.

Given two images A and B, the correct registration will be achieved when the normalized compression distance (4.3) is minimum. In this formula, C(x) (or C(y)) represents the size of the compressed image A (or B) and C(x, y) the length of the compressed fused pair (A,B). The fusion has been done superimposing the images after applying a certain transformation. To compress the images, we use both image and text compressors. In the first case, *jpeg*, *jpeg2000* and *png* compressors *are* applied (Figure 4.1). In the second case, the *bzip2* compressor is used (Figure 4.2).

The main drawback of this approach is the feasibility of the real-world compressors to capture the real compressibility of the images. The behavibour of NCD using several compressors will be analyzed in Section 4.4.

### • Shannon version

In Section 2.2, we have seen that the entropy rate of a sequence of symbols is a measure of its uncertainty, randomness or unpredictability. The entropy rate is also a measure of the compressibility of a sequence: the higher the uncertainty, the lower the compressibility. For instance, in a text, if there are strong correlations between letters (or words), knowledge of all previous letters (or words) will greatly decrease our uncertainty about the following



Figure 4.3: Entropy rate method scheme.

letters [40].

The entropy rate is now used as a measure of the compressibility of an image (see Figure 4.3). Thus, substituting the Kolmogorov complexity by the entropy rate in (4.2), the *normalized entropy rate distance* (NED) is given by

$$NED(x,y) = \frac{h(x,y) - \min\{h(x), h(y)\}}{\max\{h(x), h(y)\}},$$
(4.4)

where h(x) (or h(x)) represents the entropy rate of image A (or B) and h(x, y) the entropy rate of the pair (A,B). A similar approach has been previously proposed by Kaltchenko [73] and Dawy et al. [45] in the bioinformatics field. In this paper, the minimization of NED is introduced as a new registration criterion. The implementation of the entropy rate and the behavibour of NED will be analyzed in Section 4.5.

## 4.4 Kolmogorov Approach

In this section, the use of NID as a similarity metric for image registration is analyzed. Due to the non-computability of Kolmogorov complexity, NID is approximated by NCD. Basically, two methods are developed using, respectively, both image and text compressors.

#### 4.4.1 Image compressor-based registration

The first method consists in using the standard real-word image compressors to compute NCD between two images. For each image, the C(x) and C(y) values can be easily computed by compressing the original images and taking the size of



Figure 4.4: *Lena* image and its R, G and B channels. The similarity metric is evaluated within the dashed window.

the resulting file. The problem arises with the C(x, y) computation, since realword compressors are designed to deal with a single color image and not with two monochrome images as in our case. To overcome this limitation we propose to codify the images to be registered in a single image using the color channels.

A color image is represented as a finite number of color image planes where each color is obtained by filtering the image spectrum and by measuring the resulting luminosity energy. Usually, digital color images use three filters, corresponding to the red (R), green (G) and blue (B) channels. In Figure 4.4, we show a color image (*Lena*) with its RGB channels. Fortunately, standard medical images are monochrome and they have only one value at each pixel. Therefore, these images can be codified using only one of the RGB color channels. In Figure 4.1, where the image compressor based-registration process is illustrated, the original images A and B are represented using one channel, and the (A,B) pair using two different channels (R and G). Then, for each arrangement of images, the resulting image and the original ones are compressed. The correct registration would be achieved for the transformation that leads to the minimum NCD.

In our experiments, the most common image compression file formats are used:

- *JPEG*. This compressor subdivides the image in 8 × 8 blocks, computing for each of them the discrete cosinus transform (DCT). The terms of the DCT with low values can be omitted without a significant loss of image quality. However, several image artifacts appear in the compressed images due to the discontinuity between different blocks. This algorithm is specially designed for lossy compression, although a lossless version also exists.
- *JPEG 2000.* This compressor was created in order to improve JPEG file format. It is based on wavelet technology and it avoids the subdivision of

blocks and, therefore, its related artifacts. Nevertheless, some image details and texture disappear due to the loss of information.

• *PNG*. This compressor was created to improve the GIF file format. Its compression algorithm is based on the Lempel-Ziv-Welch (LZW) algorithm [164] and it was specially designed for lossless image compression. Since this compressor uses the intensity repetitions in order to compress, it is well-suited for synthetic images (diagrams, plots, ...), while the compression rates for natural images are not very high.

One limitation of the image-based registration method is due to the restriction of image compressors that force the images to have a rectangular shape. Such a restriction is not satisfied during the registration process since the overlap area, in general, is not rectangular. Amongst the different solutions that can be considered in our experiments, we take the maximal rectangular area contained in the overlap region for all the considered transformations. As we consider only translations along one axis, this region, represented as the dashed window of Figure 4.4, is easily feasible. However, in a real registration framework when all transformations are allowed, the selection of this region can be very complex. In this case a possible solution is to consider the whole reference image as the region to compress and fill the pixels of the moving image that the transformation has mapped outside the region with a constant value (typically zero). A drawback of this latter approach is that new information is added to the fused image.

### 4.4.2 Text compressor-based registration

The second method consists in converting the registration process into a text compression problem. In order to codify the image in a text file, each intensity value is represented as a byte. The files obtained with this process are not ASCII standard files and therefore they are not readable with a text editor.

In this approach, the critical point is the order in which the intensity values of the images have to be taken. Different scanning paths can be considered: space filling curves [81, 42] (see Figure 4.5(a)), *global line* sampling [133] (see Figure 4.5(b)) or a zigzag path (see Figure 4.5(c)) amongst others. In our case, we consider the zigzag mode for its simplicity.

Initially, input images A and B are scanned following the predefined path. From the scan path, strings  $x = (x_1, x_2, \ldots, x_n)$  and  $y = (y_1, y_2, \ldots, y_n)$  are obtained by taking in a consecutive order the intensity value of the pixels of both fixed and moving images (scaled to 0..255 if necessary), respectively. These values are codified in a text file and, C(x) and C(y) values are obtained from the length of compressed files. In a similar way, to compute C(x, y), the overlap area is scanned and the string  $\binom{x}{y} = \left(\binom{x_1}{y_1}, \binom{x_2}{y_2}, \ldots, \binom{x_n}{y_n}\right)$  is obtained by taking alternately the intensity value of the pixels of both superimposed images. Finally, this data codified in



Figure 4.5: Different scan paths on the overlap area between images X and Y: (a) space fitting curve, (b) global lines and (c) the proposed zigzag path.

a text file is compressed. In Figure 4.2 the different steps of this registration process are illustrated. For each iteration in the registration process, the three resulting string files are compressed. The notation used here has been inspired by [73].

In our experiments, we use the *bzip2* compressor which is based on the Burrows-Wheeler block-sorting text compression algorithm [20] and Huffman coding [70]. The registration position is achieved by minimizing NCD. In this case, C(x, y), C(x), and C(y) are given by the length of files compressed by *bzip2*. Note that the text compressor-based registration method has no restrictions on the shape of the overlap area, since the scan path is able to fit any shape.

## 4.5 Shannon Approach

We propose now to use the *normalized entropy rate distance* (*NED*) as a new similarity criterion. This measure has been constructed from (4.2) using the entropy rate instead of the Kolmogorov complexity. To compute the entropy rate, we assume that strings x, y and  $\begin{pmatrix} x \\ y \end{pmatrix}$  are generated by finite-order stationary Markov sources [73]. Thus, *NED* is given by

$$NED(x,y) = \frac{h\binom{x}{y} - \min\{h(x), h(y)\}}{\max\{h(x), h(y)\}},$$
(4.5)

where

$$h\begin{pmatrix}x\\y\end{pmatrix} = \lim_{L \to \infty} \frac{H\begin{pmatrix}X^L\\Y^L\end{pmatrix}}{L} = \lim_{L \to \infty} H\begin{pmatrix}X_L & X_{L-1} & \dots & X_1\\Y_L & Y_{L-1} & \dots & Y_1\end{pmatrix}$$
(4.6)

represents the entropy rate of  $\binom{x}{y}$ , i.e., the maximum compressibility for the two superimposed strings and  $H\binom{X^L}{Y^L}$  symbolizes the joint entropy of L symbols of  $\binom{x}{y}$ . With this proposal, we make use of neighbour information by considering the correspondence between blocks of pixels instead of the correspondence between individual pixels as in the classical registration methods based on MI [116].

For L = 1, the entropy rate approximation  $H\binom{X^1}{Y^1}$  is the standard Shannon joint entropy, and, then,

$$NED(x,y)_{L=1} = \frac{H(X,Y) - \min\{H(X), H(Y)\}}{\max\{H(X), H(Y)\}}$$

For higher L, the estimation of the entropy rate is better, due to the high spatial correlation between samples in medical images. This approach is similar to the one presented in Chapter 3, where blocks of intensity values are also considered, although the similarity measure is not the same.

NED can be seen from the information channel perspective. The classical registration methods based on MI consider the registration problem as an information channel,  $X \longrightarrow Y$ , where the random variables X and Y are both images, and where the mutual information of the channel must be maximized. Here, this information channel is replaced by  $X^L \longrightarrow Y^L$ , where  $X^L$  and  $Y^L$  represent the random variables built from blocks of L pixels. In this scenario, the entropy is equivalent to the entropy rate of the two-image system, and it is a measure of compressibility from the Shannon's perspective.

To compute the probabilities of the blocks of pixels, the overlap area of the two images to be registered is scanned similarly to the text compressor method (see Figure 4.5). Each block is composed of L consecutive pixels. Note that the accuracy of the computation is influenced by the size of the blocks, since the sparsity of the joint histogram and also the computational cost of the process increases with it. To overcome these limitations, reductions of the original alphabet (0..255) to 8 or 16 symbols are appropriate.

## 4.6 Results and Discussion

In this section, the performance of the compression-based registration measures is analyzed. The proposed registration methods have been tested using different synthetic and real images.

### 4.6.1 Image compression-based method

In the first experiment, the image compression-based method, comparing three different image compressors (*jpeg*, *jpeg2000* and *png*) is analyzed.

For the *jpeg* and the *jpeg2000* compressors we have considered the lossless and the lossy compression modes. Cilibrasi and Vitányi [30] define the *ideal compressor* as lossless and all the properties of *NCD* are proven for this case. For this reason, the lossless compression mode is evaluated. On the other hand, *jpeg* and *jpeg2000* compressors are specifically designed for the lossy case. With these compressors, when a low compression loss is considered, this loss is mainly due to the reduction of image noise, while the image signal is kept. In our tests, the minimum possible loss is considered. The *png* compression algorithm only performs with lossless compression.



Figure 4.6: Compression and distance results of the Lena image with the *jpeg*, *jpeg2000* and *png* compressors.

For the first test, the *Lena* image of Figure 4.4 is used. To simulate the registration process we need two images: the reference image and the floating image. We consider as the reference image the one with both G and B channels of the original color image and, as the floating one, the image corresponding to the R channel. Therefore, the fused image is the original RGB one. The R channel image is moved through the X axis from -10 pixels to 10 pixels around the origin. Since the image compressor requires images to have a rectangular shape, we consider only the rectangular window represented as a dashed square in Figure 4.4. For each translation the resulting images are compressed, obtaining C(x, y), C(x) and C(y) from the length of the compressed RGB image, the length of the GB image and the length of the R image, respectively.

In Figure 4.6 the obtained results for each compressor are depicted. For each case, the C(x, y) and the NCD measures are plotted. As it can be seen, the *jpeg* lossless approach does not lead to the desired results. This erratic behaviour is caused by the minimum variation between the size of the resulting compressed files and it can be attributed to the inherent "compression noise". Note that in all the other cases the conjecture is verified reaching the minimum of the measure at the registration position. Observe also that C(x, y) and NCD measures behave similarly.



Figure 4.7: Synthetic MR T1 and T2 (top) and real MR-CT (below) pairs of test images.





For the next experiments, two different pairs of medical images (see Figure 4.7) are used. The first pair consists of a synthetic image of a brain phantom, which simulates the T1-weighted (left image) and the T2-weighted (right image) MR scanning process [32]. The second pair of images consists of multimodal images from the same patient. The left image is a magnetic resonance (MR) image, which gives accurate information of the soft tissue, and the the right image is a computed tomography (CT), which provides precise anatomical characteristics of the brain. The registration of these modalities is of special interest for medical diagnosis.

For these tests, the reference image is assigned to the R channel and the moving one to the G channel. Again, we consider the lossless and the lossy mode of the *jpeg* and the *jpeg2000* compressors, and the *png* compression. To simulate the registration process, the moving images are translated through the X axis from -10



Figure 4.9: Compression and distance results of the synthetic MR T1 and T2 images with the *jpeg*, *jpeg2000* and *png* compressors.

pixels to 10 pixels around the registration position. A rectangular window in all the test images in order to obtain a rectangular overlap area is defined. In Figure 4.8 some images obtained during the process are presented: the two first images show the resulting images at the registration position and the two last ones when the floating image is translated 10 pixels from the registration position.

In Figure 4.9 we plot the C(x, y) and the NCD obtained with the synthetic MR images. In this case, the T1-weighted image has been taken as a fixed image and the T2-weighted image as the moving one. Observe that the lossless *jpeg* compressor does not obtain the desired results. For all other image compression approaches the assumption that minimum compression is reached at the registration position is true. The C(x, y) and the NCD plots obtained for the real MR-CT image pairs are shown in Figure 4.10. In this case the MR image has been taken as the moving one. Here, as in the previous experiments, the lossless *jpeg* compressor does not obtain the desired results. In all the other image compression approaches, the assumption that minimum compression is reached at the registration position is true.

Note that with these medical images the achieved results are very similar. Also a similar behaviour is observed in the C(x, y) and NCD plots, although the NCDones are slightly smoother. Such performance is probably due to the similar amount of information of all situations. Probably, this difference should be greater if the studied transformations lead to a position with less overlap area, since then the effect of the normalization would be more remarkable.

### 4.6.2 Text compression-based method

The second experiment was designed to evaluate the text compression-based method. In this case the used compressor is the bzip2 and the tested images are the T1 and T2 synthetic pair and the real MR-CT.

As in the previous experiment, the T1-weighted image has been taken as a fixed image and the T2-weighted image as the moving one. For the MR-CT, the MR image has been considered as the moving one. The experiment translates the moving images through the X axis from -10 pixels to 10 pixels around the registration position. In Figure 4.11, plots of the C(x, y) and NCD obtained for T1-T2 (first row) and MR-CT (second row) are presented. Observe that, for all the cases, the plots have similar behaviour. Therefore, as in the image compressor-based method, the assumption that minimum compression is reached at the registration position is verified. In our experiments with bzip2, we also observe that NCD > 1, which is a consequence of bzip2 compressor imperfections (see [30]).

If we compare Figures 4.9, 4.10 and 4.11, even though they have similar behaviour, in the case of the *bzip2* text compressor, plots have higher smoothness and less local minima. Consequently, we can say that the text compressor approach performs better than the image compressor-based one.

### 4.6.3 Entropy-based method

We now evaluate the performance of the entropy rate-based approach by reproducing the text compression-based experiment with the medical images. To evaluate the NED measure we consider three different L values, 1, 2 and 3, represented in Figure 4.12 by dash-dotted, solid and dashed lines, respectively. For comparison purposes, we also plot the standard image registration measure NMI represented in bold. Due to the high dimensionality of the joint histogram in the L=3 case, the number of bins has been reduced to 8. To preserve the testing conditions, this quantization has been kept in all cases. As we expected, entropy rate estimation decreases with L, giving us a more approximated measure of the real entropy rate and, equivalently, the string compressibility. Observe in both plots, the smoothness of the NCD curves, without local minima, and the accuracy of the registration, achieving their minimum at the correct position for both synthetic T1-T2 MR and real MR-CT pairs. From our experiments, we can conclude that the image registration using the Shannon information-based approach achieves more accuracy and robustness than real-world compressor-based approaches.



Figure 4.10: Compression and distance results of MR and CT images with the *jpeg*, *jpeg2000* and *png* compressors.



Figure 4.11: Compression (left column) and distance results (right column) of the synthetic MR T1 and T2 (first row) and the ones of the MR and CT images (second row) with the *bzip2* compressor.



Figure 4.12: Shannon-based measure results of the synthetic MR T1 and T2 (first row) and MR and CT images (second row).

## 4.7 Conclusions

We have presented a new compression-based framework for image registration using *the similarity metric*. The behaviour of this measure has been analyzed using two different perspectives: the Kolmogorov complexity and the Shannon information theory. Experimental results reveal the good performance of both versions, computed using standard real-world compressors and an estimation of entropy rate, respectively. Due to the compressor imperfections, more accurate and robust results are obtained in the second case. In contrast to traditional MI-based registration algorithms, which only use individual pixel information to perform computations, our approaches take into account spatial or structural information contained in the images.

## Chapter 5

# Multiresolution Image Registration Based on Tree Data Structures

This chapter presents a new approach for image registration based on the partitioning of the source images in binary-space (BSP) and quadtree structures, obtained with a maximum mutual information gain algorithm. Two different implementation approaches have been studied: one uses simplified images and the other deals directly with the tree data structures. With these approaches, multimodal registration experiments show an outstanding accuracy and robustness. In particular, the use of BSP partitioned images drastically reduces the grid effects compared with regular downsampled images. Another important advantage comes from the reduced size of the data structures corresponding to the simplified images, which makes this method appropriate to be applied in general multiresolution schemes and telemedicine applications.

## 5.1 Introduction

Multiresolution is a well-known strategy used in fields such as computer graphics and signal and image processing as a computational tool to enhance performance. The idea behind multiresolution consists in processing data at lower resolution reducing computational time and computational complexity to obtain a first approximation of the solution, and then, at higher resolution, use this approximation to reduce the complexity of finding a more accurate one. Examples of multiresolution can be found in many different computer graphics applications such as surface modeling, speed network transmission, volume rendering, CAD, geographic information systems, and virtual reality [125, 68, 67, 134, 28, 17]. In the image processing area, multiresolution has been used to define top-down approaches for feature detection and extraction, segmentation and image registration, amongst others [96, 124, 120, 119, 94].

The application of a multiresolution framework requires two main issues to be considered. The first one is the data structure used to maintain information at different levels of detail. Some of the structures that have been proposed are: tree based hierarchical structures, multiresolution spaces based on wavelet transforms, pyramidal representations, etc. The second issue is the method used to obtain the different level of detail representations, i.e. how to obtain simplified representations of input data. In this case, strategies such as subsampling and data compression can be used.

Focusing on image registration, where the transformation that aligns two input images has to be found, multiresolution strategies have been used to reduce the search space when looking for this transformation. Based on a coarse-to-fine hierarchical representation of the images, the process starts with both the reference and floating images on a coarser resolution and obtains a first approximation of the alignment function. Then, while going up to finer resolutions, this transformation is progressively more accurate [94, 116].

In this chapter, we propose a new multiresolution image registration process based on the partition of the images with a mutual-information-based algorithm. This algorithm is used to partition the reference and floating images to be registered in a binary-space (BSP) or in a quadtree mode. Once the hierarchical representation is created, we propose two different strategies to perform registration:

- The first one is a *pixel-based approach* that uses a simplified representation of input images such that each region of the partitioned image is represented as an homogeneous intensity area defined by the mean intensity value of the initial region. Once these images are obtained, the registration process starts.
- The second approach, denoted as a *node-based approach*, uses the hierarchical structure obtained in the partitioning process to compute the similarity metric required during the registration procedure. Instead of using the original image pixels to compute this metric, we propose to use the information stored in the nodes of the hierarchical data structure, exploiting, in this way, the coherence between simplified regions.

We evaluate the benefits of the proposed pixel-based and node-based approaches by comparing them with the classical multiresolution downsampling approach in which equally spaced pixels are used to simplify the images. The obtained results show an outstanding accuracy and robustness by using BSP partitioned images with respect to both regular downsampled and quadtree images. With BSP, grid effects are drastically reduced. We also observe that pixel-based and node-based approaches achieve similar results. In terms of complexity, the computational cost of the node-based approach depends linearly with the number of nodes while the cost of a pixel-based method remains constant with the number of regions. Therefore, for a low number of regions, it is more efficient to use the first approach.

This chapter is organized as follows. In Section 5.2, we briefly introduce the data structures and the MI-based partitioning techniques used for defining our multiresolution framework. In Section 5.3, the new image registration framework is presented and a detailed description of both proposed methods is given in Section 5.4 and Section 5.5, respectively. In Section 5.6, the different experiments that have been carried out to evaluate the methods are presented. Finally, conclusions and future work are given in Section 5.7.

## 5.2 Previous Work

In this section, the two main issues that have been taken into account to define our multiresolution framework are reviewed. First, we describe the Binary Space Partitioning (BSP) and the quadtree hierarchical data structures, and then we present the partitioning algorithm used to create the hierarchy of images with different levels of detail [122].

### 5.2.1 Hierarchical data structures

The most popular hierarchical data structures for image processing and computer vision applications are the quadtree and the BSP tree. The main advantage of these structures is that they are relatively simple to implement and inherently produce an efficient multiresolution representation of the data (see Figure 5.1).

The quadtree decomposition recursively divides the space, the image in our case, into four equal regions depending on a criterion that determines when a subdivision has to be done [131]. The root of the tree represents the original image at a zero level resolution, and the four resulting parts represent its children at a one level resolution. Each child is stored in a node of the quadtree and not decomposed nodes are called leaves.

The BSP scheme recursively subdivides the space into a set of regions. The algorithm [120] takes as input an unpartitioned region R, corresponding initially to the entire image, and a line I that partitions R into two regions is selected according to some criterion. The obtained regions are similarly partitioned in a recursive manner until a termination criterion is met. At the end of this process, we obtain a hierarchy of convex regions where the non-leaf nodes are associated with the partitioning lines and the leaf nodes with the image regions.

The main advantage of BSP-trees with respect to quadtrees is that they are not restricted to equal regions. In a quadtree, the partitioning lines are simply vertical and horizontal lines at fixed intervals and no extra computation or bits are needed to encode these lines. In our approach, the partitioning lines of the BSP-tree are chosen parallel to the borders of the image.



Chapter 5. Multiresolution Image Registration Based on Tree Data Structures

Figure 5.1: (a) A synthetic image of a polygon, (b) a quadtree and (c) a BSP representation of (a).

### 5.2.2 MI-based partitioning algorithm

As we have seen in the previous section, to create the hierarchical data structure at each node of the tree, a decision must be made to determine if the node has to be partitioned or not. Such a decision depends on a certain criterion determined by the user. In our case we will apply a criterion based on a maximum mutual information gain algorithm. This method was proposed by Rigau et al. [122] and it is based on the information bottleneck method [149]. The algorithm is constructed from the following information channel.

The information channel  $B \to R$  is defined between the random variables B (input) and R (output), which represent respectively the set of bins ( $\mathcal{B}$ ) of the histogram and the set of regions ( $\mathcal{R}$ ) of this image. Given an image  $\mathcal{I}$  of N pixels, where  $N_b$  is the frequency of bin b ( $N = \sum_{b \in \mathcal{B}} N_b$ ) and  $N_r$  is the number of pixels of region r ( $N = \sum_{r \in \mathcal{R}} N_r$ ), the three basic elements of this channel are the following:

- The conditional probability matrix p(R|B), which represents the transition probabilities from each bin of the histogram to the different regions of the image, is defined by  $p(r|b) = \frac{N_{b,r}}{N_b}$ , where  $N_{b,r}$  is the frequency of bin *b* into the region *r*. Conditional probabilities fulfill  $\forall b \in \mathcal{B}$ .  $\sum_{r \in \mathcal{R}} p(r|b) = 1$ .
- The input distribution p(B), which represents the probability of selecting each intensity bin b, is defined by  $p(b) = \frac{N_b}{N}$ .
- The output distribution p(R), which represents the normalized area of each region r, is given by  $p(r) = \frac{N_r}{N} = \sum_{b \in \mathcal{B}} p(b)p(r|b)$ .

The *mutual information* between B and R is defined by

$$I(B,R) = \sum_{b \in \mathcal{B}} \sum_{r \in \mathcal{R}} p(b,r) \log \frac{p(b,r)}{p(b)p(r)}$$
(5.1)

and represents the shared information or correlation between B and R.

For a decomposition of an image in n regions, the *ratio of mutual information* is defined by

$$M_s(n) = \frac{I(B,R)}{H(B)},$$
 (5.2)

where H(B) is the maximum value achievable for I(B, R) (when each region coincides with a pixel).

From the above reasonings, a partitioning algorithm which maximizes the gain of I(B, R) is proposed. In this algorithm, the full image is taken as the unique initial partition and is progressively subdivided according to the maximum mutual information gain for each partitioning step. This algorithm is a greedy top-down procedure which partitions an image in quasi-homogeneous regions. For more details, see Chapter 6 and [122].

## 5.3 Multiresolution Image Registration

In this section, our multiresolution image registration approach is presented. First, our method applies the MI-based partitioning algorithm to create the BSP or the quatree-based multiresolution scheme. Once the hierarchical structure is obtained, we propose two different strategies to perform the registration process. These two methods and their respective implementations are outlined below.

### • Pixel-based registration

The user enters a simplification ratio that determines the partitioning level of the images, i.e. the nodes of the hierarchy required for its representation. From this simplification process, we create two new images such that the region corresponding to each node is represented as a homogeneous intensity area (see Figure 5.2). Then, the obtained images are registered. In Section 5.4 we give a detailed description of the method and we evaluate it for the BSP and quadtree data structures for different input parameters. In Section 5.6, the proposed approach with the classical subsampling simplification strategies are also compared.

#### Node-based registration

An important component of the registration process is the metric that measures the correctness of the alignment between floating and reference images. This metric is evaluated at each iteration of the registration process until an optimum is reached. To compute the metric, input images are considered.





Figure 5.2: Main steps to obtain the simplified images.



Figure 5.3: Test images from the Vanderbilt database [104].

In this approach, we propose to compute the metric using the simplified information stored in the nodes of the hierarchical structure taking advantage of the region coherence. The implementation of the node-based registration approach and its performance will be analyzed in Sections 5.5 and 5.6, respectively.

## 5.4 Pixel-based Image Registration

The multiresolution pixel-based image registration approach is composed of two main steps: (i) the creation of a simplified representation of input images and (ii) the registration of the simplified images. These steps are detailed below.

### 1. Creation of simplified images from the hierarchical structure

Given two input images X and Y, to create the quadtree or the BSP hierarchies, the user enters a MI ratio given by  $MIR(X, \hat{Y}) = \frac{I(X, \hat{Y})}{H(X)}$ . This ratio is used as a stopping criterion and is considered as a measure of the simplification quality. See Chapter 6 for more details.

Once the BSP or the quadtree partitioned image has been obtained we construct the simplified images as follows. For each region of the partitioned image we create a homogeneous intensity region in a simplified image. This



Chapter 5. Multiresolution Image Registration Based on Tree Data Structures

Figure 5.4: (i) MR and (ii) CT images obtained from Figure 5.3(a-b). (a) Quadtree partitions with MIR = 0.7, (b) quadtree simplified images, (c) BSP partitions with MIR = 0.7, and (d) BSP simplified images.

MIR	]	MR	СТ	
	BSP	quadtree	BSP	quadtree
0.5	0.25	0.40	0.06	0.13
0.6	0.81	1.18	0.21	0.39
0.7	2.21	3.16	0.77	1.28
0.8	5.28	6.56	2.73	3.87
0.9	11.88	16.05	7.98	11.48

Table 5.1: Percentage of regions obtained with respect to the initial number of pixels corresponding to MR and CT original images of Figure 5.3(a-b) and for a given *MIR*.

intensity corresponds to the mean intensity values of the pixels of the corresponding region. The different steps that have to be applied to obtain the simplified images are illustrated in Figure 5.2.

The result of partitioning the 2D MR-CT pair of images of Figure 5.3(a-b) is shown in Figure 5.4. In the first and third column (Figures 5.4(a) and 5.4(c)) we present the quadtree and BSP partitioning of each original image. In this example, MIR = 0.7 has been used. The corresponding simplified images obtained by averaging for each region the intensity of its pixels are shown in Figures 5.4(b) and 5.4(d). The degree of simplification of the new images for different MIR is reported in Table 5.1. In each column we collect the percentage of regions obtained with the simplification with respect to the initial number of pixels corresponding to the original MR and CT images.



Figure 5.5: Registration for the MR-CT pair of Figure 5.3(a-b). (i) Quadtree and (ii) BSP subdivision methods for (a) MIR = 0.6 and (b) MIR = 0.7. The translational error  $(t_x, t_y)$  is shown for each registration.

Note that a big gain of MI is obtained with a relative low number of partitions. For instance, in the CT case, a 70% of MI (MIR = 0.7) is obtained with approximately 1% of the maximum number of partitions (number of pixels of the source image). It can also be observed that for the same MIR the number of partitions of the CT image is lower than the MR one. This is due to the fact that the higher the image homogeneity, the higher the degree of simplification. Thus, the CT image is more homogeneous than the MR image.

### 2. Registration of simplified images

In the second step of the process, the partitioned images are registered using the NMI metric. For the registration process, the NMI similarity measure and the Powell's method optimizer have been used. To illustrate the feasibility of this second step, we apply the registration to the simplified images of the MR-CT of Figure 5.3(a-b). We consider the quadtree and the BSP simplifications using MIR = 0.6 and MIR = 0.7. The registration results are shown in Figure 5.5, where (a) and (b) correspond to MIR = 0.6 and MIR = 0.7, and (i) and (ii) to the quadtree and BSP partitioned images, respectively. To evaluate the results we apply the transformation obtained from the registration of the simplified images to the original ones. In addition, for each of these images we compute the translational error  $(t_x, t_y)$  defined as the deviation in x and y translation between the transformation correspond-



Figure 5.6: Overlapping area between the floating and the reference images. (a) Information represented at pixel level (b) Information represented at node level.

ing to the registration of the original input images and the evaluated one. In all the cases, the rotational error has been omitted due to its insignificant value. Observe from the  $(t_x, t_y)$  values of each image that BSP images with MIR = 0.6 (Figure 5.5(ii.a)) achieve a lower error than quadtree images with MIR = 0.7 (Figure 5.5(i.b)). We can observe that better results are obtained with the registration of the BSP partitioned images.

## 5.5 Node-based Registration

As we have seen in Section 2.3.1, the registration process is composed of four main elements: the input images, the transform that represents the spatial mapping of points from the reference image space to points in the floating image space, the interpolator used to evaluate the moving image intensity at non-grid positions, and the metric which provides a measure of how well the reference image is matched by the transformed floating one. The metric is computed considering the pixels of the overlapping area between the images (see Figure 5.6(a)). Considering a tree simplification of these images, the overlapping area can be computed not for each pixel but for each homogeneous region (see Figure 5.6(b)). Our purpose is to exploit the coherence of homogeneous regions to compute the metric.

### 5.5.1 Metric computation

The algorithm proposed to compute the metric is described in Figure 5.7. The input parameters of the algorithm are the simplified representations of the initial reference and floating images (see Figure 5.8(a)). As in the pixel-based approach, the user provides a MIR value. Then, the quadtree or the BSP hierarchies of the

Input		
Reference Data Structure: <i>R</i>		
Floating Data Structure: F		
Output		
A similarity measure between $R$ and $F$		
Computation		
For each region $q$ in $F$ do		
Find bounding box of $q$		
Find the list $L$ of regions in $R$ inside the bounding box		
For each region p in L do		
Evaluate area overlapping between $q$ and $p$		
Update Histograms		
end		
end		
Evaluate measure		
return measure		

Figure 5.7: Algorithm of the similarity measure computation of two tree data structures.

reference and floating images are created and the corresponding reference (R) and floating (F) simplified representations are obtained (see Figure 5.2). Note that R and F representations can be defined as a list of nodes where each node represents a region of the simplified image.

To obtain the similarity measure between R and F, for each region of the list of nodes representing F, we apply the following steps:

- 1. The bounding box containing the node is defined (see Figure 5.9(a)).
- 2. All the nodes of the list of nodes representing R that intersect the bounding box are identified (see Figure 5.9(b)).
- 3. For each of these intersected nodes, the overlap area between them is computed (see Figure 5.9(c)). This step is described in detail in the next section.
- 4. The corresponding histogram entries according to the computed overlap is updated.

Finally, when all the nodes of F have been processed, the similarity measure is computed.



Figure 5.8: Reference and floating images are simplified using the *MIR* partitioning strategy.

### 5.5.2 Overlapping between nodes

The key to the metric computation algorithm is to calculate the overlapping area between nodes. Since nodes can be at different levels of the hierarchy and different transformations are applied to them, different intersection situations can be given. To consider all the possible cases, we propose a method consisting of two different steps:

• Obtaining the overlapping polygon

The polygon corresponding to the overlapping area is created by detecting all the intersection vertices between the floating and the reference regions (see Figure 5.10). These vertices are stored in the vertex polygon VP list. To identify VP points, two different cases are considered:

- The first situation is when the vertex point corresponds to the intersection of two node edges (see points *a* and *c* of Figure 5.10) or when the vertex of one node is inside the other node (see points *b*, *d* and *e* of Figure 5.10).
- The second situation, called a degenerated vertex, is when the polygon vertex does not correspond to the intersection of two node edges.
   Degenerated cases are illustrated in Figure 5.11 and correspond to the following patterns:
  - 1. *Vertex coincidence:* when a vertex of the reference region and a vertex of the floating region coincide in a point (see Figure 5.11(a)), this point has to be added to *VP* list.



Figure 5.9: (a) For each node of the floating image contained in the overlapping area a bounding box is defined. (b) The nodes of the reference image intersected by the bounding box are identified. (c) The intersection between the floating node and the reference nodes previously identified is computed.

- 2. *Vertex-edge coincidence:* when a vertex of a region coincides on a edge of the other region (see Figure 5.11(b) and 5.11(c)), this point is also added to the VP list.
- 3. *Edge-edge coincidence:* if two edges coincide (see Figure 5.11(d)), the extremes of the overlapping segment have to be added to the VP list.
- 4. *Edge intersection:* if two edges intersect in a point, this point is added to the *VP* list.
- 5. Vertex inside the other region: if a vertex of a region is inside the other region, it has to be added to the VP list.
- Computing the area of the overlapping polygon

To obtain the area of the polygon defined by the VP points, we sum the areas of the triangles that cover the polygon. To obtain these triangles we apply the following steps are applied:

- 1. Fist of all, we sort  $p_i \in VP$  with respect to the component X, i.e. if  $p_i = (x_i, y_i)$  we create the VP' list such that each  $p_i \in VP'$  satisfies that  $x_{i-1 \le x_i} \le x_{i+1}$  (see Figure 5.12(a)).
- 2. Then, we compute the first triangle considering the first three points  $p_1, p_2$  and  $p_3$  (see Figure 5.12(b)).
- 3. Being  $p_{i-2}, p_{i-1}, p_i$  the last generated triangle and  $p_{i+1}$  the next point of VP', to compute the next triangle we take  $p_i$  as a first vertex. Then,



Figure 5.10: Region overlapping.

to determine the next vertex, we create the segment between  $p_{i-1}$  and  $p_{i+1}$ . If this segment does not intersect the segment between  $p_{i-2}$  and  $p_i$ ,  $p_{i-1}$  is considered as the last point of the new triangle (see Figure 5.12(c)). Conversely, if the two segments intersect (see Figure 5.12(d)), the final point of the new triangle is  $p_{i-2}$  (see Figure 5.12(e)). This process is repeated until all VP' points have been considered.

For each of the triangles, its area is computed and, at the end, all areas are added (see Figure 5.12(f)). This algorithm requires polygons to be convex. For our requirements, this restriction is always satisfied since the overlapping polygon is always convex.

## 5.6 Results and Discussion

In this section, the performance of the proposed multiresolution image registration approach is evaluated. The pixel and the node-based registration of simplified images are considered and compared with the classical subsampling based simplification strategy. In our experiments, the MR-CT (Figure 5.3(a-b)) and MR-PET (Figure 5.3(c-d)) images, and different *MIR* values are taken. In all the experiments, the pair of images to be registered are simplified by using the same degree of simplification, i.e., an MR quadtree (or BSP) image with MIR = 0.7 is registered with a CT quadtree (or BSP) with the same MIR.

### 5.6.1 MR-CT tests

Our goal is to evaluate the robustness and accuracy of the pixel and node-based registration approaches. Two different tests are designed. In the first test, the floating image is moved through the X axis from -100 to 100 pixel units around the origin. At each step of the process a translation of one pixel is considered.



Figure 5.11: Degenerate cases: (a) vertex coincidence, (b) reference region vertex coincides with a segment, (c) floating region vertex coincides with a segment, and (d) segment overlapping.

The results of this first test applied on the MR-CT image pair are presented in Figure 5.13. The first row corresponds to the pixel-based approach and the second one to the node-based. From left to right, the columns correspond to the downsampling, quadtree and BSP multiresolution strategies. In all the plots, the bottom bold curve corresponds to the NMI registration of the source images.

If we compare the results of the pixel-based and the node-based registration approaches represented in the first (i) and second (ii) rows of Figure 5.13, respectively, it can be observed that no interpolation artifacts appear since there is no pixel interpolation. Note that all the plots are equal and hence both approaches achieve the same results demonstrating the correctness of the node-based registration. At a subvoxel level, the results of the node-based registration are equivalent to the partial volume ones [33].

To compare the results obtained with the three multiresolution strategies (subsampling, quadtree and BSP), the columns (a), (b) and (c) of Figure 5.13 are analyzed. In column (a), we illustrate the NMI measure obtained with different downsampling factors of the original images. From bottom to top, the NMI curves correspond to the downsampling of  $2 \times 2$ ,  $4 \times 4$ ,  $8 \times 8$  and  $16 \times 16$  pixels, respectively. Note that, in Figure 5.13, high artifacts appear at every *n* pixels coinciding with the downsampling factor. In the columns (b-c), we illustrate the NMI values for the quadtree and BSP partitioned images, respectively. Each curve corresponds to a different degree of simplification. From bottom to top, *MIR* ranges from 0.9 to 0.5. Observe in Figure 5.13(b) that the quadtree partition also produces correlation



Figure 5.12: Steps of the polygon area computation.

artifacts due to the regularity of its partitions. However, these artifacts are slightly reduced with respect to the downsampling case. This is due to the fact that, although the registered images have the same degree of simplification, the number and the position of the generated quadtree partitions are not the same. Finally, in Figure 5.13(c) we analyze the BSP partition. In this case, the grid artifacts are nearly completely eliminated since neither the position nor the number of partitions of the images coincide. Registration is more robust since the probability of finding a local maximum is lower as is shown by the smoothness of BSP plots. Taking into account that perfect registration is given by the maximum bottom curve, observe the high accuracy, i.e., the coincidence of the curve maxima, of the registration is achieved with MIR = 0.7, which represents an approximate reduction of 99% of the original number of pixels.

In the second test, the floating image is rotated from  $-180^{\circ}$  to  $180^{\circ}$  considering the center of the fixed image as the center of rotation. The obtained results are represented in Figure 5.14. As in the previous test, the first row corresponds to the pixel-based approach and the second one to the node-based. From left to right, the columns correspond to the downsampling, quadtree and BSP multiresolution strategies. In all the plots, the bottom bold curve corresponds to the NMI registration of the source images. Note that in this case the pixel-based and the node-based approaches do not produce the same results. If we compare the different multiresolution strategies we can see that the BSP approach (see Figure 5.14(c)) also achieves better results. Note that with the downsampling and quadtree schemes, Figure 5.14(a) and Figure 5.14(b), the regions of both images coincide at the 90°


Figure 5.13: MR-CT (Figure 5.3(a-b)) registration results corresponding to (a) downsampled, (b) quadtree, and (c) BSP images. The horizontal axis represents the slice translation on the X-axis (in pixels) and the vertical axis the value of NMI. For each plot, the NMI measure for different degrees of downsampling (a) and simplification (b-c) of the images are shown.

position (and  $-90^{\circ}$ ) and they cause a great artifact due to less dispersion on the joint histogram. In the BSP case, this artifact disappears due to less dependence on the partition positions. From these results we can see that the pixel-based approach achieves better results than the node-based one since it is less sensible to these artifacts.

#### 5.6.2 MR-PET tests

The same experiments with the MR-PET pair of images are performed. In these experiments, the images have been simplified with a MIR value ranging from 0.7 to 0.4. Greater values have not been considered since they generate to bigger hierarchical data structures.

The results obtained with the first test, when the floating image is translated through X-axis are shown in Figure 5.15. Comparing the pixel-based (first row) and the node-based (second row) approaches, we can see that both approaches achieve the same results. All the plots are similar. If we evaluate the different multiresolution strategies represented in the columns, we can see that downsampling and quadtree strategies have a lot of artifacts. The BSP case behaves better since artifacts are considerably reduced.

Although the rotational experiments shown in Figure 5.16 behave similarly to the ones in Figure 5.14, the mentioned artifacts for the downsampling and the quadtree approaches are much more emphasized. From these experiments we can



Figure 5.14: MR-CT (Figure 5.3(a-b)) registration results corresponding to (a) downsampled, (b) quadtree, and (c) BSP images. The horizontal axis represents the rotation (in degrees) and the vertical axis the value of NMI. For each plot, the NMI measure for different degrees of downsampling (a) and simplification (b-c) of the images are shown.

Number of Regions	200	300	400	500
Node-based	28.6	42.3	55.8	69.2
Pixel-based	52.3			

Table 5.2: Time required in milliseconds for the computation of the joint histogram using the pixel-based approach and the node-based approach depending on the number of regions.

conclude that the BSP simplification is more robust and accurate.

In Table 5.2, we show the mean time required to compute the joint histogram for a different number of regions. It is clear that, with the node-based approach, the time grows linearly with the number of regions, while in the pixel-based approach it is independent, since it only depends on the number of pixels, which is constant. From these results we can conclude that the node-based approach is more efficient for less than 400 regions. Such a simplification is obtained with a MIR = 0.56 for the MR image and MIR = 0.32 for the PET image.

# 5.7 Conclusions

In this chapter, we have presented a new technique for image registration based on the partitioning of the source images. The partitioning algorithm relies on the maximization of the mutual information gain for each refinement decision. The pre-





Figure 5.15: MR-PET (Figure 5.3(c-d)) registration results corresponding to (a) downsampled, (b) quadtree, and (c) BSP images. The horizontal axis represents the slice translation on the X-axis (in pixels) and the vertical axis the value of NMI. For each plot, the NMI measure for different degrees of downsampling (a) and simplification (b-c) of the images are shown.

sented method is a first step towards a full multi-resolution registration approach. Two alternatives (binary space partition and quadtree simplifications) have been analyzed and compared with a usual regular downsampling technique. The quality of the subdivision has been investigated in terms of the accuracy in registration. Results have shown the superior quality of the BSP subdivision, which allows a smoother registering function. The BSP approach also performs better than regular downsampling.



Figure 5.16: MR-PET (Figure 5.3(c-d)) registration results corresponding to (a) downsampled, (b) quadtree, and (c) BSP images. The horizontal axis represents the rotation (in degrees) and the vertical axis the value of NMI. For each plot, the NMI measure for different degrees of downsampling (a) and simplification (b-c) of the images are shown.

# **Chapter 6**

# Image Segmentation using Information Bottleneck Method

In this chapter, new image segmentation algorithms based on a hard version of the information bottleneck method are presented. The objective of this method is to extract a compact representation of a variable, considered as the input, with minimal loss of mutual information with respect to another variable, considered as the output. In the first place, we introduce a split-and-merge algorithm based on the definition of an information channel between a set of regions (input) of the image and the intensity histogram bins (output). From this channel, the maximization of the mutual information gain is used to optimize the image partitioning. Then, the merging process of the obtained regions in the previous phase is carried out by minimizing the loss of mutual information. From the inversion of the above channel, we also present a new histogram clustering algorithm based on the minimization of the mutual information loss, where now the input variable represents the histogram bins and the output is given by the set of regions obtained from the above split-and-merge algorithm. Finally, we introduce two new clustering algorithms which show how the information bottleneck method can be applied to the registration channel obtained when two multimodal images are correctly aligned. From this channel, the histogram bins of each image are clustered by maximally preserving the shared information. Different experiments on 2D and 3D images show the behavibour of the proposed algorithms.

# 6.1 Introduction

The main objective of image segmentation is to divide an image into regions that can be considered homogeneous with respect to a given criterion such as color or texture. Image segmentation is one of the most widely studied problems in image analysis and computer vision and it is a significant step towards image understanding. Many different methods, such as thresholding, region growing, region splitting and merging, active contours, and level sets, have been proposed. Each one of these methods considers the segmentation problem from a different perspective and is suitable for solving a limited number of situations. For a survey of segmentation algorithms see [55].

The purpose of this chapter is to introduce new segmentation algorithms using a hard version of the *information bottleneck method* [149]. The use of this method requires the definition of an information channel where a random variable *controls* the clustering of the other by preserving the maximum mutual information between them. That is, the objective of this method is to extract a compact representation of a random variable with minimal loss of mutual information with respect to another variable.

In this chapter, the information bottleneck method will be applied to two different channels: (i) the channel defined between the set of regions of a given image and its histogram bins, and (ii) the channel built between the histogram bins of two multimodal registered images. From the first channel, both split-and-merge and histogram clustering algorithms are introduced and, from the second channel, both one-sided and two-sided histogram clustering algorithms are presented. While the splitting process is guided by the maximization of the mutual information gain, all the other processes (merging and clustering) are driven by the minimization of the mutual information loss.

The following information-bottleneck-based algorithms represent the main contributions of this chapter:

- *Split-and-merge algorithm* (Section 6.3). In the first phase, a top-down strategy is applied to partition an image into quasi-homogeneous regions using a binary space partition (BSP) or a quadtree partition. In the second phase, a bottom-up strategy is used to merge the regions whose histograms are more similar.
- *Histogram clustering algorithm* (Section 6.4). neighbour bins of the histogram are clustered from a previously partitioned image. After assuming that the splitting-and-merge algorithm provides us with the structure of the image, our clustering algorithm tries to preserve the correlation between the clustered bins and the structure of the image.
- *Histogram clustering algorithms for two registered multimodal images* (Section 6.5). Two different algorithms are presented. The first one segments just one image at a time, while the second one segments both simultaneously. The clustering process works by extracting from each image the structures that are more relevant to the other one. In these algorithms, each image is used to control the quality of the segmentation of the other.

The proposed methods have several advantages. In spite of the fact that the clustering algorithms presented in this chapter are based on the image histograms, that do not retain spatial information, the spatial coherence is taken into account by the information channels. In the split-and-merge algorithm, this channel makes the correspondence between the structure of the image and the histogram bins. This spatial information makes the method robust to texture analysis, without assuming any a priori intensity or texture distribution. The proposed histogram clustering algorithm considers the spatial distribution of the intensities to achieve a good representation of the colors of the image. The obtained segmentation tries to preserve with a given number of colors the maximum spatial information of the original image. Finally, the registration-based segmentation is able to segment one image from the information of another. For instance, this algorithm enables us to segment images of low quality from the information contained in high quality images. This technique could be used to segment intraoperative images using high quality preoperative ones. A global advantage of these methods is that they do not assume any a priori information about the images (e.g. intensity probability distribution).

The results of our experiments show the feasibility of the information bottleneck method to deal with different 2D and 3D image segmentation techniques.

# 6.2 Information Bottleneck Method

The information bottleneck method, introduced by Tishby et al. [149], extracts a compact representation of the variable X, denoted by  $\hat{X}$ , with minimal loss of MI with respect to another variable Y (i.e.,  $\hat{X}$  preserves as much information as possible about the relevant variable Y). Soft [149] and hard [140] partitions of X can be adopted. In the first case, every cluster  $x \in \mathcal{X}$  can be assigned to every cluster  $\hat{x} \in \hat{\mathcal{X}}$  with some conditional probability  $p(\hat{x}|x)$  (soft clustering). In the second case, every cluster  $x \in \mathcal{X}$  is assigned to only one cluster  $\hat{x} \in \hat{\mathcal{X}}$  (hard clustering).

In this chapter, we focus our attention on the *agglomerative information bottle*neck method [140]. Given a cluster  $\hat{x}$  defined by  $\hat{x} = \{x_1, \ldots, x_l\}$ , where  $x_k \in \mathcal{X}$ , and given probability distributions  $p(\hat{x})$  and  $p(y|\hat{x})$  defined by

$$p(\hat{x}) = \sum_{k=1}^{l} p(x_k),$$
(6.1)

$$p(y|\hat{x}) = \frac{1}{p(\hat{x})} \sum_{k=1}^{l} p(x_k, y) \quad \forall y \in \mathcal{Y},$$
(6.2)

the following properties are fulfilled:

• The decrease in the mutual information I(X, Y) due to the merge of  $x_1, \ldots, x_l$  is given by



Figure 6.1: The information channel between the regions of the images, R, and the intensity bins, B, of the split-and-merge algorithm. The channel used in the histogram clustering algorithm is the same but inverting the direction.

$$\delta I_{\hat{x}} = p(\hat{x}) JS(\pi_1, \dots, \pi_l; p_1, \dots, p_l) \ge 0, \tag{6.3}$$

where  $\pi_k = \frac{p(x_k)}{p(\hat{x})}$  and  $p_k = p(Y|x_k)$ . An optimal clustering algorithm has to minimize  $\delta I_{\hat{x}}$ .

• An optimal merge of l components can be obtained by l - 1 consecutive optimal merges of pairs of components.

Dhillon et al. [46] presented a co-clustering algorithm applied to text-document clustering that simultaneously clusters X and Y into disjoint or hard clusters. An optimal co-clustering algorithm has to minimize the difference  $I(X, Y) - I(\hat{X}, \hat{Y})$ .

### 6.3 Split-and-Merge Algorithm

In this section we present an split-and-merge algorithm that is constructed from an information channel  $R \to B$  between the random variables R (input) and B (output), which represent, respectively, the set of regions  $\mathcal{R}$  of an image and the set of intensity bins  $\mathcal{B}$  (see Figure 6.1). This channel is defined by a conditional probability matrix p(B|R) which expresses how the pixels corresponding to each region of the image are distributed into the histogram bins. Throughout this chapter, the capital letters R and B as arguments of p() will be used to denote probability distributions. For instance, while p(R) will represent the input distribution of the regions, p(r) will denote the probability of a single region r.

Given an image with N pixels,  $N_r$  regions, and  $N_b$  intensity bins, the three basic elements of the channel  $R \rightarrow B$  are:

• The conditional probability matrix p(B|R), which represents the transition probabilities from each region of the image to the bins of the histogram, is defined by  $p(b|r) = \frac{n(r,b)}{n(r)}$ , where n(r) is the number of pixels of region



Figure 6.2: Test images: (a) Lena and (b) Hematoma. The two plots in (c) show the mutual information ratio  $(MIR_r)$  with respect to the number of regions for (a) and (b).

r and n(r,b) is the number of pixels of region r corresponding to bin b. Conditional probabilities fulfil  $\sum_{b \in \mathcal{B}} p(b|r) = 1, \forall r \in \mathcal{R}.$ 

- The input distribution p(R), which represents the probability of selecting each image region, is defined by  $p(r) = \frac{n(r)}{N}$  (i.e. the relative area of region r).
- The output distribution p(B), which represents the normalized frequency of each bin b, is given by  $p(b) = \sum_{r \in \mathcal{R}} p(r)p(b|r) = \frac{n(b)}{N}$ , where n(b) is the number of pixels corresponding to bin b.

From the data processing inequality (2.19) and the information bottleneck method (Section 6.2), we know that any clustering or quantization over R or B, respectively represented by  $\hat{R}$  and  $\hat{B}$ , will reduce I(R, B). Thus,  $I(R, B) \ge I(R, \hat{B})$ and  $I(R, B) \ge I(\hat{R}, B)$ .

#### 6.3.1 Splitting

The splitting phase of the algorithm is a greedy top-down procedure which partitions an image in quasi-homogeneous regions. Our partitioning strategy takes the full image as the unique initial partition and progressively subdivides it (e.g. with vertical or horizontal lines in 2D images (BSP)) chosen according to the maximum MI gain for each partitioning step. In our experiments, BSP and quad-tree strategies will be used. Note that similar algorithms have been introduced in the context of pattern recognition [135], learning [77], and DNA segmentation [14]. This splitting algorithm has been introduced by Rigau et al. in [121].

The partitioning process is represented over the channel  $\tilde{R} \to B$ , where  $\tilde{R}$  denotes that R is the variable to be partitioned. Note that this channel varies at each partition step because the number of regions is increased and, consequently, the marginal probabilities of  $\tilde{R}$  and the conditional probabilities of  $\tilde{R}$  known B also change. For a BSP strategy, the gain of MI due to the partition of a region  $\tilde{r}$  in two neighbour regions  $r_1$  and  $r_2$ , such that

$$p(\tilde{r}) = p(r_1) + p(r_2) \tag{6.4}$$

and

$$p(b|\tilde{r}) = \frac{p(r_1)p(b|r_1) + p(r_2)p(b|r_2)}{p(\tilde{r})},$$
(6.5)

is given by

$$\delta I_{\tilde{r}} = I(R, B) - I(R, B) = p(\tilde{r}) JS(\pi_1, \pi_2; p(B|r_1), p(B|r_2)), \qquad (6.6)$$

where  $\pi_1 = \frac{p(r_1)}{p(\tilde{r})}$  and  $\pi_2 = \frac{p(r_2)}{p(\tilde{r})}$ . The JS-divergence  $JS(\pi_i, \pi_j; p(B|r_1), p(B|r_2))$  between two regions can be interpreted as a measure of *dissimilarity* between them. That is, when a region is partitioned, the gain of MI is equal to the degree of dissimilarity between the resulting regions times the size of the region. In our splitting algorithm, the optimal partition is determined by the the maximum MI gain  $\delta I_{\tilde{r}}$ .

The BSP partitioning algorithm can be represented by an evolving binary tree where each leaf corresponds to a terminal region of the image [135]. At each partitioning step, the tree gains information from the original image such that each internal node k contains the information  $I_k$  gained with its corresponding splitting. At a given moment, I(R, B) can be obtained adding up the information available at the internal nodes of the tree weighted by p(k), where  $p(k) = \frac{n(k)}{N}$  is the relative area of the region associated with node k and n(k) is the number of pixels of this region. Thus, the MI of the channel is given by

$$I(R,B) = \sum_{k=1}^{T} p(k)I_k,$$
(6.7)



Figure 6.3: Partition of the Lena image (Figure 6.2(a)) with two different  $MIR_r$  for (a-b) quadtree and (c-d) BSP simplifications.

where T is the number of internal nodes. It is important to stress that the best partition can be decided locally. That is, the information gained  $I_k$  in a given node k is independent of the level of partitioning of the other regions of the image.

From the Equation (2.7), the partitioning procedure can also be visualized as H(B) = I(R, B) + H(B|R), where H(B) is the histogram entropy and I(B, R) and H(B|R) represent, respectively, the successive values of MI and conditional entropy obtained after the successive partitions. The progressive acquisition of information increases I(R, B) and decreases H(B|R). This reduction of conditional entropy is due to the progressive homogenization of the resulting regions. Observe that the maximum MI that can be achieved is the histogram entropy H(B), that remains constant along the process. The partitioning algorithm can be stopped using a ratio  $MIR_r = \frac{I(R,B)}{H(B)}$  of mutual information gain or a predefined number of regions  $N_r$ .

Figure 6.2(a) and 6.2(b) show two test images used in our experiments. The first corresponds to the well-known Lena image and the second to a CT medical brain image with a hematoma lesion. In this chapter, the segmentation of colored images is obtained using the luminance channel. The two curves in Figure 6.2(c) indicate the behavibour of  $MIR_r$  with respect to the number of partitions, which have been obtained using a BSP strategy, for both test images. These plots show the concavity of the  $MIR_r$  function. It can be clearly appreciated that a big gain of MI is obtained with a low number of partitions. Thus, for instance, a 50% of



Figure 6.4: Partition of the Hematoma image (Figure 6.2(b)) with two different  $MIR_r$  for (a-b) quadtree and (c-d) BSP simplifications.

MI is obtained with approximately 1% of the maximum number of partitions for the Hematoma test image. Observe that in the Hematoma image less partitions are needed to extract the same  $MIR_r$  than in the Lena image. Note also that the maximum  $MIR_r$  is achieved with the 50% of the regions in the Hematoma image, since the final splitting iterations do not gain any information because the final regions are completely homogeneous.

Figure 6.3 and Figure 6.4 present the results of partitioning the Lena and Hematoma test images. We show the partitioned images corresponding to two different  $MIR_r$  for quadtree and BSP simplifications. Observe that, for the same degree of simplification, the BSP partition fits better to the image structure, due to the freedom of the partition position of this simplification scheme. For instance, observe how the first BSP partitions of the Hematoma image (Figure 6.4(c)) try to separate the brain structure from the background. Despite these interesting results, they can not be used by themselves as final segmentation and a merging process is needed to achieve a correct image segmentation. This merging process is widely explained in the next section.

#### 6.3.2 Merging

From the agglomerative information bottleneck method [140] applied to the channel  $R \rightarrow B$ , we know that any clustering over R will not increase I(R, B). Analogous to the MI gain (6.6) obtained in the splitting phase, the loss of MI due to the clustering  $\hat{r}$  of two neighbour regions  $r_1$  and  $r_2$  is given by

$$\delta I_{\hat{r}} = I(R, B) - I(\hat{R}, B) = p(\hat{r}) JS(\pi_1, \pi_2; p(B|r_1), p(B|r_2)), \qquad (6.8)$$

where  $\hat{X}$  denotes that the variable X has been clustered,  $p(\hat{r}) = p(r_1) + p(r_2)$ ,  $p(b|\hat{r}) = \frac{p(r_1)p(b|r_1) + p(r_2)p(b|r_2)}{p(\hat{r})}$ ,  $\pi_1 = \frac{p(r_1)}{p(\hat{r})}$ , and  $\pi_2 = \frac{p(r_2)}{p(\hat{r})}$ . As we have seen in the splitting phase, the JS-divergence between two regions

As we have seen in the splitting phase, the JS-divergence between two regions can be interpreted as a measure of *dissimilarity* between them. The similarity will be maximum when the two regions have the same histogram: if  $p(B|r_1) = p(B|r_2)$ , then  $\delta I_{\hat{r}} = 0$ . Thus, if two regions are very similar (i.e., the JS-divergence between them is small) the channel could be simplified by substituting these two regions by their merging, without a significant loss of information. This is the principle that leads to the following merging algorithm.

From a given image partitioning, the algorithm merges successively the pairs  $(r_1, r_2)$  of neighbour regions such that  $\delta I_{\hat{r}}$  is minimum. Thus, the number of regions decreases progressively together with the MI of the channel. Similarly to the splitting algorithm, the stopping criterion can be determined by the ratio  $MIR_r = \frac{I(R,B)}{H(B)}$  or a predefined number of regions.

Note that the clustering  $\hat{R}$  of all regions would give  $I(B, \hat{R}) = 0$ . From (2.7), during the merging process  $H(B) = I(B, \hat{R}) + H(B|\hat{R})$ , where  $I(B, \hat{R})$  and  $H(B|\hat{R})$  represent, respectively, the successive values of MI and conditional entropy obtained after the successive mergings. Remember that H(B) remains constant. Note also that  $H(B|\hat{R})$  is the average entropy of the regions, given by

$$H(B|\widehat{R}) = -\sum_{r \in \mathcal{R}} p(r) \sum_{b \in \mathcal{B}} p(b|r) \log p(b|r)$$
  
=  $-\sum_{r \in \mathcal{R}} p(r) H(B|r),$  (6.9)

where H(B|r) is the entropy of the normalized histogram of region r. If two regions are clustered:

$$\delta I_{\hat{r}} = I(R, B) - I(\hat{R}, B) = H(B|\hat{R}) - H(B|R).$$
(6.10)

Thus,  $H(B|\hat{R})$  never decreases at any iteration due to the mixing of the histogram regions.

In Figure 6.5, we show the results of merging the regions of the images of Figures 6.2(a) and 6.2(b) obtained from the splitting phase with a  $MIR_r$  of 0.8 with the BSP simplification. For both images, the results of 6 and 10 different regions are shown. Observe that in this case the main structures of the image are separated, specially for the Hematoma image, where the lesion, the skull, and internal brain structures, as the ventricles, are correctly identified. In the Lena image the main structures of the images are identified, but the illumination problem



Figure 6.5: Segmentation results of the split-and-merge algorithm for the Lena image (Figure 6.2(a)) and Hematoma image (Figure 6.2(a)), where R represents the final number of regions of each image.

over the same object is not solved at all by the method. For instance, observe the uncorrect segmentation of the hat. This is due to the fact that the method only deals with local intensities and not with other image features such as gradient or texture.

In Figure 6.6, we depict the results of applying the split-and-merge algorithm to four images of the Berkeley database [97], where a given number of regions has been predetermined for each image. Note the good behavibour of our split-and-merge algorithm, which is able to detect the homogeneity of some textured regions (such as the field in Figure 6.6(a) or the sand in Figure 6.6(d)). This is because this kind of regions maintain a similar spatial distribution of the intensities along the whole region.

# 6.4 Histogram Clustering Algorithm

In this section we present a greedy histogram clustering algorithm which takes as input a partitioned image and obtains a histogram clustering based on the minimization of the loss of MI. That is, we group the bins of the histogram so that the MI is maximally preserved. From the perspective of the information bottleneck method, the binning process is controlled by a given partition of the image. This histogram clustering algorithm has been introduced by Rigau et al. in [121].

which is a result of inverting the channel of the previous section. This channel



(c) R=12

(d) R=2

Figure 6.6: Segmentation results of the split-and-merge algorithm for different images from the Berkeley database, where R represents the final number of regions of each image.

is defined by a conditional probability matrix p(R|B) which expresses how the pixels corresponding to each histogram bin are distributed into the regions of the image. Bayes' theorem, expressed by p(b)p(r|b) = p(r)p(b|r), establishes the relationship between the conditional probabilities of both channels  $B \to R$  and  $R \to B$ .

The basic idea underlying our histogram clustering algorithm is to capture the maximum information of the image with the minimum number of histogram bins. Analogous to the merging algorithm of the previous section, the loss of MI due to the clustering  $\hat{b}$  of two neighbour bins  $b_1$  and  $b_2$  is given by

$$\delta I_{\hat{b}} = I(B, R) - I(\hat{B}, R) = p(\hat{b}) JS(\pi_1, \pi_2; p(R|b_1), p(R|b_2)), \qquad (6.11)$$

where  $p(\hat{b}) = p(b_1) + p(b_2)$ ,  $p(r|\hat{b}) = \frac{p(b_1)p(r|b_1) + p(b_2)p(r|b_2)}{p(\hat{b})}$ ,  $\pi_1 = \frac{p(b_1)}{p(\hat{b})}$ , and  $\pi_2 = \frac{p(b_2)}{p(\hat{b})}$ . Thus, when two neighbour bins  $b_1$  and  $b_2$  are equally distributed in the regions of the image  $(p(R|b_1) = p(R|b_2))$ , their clustering results in  $\delta I_{\hat{b}} = 0$ . In general, if two bins are very similar, the channel can be simplified by substituting these two bins by their clustering, without a significant loss of information. Our algorithm proceeds by merging two neighbour bins so that the loss of MI is minimum. The stopping criterion is given by the ratio  $MIR_b = \frac{I(\hat{B},R)}{I(B,R)}$  or a predefined



Figure 6.7: Segmentation results of the histogram clustering algorithm for the Lena image (Figure 6.2.a) and Hematoma image (Figure 6.2.b), where C represents the final number of intensity bins of each image.

number of bins  $N_{\rm b}$ .

Note that, during the clustering process  $H(R) = H(R|\widehat{B}) + I(\widehat{B}, R)$ , where H(R) is the entropy of p(R), and  $H(R|\widehat{B})$  and  $I(\widehat{B}, R)$  represent, respectively, the successive values of conditional entropy and MI obtained after the successive clusterings. Observe also that  $H(R|\widehat{B})$  is the average entropy of the bins (i.e. a measure of the degree of dispersion of the bins in the set of regions) and increases (or remains constant) at each iteration.

In Figure 6.7 we show the segmented images obtained from the partitions achieved with the split-and-merge algorithm with  $MIR_r = 0.8$  as stopping criterion of the splitting process and 100 regions for the merging one. For each image, the results obtained using 4 and 6 clusters are shown. For instance, observe how the internal structures of the brain are approximately preserved using only 6 clusters.

#### 6.5 Registration-based Segmentation

In this section, two histogram clustering algorithms based on the channel established between two registered images A and B are introduced. The main idea behind our algorithms is that the segmentation of image A is obtained by extracting the structures that are most relevant for image B. These histogram clustering algorithms have been introduced in [8].

```
\begin{array}{l} \textbf{Input}\\ \textbf{Joint probability distribution: } p(x,y)\\ \textbf{Number of clusters: } m \in \{1..|X|\}\\ \textbf{Output}\\ \textbf{A partition of } X \text{ into } m \text{ clusters}\\ \textbf{Computation}\\ \widehat{X} \leftarrow X\\ \forall i \in \{1..|X|-1\}.\delta I_{\widehat{x}}(i) \leftarrow (p(x_i) + p(x_{i+1}))JS(\pi_i, \pi_{i+1}; p(Y|x_i), p(Y|x_{i+1}))\\ \textbf{while } |\widehat{X}| > m \textbf{do}\\ k \leftarrow \min_i(\delta I_{\widehat{x}}(i))\\ \widehat{x} \leftarrow \operatorname{merge}(x_k, x_{k+1})\\ \widehat{X} \leftarrow (X - \{x_k, x_{k+1}\}) \bigcup \{\widehat{x}\}\\ \textbf{Update } \delta I_{\widehat{x}} \text{ for the neighbours of } \widehat{x}\\ X \leftarrow \widehat{X}\\ \textbf{end while} \end{array}
```

Figure 6.8: One-sided clustering algorithm.

#### 6.5.1 One-sided clustering algorithm

We present a greedy hierarchical clustering algorithm that clusters the histogram bins of image A by minimizing the loss of MI between A and B. First of all, in a preprocessing step, images A and B have to be registered, establishing an information channel  $X \to Y$ , where X and Y denote, respectively, the histograms of A and B (see Figure 6.9). From the data processing inequality (2.19) and the information bottleneck method (see Section 6.2), we know that any clustering over X (for instance, merging neighbour histogram bins  $x_1$  and  $x_2$ ), denoted by  $\hat{X}$ , will reduce I(X, Y).

At the initial stage of our algorithm (see Figure 6.8), only one intensity value is assigned to each histogram bin of X. Then, the algorithm proceeds greedily by merging two neighbour clusters so that the loss of MI is minimum. This procedure merges the two clusters which are more similar from the perspective of B. Note the constraint that only neighbour bins can be merged. The cardinality  $|\hat{X}|$  goes from |X| to 1 in the extreme case.

The efficiency of this algorithm can be greatly improved if the reduction of MI due to the merging of bins  $x_1$  and  $x_2$  is computed by

$$\delta I_{\hat{x}} = p(\hat{x}) JS(\pi_1, \pi_2; p(Y|x_1), p(Y|x_2)), \tag{6.12}$$

where  $p(\hat{x}) = p(x_1) + p(x_2)$ ,  $\pi_i = \frac{p(x_1)}{p(\hat{x})}$ ,  $\pi_2 = \frac{p(x_2)}{p(\hat{x})}$ , and  $p(Y|x_1)$  and  $p(Y|x_2)$  denote, respectively, the corresponding rows of the conditional probability matrix of the information channel [140]. The evaluation of  $\delta I_{\hat{x}}$  for each pair of clusters is done in O(|Y|) operations and, at each iteration of the algorithm, it is only necessary to compute the  $\delta I_{\hat{x}}$  of the new cluster with its two corresponding neighbours. All the other precomputed  $\delta I_{\hat{x}}$  values remain unchanged [140].

Similar to the algorithms of Sections 6.3 and 6.4, clustering can be stopped



Figure 6.9: The information channel between the intensity bins of image A, X, and the intensity bins of image B, Y, of the registration-based segmentation algorithms.

using several criteria: a fixed number of clusters, a given ratio  $MIR_b = \frac{I(\hat{X},Y)}{I(X,Y)}$ , or a variation  $\delta I_{\hat{x}}$  greater than a given  $\epsilon$ . The  $MIR_b$  ratio is considered as a quality measure of the clustering. In Section 6.5.3, we will show the results of the normalized mutual information,  $NMI = I(\hat{X}, Y)/H(\hat{X}, Y)$ , which provides us with an efficiency coefficient of the segmentation process [21], and  $-\delta I_{\hat{x}}/I(X,Y)$ , which indicates the relative loss of information of a given clustering [149].

#### 6.5.2 Co-clustering algorithm

Let us now consider a simultaneous clustering of images A and B. Unlike the algorithm presented by Dhillon [46] for word-document clustering, which alternatively clusters the variables  $\hat{X}$  and  $\hat{Y}$ , our algorithm (see Figure 6.10) chooses at each step the best merging of one of the two images (i.e., the one that entails a minimum reduction of MI). The similarity between the two images is being symmetrically exploited. Thus, each clustering step benefits from the progressive simplification of the images. One of the main advantages of this algorithm is the great reduction of sparseness and noise of the joint probability matrix. As we will see with the experimental results, the simultaneous merging over the images A and B obtain better results than with the one-sided algorithm.

From the data processing inequality (2.19),  $I(\hat{X}, \hat{Y})$  is a decreasing function with respect to the reduction of the total number of clusters  $|\hat{X}| + |\hat{Y}|$ . Thus,  $I(\hat{X}, \hat{Y}) \leq I(X, Y)$ . Like the one-sided algorithm, the stopping criterion can be given by a predefined number of bins, a given ratio  $MIR = I(\hat{X}, \hat{Y})/I(X, Y)$  or a variation  $\delta I_{\hat{x}}$  (or  $\delta I_{\hat{y}}$ ) greater than a given  $\epsilon$ . Similarly to the above one-sided algorithm, the reduction of MI can be computed from the JS-divergence (6.12). Input Joint probability distribution: p(x, y)Number of clusters:  $m \in \{1..|X| + |Y|\}$ Output A partition of (X, Y) into m clusters Computation  $(\widehat{X}, \widehat{Y}) \leftarrow (X, Y)$  $\forall i \in \{1..|X|-1\} \\ \delta I_{\hat{x}}(i) \leftarrow (p(x_i) + p(x_{i+1})) \\ JS(\pi_i, \pi_{i+1}; p(Y|x_i), p(Y|x_{i+1})) \\ \delta I_{\hat{x}}(i) \leftarrow (p(x_i) + p(x_{i+1})) \\ \delta I_{\hat{x}}(i) \leftarrow (p(x_i) + p(x_i) + p(x_i)) \\ \delta I_{\hat{x}}(i) \leftarrow (p(x_i) + p(x_i) + p(x_i)) \\ \delta I_{\hat{x}}(i$  $\forall j \in \{1..|Y|-1\} \\ \delta I_{\hat{y}}(j) \leftarrow (p(y_j) + p(y_{j+1})) \\ JS(\pi_j, \pi_{j+1}; p(X|y_j), p(X|y_{j+1}))$ while  $|\widehat{X}| + |\widehat{Y}| > m$  do  $k \leftarrow \min_{i,j}(\delta I_{\hat{x}}(i), \delta I_{\hat{y}}(j))$ if k indexes  $\widehat{X}$  then associate (Z, V) to  $(\widehat{X}, \widehat{Y})$  else associate (Z, V) to  $(\widehat{Y}, \widehat{X})$  $\hat{z} \leftarrow \operatorname{merge}(z_k, z_{k+1})$  $\widehat{Z} \leftarrow (Z - \{z_k, z_{k+1}\}) \bigcup \{\widehat{z}\}$ Update  $\delta I_{\hat{z}}$  for the neighbours of  $\hat{z}$ Update all  $\delta I_v$  $(X,Y) \leftarrow (\widehat{X},\widehat{Y})$ end while

Figure 6.10: Co-clustering algorithm.

But in the co-clustering algorithm, for each clustering of  $\hat{X}$  (or  $\hat{Y}$ ), it is necessary to recompute all the  $\delta I_{\hat{y}}$  (or  $\delta I_{\hat{x}}$ ). Figure 6.10 shows the co-clustering algorithm where the stopping criterion is given by the total number of clusters.

#### 6.5.3 Results and discussion

To evaluate the performance of the two registration-based segmentation algorithms, we have used both synthetic and real images. The first test images are a set of synthetic magnetic resonance T1 (MR-T1) and T2 (MR-T2) image modalities from the Brainweb database [32]. These images are obtained synthetically from a phantom and they can be generated with different levels of image noise. These two image modalities are acquired exactly in the same spatial position and therefore the preprocessing registration step is not required. The second test images are real data from a patient from the Vanderbilt database [104]. This dataset is composed of MR and CT image modalities. The resolution of the MR and CT is  $256 \times 256 \times 26$  and  $512 \times 512 \times 28$ , respectively. These MR and CT images have been registered using the NMI measure [144].

Figure 7.8 shows the results of the proposed one-sided and co-clustering algorithms for the MR-T1 and MR-T2 Brainweb 3D images with a 3% of noise. These images are simulated from a synthetic atlas and they are perfectly registered since the same process is applied to achieve both images. The original MR-T2 and MR-T1 images are depicted in Figure 7.8(ii.a) and Figure 7.8(iii.a), respectively. Columns (b-d) show the segmented images with 4, 5, and 6 clusters, respectively. The results obtained with the one-sided algorithm applied on the MR-T1 and MR- T2 images are shown in Figure 7.8(i.b-d) and Figure 7.8(iv.b-d), respectively. The results obtained with the co-clustering algorithm are shown for the MR-T2 image in Figure 7.8(ii.b-d) and for the MR-T1 in Figure 7.8(iii.b-d).

Observe the good segmentation results achieved with both methods for the MR-T2 image. For both methods, the images obtained with only 4 clusters distinguish between background (black), white matter (dark gray), gray matter (light gray), and ventricles and cerebral fluids (white), which are the main structures of brain anatomy. The results are similar for the MR-T1 image and the one-sided algorithm, but they are not so satisfactory for the co-clustering one. In this case, the background is split into two clusters while gray and white matter are considered in the same cluster. This may be because the background has higher probability in comparison with any other region of the image. This undesired behavibour disappears when 5 or 6 clusters are considered.

With the next experiment we want to simulate the case where one image of poor quality is segmented considering a high quality image, similar than the preoperative and intraoperative images. In order to study this situation, we have considered the MR-T1 Brainweb image with 1% of noise to be a high quality image and a MR-T2 Brainweb image with 7% of noise to be a poor quality image. In this situation only the one-sided algorithm is considered, taking as a control variable the high quality image (MR-T1, Figure 6.12(a)) and segmenting the poor quality image (MR-T2, Figure 6.12(b)). The results of the one-sided algorithm with 4,5 and 6 clusters are plotted in Figures 6.12(c), 6.12(d), and 6.12(e), respectively.

As we can observe in these images, even the poor quality of the original one, the segmentation results try to separate correctly the main parts of the brain image: background, ventricles, white matter and gray matter. This is because the control variable of the segmentation method is very accurate and tries to achieve the maximum relationship between the input image and the achieved segmentation.

In Figure 6.13, we show the results obtained with the one-sided and co-clustering algorithms applied on the CT (Figure 6.13(ii.a)) and MR (Figure 6.13(iii.a)) original image of the Vanderbilt dataset. The composition of Figure 6.13 is similar to the one in Figure 7.8. Columns *b-d* show the segmented images with 2, 4, and 6 clusters, respectively. The results obtained with the one-sided algorithm applied on the CT and MR images are shown in Figure 6.13(i.b-d) and Figure 6.13(iv.b-d), respectively. The results obtained with the co-clustering algorithm are shown for the CT image in Figure 6.13(ii.b-d) and for the MR in Figure 6.13(iii.b-d).

If we compare the original unsegmented images with the resulting segmented images, we can see that the best results are obtained with the co-clustering algorithm (Figure 6.13(ii-iii.b-d)). There is clear evidence that hidden structures of the image are more precisely recovered. Compare, for instance, the images for an equal number of clusters of Figure 6.13(i.c) and Figure 6.13(ii.c). This better behavibour can be explained because in the co-clustering case we make use of all bidirectional information obtained with the progressive simplification of both images. For both algorithms, results appear much better when segmenting the CT images than the MR ones. This is due to the fact that the segmentation of the CT images benefits a



Figure 6.11: (a) Original images from the Brainweb database with 3% of noise. (b,c,d) Images segmented using 4, 5, and 6 bins, respectively. (i,iv) Images obtained with the one-sided algorithm. (ii,iii) Images obtained with the co-clustering algorithm.



Figure 6.12: (a) Original control image MR-T1 with 1% of noise. (b) Original image MR-T2 with 7% of noise. (c,d,e) Results of segmenting (b) using 4, 5, and 6 bins, respectively.

lot from the precise information contained in the MR histogram.

Figure 6.14(a) and Figure 6.14(c), corresponding to the dataset of Figure 6.13, plot the MIR vs the number of clusters for the one-sided and co-clustering algorithms, respectively. We can clearly observe the high quality of the resulting images with a low number of clusters. If the number of clusters decreases below a critical value, MI falls dramatically. On the contrary, to the left of this critical value, MI does not increase significantly with the number of clusters. This critical point can be detected by the stopping criterion given by the variation of MI (see Sec. 3).

In Figure 6.14(b) and Figure 6.14(d), the efficiency coefficient NMI against the number of clusters for the one-sided and co-clustering algorithms is plotted, respectively. Notice that the efficiency is maximum when the number of bins is low. Comparing both plots, we can see that, while the one-sided algorithm always increases monotonically, in the co-clustering there are fluctuations. This is due to the different decreasing rate of MI and joint-entropy for the co-clustering algorithm.

#### 6.6 Conclusions

We have presented a general framework for image segmentation based on a hard version of the information bottleneck method. Three different segmentation algorithms have been introduced: a split-and-merge, a histogram clustering and a registration-based clustering. For the two first algorithms, an information channel between the regions of the image and the histogram bins has been defined.



Figure 6.13: (a) Original dataset images. (b,c,d) Images segmented using 2, 4, and 6 bins, respectively. (i,iv) Images obtained with the one-sided algorithm. (ii,iii) Images obtained with the co-clustering algorithm.



Figure 6.14: (a) MIR and (b) NMI vs the number of clusters obtained with the one-sided algorithm applied on CT (solid line) and MR (dashed line) images of Figure 6.13. (c) MIR and (d) NMI vs the number of clusters obtained with the co-clustering algorithm applied on the same images.

Based on the preservation of mutual information, the spatial distribution and the histogram bins are maximally correlated. For the third algorithm, a channel between two multimodal images is defined, allowing to segment one image preserving the maximum information given by the other one. The main advantages of these methods are that do not assume any a priori information about the images (e.g. intensity probability distribution) and that take into account the spatial distribution of the samples. Different experiments on both natural and medical images have shown the good behavior of the proposed algorithms.

# **Chapter 7**

# **Image Segmentation using Excess Entropy**

In this chapter, we present a novel information-theoretic approach for thresholding-based segmentation that uses the excess entropy to measure the structural information of a 2D or 3D image and to locate the optimal thresholds. This approach is based on the conjecture that the optimal thresholding corresponds to the segmentation with maximum structure, i.e., in our case, maximum excess entropy. The contributions of this chapter are severalfold. First, we introduce the excess entropy as a measure of the spatial structure of an image. Second, we present an adaptive thresholding method based on the maximization of excess entropy. Third, we propose the use of uniformly distributed random lines to overcome the main drawbacks of the excess entropy computation. To show the good performance of the proposed segmentation approach different experiments on synthetic and real brain models are carried out.

## 7.1 Introduction

The main objective of image segmentation is to divide an image into regions that can be considered homogeneous with respect to a given criterion such as color or texture. Segmentation is an essential part of any image analysis system and especially in medical environments, where segmented images provide valuable information for diagnosis. Image thresholding, which segments an image by dividing its intensity histogram into a set of thresholds, is one of the most used segmentation techniques, because of its simplicity and efficiency [55, 137]. In the case of bimodal images, two main approaches of this technique can be considered: (i) the global thresholding that compares all the image pixels to a same value, which may be constant or be chosen from the image histogram [55], and (ii) the local thresholding that computes an independent threshold for each pixel over a local window whose center is the pixel being binarized. In this chapter we will focus on a global thresholding not restricted to bimodal images.

The excess entropy, first introduced by Crutchfield and Packard [41], is a measure of global correlation or structure for spatial systems in any dimension. It is commonly used and well understood in one dimension, but some difficulties are found in its extension to higher dimensions. In this chapter, we propose a thresholding technique that uses the excess entropy to find the optimal thresholds of a 2D or 3D image automatically. Our approach is based on the assumption that a medical image is structured in regions, such as tissues or organs, and the conjecture that the optimal thresholding corresponds to the segmentation with maximum structure, i.e., maximum excess entropy. Hence, we present a thresholding method that uses the maximization of excess entropy to determine the optimal thresholds. The computation of the excess entropy for a 3D dataset requires, for each voxel, an adequate selection of a sequence of neighbour voxels. Depending on how these neighbour voxels are selected, different problems such as high-dimensionality, sparsity, and non-invariance to rotation appear. To overcome these problems, excess entropy is computed using uniformly distributed random lines [9]. Experimental results analyze the behavibour of our approach for different image modalities.

This chapter is organized as follows. In Section 7.2 some background and related work are reviewed. In Section 7.3 the excess entropy is introduced as a measure of the spatial structure of an image. In Section 7.4, a new method is presented for thresholding segmentation using the excess entropy. In Section 7.5, different experiments on synthetic and real medical images are carried out. Finally, conclusions are summarized in Section 7.6.

# 7.2 Background and Related Work

We review some basic information-theoretic measures [39, 49] and the image segmentation based on thresholding.

#### 7.2.1 Information-Theoretic Measures

We review now the definitions of entropy rate and excess entropy. The notation used here is inspired by the work of Feldman and Crutchfield [49].

Given a chain  $\ldots X_{-2}X_{-1}X_0X_1X_2\ldots$  of random variables  $X_i$  taking values in  $\mathcal{X}$ , a block of L consecutive random variables is denoted by  $X^L = X_1 \ldots X_L$ . The probability that the particular L-block  $x^L$  occurs is denoted by the joint probability  $p(x^L) = p(x_1, x_2, \ldots, x_L)$ . The Shannon entropy of length-L sequences or L-block entropy is defined by

$$H(X^L) = -\sum_{x^L \in \mathcal{X}^L} p(x^L) \log p(x^L), \tag{7.1}$$

where the sum runs over all possible L-blocks. The entropy rate is defined by

$$h = \lim_{L \to \infty} \frac{H(X^L)}{L} = \lim_{L \to \infty} h(L),$$
(7.2)

where

$$h(L) = H(X^{L}) - H(X^{L-1})$$
(7.3)

is the entropy of a symbol conditioned on a block of L - 1 adjacent symbols. The two terms of Equation (7.2) are shown to be equivalent in [40, 49]. The entropy rate of a sequence quantifies the average amount of information per symbol x and the optimal achievement for any possible compression algorithm [39]. The entropy rate is always equal or lower than the Shannon entropy and is only equal when there is no correlation between consecutive symbols.

A complementary measure to the entropy rate is the *excess entropy*, which is a measure of the *structure* of a system. The *excess entropy* is defined by

$$E \equiv \sum_{L=1}^{\infty} (h(L) - h)$$
(7.4)

and captures how h(L) converges to its asymptotic value h. Figure 7.1(a) is a graphical representations of the excess entropy measure, which is represented by the shaded area, corresponding to the sum of differences between h(L) and the limit h.

If one inserts Equation (7.3) into Equation (7.4), the sum telescopes and one arrives at an alternate expression for the excess entropy [49]:

$$E = \lim_{L \to \infty} [H(X^L) - h \cdot L].$$
(7.5)

Hence, excess entropy is the y-intercept of the straight line to which  $H(X^L)$  asymptotes as indicated in Figure 7.1(b).

It is important to note that, when we take into account only a few number of symbols in the entropy computation, the system appears more random that it actually is. This excess randomness tells us how much additional information must be gained about the configurations in order to reveal the actual uncertainty h [50].

#### 7.2.2 Image Segmentation based on Thresholding

Thresholding is a widely used technique for image segmentation due to its simplicity and efficiency, especially in applications where speed is an important factor. Its goal is to find one or more thresholds from the image histogram automatically. These techniques are suitable in segmentation frameworks as a first step towards a final segmentation. To locate the thresholds, parametric and nonparametric approaches can be used. In the first case, a gray level distribution of a region is



Figure 7.1: Two different graphical representations of the excess entropy measure, corresponding to the Equations (7.4) and (7.5), respectively. Images obtained from [50].



Figure 7.2: (a) Global lines are cast from the walls of the bounding box, (b) intensity values are captured at evenly spaced positions over the global lines from an initial random offset and (c) neighbour intensity values are taken in *L*-blocks.

assumed, while, in the nonparametric approach, no assumptions are made. In this case, the thresholds are obtained in an optimal manner according to some criteria. Thresholding is best suited for bimodal distribution, such as solid objects resting upon a contrast background [137].

Information theory has been applied to define some of these criteria. A first method was proposed by Pun [118] and enhanced by Kapur et al. [74]. The histogram is separated into independent classes so that its entropy is maximized. Brink [18] extended this method to two dimensions by introducing spatial information. More recently, Rigau et al. [122] introduced an algorithm for medical image segmentation based on the mutual information optimization of the information channel between the histogram bins and the regions of the partitioned image.

# 7.3 Spatial structure of an image

In this section, the excess entropy is introduced as a measure of the spatial structure of a 2D or 3D image. Structure here is taken to be a statement which expresses the degree of correlation between the components of a system. Excess entropy, which provides us with a measure of the regularities present in an image, can also be interpreted as the the degree of predictability of a pixel (or voxel) given its neighbours. From the concepts introduced in Section 7.2.1, we analyze how the excess entropy can be computed from Equation (7.5).

In the context of an image,  $\mathcal{X}$  represents the set of clusters or bins of the image histogram and  $x^L$  is given by a set of *L neighbour* intensity values<sup>1</sup>. In order to compute the excess entropy, two main considerations have to be taken into account:

- The first is the definition of the neighbourhood concept pixel or voxel. While neighbourhood is unique and unambiguous in 1D, its extension to 2D or 3D introduces ambiguity, since a sequence of L-block neighbour pixels or voxels can be selected in different manners [49].
- The second is the computation of L-block entropies when L → ∞. In practice, L-block entropies for high L are not computable, since the number of elements of the joint histogram (required to compute joint probabilities p(x<sup>L</sup>)) is given by N<sup>L</sup>, where N is the cardinality of the system. Note that in our case, N is the number of clusters or bins of the segmented image histogram, i.e., the number of colors of the image. Thus, a tradeoff between the accuracy of the measure, given by L, and the number of clusters |X| is required.

To overcome the neighbourhood problem, uniformly distributed random lines, also called *global lines* [133] are used. Global lines sample the 3D-volume stochastically in the sense of integral geometry, i.e., invariant to translations and rotations [132]. These lines are generated from the walls of a convex bounding box containing the volume [26]. This can be done taking a random point on the surface of the convex bounding box and a cosinus distributed random direction as it is illustrated in Figure 7.2(a). The sequence of intensity values (L-block  $X^L$ ) needed to estimate the joint probabilities is captured at evenly spaced positions over the global lines from an initial random offset, that ranges from 0 to the step size (see Figure 7.2(b)). Points chosen on each line provide us with the intensities to calculate the *L*-block entropies, required to compute the excess entropy (see Figure 7.2(c)). In this manner, the 3D-neighbourhood problem is reduced to 1D, where the concept of neighbourhood is well defined. In our implementation, *N* is taken as an input parameter of the algorithm, while *L* is determined from *N* such that the computation of the joint histogram is attainable.

To illustrate the behavibour of the excess entropy as a measure of the image structure, we use the 2D images of Figure 7.3. The two first images (a) and (b) represent the same scene with the colors interchanged. In this case, the excess entropy values are the same since the structure of the image is not dependent on the colors. In the third image (c), some additional shapes are added to the original image (a), keeping the same probability for each color. Because of the higher variability of

<sup>&</sup>lt;sup>1</sup>We recall here two basic ideas already introduced in Chapter 3.



Figure 7.3: Synthetic images and their excess entropy values.

the obtained image, the excess entropy measure decreases, reflecting a lower spatial structure. The last image (d) has been generated by swapping 200000 points of image (a). Each swapping has been done by choosing two random points of the image and interchanging their intensity values. Observe that now the image has not spatial structure (no shape can be detected) and, therefore, the excess entropy is close to 0. It is important to remark that the values of the Shannon entropy of all the images of Figure 7.3 are the same, since the probabilities of each color have remained unaltered.

## 7.4 Thresholding based on Excess Entropy

In this section, a thresholding segmentation method that uses the excess entropy to obtain the optimal thresholds is presented.

We propose a nonparametric approach to segment the image without a priori assumptions about the underlying data. As it has been described in the previous section, excess entropy can be used as a measure of the structure of a system. From the assumption that an image is structured in regions, we conjecture that the optimal thresholding should provide us with the maximum structure. Consequently, the selection of thresholds will be formulated as a histogram quantization problem using the maximization of excess entropy. That is, the optimal histogram quantization should correspond to the maximum excess entropy of the resulting image. Figure 7.4 shows an example to illustrate the behavibour of our method compared to the k-means algorithm [47]. Figure 7.4(a) is a four-colored synthetic image



Figure 7.4: Preliminary experiment: (a) 4-colored image, (b) test image, (c) segmentation result of the k-means algorithm and (d) segmentation result of the proposed method.

with two nested squares on a background with two different colors randomly distributed. Figure 7.4(b) has been obtained by adding 1% of Gaussian noise to the original image. Figures 7.4(c) and 7.4(d) show the segmentation of the image (b) in three different clusters obtained by the k-means algorithm, one of the most popular clustering algorithms, and the excess entropy method applied to the noisy image, respectively. Observe that the k-means method separates the background in two colors and merges the two nested square shapes. This behavibour is due to the fact that the number of background pixels is greater than the number of pixels of the squares and this causes that the intensity error classification is lower if the background is split into two different clusters. Using our method, the spatial information is taken into account and, hence, the square shapes have been mainly preserved. Observe that the misclassified pixels are caused by the Gaussian noise of the input image (Figure 7.4(b)).

The kernel of our method is the computation of excess entropy given by the Equation (7.4), where N, the number of colors of the segmented image, and L, the block length, are input parameters of the segmentation algorithm. If C is the number of colors of the original image and N - 1 the number of thresholds, the quantity of possible combinations of different threshold levels is  $\frac{C!}{(C-(N-1))!}$ . Note that for standard medical images (with a minimum of 256 colors) this number becomes intractable when N is greater than 4 or 5. Therefore, except for the case of thresholding in two or three groups, it will be impossible to explore all the possible



Figure 7.5: Excess entropy values of the image 7.4(b) with 3 colors for different threshold levels.



Figure 7.6: Block diagram showing the excess entropy-based thresholding method.

solutions. For instance, in the segmentation process of Figure 7.4, all possible threshold levels have been checked. In the cases that the number of threshold levels is high, certain optimization criterion must be used.

In Figure 7.5, we plot the values of excess entropy values of the image 7.4(b) segmented with two different threshold values, ranging from 0 to 255. As it can be seen, the function has a smooth behavibour and therefore an optimization process can easily converge to the optimal solution. In our implementation, in order to properly locate the thresholds avoiding local minima, that can appear for a higher number of optimization parameters, a global genetic-based strategy is applied, since this is robust to these local minima. If high accuracy results are required, a second optimization process based on a gradient descent approach is applied.

Taking into account all these considerations we propose the iterative algorithm represented in Figure 7.6 and described below:



Figure 7.7: Histogram of the T1 image with 3% of noise.

- 1. *Input parameters:* the image, the number of colors (N), the block length (L) and the number of iterations (I).
- 2. *Threshold initialization*. This module computes the image histogram and distributes uniformly N 1 thresholds.
- 3. *Segmented image generation*. The input image is segmented using the last computed thresholds.
- 4. *Excess entropy computation*. This module computes the excess entropy E of the previous segmented image.
- 5. If E is maximum, the thresholds of the image are stored as the optimal thresholds.
- 6. If the number of iterations is not reached, we *recompute thresholds* by adding a noise term to the optimal thresholds. This noise is generated following a gaussian distribution. Then we go to step 3.
- 7. *Final segmentation*. The input image is segmented using the optimal thresholds.

At the end of the process, if high accurate threshold locations are needed, the gradient descent algorithm is applied.

# 7.5 Results and Discussion

The proposed segmentation approach has been developed using ITK [1] and VTK [2] libraries, and integrated in the medical image visualization and processing framework developed in our laboratory. It has been evaluated on different synthetic and real sequences of brain images.

An important limitation to be considered when segmenting brain images is the overlapping intensity values between different tissues. To illustrate such a limitation (see Figure 7.7), the histogram of a T1 image from Brainweb database [32]



Figure 7.8: T1 (first column) and T2 (third column) images of Brainweb database with different degrees of noise. In the second and fourth columns, the corresponding segmented images with N = 5.

is plotted. In this plot, each one of the ten tissues represented in the T1 image is shown with a different color. Note how different intensity values overlap and consequently thresholding methods are not capable to isolate one from the other. For instance, both cerebro spinal fluid and skin intensity values range from 20 to 80 and, hence, they are not distinguishable. Due to this limitation, the main purpose of segmentation techniques is not the separation of all real brain tissues but only the most important ones which in general correspond to background, cerebro spinal fluid, grey matter, white matter and skull. Therefore, in our experiments the number of clusters to be considered has been 4, 5 or 6 (i.e.  $4 \le N \le 6$ ). The L parameter has been set to 6 and the number of iterations to 200.

The first experiment has been designed to evaluate the performance of the proposed approach for images with different levels of noise. With this purpose a set of synthetic magnetic resonance (MR) images from the Brainweb database [32] is used. The method with N = 5 is applied to T1 and T2 image modalities with three different levels of noise 0%, 3% and 7%, respectively.

The obtained results are illustrated in Figure 7.8. In the first row, we show, from left to right, the T1 image without noise and the obtained segmentation, and the T2



Figure 7.9: Comparison of the segmented images of Figure 7.8 with the true image. T1 (first row) and T2 (second row) results with 0%, 3% and 7% of noise are plotted. Columns of each plot represent: (1) air, (2) cerebro spinal fluid, (3) grey matter,(4) white matter, (5) fatty tissue, (6) muscle and skin, (7) mostly skin, (8) skull, (9) glial tissue and (10) other tissue.

image without noise and its corresponding segmentation. In a similar way, the second and third rows represent the results obtained from the T1 and T2 images with 3% and 7% of noise, respectively. Observe that, in the case of T1 without noise, background, skull and cerebro spinal fluid (CSF) are correctly separated, while the classification gets worse when noise increases. Such a behavibour is typical of the thresholding approach. Similar results are obtained with noisy T2 images. In the case of T2 without noise, CSF and grey matter are grouped in the same cluster, while background is split into two clusters. Observe also the rectangular shape that surrounds the brain in the segmented image. This shape affects the excess entropy measure leading to an uncorrect segmentation. This undesirable effect disappears in the case of noisy images, achieving a correct segmentations. We want to remark that these patterns do not appear in real images and hence such a bad behavibour is not expected for real data.

In our second experiment, the segmented images of the first experiment are compared with the true standard that corresponds to a brain image segmented in ten tissues [32]. The goal of this comparison is to evaluate which tissues and how the tissues of the true images have been grouped into the five clusters used for the image segmentation. To illustrate the obtained results for each one of the test
images we have generated a bar plot (see Figure 7.9). Each bar represents one of the ten tissues of the true standard, from left to right, air, CSF, grey matter (GM), white matter (WM), fatty tissue, muscle and skin, mostly skin, skull, glial tissue and other tissue. To represent the five clusters used for segmentation five different colors have been considered. The bars have been filled with one or more colors according to the % of pixels of the corresponding tissue in the different clusters. For instance, in plot 7.9(a) corresponding to T1 image without noise, we can see from the first bar that air pixels have been grouped into two different clusters, 75% into cluster 1 and 25% into cluster 2.

In the first row of Figure 7.9, the T1 images with 0%, 3% and 7% of noise are plotted. Observe that in the case of T1 without noise (plot 7.9(a)) the four main tissues (CSF, GM, WM and skull) and air are well classified, in the sense that there is a good correspondence between the clusters of the segmented image and the main tissues of the true image. In particular, 90% of CSF pixels are of cluster 3 and 10% of cluster 2; 65% of GM pixels are of cluster 4, 25% of cluster 5 and 10% of cluster 1; and 75% of air is of cluster 1 and 25% of cluster 2. As it was expected the quality of the T1 segmentation decreases when the percentage of noisy increases. In spite of this, it can be seen that the method still groups the main tissues in the same clusters (see plots 7.9(b) and 7.9(c)).

In the second row of Figure 7.9, the T2 images with 0%, 3% and 7% of noise are plotted. In the case of T2 without noise (plot 7.9(d)), air is separated in two main clusters (corresponding to the two background regions represented in Figure 7.8 (T2, noise=0%), while CSF and grey matter are grouped in the same cluster. In the case of noisy T2 images (see plots 7.9(e) and 7.9(f)), it can be seen that the main tissues are well classified.

In the next experiment we evaluate the performance of the proposed approach for different values of N. For the test we use the T1 image with 3% of noise and N = 4..6. The results are listed in Table 7.1 where rows represent the ten tissues of the true standard and the columns the clusters generated for each one of the N values. The listed values represent the distribution of the tissue in each cluster. If we analyze the distribution of tissues for the different cases, it can be seen that white matter is perfectly segmented. Observe that the cluster containing white matter also contains other brain tissues. This is an expected result due to the overlapping intensity values (see Figure 7.7).

N=6	(9)			24.5	100	100	3.5			19.6	98.8
	(5)			60.1			35.7			50.6	1.2
	(4)		51.1	15.4			60.8	42.3		29.8	
	(3)	1.6	48.9					57.6	25.6		
	(2)	38.0						0.1	54.2		
	(1)	60.4							20.2		
N=5	(5)			24.5	100	100	3.5			19.0	98.8
	(4)			61.6			39.6			53.0	1.2
	(3)		72.0	13.9			56.9	60.5		27.4	
	(2)	39.6	28.0					39.5	79.8		
	(1)	60.4							20.2		
N=4	(4)			42.5	100	100	7.2			31.8	7.99
	(3)		5.8	56.9			84.4	1.4		66.8	0.3
	(2)	39.6	94.2	0.6			8.4	98.6	79.8	1.4	
	(1)	60.4							20.2		
Brain	Tissue	air	CSF	GM	WM	FT	MS	SKN	SKL	GT	other

Table 7.1: Distribution of brain tissues in the different clusters after applying our method with $N = 4$ , $N = 5$ and $N = 6$ . Rows
represent, from top to bottom: air, cerebro spinal fluid (CFS), grey matter (GM), white matter (WM), fatty tissue (FT), muscle and skin
(MS), mostly skin (SKN), skull (SKL), glial tissue (GT) and other tissues.

Finally, a set of real computed tomography and magnetic resonance brain images obtained from the Hospital Josep Trueta of Girona has been segmented. The method has been tested with N = 6 and L = 6. The obtained results are shown in Figure 7.11. In the case of MR images, represented in the first column, the main tissues have been correctly separated. When segmenting the CT image of the second column, the method not only has isolated the main tissues but it also has enhanced a hidden pathological region. Finally, in the third column, a CT image of a patient with an intracranial hemorrhage, where the image intensities have been appropriately rescaled to a better visualization, is presented. In the segmented image, the different image tissues, the lesion and the background are correctly separated, giving a good representation of the anatomical structures.

Finally, the method is tested on real medical images of different modalities and different parts of the body. First, we present some results on mammographic images in Figure 7.10 with N = 4 clusters plus the background. The segmentation of the breast and the quantification of dense tissue is a key aspect for the tumor risk diagnosis [105]. As it can be seen breast contour is detected and dense tissue is mainly segmented. This images can be used for classification purposes [150] or as a initial step for other more sophisticated segmentation algorithms. Other examples are shown in Figure 7.11. These images are a set of real computed tomography and magnetic resonance brain images obtained from the Hospital Josep Trueta of Girona. The method has been tested with N = 6 and L = 6. The obtained results are shown in the second row of Figure 7.11. In the case of MR images, represented in the first column, the main tissues have been correctly separated. When segmenting the CT image of the second column, the method not only has isolated the main tissues but it also has enhanced a hidden pathological region. Finally, in the third column, a CT image of a patient with an intracranial hemorrhage, where the image intensities have been appropriately rescaled to a better visualization, is presented. In the segmented image, the different image tissues, the lesion and the background are correctly separated, giving a good representation of the anatomical structures.

## 7.6 Conclusions

An information-theoretic measure, called excess entropy, has been introduced to quantify the image structure and to obtain a new adaptive thresholding method for image segmentation. The main novelties of this chapter are the use of excess entropy as a measure of structural information of an image, the search for optimal thresholds by maximizing the excess entropy, and the use of uniformly distributed random lines to compute this measure. Experimental results have shown the good behavibour of the presented approach.



Figure 7.10: Mammographic images (first row) with their segmentations (second row).



Figure 7.11: Original MR and CT brain images (first row) with their corresponding segmentations (second row).

## **Chapter 8**

## **Conclusions and Future Work**

Medical imaging is an important focus of research. The development of new techniques that assist and enhance visual image interpretation in a timely and accurate manner is fundamental in real clinical environments. The aim of this thesis is to contribute to the research of two of these techniques, registration and segmentation. In this chapter, the conclusions of this thesis and the directions for our future research are presented.

## 8.1 Conclusions

In this thesis, several information theoretic tools have been presented in order to improve the existing methods in the image processing area, in particular in the image registration and segmentation fields.

Next, the main contributions of this thesis are described:

• In Chapter 3, a new sampling strategy based on global lines has been presented. The global line generation method, which is based on integral geometry, ensures that all the voxels and all the scanning directions have the same probability to be taken into account. This strategy reduces the artifacts due to the grid regularities, achieving a more accurate and robust registration. The computation time can also be reduced without losing reliability.

A generalization of the standard information channel between the two images to be registered has been also proposed. While standard registration methods consider only individual voxels of both images, our method considers blocks of voxels, incorporating in this way spatial information of the intensity distribution. These blocks are obtained using global lines since in this way all the directions are taken into account. Results demonstrate that this approach is more robust and accurate than the one provided by the standard registration measures.

• In Chapter 4, a new conjecture is presented in order to solve the image registration problem. This conjecture consists in assuming that two images are correctly aligned when we can maximally compress one image given the information of the other. From this conjecture, we present *the similarity metric*, defined by Li et al., as a new image registration similarity measure. Since this measure is non computable, some approximations have to be done. Two different approaches of this measure are studied. The first uses the standard compressors to estimate the compressibility of the images. The second uses the entropy rate to estimate this compressibility. The results show that the second approach is more accurate, mainly due to the real compressor imperfections. These approaches are absolutely innovative in the image registration field.

- In Chapter 5, a study on image simplification strategies for a multiresolution image registration is presented. Two strategies are analyzed: BSP and Quadtree. The simplification criteria are based on the Information Bottleneck method. Two registration implementations of the approaches are studied. The first generates simplified images from the data structures, while the second deals directly with the regions of the data structure. These approaches are compared with the standard regular downsampling strategy. The results demonstrate that these data structures, and in particular the BSP, drastically reduce the artifacts, achieving a more robust registration.
- In Chapter 6, a new framework for image segmentation based on the Information Bottleneck method is presented. This framework consists of different applications of the method, by defining different information channels. The first approach cames from the definition of the information channel between the regions of the image and the histogram bins, and results as a split-andmerge algorithm. In the second approach, a clustering algorithm is defined from the reversion of the previous information channel, which considers the histogram bins and the regions obtained with the split-and-merge algorithm. Finally, in the third approach, a clustering algorithm for multimodal images is presented by considering the information channel within these multimodal images. The results show the good behaviour of these methods.
- In Chapter 7, a new measure to quantify the image structure is presented. This measure is based on the concept of Excess entropy which computes the mean predictability of a sample given all the previous samples. The implementation of this measure is based on the global lines method presented in Chapter 3. This measure is also proposed as a new image clustering criterion that takes into account the spatial distribution of the intensities along the image. A set of experiments shows the suitability of this method.

## 8.2 Future Work

The ideas presented in this thesis can be expanded in different directions. Some of these are:

- The increasing capabilities of graphic hardware encourage us to explore the implementation of the global line sampling strategy with the GPU. For this implementation, we would make use of further coherence provided by bundles of parallel lines, which would allow for a hardware implementation by using a z-buffer algorithm. Moreover, we would analyze the behaviour of the generalized normalized mutual information using three or four neighbours combined with a previous non-uniform quantization of the images and other strategies to deal with the high-dimensionality problem.
- The compression-based registration approach does not achieve as precise results as the classical approach, mainly due to the compression imperfections. The specific implementation of a compression algorithm for image registration could improve these results. Moreover, different scan spacefilling curves and the global lines method can be used in order to improve the capture of structural information in the Shannon version of the similarity metric.
- Since the tests with the BSP simplification have provided very good results, a next step in our research would consist in developing and evaluating a full multiresolution framework using the BSP subdivision. In addition, a more efficient implementation of the node-based approach would reduce the computational time.
- A general framework for the information bottleneck method applied to image segmentation have been presented. Further investigation on stopping criteria is needed to determine the optimal number of both regions and clusters. On the other hand, new segmentation channels could be tested within this framework, taking into account other kinds of information, such as color or gradient. We also plan to explore the application of these methods to image fusion and level-of-detail applications.
- A new excess entropy-based measure to compute the image structure has been presented. Further research would be done in the application of the excess entropy to other image processing areas, such us image restoration or image retrieval. Moreover, the feasibility of new information theoretic measures, such as the erasure entropy, would be studied in the context of image registration and segmentation.

## 8.3 Publications

Publications that support the contents of this thesis:

 "Normalized similarity measures for medical image registration", A. Bardera, M. Feixas and I. Boada. In Medical Imaging SPIE 2004, San Diego, USA. Vol. 5370, pp. 108-118. February 2004.

- "Medical Image Segmentation based on Mutual Information Maximization", J. Rigau, M. Feixas, M. Sbert, A. Bardera and I. Boada. In Proceedings of the 7th International Conference on Medical Image Computing and Computed Assisted Intervention MICCAI 2004, Rennes-Saint Malo, France. Vol. 1, pp. 135-142. September 2004.
- "Development of a new coregistration method to asses the volume of ischemic penumbra", S. Pedraza, I. Boada, A. Bardera, M. Feixas, J. Méndez, J. Astarloa and A. Dávalos. In XXIX Congress of the European Society of Neuroradiology, Aachen, Germany. September 2004.
- "Medical Image Registration Based On Random Line Sampling", A. Bardera, M. Feixas, I. Boada and M. Sbert. In Proceedings of the IEEE International Conference on Image Processing (ICIP'05), Genoa, Italy. Vol. 1, pp. 1157-1160, September 2005.
- "A Framework to Assist Acute Stroke Diagnosis", A. Bardera, M. Feixas, I. Boada, S.Pedraza and J.Rodríguez. In Proceedings of the Vision, Modeling and Visualization 2005 (VMV'05), Erlangen, Germany. Vol. 1, pp. 359-366, November 2005.
- "Compression-based Image Registration", A. Bardera, M. Feixas, I. Boada and M. Sbert. In Proceedings of the IEEE International Symposium on Information Theory 2006 (ISIT'06), Seattle, USA. pp. 436-440, July 2006.
- "High-Dimensional Normalized Mutual Information for Image Registration using Random Lines", A. Bardera, M. Feixas, I. Boada and M. Sbert. In Proceedings of the Third International Workshop on Biomedical Image Registration 2006 (WBIR'06), Utrecht, Netherlands. Lecture Notes in Computer Science (LNCS) series. Vol. 4057, pp. 264-271, July 2006.
- "Medical Image Registration Based on BSP and Quad-tree Partitioning", A. Bardera, M. Feixas and I. Boada. In Proceedings of the Third International Workshop on Biomedical Image Registration 2006 (WBIR'06), Utrecht, Netherlands. Lecture Notes in Computer Science (LNCS) series. Vol. 4057, pp. 1-8, July 2006.
- "Registration-Based Segmentation Using the Information Bottleneck Method", A. Bardera, M. Feixas, I. Boada, J.Rigau and M. Sbert. In Proceedings of the Third Iberian Conference on Pattern Recognition and Image Analysis 2007 (IbPRIA'07), Girona, Spain. Lecture Notes in Computer Science (LNCS) series. Vol. 4478, pp. 130-137, June 2007.
- "Image Segmentation Using Excess Entropy", A. Bardera, I. Boada, M. Feixas and M. Sbert. In Journal of Signal Processing Systems, Springer New York, published online, May 2008.

 "Breast density segmentation: a comparison of clustering and region based techniques", A. Torrent, A. Bardera, A. Oliver, J. Freixenet, I. Boada, M. Feixas, R. Martí, X. Lladó, J. Pont, E. Pérez, S. Pedraza, and J. Martí. In Proceedings of the International Workshop on Digital Mammography 2008 (IWDM'08), Tucson, USA. Lecture Notes in Computer Science (LNCS) series. Vol. 5116. July 2008.

Submitted publications that support the contents of this thesis:

- 1. "Automated Method for Brain Hematoma and Edema Quantification using CT", A. Bardera, I. Boada and M. Feixas. Submitted to Computerized Medical Imaging and Graphics.
- 2. "*Image registration by compression*", A. Bardera, M. Feixas, I. Boada and M. Sbert. Submitted to Information Sciences.
- "Image Segmentation using Information Bottleneck Method", A. Bardera, J. Rigau, I. Boada, M. Feixas and M. Sbert. Submitted to IEEE Transactions on Image Processing.

Additional publications non-related with this thesis:

- "Atlas-Based Segmentation of Pathological Brain MR Images", M. Bach Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J. Villemure and J. Thiran. In Proceedings of the IEEE International Conference on Image Processing (ICIP'03), Barcelona, Spain. Vol. 1, pp. 573-576. September 2003.
- "Atlas-based Segmentation of Pathological MR Brain Images using a Model of Lesion Growth", M. Bach Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J. Villemure and J. Thiran. In IEEE Transactions on Medical Imaging, Volume 23, No 10, pp. 1301-1314, October 2004.
- "A Monte Carlo-based Fiber Tracking Algorithm using Diffusion Tensor MRI", F.Prados, A. Bardera, M. Feixas, I. Boada and M. Sbert. In Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems 2006 (CBMS'06), Salt Lake City, USA. pp. 353-358. June 2006.

# Appendix: A Framework to Assist Radiologists in Diagnosis

In the last years, medical imaging has assumed a major role in clinical diagnosis. The development of software platforms and environments that can support medical image processing and visualization has become critical. In this context, development teams where engineers and practitioners combine their knowledge are very important. Motivated by a common interest, the medical imaging group of our university and medical researchers from the Hospital Josep Trueta of Girona created a working group. We aimed to develop a tool that integrates, in the same environment, basic and advanced image analysis techniques to assist radiologists in diagnosis. With this idea the Starviewer platform was created. Starviewer is a user-friendly diagnosis application with a modular design that supports the integration of new functionalities designed to diagnose any kind of pathology.

In this appendix two of the specialized modules of the platform which have been developed in the scope of this thesis are presented. The first one is StarStroke, a module developed to assist and support treatment and diagnosis of acute stroke patients and the second one is StarHematoma, a module that integrates an automated method for brain hematoma and edema segmentation, and volume measurement using computed tomography imaging. Moreover, this platform has been used for testing several methods presented in this thesis, but, since these methods are still in a research stage, they have not been integrated into a specific module for daily clinical practice.

## **Appendix A**

# **StarStroke: a tool to Assist Acute Stroke Diagnosis**

In this chapter, a framework developed to assist and support treatment and diagnosis of acute stroke patients is presented. We propose to substitute the current clinical study based on manual image editing and manual segmentation of diffusion-weighted imaging (DWI) and perfusion weighted imaging (PWI) by an integrated framework with registration, segmentation, manipulation and visualization strategies specifically designed for the DWI/PWI processing. We describe a robust and accurate technique to register DWI/PWI data which uses an acceleration strategy based on uniformly distributed random lines that reduces considerably the processing time. We also propose a segmentation method based on PWI-DWI fused data to detect stroke and hypoperfusion areas that reduces the manual intervention of the radiologist. With our approach, the clinical study of an acute stroke patient is reduced from 15-20 minutes to 5 minutes, including the manual editing.

## A.1 Introduction

Cerebral ischemia constitutes the third cause of death in developed countries. It is believed that early intervention is the key to a successful therapeutic outcome in strokes. Therefore, the ability to rapidly diagnose the status of brain perfusion and the extension of tissue that is at risk versus that which is already irreversibly damaged is critical.

In recent years, two new magnetic resonance imaging (MRI) modalities have been increasingly used to evaluate acute stroke patients. Diffusion-weighted imaging (DWI), in which the images reflect microscopic random motion of water molecules, and perfusion weighted imaging (PWI), in which hemodynamically weighted MR sequences are based on the passage of MR contrast through brain tissue. In the study of acute strokes, DWI abnormalities are markers of critical ischemia and typically evolve into infarction. In the first few hours after the onset of the stroke, PWI abnormalities are often larger than the DWI lesions. This pattern "*PWI deficit* > *DWI lesion*" is frequently associated with subsequent lesion growth into the PWI/DWI mismatch region, indicating that the mismatch may represent tissues at risk. The area with normal diffusion and altered perfusion (mismatch area) is denoted penumbra. The penumbra is the target of acute stroke therapy [4, 109, 43].

Currently, there are no computational tools to easily measure the penumbra. For each clinical case, the analysis of DWI-PWI datasets is based on manual image editing and segmentation techniques provided by available commercial medical visualization workstations. A complete study takes from 15 to 20 minutes of user interaction. The complete automatic processing of DWI-PWI is not possible due to the diversity of data. However, our purpose is to automate and speed up the volume measurement of both infarction and mismatch regions.

We have developed a tool specifically designed for the PWI-DWI data processing that integrates registration, segmentation, manipulation, and visualization techniques. With this framework, the study of a clinical case is split into three different steps. First, the DWI and PWI data are integrated into a common coordinate system by applying a registration technique. For this process, the similarity metrics proposed in [6] are used. These metrics have been implemented using uniformly distributed random lines (see Chapter 3). Second, DWI data are processed to identify and measure the infarct area. A semi-automate segmentation technique based on region growing is applied. Third, by using the information of both previous steps, PWI data are processed to identify the lesion. Due to the difficulty of processing PWI data, we have developed a two step segmentation process that starts by solving PWI ambiguities and then, supervised by the radiologist, applies a segmentation process, based on region growing and morphological operations, to identify and measure the injured area. With our framework, the overall exploration time for each clinical case is reduced from 15-20 minutes to 5 minutes, including the manual editing.

This chapter is organized as follows. Section A.2.1 describes how registration problems are solved within the proposed framework. Techniques that have been developed for DWI and PWI segmentation are presented in Section A.2.2. Volume assessment and visualization problems are described in Sections A.2.3 and A.3, respectively. The evaluation of the method is presented in Section A.4. Finally, some conclusions are given in the last section.

## A.2 Method

In Figure A.1 a block diagram of the proposed approach is shown. It is clear that, the method consists of four basic steps: PWI-DWI registration, DWI segmentation,



Figure A.1: Block diagram of the proposed method for acute stroke assessment.

PWI black points estimation and PWI segmentation. The next sections detail each of these steps.

#### A.2.1 PWI-DWI registration

DWI and PWI show different, complementary, and partially overlapping aspects of the pathology. The integration of these image modalities in a common model will facilitate the work of the radiologist in stroke diagnosis. Therefore, the first process to be considered is the registration of the PWI-DWI data.

To perform this registration process, we proposed a new similarity measure presented in [6], using a sampling strategy based on uniformly distributed random lines (see Chapter 3). This measure is based on the normalization of Jensen's difference applied to Tsallis-Havrda-Charvrat entropy and has a tuning parameter  $\alpha$  which depending on the kind of images to be registered is set to one value or another. The measure can be seen as an extension of the NMI measure since they coincide when  $\alpha = 1$ . Our previous experiments with various registration modalities showed that the new measure was more robust than the NMI for some modalities and a determined range of the entropy parameter. Also, a certain improvement on accuracy can be obtained for a different range of  $\alpha$ . For more details see [6]. It is important to remark that this was the first method implemented in our framework, but, a posteriori, other registration methods developed in this thesis have been incorporated in this framework.

To analyze the robustness and accuracy of our measures for the PWI-DWI



Figure A.2: (a) Percentage of improvement of our measure with respect to the MI over different  $\alpha$  values. (b) NMI values using global line method. (c) Cost of the global line method over the sampling points. Dashed and solid lines correspond to Collignon's and global line method, respectively.

registration, a prospective study of 16 patients from the Josep Trueta Hospital in Girona was carried out. The DWI sets have a resolution of  $256 \times 256$  voxels in the X-Y plane and the number of acquired slices along the Z axis is 20. The PWI sets have a resolution of  $128 \times 128$  voxels in the X-Y plane and the number of acquired slices along the Z axis is 12. After some experiments detailed below, we have set  $\alpha$  to 1.5 as a tradeoff between accuracy and robustness.

The robustness of these measures has been evaluated in terms of the partial image overlap. This has been done using the parameter AFA (Area of Function Attraction) introduced by Capek et al. [24]. This parameter evaluates the range of convergence of a registration measure to its global maximum, counting the number of pixels, i.e. x-y translations in image space, from which the global maximum is reached by applying a maximum gradient method. The AFA parameter represents the robustness with respect to the different initial positions of the images to be registered and with respect to the convergence to a local maximum of the similarity measure that leads to an incorrect registration. Our measure is compared with MI. The results of this study are summarized in Figure A.2(a) where we plot the percentage of improvement in AFA terms of our measure with respect to the MI over different  $\alpha$  values.

To evaluate the accuracy of the measure, the registration results were evaluated by an expert classifying them into five categories: perfect (if all the structures of the brain are well registered), very good (if misregistration is minimal), good (if there is a general misregistration, less than 3 mm), bad (if the misregistration is greater than 3 mm) and very bad (if there is no overlap with the images). Using our measure, the results obtained range from *perfect* and *very good* in 63% of the cases, while MI only obtains these results in 17% of the cases.

As the sampling strategy, we propose to use the global lines method, presented in Chapter 3. A comparison between this method and the standard one proposed by Collignon [35] is illustrated in Figures A.2(b) and A.2(c). In Figure A.2(b), the NMI value has been computed in the classical way, i.e., considering all the voxels, and with the global line method, moving the floating image through the X axis from -10 mm to 10 mm. From top to bottom, the plots represent the values of the similarity measure using 50, 75, 100, 125 and 150 thousand points, and the bold plot represents the measure computed in the classical way. Observe that the value of the measure decreases when the number of points increases and it converges to the value measured in the classical way. This behavibour is due to the histogram dispersion when the sampling points increase. For a high number of points, the measure tends to the classical method value. In Figure A.2(c), the costs of Collignon's method and the global line method are compared. Note that for our method the time increases linearly with the number of points. Therefore, using global lines, the time can be outstandingly reduced. For example, an acceptable estimation of the measure is obtained with 50000 points and a processing time almost five times lower.

#### A.2.2 Segmentation

Image segmentation is one of most important medical image processing problems. In the particular case of MRI, the objective is to differentiate tissue types for purposes of volume measurement and visualization [31]. Although MRI segmentation methods have been quite successful on normal tissues, the current methods of MRI segmentation are still in the developmental stages for pathological tissues with some success recorded for specific disease processes.

In this section we are going to present the methods implemented to differentiate the pathological tissues in DWI and PWI images. It has to be taken into account that currently, in most hospitals, PWI and DWI images are segmented using manual editing methods. This is a laborious task and the results are operator dependent. On the other hand, the fully automatic segmentation of DWI and PWI is not possible due to the diversity of data. Therefore, our purpose is to segment the images with minimum intervention of the operator which is quicker and more reproducible than the currently applied techniques.

#### A.2.2.1 DWI segmentation

DWI images have their foundation in the mobility of water molecules in the tissue. Such mobility generates different contrasts in the image. If water is freely diffusible, then the MRI signal is attenuated, and if the movement of the water molecules is impeded in some way, for example along nerve bundles, then the MRI signal is larger. Acute cerebral ischaemia causes cell damage which results in edema due to the accumulation of intracellular water. This causes a restriction in the movement of water molecules showing up the infarct as a region of increased signal intensity. The result is an image with three distinct regions: air, which has a signal intensity close to zero; healthy brain, which has an intermediate grey scale value; and infarcted tissue, which has the highest signal intensity and is typically more heterogeneous than the other two regions (see Figure A.3).



Figure A.3: The original DWI image before and after segmentation

We have implemented a thresholding region growing segmentation technique.

In most cases, the three different regions are perfectly distinguished. Threshold values have been obtained experimentally. A display showing the threshold contours superimposed on the DWI images, together with the intensity histogram are presented to the operator (see Figure A.5). These thresholds could be modified manually by defining new threshold levels on the histogram. To distinguish the background and CSF region, a thresholding technique is sufficient, but for the infarcted tissue the problem is more difficult. In some cases, the presence of high intensity regions in the corpus callosum (due to the presence of nerve tissue) and in the base of the brain (due to the susceptibility of artifacts), which would be incorrectly classified as infarct, forces us to implement a region growing algorithm based on these thresholds, reducing these kinds of errors. For other possible inaccuracies, some manual editing tools are also supplied to modify the segmented region. However, even in the case that the thresholds have to be manually modified, the DWI segmentation is faster than the current slice by slice manual editing methods.

#### A.2.2.2 PWI segmentation

PWI images are based on the passage of MR contrast through brain tissue. PWI readily provides relative regional cerebral blood volumes and bolus peak arrival times. The quality of PWI data is greatly influenced by the duration of the intravenous contrast bolus. Furthermore, as the dynamic contrast changes (imaged in PWI) occur over only 30 to 60 seconds and must be sampled with high temporal resolution, the number of slices that can be acquired is limited and generally lower than the number of DWI images. This may be a significant limitation when the location of the stroke is unknown. Since DWI and PWI show complementary aspects of the patient, to overcome all these limitations we propose a segmentation strategy that exploits the information of the segmented DWI.

The proposed segmentation process is composed of two steps. First, it starts with an automatic pre-processing phase that removes PWI ambiguities. These ambiguities are caused by black regions which may correspond either to background, cerebrospinal fluid, pathology, or noise (see Figure A.4). In order to label each



Figure A.4: The original PWI image before and after segmentation.

one of these regions properly, we use the information of the DWI-PWI fused data and assign an intensity value to the PWI ambiguous region. This assignment is performed as follows. Background and cerebrospinal fluid areas in DWI have low intensity values, therefore, once they have been identified in the DWI, we locate them in the PWI, assigning the minimum intensity value to their ambiguous regions. Lesion areas of the DWI, with high intensity values, correspond to pathological areas of PWI. Then, we assign the maximum intensity value to ambiguous pixels of these areas. Finally, the PWI ambiguities that are not classified in the DWI are considered noise. Their intensity values are obtained from interpolation of neighbour intensities. At the end of this process we obtain an estimated PWI image with no ambiguities.

In the second step, the new PWI image is segmented proceeding as in the DWI segmentation. A segmentation based on the thresholding region growing is applied. A first approximation of threshold values is estimated from the information of the DWI image taking into account the DWI stroke zone. Since images are noisy, morphological operators are applied to improve the final segmentation results.

The results of these processes are displayed showing the threshold contours superimposed on the PWI images, together with the intensity histogram. The thresholds could be modified manually by defining new threshold levels on the histogram. In some cases manual editing is also required (see Figure A.5).

#### A.2.3 Volume assessment

The area with normal diffusion and altered perfusion is the mismatch area, also denoted penumbra. The penumbra represents the tissue at risk surrounding the irreversibly damaged ischemic core. Typically, the ischemic penumbra is partially recruited into the ischemic core during the first hours after symptom onset. Since this process might be prevented pharmacologically it has become the key point in acute stroke treatment [4].

The mismatch volume assessment is obtained from the volume assessment of DWI stroke and PWI hypoperfusion regions. In our framework the computation of all these volumes is straightforward since the DWI and PWI segmentation strategies have delimited them. Therefore, for each image modality we have to sum the



Figure A.5: Captures of the DWI and PWI segmentation processes in the first and second row, respectively.

pathological area of each slice and multiply it by the slice thickness plus interslice gap. The penumbra is obtained by the difference between the DWI stroke region and the PWI hypoperfused one. The results of the three computed volumes are displayed to the user as it is shown in Figure A.6.

## A.3 Visualization

PWI and DWI are slices taken progressively through the head. In our framework, they are stored as volumes. Nevertheless, doctors are in general more confident with 2D visualizations rather than 3D. Their main interest is the delimitation of stroke and hypoperfusion areas as opposed to 3D visualizations. Therefore, the results obtained in the different processes that are carried out are shown as series of 2D images representing slices through the volume. The user can select the orientation of the slice and also the interslice step which is limited by the resolution of the model. This capability enables the user to cross the volume in any direction.

To reduce user interaction and automate the whole process as far as possible there is a default visualization that has all the rendering parameters fixed. Obviously, the user can change the parameters by selecting the appropriate icons or by interacting directly into the view.

The rendering of the fused DWI-PWI model is based on the superimposition of PWI-DWI images. By default DWI is maintained opaque (the alpha value representing opacity is set to 1) and the PWI corresponding alpha is set to a value lower than 1 (see Figure A.6).



Figure A.6: Visualizations obtained from the DWI-PWI registered model. On the right, the Volume Assessment window where the measures of stroke, hypoperfusion, and mismatch volumes are shown.

## A.4 Evaluation

To evaluate the proposed framework a prospective study of 16 patients (64 percent male) was carried out over a six months period. All patients had a MRI study on admission within the first 12 hours after the onset of the stroke. For the purpose of this investigation, only patients with an acute MCA infarction were included. Clinical exclusion criteria were coma on admission, pure lacunar syndromes, transient ischemic attack, previous cerebral infarction impeding the clinical, and neuroradiological evaluation. The ethics committee approved the study and written informed consent was obtained from all the patients or relatives before imaging. The proposed framework has been implemented using the National Library of Medicine Insight Segmentation and Registration Toolkit (ITK) [1], which also guarantees the code quality for the library methods.

All the patients underwent MRI on a 1.5 T (Tesla) system (Philips Intera) with echoplanar capabilities of 25 mT/m gradients and 300 to 350  $\mu s$  rise times. The MRI protocol includes DWI, PWI, FLAIR, and MRA. The diffusion sequence was obtained with a single-shot spin-echo-planar pulse with a diffusion gradient b value of 0 and 1000  $s/mm^2$  along 3 axes. The other parameters were: 20 slices, 7 mmslice thickness, 0 gap, 134 ms (TE), 6000 ms (TR), 67 (epi factor) and 36 seconds of duration. Abnormalities were analysed in the trace image to avoid anisotropy. The PWI sequence was acquired after administration of 0.2 ml/kg of Gadolinium DTPA (Gd-DTPA) with a bolus technique (5 ml/s). The parameters were 260, 30 [TR/TE], 12 slices of 10 mm slice thickness, 0 gap, 60 dynamic scans and 1 minute 4 seconds of duration. The volume of hypoperfused tissue was measured in the MTT map using a threshold of 4 seconds.

The ground truth for the volumetric analyses was performed with a manual segmentation method, also called planimetric method. First, the perimeter of the area of abnormal signal intensity was traced on each DWI or PWI map and, sub-sequently, the volumetric software estimated the total volume using the thickness and the traced area on each slice. Each volume calculation was done three times,

	DWI	PWI	Mismatch	
Automotic Volume(AM)	20.01	129.32	90.62	
Automatic Volume(AWI)	(0.57,45.35) (92.51,191.23)		(57.48,109.25)	
Planimatria Mathad (PM)	13.23	132.21	86	
Fiammetric Method (FMI)	(8.29,44.73)	(67.44,182.83)	(41.4,144.73)	
Difforman (AM DM)	1.09	-11.5	-10.72	
Difference (Alvi-1 WI)	(-9.64,10.17)	(-41.55,29.97)	(-41.69,44.52)	
<b>Bolation</b> (AM/DM)	1.04	0.9	0.82	
Kelation (AW/FWI)	(0.62,1.84)	(0.72,1.12)	(0.6,1.83)	

Appendix A. StarStroke: a tool to Assist Acute Stroke Diagnosis

Table A.1: All the variables are presented with the values of median, percentil 25 and percentil 75.

and the mean value was taken as definitive.

Table A.1 shows the comparison between the volume assessment of the DWI, PWI, and mismatch area with the planimetric method (PM) and our proposed automatic method (AM). All the variables are presented with the values of median, percentil 25, and percentil 75 in  $cm^3$ .

The preliminary results show a good correlation between the measurement of the new method with respect to the PM gold standard method with ratios of 1.04-0.82. AM produces a mild overestimation of the volume of the ischemic lesion in the DWI sequence and a mild underestimation of the volume of the perfusion alteration in the PWI sequence.

### A.5 Conclusions

A framework to assist and support treatment and diagnosis of acute stroke patients has been presented. The framework integrates registration, segmentation, manipulation and visualization strategies specifically designed for the DWI-PWI processing. Empirical results demonstrate that the proposed strategy is a reliable technique and a promising approach in the diagnosis and management of acute stroke. Moreover, the clinical study time of an acute stroke patient is reduced considerably.

## **Appendix B**

# A New Method for Hematoma and Edema Quantification using CT

In this chapter, an automated method for brain hematoma and edema segmentation, and volume measurement using computed tomography imaging is presented. The method combines a region growing approach to segment the hematoma and a level set segmentation technique to segment the edema. The main novelty of this method is the strategy applied to define the propagation function required by the level set approach. To evaluate the method, 18 patients with brain hematoma and edema of different size, shape, and location were selected. The obtained results demonstrate that the proposed approach provides objective and reproducible segmentations that are similar to the results obtained manually. Moreover, the processing time is 4 minutes compared to the 10 minutes required for manual segmentation.

## **B.1** Introduction

In the last twenty years, medical imaging has assumed a major role in the diagnosis of head-injury patients. To help diagnosis, different visualization and image processing techniques have been developed. However, since the requirements of these techniques vary according to the image modality and the pathology to diagnose, the development of specialized methods is required.

In this chapter, we focus our interest on brain hematoma and edema diagnosis. Hematoma is a localized collection of blood, usually clotted, in a tissue or organ. It is caused by a break in the wall of a blood vessel and when it is in or near the brain it is particularly dangerous. In particular, we deal with the hematomas caused by an intracerebral hemorrhage (ICH). Cerebral edema is an excess of water accumulation in the intra- and/or extracellular spaces of the brain. Vasogenic edema is the most common type of edema associated with ICH. A hematoma always has a vasogenic edema surrounding it. The location and the volume of both hematoma and edema are the key prognostic factors for the treatment and the course of the patient's recovery [59, 143].

In daily clinical practice, computed tomography (CT) images are acquired from the patients to diagnose different pathologies. Although CT images accurately describe the size and the location of the hematoma, the edges of the edema are usually not well defined and the range of intensities of the healthy tissue and the edema are not easily distinguishable. This makes the segmentation process required for its quantification a difficult task. For a correct diagnosis we propose a two step automated method to segment and quantify the hematoma and edema volumes. The method combines the connected threshold region growing and the level set segmentation techniques. In the first step the hematoma is segmented and then, using the information from this segmentation, the level set method is applied to isolate the edema.

To analyze the performance of the proposed approach, 3D CT images of 18 patients with brain hematomas of different size, shape and location have been used. Our method has been compared with a manual segmentation of these images carried out by trained personnel. The obtained results show that the automated method has a higher reproducibility than the manual method, obtaining a similar accuracy. In terms of speed, while the operator time of the manual segmentation is approximately 10 minutes, our technique needs an average of 4 minutes.

This chapter is organized as follows. In Section B.2, we briefly describe the related work. In Section B.3, the proposed approach is introduced. In Section B.4, the segmentation results obtained with our method and manual segmentation are compared. Finally, conclusions are presented in Section B.5.

### **B.2 Related Work**

Image segmentation consists in subdividing the image in its constituent parts, a significant step towards image understanding. It is applied in many different areas, such as medicine, robotics, visualization, etc., and many different segmentation approaches have been proposed [55, 112, 53, 44].

Focusing on brain hematoma and edema segmentation, the majority of related bibliography is referred to Magnetic Resonance Imaging (MRI), mainly for its superior sensitivity for detecting brain tissue anomalies. In particular, Ho et al. [65] proposed a level set technique to segment the edema region of a tumor brain properly. Recently, Corso et al. [37] introduced a new method based on bayesian classifiers to segment the brain edema using multimodal images. However, despite the good performance of MRI, CT is the gold standard for ICH imaging because of its sensitivity in detecting this pathology and its rapid scanning capability [36]. In



Figure B.1: Block diagram of the segmentation process.

this context, MRI has three main limitations. First, the acute hemorrhage may be isointense and difficult to see on the MR images. Second, the patients affected by ICH require various support equipment, many of which cannot be taken into the magnet room. Third, MRI requires more time, which is a critical factor in patients with ICH. In CT brain image segmentation, the main contributions were proposed by Loncaric et al. First, they proposed a non-fuzzy method to segment the edema [38] and, then, they presented the fuzzy version of the previous approach [89]. They also introduced a system for 3-D quantitative analysis of human spontaneous ICH [88] with the aim of performing quantitative 3-D measurements of the parameters of the ICH region to correlate them to patient morbidity and mortality.

## **B.3** Segmentation Framework

The framework proposed to segment both the hematoma and the edema from CT images is divided into two stages: the hematoma segmentation step, based on a region growing approach, and the edema segmentation step, based on a level set strategy. The block diagram of the framework is shown in Figure B.1 and described in detail below.

#### **B.3.1** Hematoma segmentation

Despite the wide range of sizes, shapes and locations that a hematoma may have, in a CT image it is readily visible since the contrast between health tissue and the



Figure B.2: CT scans with a hematoma and the surrounding edema at different brain locations.

hematoma region is very high (see Figure B.2). However, the fully automatic computerized scheme for segmenting brain hematoma is difficult due to the overlapping intensities of the damaged region with other anatomical structures such as the skull or the dura mater. To overcome all these limitations, the method described below, which requires minimal manual intervention, is proposed.

As illustrated in the block diagram of Figure B.1, to segment the hematoma we use a region growing strategy. The basic approach of a region growing algorithm is to start from a seed region, taken inside the object to be segmented, and then to evaluate, using a homogeneity criterion, whether the voxels neighbouring this region are part of the object. In our case we use the connected threshold criterion, that includes, in the segmented region, all the voxels whose intensities are inside a predefined interval. The proposed segmentation process has three main stages:

- 1. *Initialization*. To start the process, the seed and the threshold values have to be given. The seed point is fixed by the user, picking a point on a 2D view of the 3D volume generated from the CT images. The threshold interval values have a default value obtained experimentally by considering the Hounsfield values of the damaged regions of different patients. In Figure B.3(a) we can see an example of a real case, and in Figure B.3(b) the result at the end of this initialization step. The threshold values can be modified if necessary.
- 2. Connected Threshold Segmentation. Once all required values have been initialized, the connected threshold procedure starts and creates a mask containing all the hematoma voxels. A display showing the generated mask superimposed on the CT images is presented to the radiological expert (see Figure B.3(c)). In the case that the obtained segmentation is not the expected one, the user performs a tuning process.
- 3. Tuning. This step includes all the operations performed by the user to adjust



Figure B.3: (a) Original image, (b) parameter setting, and (c) connected threshold segmentation.

the final result. To carry out the tuning, two different sets of tools are provided by the user. The first is designed to modify the predefined threshold values used in the first step of the process. The second is a set of morphological operations such as erosion and dilation designed to solve erroneous segmentation results caused by the overlap between the intensity values of the hematoma and the skull. This tool is used only when the hematoma is located very close to the skull. This problem can also be solved using a skull pre-segmentation process [89], however such a solution is a more time consuming task and for this reason it has been discarded. All the operations required to tune the results have been implemented with ITK [1] and VTK [2] functionalities.

At the end of this process, the boundary of the hematoma is correctly defined. Then, the total volume is obtained by counting the number of voxels inside the hematoma and multiplying it by the voxel volume. This is calculated by multiplying the area of a slice pixel and the CT slice thickness plus the interslice gap.

#### **B.3.2** Edema segmentation

The most challenging part of the proposed approach is the second phase, where the edema segmentation is carried out (see Figure B.1). The edema appears as a region surrounding the hematoma with subtle intensity variations between the edema and the health brain tissue (see Figure B.2). The borders of the edema are not well defined, being a limiting factor for the application of gradient-based segmentation techniques. Moreover, the application of simple intensity thresholding techniques is not possible due to the similarity of the intensity values of the damaged and healthy regions, and the CT image noise. To overcome all these limitations we propose a segmentation strategy based on the level set approach [136] which is effective in dealing with both small-scale noise and smoother intensity fluctuations

in volume data.

Level set segmentation relies on a surface-fitting strategy. The object to be segmented is represented by an evolving closed 3D surface embedded as the zero level of a 4-dimensional scalar function called the level set function,  $\Psi(X,t)$ . This function is evolved under the control of a partial differential equation governed by image-based features, such as mean intensity, gradient and edges, or by userdefined terms. The function is iteratively deformed and the position of the evolving contour can be deduced at each iteration by the relation  $\Psi(X(t), t) = 0$ . The main advantage of using level sets is that arbitrarily complex shapes can be modelled and topological changes such as merging and splitting are handled implicitly. Many different implementations and variants of this basic concept have been published in the literature [136, 147, 95, 163, 71].

To design the edema segmentation method we take into account the prior knowledge that the edema is always contiguous with the hematoma and it has low intensity values with respect to the rest of the voxels. Moreover, since the intensity values and the regional extension of the edema are variable, we also compute some image statistics to reduce the parameter variability between different cases.

Our implementation is also based on ITK libraries and makes use of a generic level set equation to compute the update to the solution  $\Psi$  of the partial differential equation. Practically, it is computed as

$$\frac{d}{dt}\Psi = -\beta P(x) \mid \nabla \Psi \mid +\gamma Z(x)\kappa \mid \nabla \Psi \mid,$$

where P is a propagation term, and Z is a spatial modifier term for the mean curvature  $\kappa$ . In other words, P represents the inflation and deflation of the evolving curve, while Z represents the smoothness of the curve. The scalar constants  $\beta$  and  $\gamma$  weight the relative influence of each of the terms on the movement of the interface.

The definition of  $P(\mathbf{x})$  is the cornerstone of the process. Generally, P is defined as a function depending on input gradient magnitudes, but in our case such an approach is not suitable due to the low contrast between the healthy and pathologic brain. Hence, we propose a  $P(\mathbf{x})$  intensity-based definition. At the beginning, we classify the voxels of an input model (see Figure B.4(a)) in three different groups:

- 1. First, we consider background, cerebro spinal fluid (CSF) and skull voxels which can be obtained by a thresholding strategy (see Figure B.4(b)).
- 2. Second, we take the hematoma voxels detected in the hematoma segmentation phase (see Figure B.4(c)).
- 3. Third, we consider the rest of the voxels, i.e. soft tissue with no hematoma voxels, including the edema voxels.

P has to be defined for each one of the groups. For the first and second groups, constant values -1 and 1 are assigned, respectively. In the last group, the intensity



Figure B.4: (a) Input CT scan, (b) voxels of the first group obtained from background, CSF and skull segmentation, (c) voxels of the second group obtained from the hematoma segmentation, (d) propagation function.

distribution of their voxels are analyzed. We assume that they follow a Gaussian distribution and we characterize them in terms of mean and standard deviation values (Figure B.5(a)). Since the edema has low intensity values, P has to assign a fast propagation when these values are encountered. This behavibour can be modeled by the *erfc* function, which is complementary to the error function *erf*. The *erf* function is encountered in integrating the normalized gaussian function. Then, we can define the *erfc* function as

$$erfc(x,\mu,\sigma) \equiv 1 - erf(\frac{x-\mu}{\sigma\sqrt{2}}) = \frac{2}{\sigma\sqrt{2\pi}} \int_x^\infty e^{\frac{-(t-\mu)^2}{2\sigma^2}} dt$$

where  $\mu$  is the mean and  $\sigma$  the variance of the integrated gaussian function. A plot of the function is presented in Figure B.5(b). Using this function, we define the propagation term for the voxels of the third group as

$$P(x, y, z) = \operatorname{erfc}(I(x, y, z), \mu - k \cdot \sigma, \sigma) - 1,$$

where I(x, y, z) is the intensity at voxel (x, y, z) and k is a user-defined parameter that sets the mean value (Figure B.5(c)). The darker the edema, the greater the k parameter (empirically k takes values close to 0.3). This parameter modifies the propagation term as a function of the image statistics and not directly from its

Appendix B. A New Method for Hematoma and Edema Quantification using CT



Figure B.5: (a) Gaussian distribution, (b) erfc function and (c) the proposed propagation term.



Figure B.6: Contours for different k values (0.1 (yellow), 0.2 (red), 0.3 (green), 0.4 (blue)).

intensity or gradient values. In this way, the definition of this parameter becomes less dependent on the image features and remains more stable for different patients. In Figure B.6, the obtained contours for different k values, ranging from 0.1 to 0.4, are illustrated.

Taking into account all these considerations, the resulting P(x, y, z) is defined as follows

$$P(x,y,z) = \left\{ \begin{array}{ll} -1 & (x,y,z) \in \operatorname{group} 1 \\ 1 & (x,y,z) \in \operatorname{group} 2 \\ \operatorname{erfc}(I(x,y,z), \mu - k \cdot \sigma, \sigma) - 1 & (x,y,z) \in \operatorname{group} 3 \end{array} \right.$$

In Figure B.4(d), the propagation map obtained from the function introduced to detect the edema of Figure B.4(a) is illustrated. Finally, the edema volume can be computed using the same strategy as the one used with the hematoma volume.

## **B.4** Results and Discussion

The proposed segmentation approach has been developed using ITK and VTK libraries, and has been integrated in the Starviewer platform, a medical image visualization and processing framework developed in our laboratory. To evaluate the method, 3D CT images were taken within 72 hours from 18 patients who had had a brain hemorrhage. The CT scans were acquired on a Philips mx8000 IDT with 10 detectors using a  $512 \times 512$  matrix, a 250-mm field of view and a slice thickness of 3mm with a 3mm interslice gap. A total of 45 sections per patient were acquired. For each patient the damaged area has been measured using a conventional planimetric method and the proposed one. The experiments were performed by two radiological experts (operators 1 and 2, respectively) to obtain a more objective evaluation of the method accuracy and reproducibility.

The key aspect to determine patient treatment is the knowledge of the hematoma and edema volumes [59, 143]. For this reason, we compared the volume measurements performed by our experts using the manual tracing method and the proposed automated method. For the manual method, the experts trace independently both the hematoma and the edema outline on each of the CT slices. The obtained results are illustrated in Figure B.7. In Figure B.7(a) the volumes obtained from the edema manual segmentation of both operators are compared. Axis X and Y represent the volume obtained by operator 1 and operator 2, respectively. In this plot, the ideal line (y = x) is also represented and, as can be seen, the values remain close to it. The Pearson linear correlation coefficient (0.994) and the P-value (0.000) have been computed with the Minitab software [3]. Similarly, in Figure B.7(b) the results of the automated edema segmentation are plotted on axis Y and compared with the manual segmentation volumes on axis X, plotting the results of operator 1 with a circle and the results of operator 2 with a cross. We have also computed the Pearson linear correlation coefficient for operators 1 and 2 (0.980 and 0.970, respectively) and the P-values (0.000 in both cases).

In order to measure the interoperator reliability of both methods, we define the interoperator variance with manual tracing  $(I_M)$  and automated  $(I_A)$  segmentation algorithms as

$$I_M = 1 - \frac{A_{M1} \bigcap A_{M2}}{A_{M1} \bigcup A_{M2}}$$

and

$$I_A = 1 - \frac{A_{A1} \bigcap A_{A2}}{A_{A1} \bigcup A_{A2}},$$

where,  $A_{M1}$  and  $A_{M2}$  are the manually segmented regions obtained by operator 1 and operator 2, respectively, on the same image and, in the same way,  $A_{A1}$  and  $A_{A2}$  are the regions obtained with the automated method. A value of 0 shows perfect reliability, while a value of 1 shows no reliability. The obtained results are reported in Table B.1. The inter-operator consistency with the automated approach



Figure B.7: Hematoma plus edema volume comparison. (a) Manual interobserver comparison and (b) automated-manual comparison.

	$I_M$			$I_A$		
	hematoma	edema		hematoma	edema	
Mean	0.31	0.36	Mean	0.17	0.42	
Std. dev.	0.13	0.10	Std. dev.	0.13	0.15	
Maximum	0.56	0.58	Maximum	0.52	0.67	
Minimum	0.15	0.21	Minimum	0.02	0.16	

Table B.1: Interoperator variance with manual tracing  $(I_M)$  and automated  $(I_A)$  segmentation algorithms.

is much better than the manual case for the hematoma. Note that for the hematoma segmentation the mean inter-operator variation of I with the automated method is 0.17 while with the manual method the achieved value is 0.31. In the measurements of edema, there is no significant difference in using the automated method with respect to the manual segmentation. This is due to the badly defined edema borders which are difficult to detect even for an expert operator.

The first study takes into account the volume of the segmented region, but not the matching of the involved regions. In the next experiment, we want to analyze how the segmented regions overlap. Due to the lack of a realistic phantom to be used as the ground truth (GT), we defined this from the manual segmentations performed by our experts on each volume data set. For that reason, we classify the voxels in four different groups: *true positives* (TP), hematoma voxels found by both automated and manual methods, *false positives* (FP), hematoma voxels isolated by the automated method but not by the manual one, *false negatives* (FN), hematoma voxels isolated by the manual segmentation but not by automated, and *true negatives* (TN), no hematoma voxels found by both methods. The same classification

	hematoma		edema			hematoma		edema	
Op1	CR	MR	CR	MR	Op2	CR	MR	CR	MR
Mean	0.69	0.76	0.56	0.65	Mean	0.74	0.80	0.62	0.72
Std. dev.	0.12	0.12	0.15	0.15	Std. dev.	0.10	0.08	0.11	0.14
Max.	0.87	0.96	0.74	0.81	Max.	0.86	0.90	0.77	0.96
Min.	0.41	0.56	0.21	0.29	Min.	0.48	0.54	0.42	0.50

Table B.2: Matching ratio and correspondence ratio comparing manual tracing and automated segmentation of brain hematoma and edema for each operator.

has been used for the edema segmentation.

Two measures are used for the evaluation of the results, the matching ratio (MR) and the correspondence ratio (CR), defined, respectively, by

$$MR = \frac{n_{TP}}{n_{GT}}$$

and

$$CR = \frac{n_{TP} - 0.5n_{FP}}{n_{GT}}$$

where  $n_{TP}$ ,  $n_{FP}$ , and  $n_{GT}$  represent the number of voxels of the regions TP, FP, GT, respectively.

MR is calculated as the direct ratio of the number of TP voxels to the number of GT voxels. The ideal value of MR is 1.0 while a value 0.0 indicates that there is a complete absence of any GT voxel. The CR allows us to discuss the way in which the isolated region corresponds in size and location to the GT while weighting the importance of FP and FN. When the value of CR is 1.0, the system achieves a perfect matching without any false positives. Note that CR can take negative values.

In Table B.2, the MR and CR obtained from the automated segmentation are given by considering the manual tracing of each operator as the GT. From this table, we can observe that the hematoma segmentation results achieve very good results for both operators. In the edema case, the results are not so good, due to its intrinsic difficulty in differentiating its intensity values. The achieved results are comparable to those in the studies of Fletcher et al. [52] and Chong et al. [27].

Finally, the CR and MR values for each patient with the corresponding manual segmentation are compared. The obtained results are plotted in Figure B.8, where axis X represents the manual volume and axis Y the CR parameter in the Figure B.8(a) and the MR parameter in Figure B.8(b), respectively. In both figures, the results of operator 1 are represented with a circle and the results of operator 2 with a cross. Note that the worst results are achieved for the smallest lesions. Considering the relative nature of these measures, the obtained behavibour is the desired one, since it leads to a low absolute error for the worst cases.



Figure B.8: (a) CR ratio and (b) MR ratio versus the edema plus hematoma volume.

We have computed the time required to carry out a complete study using both manual and automated methods. Manual segmentation requires an operator time average of 10 minutes while the proposed approach requires approximately 4 minutes, from which 100 seconds are spent to the slice by slice verification process.

The obtained results demonstrate that the proposed approach provides objective and reproducible segmentations that are close to the manual results. Moreover, the processing time required to evaluate a patient is reduced considerably.

### **B.5** Conclusions

The measurement of hematoma and edema volume from brain CT scans is a key prognostic factor in its treatment. In clinical practice this task is usually done manually which is very time consuming and operator dependent. We have presented an automated segmentation approach to measure the volume of these regions using a CT image. We have proposed a new level set approach which considers some image statistics in order to set the parameters. The method has been tested on a sample set of 18 patients and compared with the manual segmentation process. The obtained results are very promising, demonstrating its reliability and accuracy. The suitability of the level set approach for the edema segmentation on the CT brain scans has also been proved. The proposed approach reduces the operator processing time for each study from 10 minutes to 4 minutes.

# **Bibliography**

- [1] ITK Insight Toolkit, http://www.itk.org.
- [2] VTK Visualization Toolkit, http://www.vtk.org.
- [3] Minitab Inc., http://www.minitab.com.
- [4] P. Barber, D. Darbyand, P. Desmond, Q. Yang, R. Gerraty, D. Jolley, G. Donnan, B. Tress, and S. Davis. Prediction of stroke outcome with echoplanar perfusion- and diffusion-weighted MRI. *Neurology*, 51(1):418–426, 1998.
- [5] A. Bardera, I. Boada, M. Feixas, S. Pedraza, and J. Rodríguez. A framework to assist acute stroke diagnosis. In *Proceedings of the Vision, Modeling, and Visualization*, volume 1, pages 359–366, Erlangen, Germany, November 2005.
- [6] A. Bardera, M. Feixas, and I. Boada. Normalized similarity measures for medical image registration. In *Proceedings of Medical Imaging SPIE 2004*, volume 5370, pages 108–118, San Diego, USA, February 2004.
- [7] A. Bardera, M. Feixas, I. Boada, J. Rigau, and M. Sbert. Medical image registration based on bsp and quadtree partitioning. In *Proceedings of the Third International Workshop on Biomedical Image Registration (WBIR 2006)*, volume 4057 of *Lecture Notes in Computer Science*, pages 1–8, Utrecht, Netherlands, July 2006.
- [8] A. Bardera, M. Feixas, I. Boada, J. Rigau, and M. Sbert. Registration-based segmentation using the information bottleneck method. In *Iberian Conference on Patern Recognition and Image Analisys (IbPRIA 2007), Proceedings*, volume 4478 of *Lecture Notes in Computer Science*, pages 190–197, Girona, Spain, June 2007.
- [9] A. Bardera, M. Feixas, I. Boada, and M. Sbert. Medical image registration based on random line sampling. In *IEEE International Conference on Image Processing (ICIP'05), Proceedings*, pages 1157–1160, Genova, Italy, September 2005.
- [10] A. Bardera, M. Feixas, I. Boada, and M. Sbert. Compression-based image registration. In *IEEE International Symposium on Information Theory*, pages 436–440, Seattle, USA, July 2006.
- [11] A. Bardera, M. Feixas, I. Boada, and M. Sbert. High-dimensional normalized mutual information for image registration using random lines. In *Proceedings of the Third International Workshop on Biomedical Image Registration (WBIR 2006)*, volume 4057 of *Lecture Notes in Computer Science*, pages 264–271, Utrecht, Netherlands, July 2006.
- [12] A. Bardera, M. Feixas, I. Boada, and M. Sbert. Image segmentation using excess entropy. *Journal of Signal Processing Systems*, May 2008. Published online.
- [13] C. Bennett, P. Gács, M. Li, P. Vitányi, and W. Zurek. Information distance. *IEEE Transactions on Information Theory*, 44(4):1407–1423, July 1998.
- [14] P. Bernaola, J. L. Oliver, and R. Román. Decomposition of DNA sequence complexity. *Physical Review Letters*, 83(16):3336–3339, October 1999.
- [15] J. Beutel, H. L. Kundel, and R. L. V. Metter, editors. Handbook of Medical Imaging. Volume 1: Physics and Psychophysics. SPIE Press, 2000.
- [16] R. E. Blahut. Principles and Practice of Information Theory. Addison-Wesley, Reading (MA), USA, 1987.
- [17] I. Boada. Towards Multiresolution Integrated Surface and Volume Data Representations. PhD thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, December 2001.
- [18] A. Brink. Minimum spatial entropy threshold selection. *IEE Proceeding-Vision, Image, and Signal Processing*, 142(3):128–132, June 1995.
- [19] J. Burbea and C. R. Rao. On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory*, 28(3):489–495, May 1982.
- [20] M. Burrows and D. J. Wheeler. A block-sorting lossless data compression algorithm. Technical Report 124, Digital Systems Research Center, 1994.
- [21] T. Butz, O. Cuisenaire, and J.-P. Thiran. Multi-modal medical image registration: From information theory to optimization objective. In *Proceedings of the International Conference on Digital Signal Processing*, pages 407– 414, Santorini, Greece, July 2002.
- [22] T. Butz and J. Thiran. Affine registration with feature space mutual information. In *International Conference on Medical Image Computing and Computed Assisted Intervention (MICCAI 2001), Proceedings*, Lecture Notes in Computer Science, pages 549–556. SPIE, 2001.

- [23] B. Cabral, N. Can, and J. Foran. Accelerated volume rendering and tomographic reconstruction using texture mapping hardware. In *Proceedings of the Symposium on Volume Visualization (VVS '94)*, pages 91–98, October 1994.
- [24] M. Capek, L.Mroz, and R.Wegenkittl. Robust and fast medical registration of 3D-Multi-modality data sets. In *Proceedings of Medicon*, pages 515–518, Pula, Croatia, June 2001.
- [25] V. Caselles, F. Catte, T. Coll, and F.Dibos. A geometric model for active contours. *Numerische Mathematik*, 66:1–31, 1993.
- [26] F. Castro and M. Sbert. Application of quasi-monte carlo sampling to the multipath method for radiosity. In *Proceedings of 3rd International Conference on Monte Carlo and Quasi-Monte Carlo methods in Scientific Computing*, Claremont (CA), USA, June 1998.
- [27] V. F. H. Chong, J.-Y. Zhou, J. B. K. Khoo, J. Huang, and T.-K. Lim. Nasopharyngeal carcinoma tumor volume measurement. *Radiology*, 231(3):914–921, 2004.
- [28] P. Cignoni, C. Montani, and R.Scopigno. A comparison of mesh simplification algorithms. *Computer and Graphics*, 22(1):37–54, 1998.
- [29] R. Cilibrasi, R. de Wolf, and P. Vitányi. Algorithmic clustering of music based on string compression. *Computer Music Journal*, 28(4):49–67, 2004.
- [30] R. Cilibrasi and P. M. B. Vitányi. Clustering by compression. *IEEE Trans*actions on Information Theory, 51(4):1523–1545, April 2005.
- [31] L. Clarke, R. Velthuizen, M. Camacho, J. Heine, M. Vaydianathan, L. Hall, R. Tatcher, and M. Silbiger. MRI segmentation. Methods and applications. *Magnetic Resonance Imaging*, 13:343–368, 1995.
- [32] C. Cocosco, V. Kollokian, R.-S. Kwan, and A. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5(4):S425, 1997.
- [33] A. Collignon. Multimodality medical image registration by maximization of mutual information. PhD thesis, Catholic University of Leuven, Leuven, Belgium, May 1998.
- [34] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. 3D multimodality medical image registration using feature space clustering. In S. Verlag, editor, *First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, volume 905 of *Lecture Notes in Computer Science*, pages 195–204, Nice, France, April 1995.

- [35] A. Collignon, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. *Computational Imaging and Vision*, 3:263–274, 1995.
- [36] D. Cooper, E. Jauch, and M. Flaherty. Critical pathways for the management of stroke and intracerebral hemorrhage: A survey of US hospitals. *Critical Pathways in Cardiology*, 6(1):18–23, 2007.
- [37] J. J. Corso, E. Sharon, and A. Yuille. Multilevel segmentation and integrated bayesian model classification with an application to brain tumor segmentation. In *International Conference on Medical Image Computing and Computed Assisted Intervention (MICCAI 2006), Proceedings*, volume 2, pages 790–798, Copenhagen, Denmark, 2006.
- [38] D. Cosic and D. Loncaric. Computer system for quantitative analysis of ICH from CT head images. In *Proceedings of the 19th Annual International Conference of the IEEE EMBS*, Chicago, USA, 1997.
- [39] T. M. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons Inc., 1991.
- [40] J. P. Crutchfield and D. P. Feldman. Regularities unseen, randomness observed: The entropy convergence hierarchy. *Chaos*, 15:25–54, 2003.
- [41] J. P. Crutchfield and N. Packard. Symbolic dynamics of noisy chaos. *Phys-ica 7D*, pages 201–223, 1983.
- [42] R. Dafner, D. Cohen-Or, and Y. Matias. Context-based space filling curves. *Computer Graphics Forum*, 19(3):209–217, 2000.
- [43] A. Davalos, M. Blanco, S. Pedraza, R. Leira, M. Castellanos, J. Pumar, Y. Silva, and J. Serena. The clinical-dwi mismatch. a new diagnostic approach to the brain tissue at risk of infarction. *Neurology*, 62(1):2187–2192, 2004.
- [44] B. M. Dawant and A. P. Zijdenbos. Image segmentation. In M. Sonka and J. M. Fitzpatrick, editors, *Handbook of Medical Imaging. Volume 2: Medical Image Processing and Analysis*, pages 71–127. SPIE Press, 2000.
- [45] Z. Dawy, J. Hagenauer, P. Hanus, and J. C. Mueller. Mutual information based distance measures for classification and content recognition with applications to genetics. In *IEEE International Conference on Communications (ICC 2005)*, volume 2, pages 820–824, Seoul, Korea, May 2005.
- [46] I. S. Dhillon, S. Mallela, and D. S. Modha. Information-theoretic coclustering. In *Proceedings of The 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2003)*, pages 89–98, New York (NY), USA, 2003. ACM Press.

- [47] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley and Sons Inc., 2nd edition, 2001.
- [48] M. Feixas. An Information-Theory Framework for the Study of the Complexity of Visibility and Radiosity in a Scene. PhD thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, December 2002.
- [49] D. Feldman and J. Crutchfield. Structural information in two-dimensional patterns: Entropy convergence and excess entropy. *Physical Review E*, 67, 2003.
- [50] D. P. Feldman. A brief introduction to: Information theory, excess entropy and computational mechanics. Technical report, University of California, October 2002.
- [51] D. P. Feldman and J. P. Crutchfield. Discovering noncritical organization: Statistical mechanical, information theoreticand computational views of patterns in one-dimensional spin systems. Working Paper 98–04–026, Santa Fe Institute, Santa Fe (NM), USA, April 1998.
- [52] L. M. Fletcher-Heath, L. O. Hall, D. B. Goldgof, and F. R. Murtagh. Automatic segmentation of non-enhancing brain tumors in magnetic resonance images. *Artificial Intelligence in Medicine*, 21:43–63, 2001.
- [53] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí. Yet another survey on image segmentation: Region and boundary information integration. In *European Conference on Computer Vision*, pages 408–422, Copenhagen, Denmark, May 2002.
- [54] R. Gan and A. Chung. Multi-dimensional mutual information based robust image registration using maximum distance-gradient-magnitude. In 19th International Conference on Information Processing in Medical Imaging 2005 (IPMI'05), volume 3565 of Lecture Notes in Computer Science, pages 210– 221, July 2005. Glenwood Springs, CO, USA.
- [55] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, Upper Saddle River (NJ), USA, 2002.
- [56] P. Grassberger. Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics*, 25(9):907–938, 1986.
- [57] B. F. Green. The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, 17:429–440, 1952.
- [58] J. A. Hartigan and M. A. Wong. A k-means clustering algorithm. *Applied Statistics*, 28(1):100–108, 1979.

- [59] J. Hemphill, D. Bonovich, L. Besmertis, G. Manley, and S. Johnson. The ich score: a simple, reliable grading scale for intracerebral hemorrhage. *Stroke*, (32):890–897, 2001.
- [60] A. Herbulot, S. Jehan-Besson, M. Barlaud, and G. Aubert. Information theory for image segmentation using shape gradient. Technical Report ISRN I3S/RR-2004-43-FR, 2004.
- [61] A. Herbulot, S. Jehan-Besson, M. Barlaud, and G. Aubert. Shape gradient for image segmentation using information theory. In *International Conference on Acoustics, Speech, and Signal Processing*, volume 3, pages 21–24, Montreal, Canada, 2004.
- [62] D. L. Hill. Combination of 3D medical images from multiple modalities. PhD thesis, University of Londonn, Image Processing Group, Radiological Sciences, December 1993.
- [63] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46:R1–R45, 2001.
- [64] D. L. G. Hill, D. J. Hawkes, M. J. Gleeson, T. C. S. Cox, A. J. Strong, W. L. Wong, C. F. Ruff, N. D. Kitchen, D. G. T. Thomas, J. E. Crossman, C. Studholme, A. J. Gandhe, S. E. M. Green, and G. P. Robinson. Accurate frameless registration of MR and CT images of the head: Applications in surgery and radiotherapy planning. *Radiology*, 191:447–454, 1994.
- [65] S. Ho, E. Bullitt, and G. Gerig. Level set evolution with region competition: Automatic 3D segmentation of brain tumors. In *Proceedings of the 16th International Conference on Pattern Recognition*, pages 532–535, Quebec, Canada, 2002.
- [66] M. Holden, L. D. Griffin, and D. L. G. Hill. Multi-dimensional mutual information image similarity metrics based on derivatives of linear scale space. In *Proceedings of the APRS Workshop on Digital Image Computing*, pages 55–60, Brisbane, Australia, February 2005.
- [67] H. Hoppe. Progressive meshes. In SIGGRAPH 96 Cenference Proceedings, pages 99–108, 1996.
- [68] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stueetzle. Mesh optimization. In SIGGRAPH 93 Cenference Proceedings, pages 19–26, 1993.
- [69] S. L. Horowitz and T. Pavlidis. Picture segmentation by a directed splitand-merge procedure. In 2nd International Joint Conference on Pattern Recognition, Proceedings, pages 424–433, Copenhagen, Denmark, 1974.

- [70] D. Huffman. A method for the construction of minimum-redundancy codes. In *Proceedings of the IRE*, volume 40, pages 1098–1101, September 1952.
- [71] L. Jonasson, P. Hagmann, C. Pollo, X. Bresson, C. Richero Wilson, R. Meuli, and J. Thiran. A level set method for segmentation of the thalamus and its nuclei in DT-MRI. *Signal Processing*, 87(2):309–321, 2007.
- [72] J. T. Kajiya. The rendering equation. SIGGRAPH Comput. Graph., 20(4):143–150, 1986.
- [73] A. Kaltchenko. Algorithms for estimating information distance with application to bioinformatics and linguistics. *CoRR*, cs.CC/0404039, April 2004.
- [74] J. Kapur, P. Sahoo, and A. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics and Image Processing*, 29(3):273–285, March 1985.
- [75] M. Kass, A. Witkin, and D. Terzopoutlos. Snakes: active countour models. *International Journal Computer Vision*, 1(4):321–331, 1987.
- [76] A. Kaufman. Volume visualization. IEEE Computer Society Press Tutorial, 1991.
- [77] S. R. Kulkarni, G. Lugosi, and S. S. Venkatesh. Learning pattern classification – a survey. *IEEE Transactions on Information Theory*, 44(6):2178– 2206, 1998.
- [78] P. Lacroute and M. Levoy. Fast volume rendering using a shear-warp factorization of the viewing transformation. In *Proceedings of SIGGRAPH '94*, pages 451–458, July 1994.
- [79] S. Lavallee. Registration for computed-integrated-surgery: Methodolgy, state of the art. *Computer Integrated Surgery: Technology and Clinical Applications*, pages 77–97, 1995. MIT Press, Cambridge, Massachusettes.
- [80] T. M. Lehmann, C. Gonner, and K. Spitzer. Registration for computedintegrated-surgery: Methodolgy, state of the art. *IEEE Transactions on Medical Imaging*, 18(11):1049–1074, November 1999.
- [81] A. Lempel and J. Ziv. Compression of two-dimensional data. *IEEE Trans*actions on Information Theory, 32(1):2–8, January 1986.
- [82] M. Levoy. Display surfaces from volume data. *IEEE Computer Graphics and Applications*, 3:29–37, 1988.
- [83] M. Li, J. Badger, X. Chen, S.Kwong, P. Kearney, and H. Zang. An information-based sequence distance and its applications to whole mitochondrial genome phylogeny. *Bioinformatics*, 17(2):149–154, 2001.

- [84] M. Li, X. Chen, X. Li, B. Ma, and P. Vitányi. The similarity metric. *IEEE Transactions on Information Theory*, 50(12):3250–3264, December 2004.
- [85] M. Li and P. Vitanyi. An Introduction to Kolmogorov Complexity and Its Applications. Springer-Verlag, 1997.
- [86] S. Li. Markov random field modeling in computer vision. Springer-Verlag, 1995.
- [87] D. Lloret. Medical image registration based on a creaseness measure. PhD thesis, Universitat Autònoma de Barcelona, Barcelona, Spain, September 2002.
- [88] S. Loncaric, A. P. Dhawan, D. Kovacevic, D. Cosic, J. Broderick, and T. Brott. Quantitative intracerebral brain hemorrhage analysis. In *Proceedings of SPIE Medical Imaging*, volume 3661, pages 886–894, San Diego, USA, 1999.
- [89] S. Loncaric, D. Kovacevic, and D. Cosic. Fuzzy expert system for edema segmentation. In *Proceedings of the IEEE 9th Mediterranean Electrotechnical Conference*, pages 1476–1479, Tel-Aviv, Israel, 1998.
- [90] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. ACM SIGGRAPH Computer Graphics, 21(4):163–169, July 1987.
- [91] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.
- [92] F. Maes, D. Vandermeulen, G. Marshal, and P. Suetens. Multimodality image registration by maximization of mutual information. In *IEEE Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 187–198, June 1996.
- [93] F. Maes, D. Vandermeulen, G. Marshal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Medical Image Analysis*, 3(4):373–386, 1998.
- [94] F. Maes, D. Vandermeulen, and P. Suetens. Comparative evaluation of multiresolution optimization strategies for multimidality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4):373– 386, 1998.
- [95] R. Malladi and J. Sethian. A unified approach to noise removal, imageenhancement, and shape recovery. *IEEE Transactions on Image Imaging*, 5(11):1554–1568, 1996.

- [96] F. Marqués. Multiresolution Image Segmentation Based On Compound Random Fields: Application To Image Coding. PhD thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, December 1992.
- [97] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.
- [98] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank. Nonrigid multimodality image registration. In SPIE Medical Imaging 2001: Image Processing, volume 4322, pages 1609–1620, 2001.
- [99] C. R. Maurer, J. M. Fitzpatrick, M. Y. Wang, R. L. Galloway, R. J. Maciunas, and G. S. Allen. Registration of head volume images using implantable fiducial markers. *IEEE Transactions Medical Imaging*, 16:447–462, 1997.
- [100] N. Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995.
- [101] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1(2):91–108, 1996.
- [102] M. Meissner, H. Pfister, R. Westermann, and C. Wittenbrink. Volume visualization and volume rendering techniques. In *Tutorials 6, Eurographics* '00, 2000. Interlaken, Switzerland.
- [103] H. Müller, N. Shareef, J. Huang, and R. Crawfis. High-quality splatting on rectilinear grids with efficient culling of occluded voxels. *IEEE Transactions* on Visualization and Computer Graphics, 5(2):116–134, 1999.
- [104] National Institutes of Health. Retrospective Image Registration Evaluation. Vanderbilt University, Nashville (TN), USA, 2003. Project Number 8R01EB002124-03, Principal Investigator J. Michael Fitzpatrick.
- [105] A. Oliver. *Automatic mass segmentation in mammographic images*. PhD thesis, Universitat de Girona, Girona, Spain, July 2007.
- [106] S. Osher and J. A. Sethian. Fronts propagation with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *Journal of Computiacional Physics*, 79:12–49, 1988.
- [107] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
- [108] S. Pedraza, I. Boada, A. Bardera, M. Feixas, J. Mendez, J. Astarloa, and A. Davalos. Development of a new coregistration method to asses the volume of ischemic penumbra. In XXIX Congress of the European Society of Neuroradiology, 2004. Aachen, Germany.

- [109] S. Pedraza, Y. Silva, J. Mendez, L. Inaraja, J. Vera, J. Serena, and A. Davalos. Comparison of preperfusion and posrperfusion magnetic resonance angiography in acute stroke. *Stroke*, 35(1):22–27, 2004.
- [110] C. Pelizzari, G. Chen, and D. Spelring. Accurate three-dimensional registration of CT, PET and/or MR images of the brain. *Journal of Computer Assisted Tomography*, 13(1):20–26, 1989.
- [111] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.
- [112] D. L. Pham, C. Xu, and J. L. Prince. A survey of current methods in medical image segmentation. Technical Report JHU/ECE 99-01, 1998.
- [113] J. P. Pluim. Mutual Information Based Registration of Medical Images. PhD thesis, Image Sciences Institute, Utrecht, The Netherlands, 2001.
- [114] J. P. Pluim, J. Maintz, and M. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE, Transactions on Medical Imaging*, 19(8):809–814, 2000.
- [115] J. P. Pluim, J. Maintz, and M. Viergever. Interpolation artefacts in mutual information based image registration. In *Computer Vision and Image Understanding*, volume 77, pages 211–232, 2000.
- [116] J. P. Pluim, J. Maintz, and M. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22:986–1004, 2003.
- [117] W. Press, S. Teulokolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992.
- [118] T. Pun. A new method for gray-level picture thresholding using the entropy of the histogram. *Signal Processing*, 2:223–237, 1980.
- [119] H. Rabiee, R. Kashyap, and H. Radha. Multiresolution image compression with BSP trees and multilevel BTC. 3:600–603, 1995.
- [120] H. Radha, R. Leonardi, M. Vetterli, and B. Naylor. Binary Space Partitioning (BSP) tree representation of images. *Journal of Visual Communication* and Image Representation, 2(3):201–221, September 1991.
- [121] J. Rigau, M. Feixas, and M. Sbert. An information theoretic framework for image segmentation. In *IEEE International Conference on Image Processing (ICIP'04)*, volume 2, pages 1193–1196, Singapore, Republic of Singapore, October 2004.

- [122] J. Rigau, M. Feixas, M. Sbert, A. Bardera, and I. Boada. Medical image segmentation based on mutual information maximization. In *International Conference on Medical Image Computing and Computed Assisted Intervention (MICCAI 2004)*, volume 3216 of *Lecture Notes in Computer Science*, pages 135–142, Rennes-Saint Malo, France, September 2004.
- [123] J.-F. Rivest, P. Soille, and S. Beucher. Morphological gradients. In E. Dougherty, J. Astola, and G. Boncelet, editors, *SPIE Nonlinear image processing III*, volume 1658, pages 139–150, February 1992.
- [124] A. Rosenfeld, editor. *Multiresolution Image Processing and Analysis*. Springer, Berlin, 1984.
- [125] J. Rossignac and P. Borrel. Multi-resolution 3D approximations for rendering complex scenes. In B. Falcidieno and T. Kunii, editors, *Modeling in Computer Graphics: Methods and Applications*, pages 455–465, Berlin, 1993. Springer-Verlag. Proc. of Conf., Genoa, Italy, June 1993. (Also available as IBM Research Report RC 17697, Feb. 1992, Yorktown Heights, NY 10598).
- [126] D. Rueckert, M. J. Clarkson, D. L. J. Hill, and D. J. Hawkes. Non-rigid registration using higher order mutual information. In *Medical Imaging: Image Processing*, pages 438 – 447, 2000. Ed. Bellingham, SPIE Press.
- [127] D. B. Russakoff, C. Tomassi, T. Rohlfing, and C. R. Maurer. Image similarity using mutual information of regions. In 8th European Conference on Computer Vision, volume 3, pages 596 – 608, Prague, Czech Republic, May 2004.
- [128] M. R. Sabuncu and P. J. Ramadge. Spatial information in entropy-based image registration. In *Proceedings of 2nd Workshop in Biomedical Image Registration (WBIR'03)*, volume 2717 of *Lecture Notes in Computer Science*, pages 132–141, Philadelphia, USA, May 2003.
- [129] P. K. Sahoo, S. Soltani, A. K. Wong, and Y. C. Chen. A survey of thresholding techniques. *Comput. Vision Graph. Image Process.*, 41(2):233–260, 1988.
- [130] O. Salvado and D. Wilson. Removal of interpolation induced artifacts in similarity surfaces. In *Proceedings of the Third International Workshop on Biomedical Image Registration (WBIR 2006)*, volume 4057 of *Lecture Notes in Computer Science*, pages 43–49, Utrecht, Netherlands, 2006. Springer Berlin / Heidelberg.
- [131] H. Samet. Application of Spatial Data Structures: Computer Graphics, Image Processing, and CIS. Addison Wesley, Reading, MA, 1990.

- [132] L. A. Santaló. Integral Geometry and Geometric Probability. Cambridge University Press, 1976.
- [133] M. Sbert. An integral geometry based method for fast form-factor computation. *Computer Graphics Forum (Proceedings of Eurographics'93)*, 12(3):409–420, 1993.
- [134] W. J. Schroeder, J. A. Zarge, and W. E. Lorensen. Decimation of triangle meshes. *Computer Graphics (SIGGRAPH '92 Proc.)*, 26(2):65–70, July 1992.
- [135] I. K. Sethi and G. Sarvarayudu. Hierarchical classifier design using mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelli*gence, 4(4):441–445, July 1982.
- [136] J. A. Sethian. Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Material Science. Cambridge University Press, 1999.
- [137] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–168, January 2004.
- [138] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, July, October 1948.
- [139] R. Shaw. The Dripping Faucet as a Model Chaotic System. Aerial Press, Santa Cruz (CA), USA, 1984.
- [140] N. Slonim and N. Tishby. Agglomerative information bottleneck. In Proceedings of NIPS-12 (Neural Information Processing Systems), pages 617– 623. MIT Press, 2000.
- [141] N. Slonim and N. Tishby. Document clustering using word clusters via the information bottleneck method. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 208–215, Athens, Greece, 2000. ACM Press.
- [142] M. Sonka and J. M. Fitzpatrick, editors. *Handbook of Medical Imaging*. *Volume 2: Medical Image Processing and Analysis.* SPIE Press, 2000.
- [143] H. Strik, H. Borchert, C. Fels, M. Knauth, O. Rienhoff, M. Bahr, and J. Verhey. Three-dimensional reconstruction and volumetry of intracranial haemorrhage and its mass effect. *Neuroradiology*, 47:417–424, 2005.
- [144] C. Studholme. Measures of 3D medical image alignment. PhD thesis, Computational Imaging Science Group, Division of Radiological Sciences, United Medical and Dental school's of Guy's and St Thomas's Hospitals, 1997.

- [145] C. Studholme, D. Hill, and D. Hawkes. Using voxel similarity as a measure of medical image registration. In *Proceedings of the British Machine Vision Conference*, pages 235–244, 1994.
- [146] C. Studholme, D. Hill, and D. Hawkes. Multiresolution voxel similarity measures for mr-pet registration. In *Proceedings of the Information Processing in Medical Imaging Conference*, pages 287–298, Ile de Berder, France, June 1995.
- [147] J. S. Suri, S. Singh, and S. Laxminarayan. Medical image segmentation using level sets. In J. Suri and S. Laxminarayan, editors, *PDE and Level Sets: Algorithmic Approaches to Static and Motion Imagery*, pages 301– 340. Kluwer Academic/Plenum Publishers, 2002.
- [148] P. Szépfalusy and G. Györgyi. Entropy decay as a measure of stochasticity in chaotic systems. *Physical Review A*, 33(4):2852, 1986.
- [149] N. Tishby, F. C. Pereira, and W. Bialek. The information bottleneck method. In Proceedings of the 37th Annual Allerton Conference on Communication, Control and Computing, pages 368–377, 1999.
- [150] A. Torrent, A. Bardera, A. Oliver, J. Freixenet, I. Boada, M. Feixas, R. Martí, X. Lladó, J. Pont, E. Pérez, S. Pedraza, and J. Martí. Breast density segmentation: a comparison of clustering and region based techniques. In *Proceedings of the International Workshop on Digital Mammography (IWDM'08)*, volume 5116 of *Lecture Notes in Computer Science*, Tucson, USA, July 2008.
- [151] J. Tsao. Interpolation artifacts in multimodal image registration based on maximization of mutual information. *IEEE Transactions on Medical Imaging*, 22:854–864, November 2003.
- [152] A. Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42:230– 265, 1937.
- [153] M. Unser. Splines. A perfect fit for signal and image processing. IEEE Signal Processing Magazine, 16:22–38, 1999.
- [154] M. Unser and P. Thévenaz. Stochastic sampling for computing the mutual information of two images. In 5th International Workshop on Sampling Theory and Applications, 2003, Proceedings., pages 102–109, Strobl, Austria, May 2003.
- [155] J. C. van der Lubbe. Information Theory. Cambridge University Press, Cambridge, UK, 1997.

- [156] P. A. Viola. Alignment by Maximization of Mutual Information. PhD thesis, MIT Artificial Intelligence Laboratory (TR 1548), Massachusetts (MA), USA, 1995.
- [157] W. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information, 1996.
- [158] L. Westover. Interactive volume rendering. In *Chael Hill Volume Visualiza*tion Workshop, pages 9–16, 1989.
- [159] L. Westover. Footprint evaluation for volume rendering. In *Proceedings of Siggraph*, pages 367–376, 1990.
- [160] S. Wolfram. A New Kind of Science. Wolfram Media, 2002.
- [161] R. Woods, J. Mazziotta, and S. Cherry. MRI-PET registration with automated algorithm. *Computer Assisted Tomography*, 17(4):536–546, 1993.
- [162] R. W. Yeung. A First Course in Information Theory. Springer, 2002.
- [163] A. Yezzi, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum. A geometric snake model for segmentation of medical imagery. *IEEE Transaction in Medical Imaging*, 16:199–209, 1997.
- [164] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, 23(3):337–343, May 1977.
- [165] S. Zucker. Region growing: Childhood and adolescence. *Computer Graphics and Image Processing*, 5:382–399, 1976.