Automated Sagmentation of Cerebral Structures Incorporating Explicit Knowledge

Alain Pitiot

École des Mines de Paris

Segmentation Automatique des Structures Cérébrales s'appuyant sur des Connaissances Explicites

THÈSE

présentée et soutenue publiquement le 26 Novembre 2003

pour l'obtention du

Doctorat de l'École des Mines de Paris

(spécialité Informatique temps-réel, robotique, automatique)

par

Alain PITIOT

Co-Directeurs de thèse : Nicholas AYACHE, Paul M. THOMPSON

Composition du jury

Président :	Jean-Paul Marmorat
Rapporteurs :	Jean-François Mangin
	Dirk Vandermeulen
Examinateurs :	Nicholas Ayache
	Paul M. Thompson
	Hervé Delingette
Invité :	Grégoire Malandain

Institut National de Recherche en Informatique et Automatique

Alain Pitiot

Automated Segmentation of Cerebral Structures Incorporating Explicit Knowledge

Segmentation Automatique des Structures Cérébrales s'appuyant sur des Connaissances Explicites

Ph.D. manuscript

Acknowledgment

I will start this thank-you note with a little disclaimer: it is quite possible that you will not find your name on this page. This could be because: (a) we have never met; (b) we met under circumstances that I could not confess here; (c) I could not summon your name when I wrote these a few lines (even though both your personality and your advice truly made a lasting impression on me), or (d) I knew you would not read these 200+ pages so I'd rather thank you in person (I hope I did).

That being said, I now have the difficult task of acknowledging the influence of the many people who, by their mere presence, after countless hours of discussions (read, arguments), or through their encouragements, shaped this Ph.D. thesis. Sharing my time between two laboratories has been a truly rewarding experience. Unfortunately, it also means I have twice as many people to thank; that is, I'm twice more likely to have forgotten you.

First and foremost, I would like to send my gratitude to my Ph.D. co-directors: Nicholas Ayache from the EPIDAURE laboratory and Paul Thompson from the LONI laboratory. I've immensely learnt under their leadership, and certainly hope I will again have the opportunity to collaborate with them in the future.

Working with Hervé Delingette, Grégoire Malandain, Jacopo Annese and Eric Bardinet as been both an honor and a pleasure. Their (sometimes unusual) sense of humor and (extensive) scientific experience will be greatly missed.

I will extend my gratitude to the reviewers ("rapporteurs") of this thesis, Jean-François Mangin and Dirk Vandermeulen, who had the ungrateful task of reading through these pages to evaluate my work. Thank you also to Jean-Paul Marmorat, the president of the defense jury, for his insights.

Maria Vallenilla, arguably the kindest soul in the western world, too deserves to be thanked for without her there would not have been a Ph.D in the first place.

Many thanks to the EPIDAURE and LONI teams for their varied personalities, for the deep and frivolous conversations that I so very much enjoyed all through these years, for the food we cooked and the wine we drank. Thank you then to Allan, Andrew, Céline, Christine, David, Emma, Guillaume, Guillaume, Heike, Isabelle, Jonathan, Karen, Linda, Marc, Maxime, Marius, Mauricio, Miguel, Nicolas, Olivier, Olivier, Pascal, Radu, Rasmus, Sebastien, Thibaut, Valérie, Vincent and Xavier. We will meet again.

Associated Team Program

This Ph.D. thesis has been conducted within the framework of the INRIA sponsored associated team program. The objective of this program is to strengthen and expand established relationships between a team of INRIA researchers and that of a foreign laboratory. In our case, our endeavour (and time) was shared between the EPIDAURE laboratory at INRIA Sophia Antipolis and the LONI laboratory at the UCLA School of Medicine, Department of Neurology.

The INRIA EPIDAURE team focuses on the design and development of tools adapted to the analysis of multidimensional and multimodal medical images, with the aim to enhance the available diagnosis. A non-exhaustive selection of the main research axes would encompass the extraction of quantitative parameters for automated/aided diagnosis (shape, texture, displacement), the spatial and temporal registration of 2-D and 3-D medical images, multimodal image fusion problems, the analysis of deformable displacements, or the design and manipulation of digital atlases with, in the longer run, the goal of automating the morphometric and functional analysis of the brain.

A member of the UCLA School of Medicine, the LONI laboratory is dedicated to the development of scientific approaches for the comprehensive mapping of brain structure and function. Use of fluid warping approaches has enabled the development of disease-specific digital brain atlases (for Alzheimer's disease and schizophrenia, among other diseases). Based on large human populations, and containing thousands of 3-D structure models, these atlases encode patterns of anatomical variation and can be used to detect group-specific patterns of anatomic or functional alterations.

This close collaboration primarily focuses on the building and analysis of brain atlases.

> The explosive growth in brain imaging technologies (such as magnetic resonance imaging) has only been matched by a tremendous increase in the number of investigations focusing on the structural and functional organization of the human brain. However, the formidable complexity and variability of cerebral structures present a substantial challenge to their effective analysis and modeling, and require the conjunction of a number of techniques drawn from a large panel of scientific domains: computer vision, image analysis or statistical modeling to name just a few. A natural way to condense, analyse and communicate brain data, digital brain atlases consequently stem from the combination of many algorithms whose underlying hypotheses must be validated *a posteriori*.
>
> The association of both teams within this program enables a better synthesis of the experience and knowledge they have both acquired, and makes it possible to jointly tackle the restoration, segmentation or image registration issues raised by the development of digital brain atlases.

With the automated segmentation of anatomical structures in MRI as its central theme, this Ph.D. thesis was part of the effort towards the development of better atlas building tools. Our presence at the LONI laboratory also offered the opportunity to adapt the developed methodologies to other types of applications or data, and to tackle a range of side projects, not necessarily in direct relation to the central theme. In particular, we took part in the Mouse Atlas Project (see Section 2.4.3.3) which aims to produce an imaging framework to correlate gene expression with anatomic and molecular information, and in the Visual Cortex project (see Section 3.3.4.5) whose goal is the analysis of the histological structure of the cerebral cortex in the visual areas of the human brain.

Contents

Chapter 1

Introduction
Introduction[1]

Since we cannot know all that there is to be known
about anything, we ought to know a little about everything.
Blaise Pascal.

Après un rapide aperçu des tenants de l'imagerie médicale, ce premier chapitre d'introduction motive les objectifs de notre étude et précise les hypothèses sous-jacentes. Confronté à la nécessité de délinéer avec une précision contrôlable un nombre toujours plus important de structures cérébrales dans les image IRM (comme aide au diagnostique, afin de suivre dans le temps les dysfontionements du cerveau ou encore d'en étudier la variabilité anatomique) la segmentation automatique apparaît comme un passage obligé. Nous avançons que l'utilisation d'un maximum de connaissances a priori, *sous forme à la fois explicite (c'est à dire établie par un opérateur humain) et implicite (telle que découverte par un algorithme automatique ou semi-automatique) est requise pour faire face à la très grande variabilité de forme et d'apparence des structures anatomiques. Le paradigme de déformation retenu est celui des modèles déformables, qui fournissent un cadre propice à l'incorporation d'expertise médicale* a priori. *Nous appréhandons la problématique de segmentation des structures cibles (corps calleux, noyaux caudés, ventricules et hippocampes) sous l'angle du processus d'appariement d'un groupe de patrons déformables à leurs contours. Ces patrons évoluent en parallèle sous la supervision d'une série de règles issues à la fois de l'analyse de leur dynamique de déformation et de l'expertise médicale. Nous soumettons également les modèles à une palette de contraintes elles aussi conçues sur la base des connaissances* a priori *des propriétés de texture et de forme des structures à segmenter. Ces contraintes seront détaillées dans les chapitres 2 et 3.*

[1] Non-French speakers may want to skip this page which summarizes the content of the chapter in French.

'Tis written: "*In the beginning was the Word!*"
Here now I'm balked! Who'll put me in accord?
It is impossible, the Word so high to prize,
I must translate it otherwise
If I am rightly by the Spirit taught.
'Tis written: *In the beginning was the Thought!*
Consider well that line, the first you see,
That your pen may not write too hastily!
Is it then Thought that works, creative, hour by hour?
Thus should it stand: *In the beginning was the Power!*
Yet even while I write this word, I falter,
For something warns me, this too I shall alter.
The Spirit's helping me! I see now what I need
And write assured: *In the beginning was the Deed!*
Johann W. von Goethe, *Faust*, 1831

Asger Jorn, *In The Beginning Was The Image*, 1965

Images are representations. They contain descriptive information about the entities (objects, collections of objects, other images, etc.) they represent. However, since they do not, and cannot, encompass all the characteristics of these entities, they are in fact only an abstraction of them. An image is then inherently linked to the choice of the abstracted characteristics, a choice made both by the device which produces the image, and by the viewer who observes it.

A restricted type of image, digital pictures are organized collections of values linked via a chain of treatments and sensors to some underlying physical measures. In this respect, they are a means to analyse measurements: once the values are mapped to a pallet of colors or gray shades, the human eye can better decipher the otherwise cryptic nature of a raw list of numbers (compare Figure 1.1(a) and 1.1(b)). The necessity to understand these values then becomes that of extracting meaningful information from the associated pictures[2].

In spite of the remarkable ability of the human brain to find patterns and relations in data flows (sometimes to the point where meaning is extracted from utter chaos), algorithms are often better suited than humans, in terms of speed or accuracy or reproducibility, to carefully pre-process images and summarize their content in maps, graphs or diagrams (Figure 1.1(c)).

[2]Note that we will use in this manuscript the word "image" as an equivalent to the word "picture".

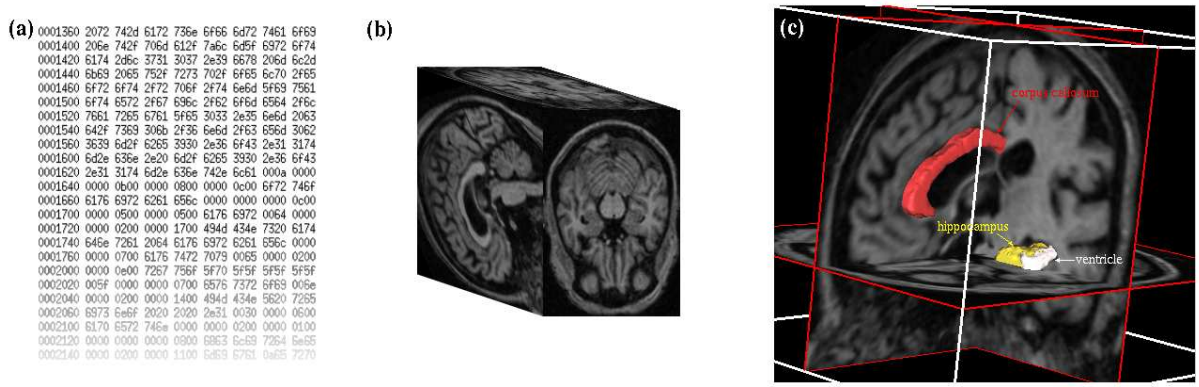


Figure 1.1: Image analysis: (a) from raw numbers: excerpt from a screen dump of the data file associated to a magnetic resonance image (MRI); (b) through 3-D image representation: 3-D view of the file's MRI; (c) to high-level semantic diagram: 3-D view of the MRI with some superimposed segmented and labelled structures.

Image analysis is then shared between the machine which collects and processes the input data to extract information, and the human operator/observer who, in addition to guiding the machine in these tasks, post-processes and more importantly interprets the produced information[3].

What is more, the development of new imaging techniques (X-ray, ultrasound, radar) has enabled the investigation of domains that are far out of reach of the naked eye. We can now observe objects or follow processes within closed sarcophagi, through walls, metals or flesh, at micron scale or at astronomical distances.

Nowhere has this ability to "see beyond" been more beneficial than in the medical world. Since the early days of anatomy, rapid development of imaging technologies (magnetic resonance imaging (MRI), computer assisted tomography (CT), positron emission tomography (PET), cryogenic systems, etc.: see Figure 1.2) now permits not only the precise analysis of dead tissues but also the non-invasive exploration of living organs and organisms. These techniques make it possible to elucidate the structures of organs and cells, enable the observation, and from there the comprehension, of the way they function, and give clinicians the means to monitor their dysfunctions, or assist in the removal of possible pathologies.

One of the most complex and least accessible organs, the human brain is the primary beneficiary of these new medical imaging techniques. Its complexity is expressed at a variety of scales. At a microscopic level, neurons, glial cells and fibers form the support tissue of the cerebral communication infrastructure (gray and white matter). They bathe in the cerebrospinal fluid (CSF), produced by the ventricular system. At a more macroscopic level, the brain is partitioned into a number of regions (brainstem, cerebellum, limbic system, diencephalon and cerebrum) which are associated with high-level mechanisms such as limb coordination, emotions or body temperature control, etc. Inside these regions, we distinguish structures (amygdala, hyppocampus, basal ganglia, ...) whose functional importance has been demonstrated by decades of anatomical and pathological studies. A deeper understanding of both the anatomical characteristics of these

[3]With the advent of artificial intelligence, difficult high-level tasks (automated speech to text translation, optical document recognition and indexation, etc.) can now sometimes be assigned to advanced cognitive systems, blurring the line between man and machine understanding.

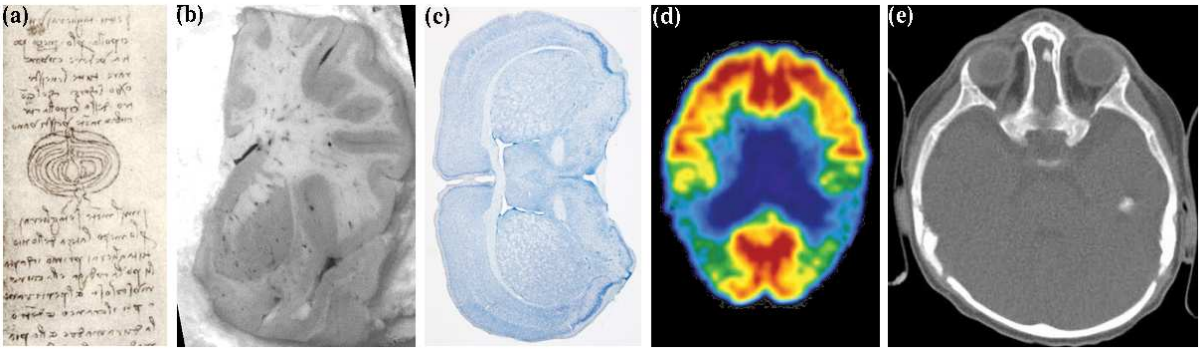


Figure 1.2: A collection of imaging modalites: (a) Leonardo Da Vinci's histological-like dissection of an onion, (b) cryosection of a human brain, (c) Nissl-stained histological section of a mouse brain, (d) PET image of the human brain, (e) CT scan of a human head.

structures and of the carefully organized web of their inter-relationships could play a crucial role in the search for more efficient cures for neurodegenerative diseases (Parkinson syndrome, Huntington's chorea, ...), to assist pre- or per-operative surgical planning and guidance, or in the design of more reliable diagnosis systems for a variety of lesions, cancers and other brain pathologies. Consequently, clarifying the correlation between brain structure and function is a major long-term goal of neuroscience.

The precise segmentation and labeling of anatomical structures is a pivotal first step along the path towards this better comprehension. Given the wide variety of shapes and appearances these structures can present, the need for a high accuracy however makes it a true challenge. While this task has traditionally been tackled by human operators, the many drawbacks of manual delineation (lack of reproducibility, strong *a priori* biases, unavailability of sufficient resources to handle ever-growing databases of images) advocate the use of automated methods.

We propose in this thesis an automated segmentation system for medical images, most especially *in vivo* brain MRI[4]. We have focused on devising a segmentation methodology, built from the combination of a variety of techniques, where maximum use is made of the available *a priori* medical expertise, as we believe it is an essential ingredient for a successful segmentation. Namely, rather than considering the input images outside of their context, a lot can be gained by incorporating medical knowledge into the segmentation algorithms. To illustrate the interest of our approach, we have selected a panel of structures, from the various systems described above: the lateral ventricles, the corpus callosum, the caudate nucleus and the hippocampus.

Furthermore, while MRI (which we will describe at greater length in Section 2.2.1) is the imaging modality of choice to monitor, mostly at a macroscopic level, processes or structures *in vivo*, much more information about the microscopic characteristics of the underlying tissues can be obtained through *post-mortem* histology (see Section 2.2.2), which additionally offers greater resolution (of an order of magnitude 100 to 1000 times bigger). Use of a histological atlases could therefore greatly increase the accuracy and robustness of an MRI segmentation

[4]In the course of our study, we have also had the opportunity to apply the developed techniques to other types of images in different contexts: the texture filters detailed in Chapter 2 also handled the segmentation of mouse brain tissues in Nissl stained/cryo- sections, our reparameterization technique (Chapter 3) served to compute mean intensity profiles for the analysis of the cortical mantle, etc. See Section 1.4 for an overview of these additional applications.

process. Building such a 3-D atlas requires that the 2-D histological slices be registered. We have accordingly developed a novel 2-D registration technique where, again, *a priori* information about the histological treatments helps achieve a better registration.

The following section presents a rapid overview of brain anatomy, with emphasis on the structures we have selected and details about their functional importance. We then motivate the interest of an automated approach in view of the problems inherent to manual delineation and advocate the use of *a priori* medical information (Section 1.2). Finally, we describe the various techniques we have developed and introduce the layout of the manuscript in Section 1.3.

1.1 A brief overview of cerebral anatomy

> If the brain were so simple we could understand it,
> we would be so simple we couldn't.
> Lyall Watson.

At a glance, and unfortunately even after decades of minute studies, the human brain is a signal processing black-box[5]. Through the senses, it receives a constant flow of stimuli that it processes via a variety of complex mechanisms. The extracted information is then carefully organized and stored in memory, ready to be used when need be. The diversity of these mechanisms is matched by the extent of the expertise of the brain: thoughts, memory and speech, limb movements and the function of many organs are elaborated, controlled and corrected in the brain. This functional complexity is largely mirrored by the intricate anatomy of the nervous system of which it is part.

The nervous system consists of two components: the *central* nervous system where we find the brain itself, its cranial nerves and the spinal cord; and the *peripheral* nervous system, composed of the spinal nerves (that branch from the spinal cord) and the autonomous nervous system (divided into the sympathetic and parasympathetic nervous system). Even though we here focus on the brain itself, it should be kept in mind that these two systems are in constant interaction.

1.1.1 A microscopic view of the brain

1.1.1.1 Neurons and glial cells

At a microscopic level, the brain is made up of two types of cells: neurons and glial cells.

Neurons are responsible for sending and receiving nerve impulses in the form of electro-chemical signals. The human brain contains about 100 billion of them. Presenting a wide variety of structures (sensory, motoneurons, interneurons), their size varies from $4\mu m$ to $100\mu m$ in diameter, while their length varies from a few millimeters to several meters.

Neuron consists of a cell body (soma), with branching dendrites, which play the role of signal receivers, and a projection called an axon, which conducts the nerve signal away from the cell body (see Figure 1.3).

[5] This is not to say that we do not know anything about the "ways of the brain": the combined use of functional and anatomical MRI has enabled the careful analysis of many internal structures, the complete mapping of some sensory paths is nearing completion etc. However, we still do not have at our disposal a satisfying functional model of the brain, even at a coarse scale.

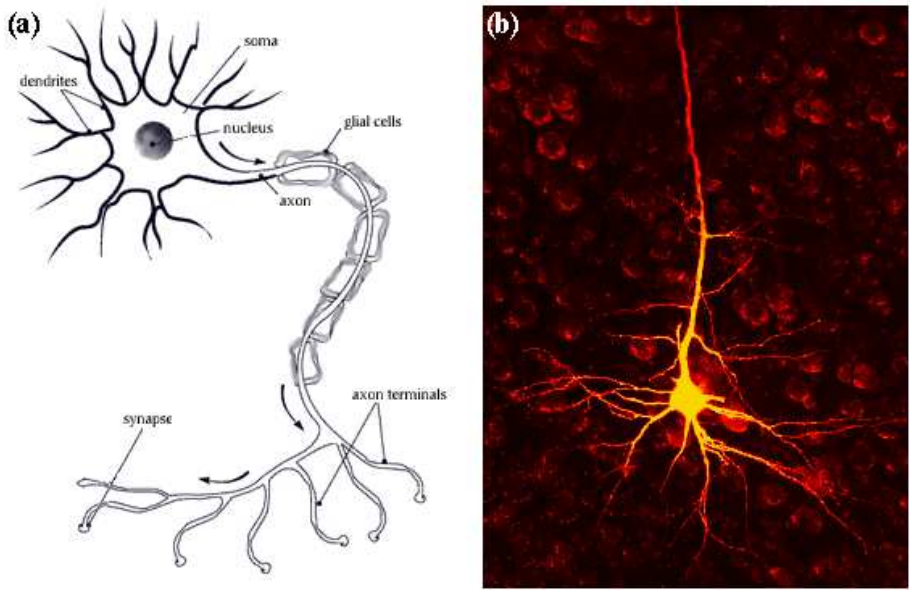


Figure 1.3: Human brain neuron: (a) general diagram, (b) microscopic photography.

- The soma contains the neuron's nucleus (with DNA and typical nuclear organelles).
- Dendrites branch from the cell body and receive messages.
- The axon terminals transmit the electro-chemical signal across synapses (gaps between the axon terminal and the receiving cell). Typical neurons have between 1,000 to 10,000 synapses.

Myelin coats, insulates and increases the transmission speed along the axon.

Glial cells (from "glia", glue) make up 90 percent of the brain's cells. They are responsible for providing physical and nutritional support for neurons, for the digestion of parts of dead neurons, for manufacturing myelin for neurons, for recycling neurotransmitters, etc.

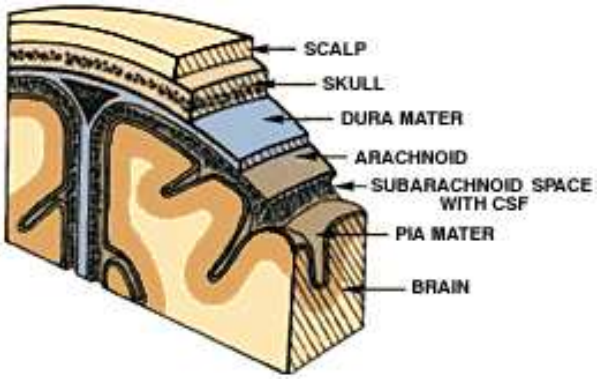


Figure 1.4: The human brain meninges.

1.1.1.2 The meninges

In addition to the skull (bones and cranium), the brain is protected by three consecutive layers of tissue: the *dura mater* (the outermost layer), the arachnoid and the pia mater (see Figure 1.4).

The *dura mater* is itself made up of two membranes (inelastic films): the periosteum and the dura. An inner layer, the dura lines the inside of the entire skull and creates little folds or compartments in which parts of the brain are protected and secured. Two particular folds can be distinguished: the falx which separates the right and left hemisphere (useful for determining the mid-sagittal section on an MRI), and the tentorium which separates the upper and lower parts of the brain.

The arachnoid is a delicate layer of elastic tissue and blood vessels. It is separated from the dura by the subdural space, and from the *pia mater* by the subarachnoid space in which the cerebrospinal fluid flows.

The pia mater, which perfectly follows the folds of the brain, contains the many blood vessels which supply the brain with oxygen and nutriments.

1.1.1.3 Cerebrospinal fluid and the ventricles

A clear watery substance, the cerebrospinal fluid (CSF) is found both within and around the brain and spinal chord and helps to cushion them from injury. It circulates through channels around the brain, where it is constantly absorbed and replaced. Inside the brain, it is produced by the ventricular system (mostly by a specialized structure inside the ventricles: the choroid plexus).

The ventricular system is divided into four cavities called ventricles which are connected by a series of holes (foramina) and tubes. The lateral (first and second) ventricles are enclosed in the cerebral hemispheres. They communicate through a separate opening called the foramen of Monro with the third ventricle (in the center of the brain), whose external walls are made up of

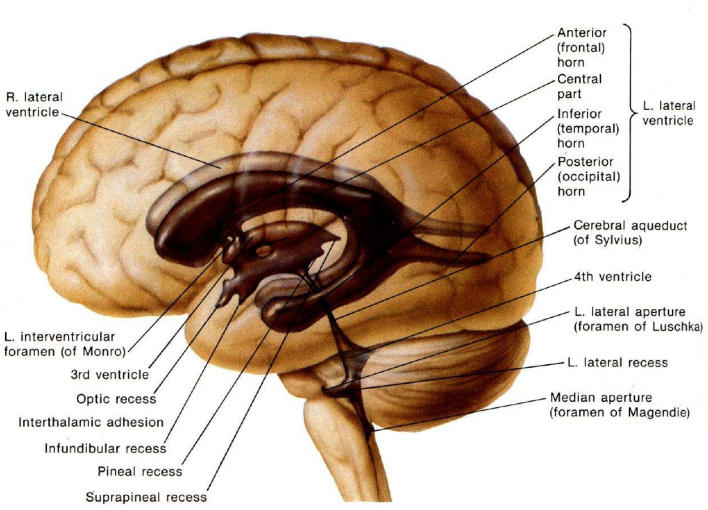


Figure 1.5: The ventricular system (courtesy of CIBA).

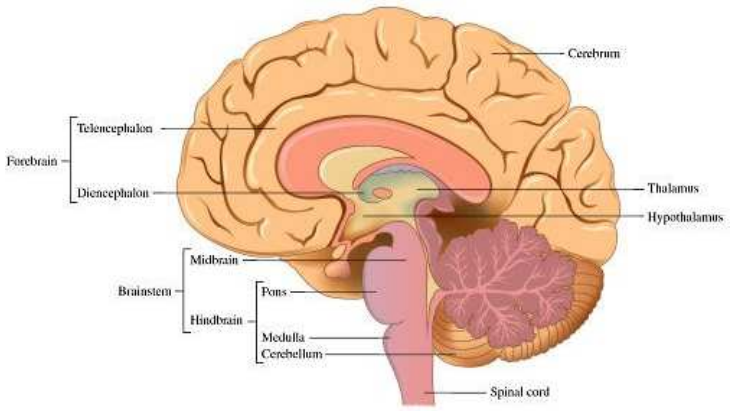


Figure 1.6: The main regions of the brain (courtesy of G. Matthews).

the thalamus and hypothalamus. This third ventricle connects with the fourth ventricle through the Aqueduct of Sylvius, while the fourth ventricle connects with the subarachnoid space via the foramen of Magendie (see Figure 1.5).

Note that the balance between the amount of cerebrospinal fluid that is absorbed and the amount that is produced is delicate and may sometimes be disrupted (leading to hydrocephalus).

1.1.2 Structures of the brain

An heterogeneous organ, the brain consists of 5 distinct regions, each responsible for a different activity: the brainstem, the cerebellum, the limbic system, the diencephalon, and the cerebral cortex (see Figure 1.6).

The brainstem connects the brain to the spinal cord. Together with the nuclei that it contains, it controls many basic body functions, such as heart rate, breathing, eating, and sleeping.

The cerebellum coordinates the brain's instructions for skilled repetitive movements and for maintaining balance and posture.

The limbic system is a collection of more evolutionarily primitive brain structures involved in many of our emotions and motivations, particularly those that are related to survival, such as fear, anger, and emotions related to sexual behavior. Parts of this system, the amygdala and hippocampus are also involved in memory, the fornix is a major fiber tract by which the hippocampus of each cerebral hemisphere communicates with the contralateral hippocampus and other brain structures.

The diencephalon, which is also located beneath the cerebral hemispheres, contains the thalamus and hypothalamus. The thalamus is involved in sensory perception and regulation of motor functions. The hypothalamus plays a major role in regulating hormones, the pituitary gland, body temperature, the adrenal glands, and many other vital activities.

The cerebrum, which forms the bulk of the brain, is divided between the left and right cerebral hemispheres. These two sides of the brain are joined at the bottom by the corpus callosum which delivers messages from one half of the brain to the other.

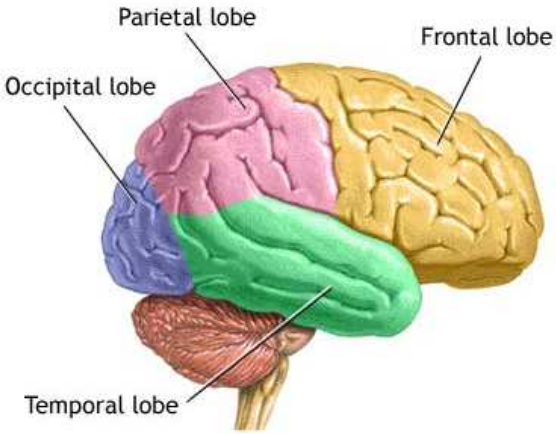


Figure 1.7: The brain lobes (courtesy of ADAM).

1.1.2.1 Gray and white matter

The surface of the cerebrum contains billions of neurons and glia that together form the cerebral cortex or gray matter, while beneath the surface, connecting fibers between neurons form the white matter. Within the white matter of the cerebrum, we find other well defined collections of gray matter: the basal ganglia (see Section 1.1.2.3) and the limbic system.

1.1.2.2 Gyri, sulci and lobes

The cerebral cortex presents a variety of small grooves (sulci), larger grooves (fissures) and bulges between the grooves (gyri). The fissures, which can be identified across individual brains, are clear landmarks which help isolate very specific regions of the brain: the lobes (see Figure 1.7).

The cerebrum may then be divided into pairs of frontal, temporal, parietal and occipital lobes. Each lobe may be divided, once again, into areas that serve specific functions.

Frontal Lobes: It contains the primary motor cortex and the precentral gyrus, which correspond to the areas that produce movement of parts of the body. The premotor cortex (a region found beside the primary motor cortex) is responsible for eye and head movements, and sense of orientation.

The prefrontal cortex plays an important part in memory, intelligence, concentration, temper and personality.

Broca's area, important in language production, is also found in the frontal lobe.

Occipital Lobes: These lobes correspond to the visual cortex: they contain regions that contribute to our visual field or how the eyes perceive the world. They are responsible for the perception of lights and objects and allow their recognition and identification.

The occipital lobe on the right interprets visual signals from your left visual space, while the left occipital lobe does the same for your right visual space. Damage to one occipital lobe may result in loss of vision in the opposite visual field.

Temporal Lobes: They are the primary region responsible for memory. They contain Wernicke's area (important for language and speech functions). They also encompass the primary auditory cortex (audio equivalent of the primary visual cortex), which processes sounds.

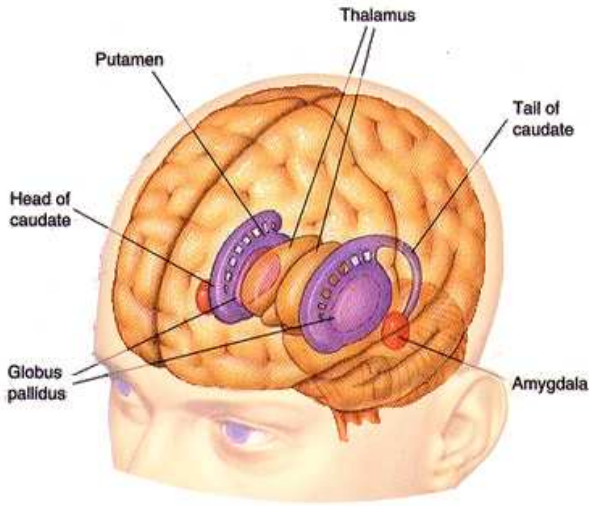


Figure 1.8: The basal ganglia (courtesy of J.L. Driesen).

Parietal Lobes: The parietal lobes interpret, simultaneously, sensory signals received from other areas of the brain such as vision, hearing, motor, sensory and memory. Together, memory and the new information that is received give meaning to objects.

1.1.2.3 Basal ganglia

The basal ganglia, like the limbic system, are embedded deep within the white matter of the cerebrum. Notwithstanding the disagreements between authors as to the number of nuclei, the basal ganglia consist, at the minimum, of the caudate nucleus and the lenticular nucleus. The nuclei of the basal ganglia are always paired structures, with one member per hemisphere (see Figure 1.8).

The caudate nucleus is bordered laterally by the fibers of the corpus callosum and medially by the lateral ventricle. It is a C-shaped mass of gray matter with an anteriorly placed bulge, the head, a middle portion, the body, and a curved posterior-inferior portion, the tail.

The lenticular nucleus is a thumb-sized mass of gray matter consisting of two divisions - globus pallidus and putamen. The putamen is continuous with and lies inferior to the head of caudate nucleus in the temporal lobe. The globus pallidus lies immediately medial to putamen in the temporal lobe.

Functionally, the basal ganglia receive input from the primary, motor and somatosensory cortices. In turn, they provide output, through the thalamus, to the primary, supplementary and premotor cortices. These connections are both excitatory (facilitating movement) and inhibitory (suppressing movements). The basal ganglia are then essential for the execution of normal volitional movements as they facilitate the execution of purposeful movement patterns and inhibit unwanted movements.

Basal ganglia lesions result in hypokinesia (reduced speed and range of movements), bradykinesia (slowness of movements), akinesia (difficulty in initiating voluntary movements), and dystonia (rigidity in muscles due to increased muscle tone). Lesions of basal ganglia also frequently

result in hyperkinesia (involuntary movements) with symptoms of athetosis (slow, writhing movements that interfere with willed movements), chorea (involuntary rhythmic movements akin to dance moves), tremor (rhythmic shaking of body parts either when at rest, resting tremor, or or when a movement is attempted, intention tremor) and ballismus (or ballism characterized by wild flinging movements).

Basal ganglia are implicated in Huntington's chorea, a fatal disease characterized by jerky, irregular, rapid and purposeless movements and dementia.

The Parkinson's disease (whose symptoms include hypokinesia, tremor, mask-like facial appearance and dysarthria) is attributed to a damage to the substantia nigra in the midbrain and the resulting disruption of a part of the extrapyramidal pathway (connections between the basal ganglia and the motor and somatosensory cortices).

1.2 Segmenting the internal brain structures

In view of the functional importance of the internal structures of the brain, their precise segmentation and labeling is a major objective of neuro-informatics. The need, shared across many levels of description, for such correlation between brain structures and functions is exemplified by the vast number of studies concerned with the analysis of cortical structures: across homogeneous populations with a particular disease [151], through the developmental cycle [17] or between normal and diseased patients [196].

While qualitative analysis may sometimes be sufficient for disease diagnosis, quantitative analysis, for which segmentation and labeling are absolute prerequisites, is necessary for a variety of applications: longitudinal monitoring of disease progression or remission [90, 169], pre-operative evaluation and surgical planning [97], radiotherapy treatment planning [154] or statistical analysis of anatomical variability [42, 195].

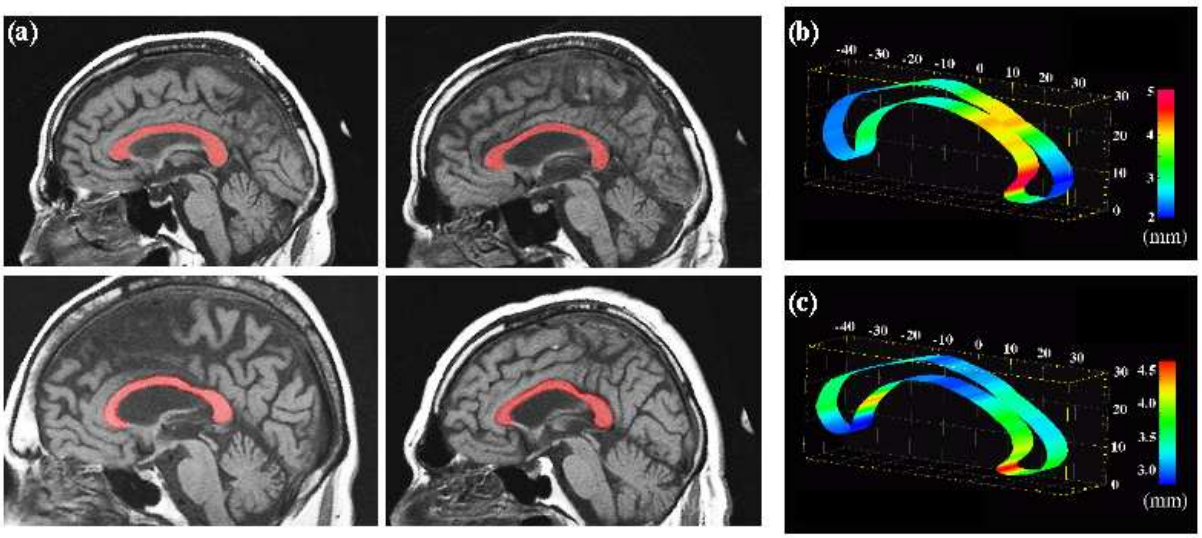


Figure 1.9: Variability of the corpus callosum: (a) four T1-weighted MRI with delineated corpus callosum (red); variability maps of the corpus callosum in Alzheimer's disease patients (b) and in elderly normal subjects (c).

Yet, effective segmentation and labeling of anatomical structures in medical images proves to be especially challenging, given the wide variety of shapes, sizes and appearances a structure can present. As an illustration, Figure 1.9(a) displays a few sample corpora callosa in the mid-sagittal slice of T1-weighted MRIs ($1mm^3$ voxel resolution). Note how the locations, shapes and relative intensities of the corpus callosum vary from one brain to the next. This figure also compares the variability of the corpus callosum in patients with Alzheimer's disease (b) and in elderly normal subjects (c). The root mean square variation of callosal points in each subject group around an average boundary representation of the callosum is displayed in color, at the mid-sagittal plane of Talairach stereotaxic space. Clearly, variations, albeit sometimes rather subtle, are observed, both within a group, and across groups [197].

In spite of the high variability of the brain structures, the delineation process calls for high precision as the quality of the medical diagnosis often depends on how accurately the various structures can be identified. For instance, as argued in Thompson *et al.* [197], given the corpus callosum's key role as the primary cortical projection system, regional analysis of its structure is a key factor in the radiological assessment of a wide range of neurological disorders (Alzheimer's disease, multi-infarct dementia, dysplasias, etc.). Nonetheless, the variations illustrated in Figure 1.9 make it difficult to detect and classify abnormal structural patterns. As a result, intense debate still rages on whether different callosal regions undergo selective changes in each of these disease processes. This controversy may be put to rest by precise and reliable segmentation.

1.2.1 Manual delineation

Because of the difficulty to accurately and reliably[6] delineate structures in medical images, this task has traditionally been assigned to human operators.

However, given the improvements achieved over the past few years by imaging tools (commercial MR scanners now routinely resolve images at millimetric resolution, digital cameras can convert histological sections into million-pixel images) the manual segmentation phase has become a labor intensive and rather time-consuming process. In the case of an MRI, a trained operator typically has to go through more than a hundred 256 x 256 images for a $1mm^3$ resolution scan, slice by slice, to extract the contours of the target structures, one after the other. This manual editing is not only tedious but also particularly prone to errors, as assessed by various intra- or inter- operator variability studies [183, 182, 150, and others].

Manual editing thus suffers from many drawbacks:

- The results are often difficult if not impossible to reproduce. Even experienced operators display significant variability for difficult structures, such as the hippocampus.
- For 3-D delineation, editing tools usually display 3-D data in the form of 3 synchronized 2-D orthogonal views (sagittal, coronal and axial) onto which the operator draws the contour of the target structure. The output data therefore consists of a series of 2-D contours, from which a continuous 3-D surface has to be extracted: this is a non-trivial post-processing task, itself prone to errors. Moreover, since the operator has to mentally reconstruct the 3-D shape of the structure from a series of 2-D views, inter-slice inconsistencies and bumps

[6] In our case, reliability refers to the capacity to both outline a structure with minimal error and in a reproducible fashion, and to detect errors or identify situations where the delineation task cannot be completed (for instance, when the input image is corrupted by too much noise or artifacts).

are inevitable. More robust segmentation methods can usually be derived from true 3-D structure models in that they can ensure globally smoother and more coherent surfaces across slices.

- The operator is always biased by his *a priori* medical experience: some structures may exhibit very poor contrast in certain slices causing the operator to rely on his knowledge of the expected shape or smoothness of the structure at that particular slice to complete the delineation. Use of this medical knowledge, while highly desirable, becomes detrimental when it conflicts with a precise segmentation of the actual instance: the trade-off between medical experience and trust in the displayed information is difficult to determine and maintain through time with human operators.

Low-level segmentation techniques (region growing [32], edge detection [226], livewire [64], etc.) may be used to assist the human operator but could hardly replace manual delineation as they require substantial amounts of expert guidance. They usually consider only local neighborhoods without a higher order comprehension of the nature of the image, that is, they operate on dimensions computed from values attached to the voxels in the image without necessarily establishing a relationship with the reality that they represent. For instance, in edge detection, a segmented contour may very well not correspond at all to the boundary of the target structure and merely follow decoy structures or noise artifacts. Those problems are rooted into the inherently numerical nature of the data manipulated by these low-level model-free algorithms. Conversely, high-level model-based approaches operate on symbolic concepts associated to a representation of the reality extracted from the image. They are linked to the interpretation and the understanding of the input data and can overcome many of the limitations of manual slice editing[7].

1.2.2 Automated segmentation with *explicit* information

Most of the studies mentioned above rely on large databases of images and structure delineations, painstakingly collected in hospitals and research laboratories. The mere size of these databases, which is expected to increase exponentially, alone mandates the use of automated methods.

Similar needs for automated segmentation arise when building brain atlases (see [192] for a review), which can guide stereotaxic neurosurgical procedures and provide a precise quantitative framework for multi-modality brain mapping. Namely, these atlases are still rather tedious to build, as many components typically have to be interactively outlined.

However, to reach the desired level of accuracy, many difficulties have to be overcome : input images are noisy, not very well contrasted and full of decoys (many structures are similar in shape and/or in intensity), the target structures are fairly variable in shape and intensity, etc.

Several automated segmentation methods have been proposed in the literature to extract anatomical structures, using an array of feature descriptors and shape models. In view of the complex nature of the problem, deformable templates seem the weapon of choice:

[7] Incidentally, such distinction between the problem of contour modeling and that of edge extraction is characteristic of Marr's vision paradigm [139]: while modeling approaches usually require a training set of segmented contours as an input, low-level edged extraction is generally performed without reference to an *a priori* contour model. The interest of that dichotomy lies in its ability to decompose the image analysis problem into a number of independent and manageable tasks. Unfortunately, it also may result in a unidirectional and somewhat irreversible cascade of errors. Furthermore, due to image noise and the image projection process, local model-free edge extraction is an ill-posed problem with no unique solution [162].

- they can adequately handle discontinuities and irregularities;
- they are compact enough to describe a wide variety of shapes with a minimum number of parameters;
- they often provide an analytical model of the structure once segmented (at least locally); and
- *a priori* knowledge on the shape, location, or appearance of the target structure can be used to guide the deformation process.

Many deformable template techniques have been presented in the literature, where statistical analysis helps introduce *a priori* knowledge about the shape or the appearance of the target structures into the segmentation method.

Yet, most of these approaches fall in the *implicit knowledge* category: from a learning set of *a priori* segmented instances of an anatomical structure, they have to automatically discover the relationships and functional dependencies of the various parameters of the model.

However, *explicit* information about the target structures is often available, in the form of medical expertise. For instance, the relative positions of most of the deep grey nuclei is fairly constant across individuals, anatomical structures should not to intersect, etc. From these observations, a series of rules can be derived to better drive the segmentation process.

We submit that the use of this *a priori* medical expertise is the key to a robust and accurate segmentation system.

1.3 Composite segmentation system for medical images

We propose an automated segmentation system for medical images. We have focused on devising a methodology, built from the combination of a variety of techniques, where maximum use is made of the available medical expertise, either in the form of *implicit* knowledge or of *explicit* information. Within this system, we distinguish four different components, to each of which we have devoted a chapter in this manuscript.

Neural texture filtering: The core of our system consists of an automated segmentation method. We approach the issue of boundary finding as a process of fitting a series of deformable templates to the contours of the target structures. This process relies on the minimization of an objective function to drive the templates towards the strongest edges of the input image, which should then correspond to those of the target structure. Unfortunately, noise and decoys contribute to a poor edge map, which might impair the deformation. Interaction with neuro-scientists prompted us to consider texture as a discriminating element for the target structures. For each of them we devise in Chapter 2, from a learning set of already delineated instances in MRIs (or in histological sections), a specific texture filter which helps build a texture constraint to bias the evolution of the deformable templates towards the most texture probable boundaries.

> We first evaluate two classical non-parametric statistical approach to texture analysis: a linear classifier (linear discriminant analysis) and a linear classifier in a non-linear projection space (support vector machines). Their inability to

properly classify all of the selected target structures lead us to design a *learnt* non-linear technique: neural networks. Our neural classifier draws on anatomical experience about both the characteristic set of textures of each target structure and that of their immediately surrounding neighbours to better extract the target voxels. It consists of two stages. A hybrid master classifier / mixture of experts architecture is first used in conjunction with adaptive spline neurons. Together with two internal feature selection schemes and a dynamic learning set, it provides a first classification of the textures of the input images. The second stage is a multi-scale neural network which increases the classification performances by introducing spatial constraints. We apply this hybrid filter to the segmentation of anatomical structures from MR images and from histological sections.

Statistical shape modeling: The same learning sets of instances enables us to add shape constraints to the deformation process. Namely, even though a given structure can present a wide variety of forms, it seems that the notion of biological shape is reasonably well explained by a statistical description over a large population of instances. Chapter 3 introduces a new local shape descriptor, the observed transport descriptor, whose use enables a better reparameterization of the instances from the structure learning sets. Once reparameterized, we build statistical shape models with a classic principle component analysis approach. These models serve to constrain the evolution of the templates (within a hybrid global/local scheme) towards shapes that are more admissible with respect to the associated shape statistics.

Our reparameterization approach computes a dense correspondence field between two instances of a structure, where the characteristics of the field bear close resemblance to those in the *a priori* learning set. From the values of our shape descriptor (whose properties are particularly amenable to the matching problem) obtained at every point of the instances to be matched, we compute a distance matrix which embeds the correspondence problem in a highly expressive and redundant construct and facilitates its manipulation. We describe a non-learning technique and the two learning strategies that we have developed, which all rely on the distance matrix, and discuss their applications to the matching of a variety of 1-D, 2-D and 3-D structures.

Knowledge-driven segmentation: More explicit information is introduced in Chapter 4, where the target cerebral structures are simultaneously segmented. A number of rules, extracted from both the analysis of the dynamic behavior of our templates and medical experience, also drive the template deformation. Finally, non-linear registration of an hybrid MRI/structure atlas (built *a priori*) to the input MRI serves to initialize the templates.

This initial registration serves to initialize a series of simplex meshes (one mesh per target structure) and makes the approach more robust. Each simplex mesh is then iteratively modified to minimize a hybrid local/global energy which incorporates an internal regularization energy, an external term which couples the models to the underlying image features and a global shape-constrained term. Those meshes evolved in parallel within a rule-controlled framework whose purpose is to maximise the achieved match over each structure while respecting the distance, position, etc. constraints derived from medical expertise.

Piecewise affine registration: A first step for the construction of a hybrid anatomical atlas, image registration is addressed in Chapter 5, where we introduce a new similarity measure (the constrained correlation coefficient) and a novel 2-D registration paradigm (piecewise affine), here also with a view to use *a priori* information to guide the registration process and yield better performances.

> The new paradigm is concerned with the registration of 2-D biological images (histological sections or autoradiographs) to 2-D images from the same or different modalities (*e.g.*, histology or MRI). The process of acquiring these images typically induces composite transformations that we model as a number of rigid or affine local transformations embedded in an elastic one. We propose a registration approach closely derived from this model, which also extends to the registration of 3-D images (MRI for instance).
>
> Given a pair of input images, we first compute a dense similarity field between them with a block matching algorithm. A hierarchical clustering algorithm then automatically partitions this field into a number of classes from which we extract independent pairs of sub-images. Our clustering algorithm relies on the Earth mover's distribution metric and is additionally guided by robust least-square estimation of the transformations associated with each cluster. Finally, the pairs of sub-images are, independently, affinely registered and a hybrid affine/non-linear interpolation scheme is used to compose the output registered image.

Figure 1.10 summarizes the various techniques developed during this thesis, and their relationships within our system.

1.4 Summary of contributions

> We are generally the better persuaded by the reasons
> we discover ourselves than by those given to us by others.
> Blaise Pascal.

We summarize here the various contributions of our work. We report both the publications directly linked to the central theme of this thesis and the additional applications developed within the framework of the associated team program[8].

Neural texture filtering. A hybrid statistical/neural classifier, closely related to a 2-D version of the neural classifier detailed in Chapter 2 was presented orally at the WCCI-IJCNN '02 conference, in the context of the segmentation of anatomical structures in 2D slices of MR data:

> *A. Pitiot*, A. Toga, N. Ayache, and P.M. Thompson, *"Texture based MRI segmentation with a two-stage hybrid neural classifier"*, proceedings of the World Congress on Computational Intelligence / INNS-IEEE International Joint Conference on Neural Networks (WCCI-IJCNN'02), 2002.

[8]These side applications will be commented all through the manuscript, wherever we describe the associated methods.

Its application to the segmentation of mouse brain tissues in histological sections was part of an oral presentation at HBM '02:

> A. Mackenzie-Graham, E. Lee, I.D. Dinov, *A. Pitiot*, G. Hu, M. Bota, Y. Ding, L. Capetillo-Cunlife, K. Crawford, B. Truong and A.W. Toga, *"Atlas of the C57BL/6 mouse brain: A multimodal, multidimensional approach"*, proceedings of Human Brain Mapping (HBM'02), 2002.

and in a full paper in Journal of Anatomy:

> A. Mackenzie-Graham, E. Lee, I.D. Dinov, *A. Pitiot*, G. Hu, M. Bota, Y. Ding, L. Capetillo-Cunlife, K. Crawford, B. Truong and A.W. Toga, *"Atlas of the C57BL/6 mouse brain: A multimodal, multidimensional approach"*, Journal of Anatomy, 2003.

Statistical shape modeling. The observed transport descriptor and the learning approach to reparameterization developed in Chapter 3 were presented orally at the IPMI '03 conference:

> *A. Pitiot*, H. Delingette, A.W. Toga and P.M. Thompson, *"Learning Object Correspondences with the Observed Transport Shape Measure"*, proceedings of Information Processing in Medical Imaging (IPMI'03), 2003.

The application of our descriptor to the computation of a mean cortical profile in the framework of the analysis of the myelo-architecture of the human cerebral cortex was part of the methodology we fully devised, implemented and reported in a NeuroImage journal paper, whose neuro-anatomical content was managed by J. Annese from the LONI laboratory:

> J. Annese, *A. Pitiot*, I.D. Dinov, and A.W. Toga, *"A myelo-architectonic method for the structural classification of cortical areas"*, to appear in NeuroImage, 2003.

Two early versions and an extension of this latter work were presented as posters respectively at HBM'01, HBM'02 and HBM'03:

> J. Annese, *A. Pitiot*, and A.W. Toga, *"Complex topological analysis of the human striate cortex (V1)"*, proceedings of Human Brain Mapping (HBM'01), 2001.

> J Annese, *A. Pitiot*, and A. Toga, *"3D cortical thickness maps from histological volumes"*, proceedings of Human Brain Mapping (HBM'02), 2002.

> J. Annese, *A. Pitiot*, I.D. Dinov, and A.W. Toga, *"A myelo-architectonic method for the structural classification of cortical areas"*, proceedings of Human Brain Mapping (HBM'03), 2003.

Knowledge driven segmentation. The rule-based segmentation system described in Chapter 4 was presented as a poster at MICCAI'03:

> *A. Pitiot*, H. Delingette, N. Ayache, and P.M. Thompson, *"Expert-Knowledge-Guided Segmentation System for Brain MRI"*, proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI'03), 2003.

A 2-D version of the hybrid evolutionary programming framework within which we have tested our shape constraints was detailed, along with a complete 2-D segmentation system for anatomical structures, in a TMI journal paper:

> *A. Pitiot*, A.W. Toga and P.M. Thompson, *"Adaptive Elastic Segmentation of Brain MRI via Shape Model Guided Evolutionary Programming"*, IEEE Transactions on Medical Imaging, 21(8):910-923, 2002.

Piecewise affine registration. The piecewise affine registration methodology for biological images detailed in Chapter 5 was presented orally at WBIR'03:

> *A. Pitiot*, G. Malandain, E. Bardinet, and P. Thompson, *"Piecewise Affine Registration of Biological Images"*, proceedings of the Second International Workshop on Biomedical Image Registration (WBIR'03), 2003.

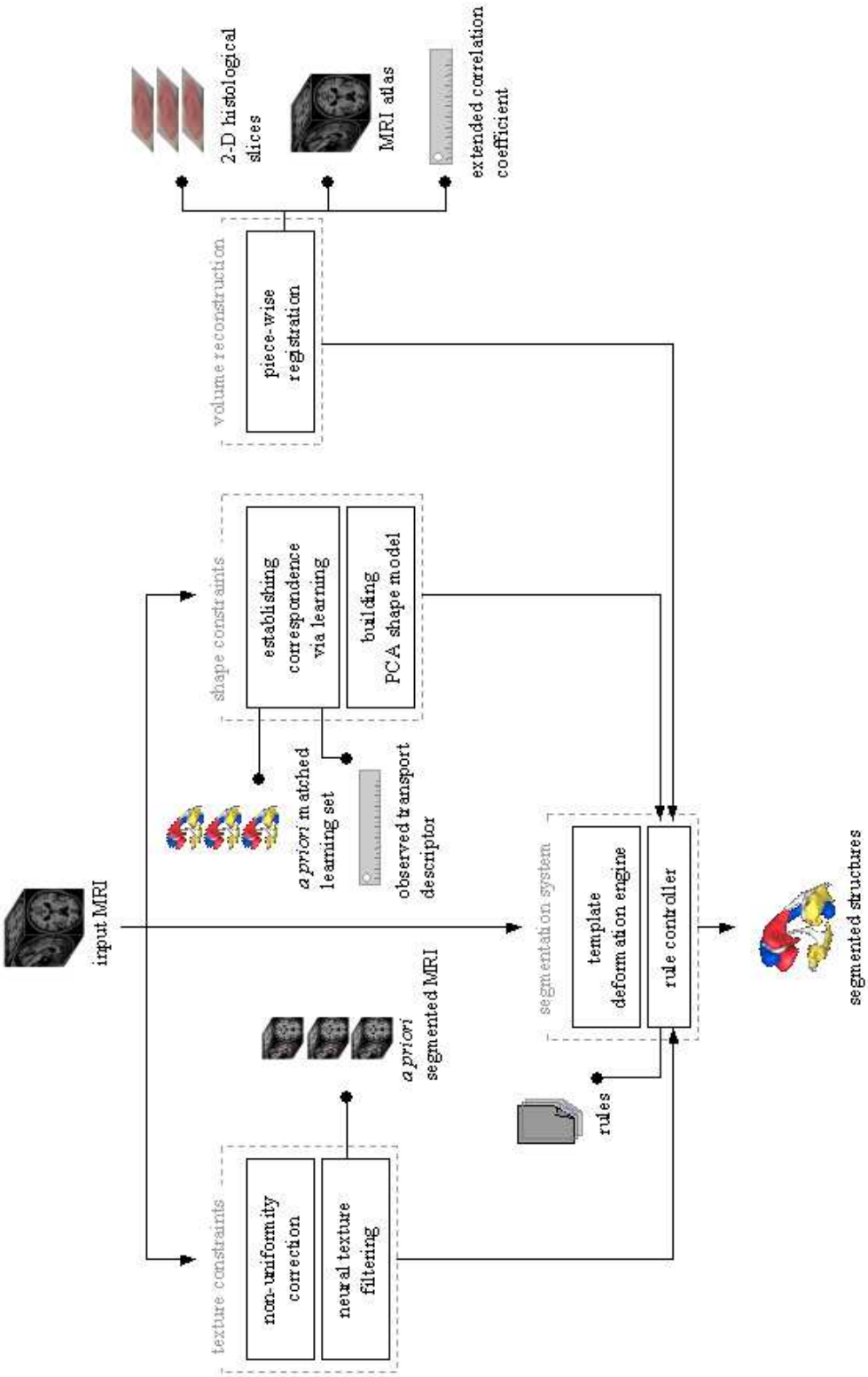


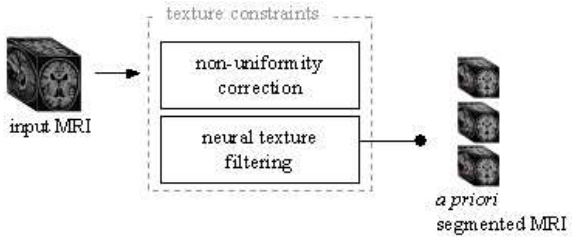
Figure 1.10: General overview of the segmentation system.

Chapter 2

Neural Texture Filtering
Filtrage de Texture

There is nothing worse than a sharp image of a fuzzy concept.
Ansel Adams.

Ce chapitre introduit les deux principales modalités sur lesquelles nous nous sommes concentrés: l'imagerie par résonance magnétique (IRM) et l'histologie. Une analyse de leurs characteristiques nous amène à considérer la notion de texture comme élément discriminant pour les structures cibles. A partir d'une classification texturale de régions d'intérêt définies dans l'image en entrée, nous dériverons au chapitre 4 des contraintes de texture, lesquelles iront biaiser la déformation des patrons en direction des contours les plus probables au sens de la texture. Nous décrivons trois classificateurs de texture: deux approches classiques sous la forme d'une analyse linéaire discriminante et de support vector machines linéaires, et un classificateur non-linéaire sous la forme d'une architecture neuronale originale, soit une progression à la fois en terme de complexité (depuis linéaire et linéaire dans un espace projectif non-linéaire vers non-linéaire avec apprentissage) et de performances. Notre approche neuronale consiste en deux étapes de classification, avec une deuxième étapes dont la tache est de corriger les erreurs de la première et d'incorporer de l'information spatiale. Elle s'appuie par ailleurs sur une phase d'apprentissage dynamique pour traiter le problème de représentation de l'ensemble d'apprentissage de textures. C'est par l'intermédiaire de ce dernier que nous introduisons de l'information a priori *sous forme semi-implicite.*



Texture constraint.

Our deformable template approach, like most of those presented in the literature, relies on the minimization of an objective function to segment the boundaries of the target structures in the input image. At a glance, this function (which will be detailed later in Chapter 4) is a compound energy made up of two terms: an *internal* energy which characterizes the possible template deformations and an *external* energy (an image energy) which couples the templates to the underlying image. Classically, this coupling energy is based on a measure of intensity gradient and leads the templates towards the strongest edges of the input image. Unfortunately, the nature of the imaging modalities used (MRI or histology in our case) is such that gradient information alone is often not sufficient to ensure a reliable final segmentation: noise, imaging artifacts or inhomogeneities all contribute to a poor gradient map. Consequently, we need to extract from the textural characteristics of the input images some deformation constraints that will drive the moving templates towards more probable boundaries.

From intensity homogenization to texture based classification, we detail in this chapter the chain of pre-treatments we have devised to cope with this classification task.

2.1 Introduction

Cerebral structures are not equally well defined in brain imaging modalities. The corpus callosum for instance (see Figure 2.1) is remarkably smooth and contrasted in the mid-sagittal section of a typical T1-weighted MRI. Its mean intensity is clearly larger than that of most of the immediately surrounding tissues. Similarly, the intensity distribution of the lateral ventricles (Figure 2.2) clearly distinguishes them from nearby structures (the caudate nucleus for instance). However, the borders of the hippocampus (Figure 2.13) are significantly harder to outline in some areas.

Furthermore, most of these structures present fuzzy boundaries. This, together with partial volume effects, explains the bi-modal nature of the corpus callosum intensity distribution (Figure 2.1(b)) : in addition to a clear peak with small variance centered around $I = 200$, we find another mode with lower intensity values and a much larger dispersion. The somewhat long tails of the intensity distribution of both the caudate nuclei and the ventricles in Figure 2.2 is another illustration of these combined phenomena.

Additionally, these structures are often surrounded by decoy elements with similar intensity distribution. In Figure 2.1 for instance, the intensity distribution of the fornix (in yellow) substantially overlaps that of the corpus callosum (in red).

Finally, noise, partial volume effects and bias fields also deteriorate the quality of the input images.

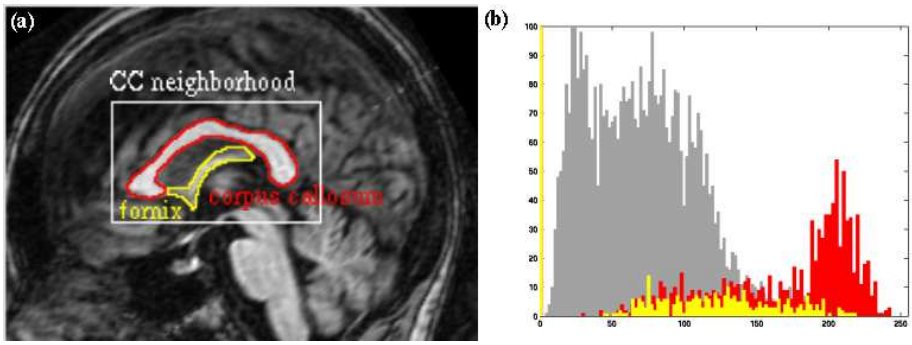


Figure 2.1: The corpus callosum and its neighbours in a T1-weighted MRI section (a) and their associated intensity distributions (b).

Histological sections are not exempt from problems either (see Figure 2.4). Staining density may vary across the image, insufficiently controlled lighting conditions may create intensity inhomogeneities in the images, the glass slide onto which the stained tissue is mounted may not be perfectly clean (dust particles or fingerprints may show in the photographs), etc.

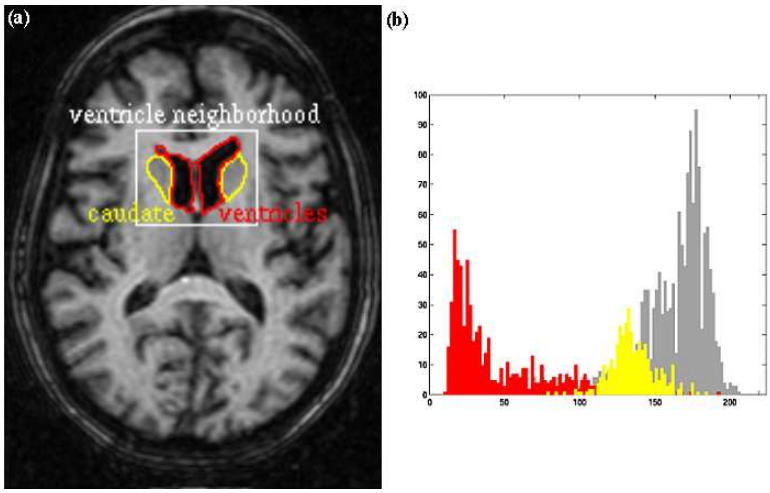


Figure 2.2: The ventricles and caudate nuclei in a T1-weighted MRI section (a) and their associated intensity distributions (b).

In fine, the combination of these various difficulties can substantially hamper the segmentation process. Rather than complicating the latter to make it more robust, a careful analysis of the properties of the former helped us derive a number of filters whose task is to extract, from the input image, the voxels that form the target structure. These target voxels will then be used as a constraint imposed on the deformable templates of Chapter 4 to bias the deformations towards the "texture-probable" boundaries of the target structures.

After a brief review of the principles of MRI and histology and of their inherent potential defects (Section 2.2), we introduce the use of texture (Section 2.3) as a means to characterize the target structures before detailing the various supervised texture classification algorithms we have devised (Section 2.4).

2.2 Acquisition modalities

We present here the two most commonly used modalities for *in vivo* and *post mortem* analysis: magnetic resonance imaging and histology, respectively.[1]

2.2.1 Magnetic Resonance Imaging (MRI)

Since the independent discovery in 1946 by Felix Bloch and Edward Purcell of the interaction of atom nuclei and radio waves, and the work of Dr. Damidian and his team to build the first scanners to exploit the abundance of water in the human body to record a radio signal, Nuclear Magnetic Resonance (NMR), later to become Magnetic Resonance Imaging, has gained momentum to become the favoured tool in the non-invasive exploration and observation of the human brain.

2.2.1.1 The physics of MRI

MRI is an extension of Nuclear Magnetic Resonance (see [27] for an in-depth description of MRI). In biomedical imaging applications, this resonance most preeminently takes place in the nuclei of the hydrogen atoms in water, given its high concentration of protons and dynamical properties.

Protons exhibit a spin, that is, they rotate about their axis. In conjunction with their charge, this gives them the magnetic property of a dipodal magnet. In conjunction with their mass, they acquire an angular momentum. These two properties are responsible for the precession of protons when placed in a magnetic field. A proton's precessional frequency (the Larmor frequency) is proportional to the strength of this magnetic field. The scaling factor between the Larmor frequency and the magnetic field is called the gyromagnetic ratio.

In a tissue that contains a large number of nuclei, the Brownian motion of protons does not yield any net magnetization. However, when a homogeneous magnetic field B_0 is applied, the magnetic moments line up with the field and yield a net magnetization, parallel and proportional to this magnetic field. If we apply another magnetic field B_1 tuned to the Larmor frequency, and perpendicular to B_0, the protons become excited which induces the precession of the net magnetization vector about the direction of B_0. When B_1 is stopped, the net magnetization decays to its equilibrium alignment (so-called de-phasing mechanism) which causes the emission of a radio-frequency (RF) signal at the Larmor frequency. A coil of wire, connected to a sensitive amplifier and tuned to that Larmor frequency can record this signal.

The time constant for this decay is referred to as $T2^*$. This constant however depends on both fixed and time varying field changes. While fixed inhomogeneities in the magnetic field are uninteresting for MRI, time varying field changes are associated with the characteristics of the tissues being imaged. Using a carefully crafted sequence of excitation pulses, the decay time ($T2$) of the sole time varying changes (spin-spin relaxation mechanism) can be recorded. A so-called spin-lattice time constant ($T1$) can also be recorded if we wait a little longer for the distribution of the energy deposited by the RF excitation throughout the neighborhood.

However, recording this NMR signal is not sufficient to obtain an image: the signal must be associated with spatial information. 2-D slice selection is achieved by super-imposing a field

[1] Note that this chapter focuses mostly on the texture filtering of MR images. The additional application of the developed filters to histological data was driven by one of the projects conducted at the LONI laboratory within the "associated team" framework. It was also the opportunity to assess the generalizability of our texture approach.

gradient over B_0. A complex sequence of pulses and use of Fourier transforms then enables the extraction of a 2-D position from this 2-D slice.

2.2.1.2 MRI defects

In spite of its evident qualities, MR imaging certainly is not a panacea and suffers from many shortcomings. While some of them are linked to the nature of the image acquisition procedure (patient motions, variable imaging parameters), others stem from the underlying physical phenomena (chemical shift, bias fields) or are typical of sampled data (sampling and quantization effects, spatial aliasing, partial volume effect, ...). However, they do not affect images to the same extent. "Aliasing" for instance is easily detectable as affected images present wrap-arounds of brain parts on to others and can thus be visually (manually) discarded. On the other hand, RF inhomogeneities are not only a lot less recognizable but also affect most if not all MR acquisitions, and must then be corrected.

We briefly present here a selection of the most representative artifacts and how we deal with them (see also Figure 2.3 for a few illustrations).

2.2.1.2.1 variable imaging parameters: MR images of the same patient obtained from different scanners, or even from the same machine at different times, and with different sets of parameters yield different absolute intensities, contrasts and noise levels for tissues and structures, even within the same "modality" (T1 or T2).

The various supervised classifiers we have devised consequently require a learning set of labeled images whose characteristics must bear reasonable resemblance to those we want to process. Their sensitivity to MRI imaging characteristics is studied in Section 2.4.3.4.

2.2.1.2.2 motion noise: Motion artifacts can be either periodical (due to blood or CSF flows, cardiac beats, ordinary breath) or random (patient head moves). Although faster scanners tend to reduce their overall impact, they can at time significantly alter the quality of the produced images. A number of methods have been devised to take them into account using a variety of movement models (see [7] and [227] for a review): constant speed linear moves, periodical moves along *a priori* set axis ... A careful review of the database and test MR images did not provide evidences that such careful modelings were required for our applications. Global head motions were corrected using a robust rigid registration algorithm [156].

2.2.1.2.3 physical and sampling artifacts: The complexity and sensitivity of the many physical phenomena and production/reception components that interact to produce an MR image are the sources of an equally large number of "physics" based artifacts. From chemical shift and aliasing to Moiré fringes or RF overflow to name just a few (see [157] for an extensive review), these all contribute to various degrees to the deterioration of the final image. The induced noises are also intimately linked to the chosen imaging parameters and the characteristics of the scanner: the stronger the magnetic field, the better the contrast for instance. Again, only the main artifacts are introduced here.

quantization: An MR image consists of a set of 3-D voxels organized within a parallelepiped, with an intensity measure associated to each voxel. Those measures are coded on a given number of bits (usually 16 in our case, but sometimes only 8). However, not all images exploit the full dynamics allowed by those bits, which results in a loss of radiometric precision. Expanding the dynamic range of images acquired on less than 16 bits is not a trivial

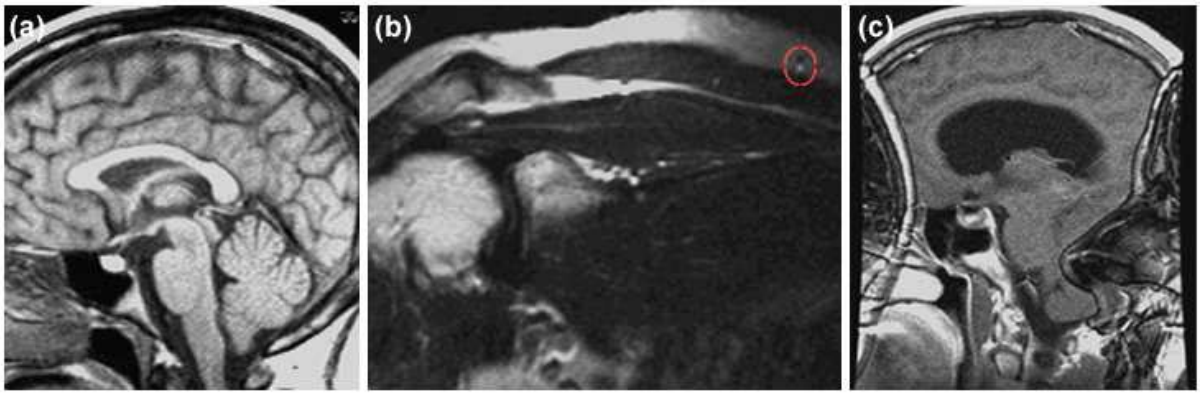


Figure 2.3: A selection of MRI artifacts: (a) patient head movement towards the end of an acquisition; (b) chemical shift: while the fluid in the vessel (white spot in the red circle) was mapped correctly, the surrounding fat has artificially been moved upward; (c) phase wrap (aliasing).

issue as many quantization algorithms can be used: one could simply apply a linear transform to the original image histogram to scale its dynamic range, or a variable-bin strategy could be favored. We chose here to use a simple linear quantization algorithm. Empirical experiments demonstrated that finer strategies did not impact the overall performances of our texture classifiers.

chemical shift: As mentioned above, the spatial position of protons is extracted from the frequency at which they precess about the B_0 axis. However, this frequency varies depending on the characteristics of the scanned material: for instance, fat protons precess at a frequency slightly lower than water protons (73 Hz slower, when a 0.5T magnetic field is applied). Yet, during the spatial extraction phase, the MR scanner assumes that all protons have the same Larmor frequency, which causes fat protons to be mapped to the wrong position. Use of particular pulse sequences (Carr-Purcell-Meiboom-Gill) can alleviate this issue.

partial volume effect: These effects appear when more than one type of tissue occupies the same voxel in an MRI [128]. The resulting intensity is then a mixture of the values associated with these tissues. Given the complexity of the human brain, and the resolution of standard MR scanners (1 mm^3 on average), partial volume voxels may account for a large percentage of the total number of voxels of a given structures (see [80] for quantitative evaluations). Our learning approach to texture classification does take these effects into account.

RF bias field: The inhomogeneities of the B_0 magnetic field, the presence of a recording coil, etc. create biases during the acquisition process and alter the recorded intensities in an MR image. The removal of these intensity non-uniformities proves to be an essential prerequisite for the quantitative analysis of MR data. Clearly, spatial variations of the image signal unrelated to anatomic information yield spurious changes of contrasts and may often significantly reduce the accuracy of subsequent computational procedures. A number of techniques have been discussed in the literature [187, 25, 38]. Following the results of the comparative study reported by Arnold *et al.* [5], we chose the N3 method [187] to treat

our input MR data. It eliminates the dependence of the field estimate on anatomy with an iterative approach which estimates both the multiplicative bias field and the distribution of the true tissue intensities. It essentially consists of a nonparametric nonuniform intensity normalization (N3), and is reported as independent from pulse sequence and insensitive to pathological data that might otherwise violate model assumptions.

2.2.2 Histology

2.2.2.1 Techniques of histology

One of the many microscopic techniques used to study in detail the characteristics of tissues, histology encompasses a series of processes that together make a microscopic glass slide from a slab of tissue, extracted from the body. It involves fixation, tissue processing, sectioning, staining, and, in our digital case, photography.

Fixation: The first step of the histological process, fixation is used to preserve tissues permanently, in a state as life-like as possible. It is consequently carried out as soon as possible after tissue extraction to avoid nuclear shrinkage and loss of cellular organelles. A variety of fixatives are being used, depending on the type of the studied tissue or on the objectives of the histological analysis. According to their mechanism of action, fixatives are organized as five families:

- Aldehydes (formaldehyde in particular) fix tissue by forming cross-linkages in the proteins. As cross-linkages only minimally alter the structure of protein, they preserve the antigenicity. Although a slow process, formaldehyde (formalin) penetration is very good.
- Among the mercurial fixatives, we find B-5, and Zenker's. Their fixation mechanism is still unknown. They are fast, provide good nuclear details but penetrate poorly.
- Alcohols (methanol, ethanol) are very quick fixative and give good nuclear detail. However, they are protein denaturants and may cause brittleness and hardness.
- Oxidizing agents (potassium permanganate, potassium dichromate, osmium tetroxide) too fix by proteinic cross-linkage. Yet, they cause severe denaturation of tissues and are thus not commonly used.
- Picrates (Bouin's solution, for instance) all contain picric acid, whose mechanism of action is still unknown. They give nuclear details in the range of those obtained with mercurial solution with a reduced hardness.

Fixation is a delicate process, whose quality depends on a number of factors: pH should be controlled (in the range 6 to 8) to prevent excessive acidity, penetration must be thorough (and can be enhanced by cutting thinner slices), a 10:1 ratio of fixative volume to tissue must be maintained, high temperatures enhance fixation but can cook the tissues, etc.

Tissue Processing: This technique embeds the fixed tissue in a material with similar density (often paraffin, even if plastics can sometimes be used) in order to facilitate its manipulation and cutting into thin microscopic sections. It consists of two steps: dehydration and clearing.

- Dehydration is necessary to remove water from the wet fixed tissues which could not otherwise be infiltrated with paraffin. This operation is usually conducted through a series of alcoholic baths.
- The clearing process removes the dehydrant with a chemical miscible with paraffin, usually xylene.

Paraffin embedded tissues are then snap frozen in a cold liquid (-20^oC to -70^oC) to turn them into a solid block and facilitate the cutting process.

Sectioning: Once the tissues processed, a microtome (basically a very sharp knife with a control mechanism to advance the paraffin embedded block across the knife with micrometric precision) is used to cut the sections, which are then floated on a warm water bath and unfolded on a microscopic slide with a brush. Slides are then warmed up to help the tissue adhere (glue-coated slides can also be used when heat is to be avoided).

De-embedding: Once the tissues cut and adhered to a glass slide, their embedding paraffin must be removed to allow the penetration of water soluble dyes. This is achieved by running the sections through xylene to alcohols to water.

Staining: Depending on the objectives of the histological study, various dyes can be chosen, as each dye stains a particular cellular components. In our case, myelin and Nissl stains were used.

- Myelin stains color the myelin sheath of nerve fibers, and serve to recover the fibre bundle pathways. The most common techniques are Weil method, Weigert or Marchi myelin stains.
- Nissl stains use cresyl violet to reveal cell bodies and highlight nuclear structures.

Note that the staining timing must be tightly controlled as the obtained tint (and contrast) depends on the time the section spent in contact with the dye.

Coverslipping: The stained section on the slide is then covered with a thin piece glass to protect the tissue, provide better optical quality for viewing under the microscope, and ensure long-term preservation. The stained slide is then taken through a series of alcohol solutions to remove the water, then through clearing agents. Finally, a permanent resinous substance is placed over the section.

Image acquisition: The protected slides are then placed under a microscope and a digital picture is taken, at an operator specified magnification and under controlled lightening conditions. From the care with which the entire imaging system was set up depends the quality of the final picture. Choosing the correct parameters is a matter of compromises, which vary with the goal of the histological studies: for instance, increasing the gain of the digital camera results in a brighter image but increases the noise, color interpolation is necessary to compensate the color masking techniques used in digital CCDs, the depth of field is constrained by the chosen magnification, etc.

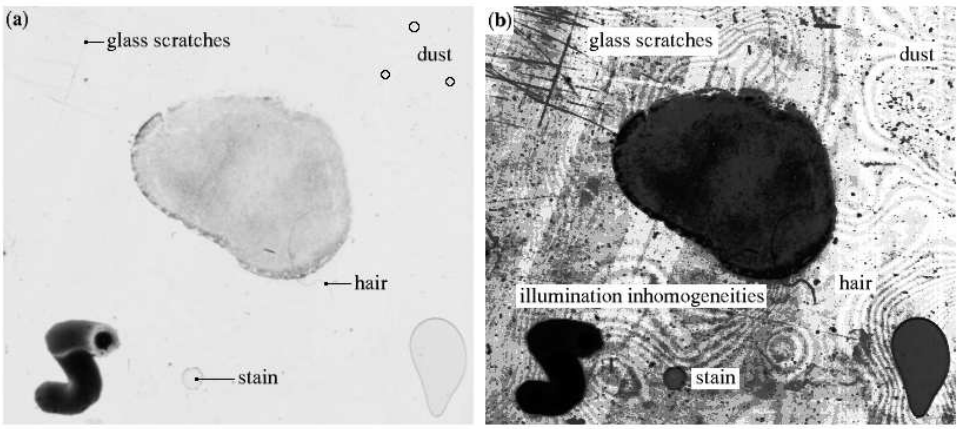


Figure 2.4: Typical artifacts in a myelin-stained 50 micron section of the human brain: (a) original image with highlighted scratches, dust and stains; (b) after intensity equalization to stress the intensity inhomogeneities.

2.2.2.2 Histological artifacts

Improper fixation (lengthy fixation time, wrong type of fixative), poor dehydration or sectioning difficulties result in artifacts in the final image. While particular care during the histological preparation may be sufficient to avoid some of them (the fixation issues for instance), others (tearing problems and "floaters") will inevitably make their way to the microscopic images and must be tackled *a posteriori*. We will detail in Chapter 5 the techniques we have developed to deal with some of these problems.

Fixation difficulties: Improper buffer concentration or fixation time may result in the formation of fine black precipitates on the slides, with little or no relation with the actual underlying tissue. These are formalin-heme pigments, which tend to appear when the tissue becomes acidic as the formalin buffer comes to exhaustion. They are especially likely in thicker slices of very cellular or bloody tissues.

Sectioning issues: Insufficient dehydration prior to clearing or paraffin wax infiltration make tissues hard to section on the microtome. This yields tearing artifacts and holes in the sections.

Use of alcohols such as ethanol, although excellent fixatives for cytologic smears, tend to make tissue sections brittle, which also result in microtome sectioning artifacts with chattering and a "venetian blind" appearance.

Too warm or blunt a blade will also tear apart delicate tissues.

Mounting problems: Air might be sucked under the coverslip of the mounting media, as it dries, and create bubbles if it is too thin. Bubbles can also result from contamination of the clearing agents or of the coverslipping media.

Dust particles or hair can also appear under the microscope when the mounting media was not handle with sufficient care, or exposed unprotected.

Pieces of tissue ("floaters") detached during the cutting process, whether they belonged to disjoint anatomical components or were torn apart by the blade, are re-assembled on the

glass slide, by the operator. Their respective positions and orientations in the final image may then not perfectly reflect the original anatomical situation.

Staining difficulties: As the obtained tint depends on the time the sections spend in contact with the dyes, differences in contrast can be observed between different histological batches. Additionally, the dye might penetrate deeper in certain parts of the tissue and more superficially elsewhere, which creates artificial changes of contrast. Histogram control techniques ([134] for instance) can be used to homogenize the contrast between consecutive slices.

Photographic issues: Poorly controlled lightening conditions during the acquisition of microscopic pictures create intensity inhomogeneities which may alter the contrast of the observed tissue. Here also, intensity homogenization and histogram control algorithms can alleviate this issue.

2.3 Texture analysis

Even though intensity homogenization can remove some of the artifacts inherent to the physics of the acquisition modalities, many of the difficulties detailed in Section 2.2 remain: the probability densities of the target structures still overlap with those of the surrounding tissues, the target edges are fuzzy, noise is still present etc. Consequently, we need a better descriptor than the sole intensity gradient measure to characterize the targets. Interaction with neuro-scientists prompted us to consider texture as a discriminating element.

Despite the fact that it is commonly referred to in the computer vision literature, no common definition of the concept of "texture" has yet emerged, as exemplified by the long list of definitions compiled by Coggins [35]. Among others, we find:

- "A region in an image has a constant texture if a set of local statistics or other local properties of the picture function are constant, slowly varying, or approximately periodic." [186]

- "Texture is defined for our purposes as an attribute of a field having no components that appear enumerable. The phase relations between the components are thus not apparent. Nor should the field contain an obvious gradient. The intent of this definition is to direct attention of the observer to the global properties of the display *i.e.*, its overall coarseness, bumpiness, or fineness. Physically, nonenumerable (aperiodic) patterns are generated by stochastic as opposed to deterministic processes. Perceptually, however, the set of all patterns without obvious enumerable components will include many deterministic (and even periodic) textures." [170]

- "Texture is an apparently paradoxical notion. On the one hand, it is commonly used in the early processing of visual information, especially for practical classification purposes. On the other hand, no one has succeeded in producing a commonly accepted definition of texture. The resolution of this paradox, we feel, will depend on a richer, more developed model for early visual information processing, a central aspect of which will be representational systems at many different levels of abstraction. These levels will most probably include actual intensities at the bottom and will progress through edge and orientation descriptors to surface, and perhaps volumetric descriptors. Given these multi-level structures, it seems

clear that they should be included in the definition of, and in the computation of, texture descriptors." [230]

- "The notion of texture appears to depend upon three ingredients: (i) some local order is repeated over a region which is large in comparison to the orders size, (ii) the order consists in the nonrandom arrangement of elementary parts, and (iii) the parts are roughly uniform entities having approximately the same dimensions everywhere within the textured region." [92]

In short, it can be described as a function of the spatial variation of voxel intensities (see [33] for an in-depth review of texture analysis).

Here, we are concerned with *texture classification*. That is to say, we want to partition the input image, or more precisely a region of interest defined in the input image, in order to single out the group of voxels which belong to the target structure in a unique separate class. In Figure 2.1 for instance, even though both the fornix and the corpus callosum consist of white matter, a different orientation of the fiber yields two different textures (despite their overlapping intensity distributions). Similarly, even though the mean intensity of the hippocampus (Figure 2.13) is close to that of its neighbour, it exhibits a characteristic set of textures.

A number of texture descriptors have been discussed in the literature (see [200] and [33] for a review). These make use of statistical measures, time-space-frequency decompositions or fractal parameters for instance. As every descriptor has weaknesses and strengths, one usually combines them to improve the classification performances.

We present in the following Section 2.3.1 a few commonly used texture descriptors before detailing a means to select the most promising ones (Section 2.3.2) which are then combined in a knowledge-driven texture classifier (Section 2.4).

2.3.1 Texture descriptors

Since modelling the physical phenomena that produced variations of intensity in the acquired image is a difficult task[2], texture is more easily characterized by the 2-D (3-D) variations of intensity in the image. It should be noted that texture is only meaningful as a characteristics of areas (volumes): it is a contextual property which must be defined on a spatial neighborhood (whose size and shape depends upon the texture type), not just a single point. Texture then involves the spatial distribution of gray values over this neighborhood. From a physiological point of view, texture has a number of perceived qualities. For instance, Laws [123] singles out: uniformity, density, coarseness, roughness, regularity, linearity, directionality, direction, frequency, and phase. In spite of the functional dependencies between these qualities, they still define a high dimensional space, whose dimensionality probably explains why there is not a unique way to characterize texture which works well for every texture.

2.3.1.1 Statistical Methods

These methods rely on a definition of texture as a spatial distribution of gray values.

[2] Difficult, but not impossible as many teams are developing MR simulators (see [122] for instance).

2.3.1.1.1 Co-occurence matrices and Haralick descriptors. Co-occurence matrices embed the second order spatial statistics of an image region in a double-entry integer matrix.

Let $I : \left| \begin{array}{ccc} \Omega & \rightarrow & \mathbb{N} \\ s & \mapsto & I[s] \end{array} \right.$ be a pictorial region defined in a n-dimensional discrete gray-scale image. Given a translation $u \in \mathbb{R}^n$, the co-occurrence matrix CM_u of region I is a $g \times g$ matrix where g is the number of gray levels in which I is quantified (in practice, only a limited number of gray levels is used: 16 in our case). It is written as follows:

$$\forall (a,b),\ CM_u(a,b) = \text{card}\,\{(s, s+u) \in \mathbb{R}^n \,|I[s] = a, I[s+u] = b\} \tag{2.1}$$

where (a, b) is a pair of grayscale values.

$CM_u(a, b)$ is then the number of pairs of sites $(s, s + u)$ in I such that s has a gray level a, and $s + u$ has a gray level b, that is, the joint probability of occurrence of gray level a and b for two sites in I with the spatial relationship defined by the translation u.

While these matrices are particularly instructive about the textural content of a region, they are difficult to handle as they contain a lot of values ($g \times g$ integers) among which we find a large number of zeros. Haralick suggested in [91] to summarize the information content of the matrices in a few real valued descriptors. Among the most commonly used ones, we find (in their normalized form):

- energy: $e_u(I) = \frac{1}{T(u)} \sum_a \sum_b CM_u(a,b)^2$, where $T(u)$ is the total number of pairs $(s, s+u)$ in I. It is a measure of both the grayscale homogeneity and the periodicity with respect to u.
- contrast: $c_u(I) = \frac{1}{T(u)(g-1)^2} \sum_{k=0}^{g-1} k^2 \sum_{|a-b|=k} CM_u(a,b)$. The contrast is all the more high when the difference in gray value between two u-related site is large.
- entropy: $t_u(I) = 1 - \frac{1}{T(u) \ln T(u)} \sum_a \sum_b CM_u(a,b).\ln CM_u(a,b)$. It is an indicator of the disorder exhibited by a texture: it is low when the same pair of grayscale values is repeated multiple times, and high when most pairs are scarcely represented in the matrix.
- correlation: $o_u(I) = \frac{1}{T(u)\sigma_x\sigma_y} |\sum_a \sum_b (a - \mu_x)(b - \mu_y) CM_u(a,b)|$. It measures the correlation between the lines and the columns of the co-occurrence matrix.

2.3.1.1.2 Autocorrelation features. The autocorrelation function relies on the repetitive nature of the placement of texture primitives and on the characteristic size of these primitives to measure the amount of regularity of the texture and its coarseness, respectively. It basically assesses the spatial linear relationship between the texture primitives. It is written:

$$\forall s \in \Omega,\ \rho(s) = \frac{\sum_{u \in \Omega,\, s.t.\, s+u \in \Omega} I(u) I(s+u)}{\sum_{u \in \Omega} I(u)^2} \tag{2.2}$$

2.3.1.2 Structural methods

Structural models assume that a given texture is made up of a number of texture primitives, organized according to a series of placement rules. The textural analysis then consists of two steps: texture primitive extraction and placement rule inference. A variety of techniques have been proposed in the literature to extract texture primitives (where a primitive is usually defined

as a region with uniform gray levels) images. Voorhees *et al.* [211] for instance, suggest to first filter the textured region with Laplacian of Gaussian (LoG) masks, at different scales before combining the extracted information to extract the image "blobs", which, they argued, are pivotal to the perception of texture. Blostein *et al.* [18] combine a similar LoG filtering surface shape computation to improve the overall performances. Tomita *et al.* [199] first used a medial axis transform on the connected components of a segmented image and then compute the mean intensity and shapes of the extracted primitives. In [229], Zucker considers actual textures as distorted versions of ideal textures. The graph used to model the placement of the ideal texture primitives is then transformed to generate the actual texture. A tree grammar is proposed by Fu [71] to model the placement rules. Textures are then described by the string generated by the production rule of the grammar, and the vocabulary defined by the texture primitives.

2.3.1.3 Model based methods

These approaches summarize the essential perceived characteristics of a texture into an image model, which can be used both for the analysis (the parameters of the model then serve as a texture descriptor) and the synthesis of the considered texture.

Markov random fields. A favored means to model images, markov random fields (MRF) are particularly well-suited to capture the local textural characteristics of images as they provide a consistent way to model context-dependent entities such as image pixels and correlated features. They rely on the assumption that the intensity of any given voxel is solely determined by that of its neighbours. Mutual influences between pixels and textures are then characterized using conditional MRF distributions.

We first need to define a neighborhood system to impose on the voxels of the input image. First-order (four-connected in2-D) or second-order (eight-connected in 2-D) neighborhoods can be used. Within these neighborhoods, cliques (subset of sites in which every pair of distinct sites are neighbors) are often used to define the conditional MRF distributions.

More precisely, given a neighborhood system $N = \{N_i, i \in L\}$ defined on a given lattice L, with N_i the set of sites neighbouring i,($i \notin N_i$, $i \in N_j \Leftrightarrow j \in N_i$), a random field X is a Markov Random Field with respect N if and only if:

$$\begin{aligned} &\forall x \in X,\ P(x) > 0 \\ &\forall i \in N,\ P(x_i | x_{L-\{i\}}) = P\left(x_i | x_{N_i}\right) \end{aligned} \tag{2.3}$$

In accordance with the Hammersley-Clifford theorem [15], an MRF can also be represented by a Gibbs distribution:

$$P(x) = Z^{-1}.e^{-U(x)} \tag{2.4}$$

where

$$Z = \sum_{x \in X} e^{-U(x)} \tag{2.5}$$

is the so-called partition function with acts as a normalizing constant, and $U(x)$ is an energy function:

$$U(x) = \sum_{c \in C} V_c(x) \tag{2.6}$$

that is, a sum of clique potentials $V_c(x)$ over all possible cliques C.

Markov random fields can obviously be used to model texture. In the Derin-Elliot model [60] for instance, or in the auto-binomial model [15], only single pixel and pairwise pixel cliques in the second order neighbors of a given site are considered. The derived conditional probabilities are written:

$$P(x_i \,|R_i) = \frac{1}{Z_t} e^{-w(x_i|R_i)^{T_\theta}} \tag{2.7}$$

where θ is the set of parameters which model the textural properties of the image. This yields the following Gibbs random field:

$$U(x) = \frac{1}{2} \sum_{i=1}^{MN} w(x_{i,} R_i)^{T_\theta} \tag{2.8}$$

where $\theta = [\theta_1\, \theta_2\, \theta_3\, \theta_4]^T$ and $w(x_i, R_i) = [w_1(x_1)\, w_2(x_2)\, w_3(x_3)\, w_4(x_4)]^T$ is defined as follows:

- Derin-Elliott model: $w_r(x_i) = 1(x_i, x_{i-r}) + 1(x_i, x_{i+r})$, $1 \leq r \leq 4$
- Auto-binomial model: $w_r(x_i) = 1(x_{i-r} + x_{i+r})$, $1 \leq r \leq 4$

with 1(a,b)= $\left| \begin{array}{l} -1 \text{ if } a = b \\ 1 \text{ otherwise} \end{array} \right.$.

Fractals. The statistical properties of natural textures in general and their self-similarity at different scales in particular makes of fractal analysis a useful modelling and analysis tool. As argued by Pentland [69], images of most natural surfaces can be modelled as spatially isotropic fractals. Many fractal descriptors and approaches have been described in the literature (see [33, 207] for a review and [153], for instance, for an evaluation of classification performances with respect to non-fractal descriptors).

Among others, fractal dimension has proved a popular and effective tool. At a glance, the fractal dimension is a measure of the roughness of a surface. The box method approach [179] estimates the fractal dimension by computing the slope of the least square regression of the number of disjoint boxes, with increasing sizes, required to cover the image, considered as a surface.

A textural descriptor from multifractal analysis[3], the (point-wise) Hölder exponent [207] too corresponds to an intuitive perception of regularity. It is computed by comparing the total variation of gray levels with functions of the form $\epsilon \rightarrow \epsilon^\alpha$, that is, via regression over windows with increasing size centered at the point at which the exponent is evaluated. Within a sub-window defined in the input image, several descriptors (or capacities [207]) can then be evaluated once the exponent of each voxel in the subwindow has been computed: the minimum value ("min"), the maximum value ("max"), the most frequent value or the cardinality of the largest subset of pixels with the same value ("iso"), the most frequent value when noise is taken into account ("adaptive iso"), etc.

[3]The better comprehension of the intimate relationship between fractal dimension and fractal spectrum is a central issue of multifractal theory [26].

2.3.1.4 Signal processing methods

In view of the statistical properties of texture, frequency analysis is a powerful way to characterize it. It is also the natural way, following the evidence put forth by psychophysics and electrophysiology [19] that the human visual system decomposes incoming images into their frequency and orientation components.

Spatial domain filters. The earliest spatial domain filtering methods relied on a measure of the edge density per unit area to capture the textural properties of a region. Indeed, the finer the texture, the higher the density of edges per unit area. The magnitude of the response of a straight-forward edge mask, the Robert or the Laplacian operator, for instance, gives a measure of edgeness [47].

Spatial masks can also serve to compute spatial moments [78]. The $(p+q)^{\text{th}}$ moment of an image region I, computed over a sub-window W defined around a site in I is given by:

$$m_{pq}(W) = \sum_{(x,y,z)\in W} x^p y^q z^r I(x,y,z) \tag{2.9}$$

By centering and normalizing the coordinates x, y and z over the sub-window W, spatial masks are derived, with which the input image can be convoluted to compute a texture feature map. Note that, as demonstrated by Gagalowicz [72], textures can be build between which the human eye can discriminate although they exhibit identical first, second, third, fourth and even fifth order spatial moments.

In [75], Malik *et al.* modeled the preattentive texture perception in human visual system using spatial domain filtering. The input image is convoluted with a bank of even-symmetric filters (difference of offset Gaussian functions) with half-wave rectification. Those are ways to introduce non-linearity into the computation of texture features so as to discriminate between texture identical up to second-order statistics. This method is particularly amenable to characterize synthetic, as opposed to natural textures.

Fourier filtering. Obviously, frequency analysis of a given texture is more naturally carried on in the frequency domain than in the spatial domain. In [79], Coggins *et al.* used a set of frequency and orientation selection filters to segment and classify textured image. They use 4 orientation filters (0, 45, 90 and 135 degrees) and a number of frequency selective filters (which depends on the size of the input image, usually they use 1,2,4,8,16,32 and 64 cycles/image).

Gabor and Wavelet models. While the Fourier transform performs a global frequency analysis of the input images, introducing spatial dependency into the Fourier formulation leads to a more localized analysis in the spatial domain. This is typically done through a windowed Fourier transform:

$$F_w(u,\xi) = \int_{-\infty}^{+\infty} f(x) w(x-\xi) e^{-j2\pi ux} dx \tag{2.10}$$

The Gabor transform is a special case of the windowed Fourier transform when the window function w is Gaussian. It allows the computation of local estimates of the frequency content of an image (see [33] for details).

Clearly, Gabor filters are frequency and orientation selective (as can be seen from their Fourier domain analysis [33]) and take the form, in the spatial domain, of a sinusoidally modulated Gaussian profile, rotated at a particular orientation. They are also conjointly optimal in providing

the maximum possible resolution for information in both the spatial and frequency domain, and a good approximation of the receptive field profiles of simple cells in the visual cortex of some mammals [51]. This optimality makes them particularly successful in textural analysis as they achieve the best possible compromise between localization in the spatial domain (required to accurately determine the boundary of a textured area) and small bandwidth in the frequency domain (required to allow finer distinction between closely related textures).

Often, a number of Gabor filters, selected in an ad-hoc or statistical manner, are combined in a filter bank. In [105] for instance, Jain *et al.* used a bank of Gabor filters at multiple scales and orientations to perform unsupervized texture segmentation.

Although often used in 2-D, Gabor filters can be extended to 3-D in a straightforward manner, as described in [11].

2.3.1.5 Remark on anisotropic images

The models described above are usually meant to be used with isotropic images. When dealing with anisotropic images, one must either resample the data to establish isotropic conditions (which might substantially alter textures, most especially if the ratio between the largest and the smallest voxel dimension is high), or modify the descriptors to take anisotropy into account (recommended technique). We did not have to deal with this problem here as all of our images were isotropic.

2.3.1.6 Feature/descriptor families

We call a descriptor family a set of related descriptors, in that they are derived from the same underlying principles with varying parameters and should thus exhibit similar characteristics. In our case, we define the following families:

- Haralick family with translation u: the set of the most common Haralick descriptors (energy, entropy, contrast and correlation) for a given translation u to which we add the mean intensity and the variance of the intensity of the region considered.
- Gabor family with scale n: the set of Gabor descriptors along a series of fixed orientations at the specified scale. The orientations are computed to obtain an even coverage of the spectrum by setting (1) the ratio of the standard deviation of the Gaussian describing the filter's transfer function in the frequency domain to the filter center frequency and (2) the ratio of angular interval between filter orientations and the standard deviation of the angular Gaussian function used to construct the filters in the frequency domain.
- Hölder family with capacity *cap*: the set of the chosen capacity values ("min", "max", "iso" or "adaptive iso") for the ranges of window sizes over which the exponents are evaluated.

2.3.2 Choosing a reasonably small set of descriptors

While a given descriptor can be particularly well suited for a given task (Fourier filters can easily distinguish between low frequency and high frequency textural patterns, for instance), very few are equally efficient in a variety of situations. Rather than looking for both the most specific and most robust descriptor (which, almost by definition, does not exist), a more reasonable endeavour consists of selecting a set of descriptors, which, when merged together, improve the classification performances. A number of such approaches have been presented in the literature for MR images or histological sections (please refer to [141] for a more in-depth review):

MRI: Kovalev *et al.* introduced in [121] the extended multisort co-occurrence matrices (which incorporate elementary image features such as intensity gradient magnitude and mutual orientation) and applied them to the automated diagnosis of Alzheimer's disease and to the segmentation of diffuse brain lesions, both in T1-weighted MRI. They selected intensity, gradient magnitude and relative orientation of gradient vectors as discriminant descriptors. In [176], a Markov Random Field model of texture was used in conjunction with a D'Agostino-Pearson normality test to classify the tissues of T1-weighted MRI. They put a premium on carefully modelling partial volume effect. Use of the Hölder exponent enabled them to better discriminate between uniform, convex and concave regions. First order statistics (intensity value, mean and standard deviation intensity) served as feature descriptors in [189] where 3 types of classifiers (self-organizing map, feed-forward neural network, k-nearest neighbor) where evaluated as a means to characterize tissue types in MR images (T1-weighted, T2-weighted and PD).

Histology: In Van de Wouwer *et al.* [57], wavelets served as chromatin texture descriptors in the automated identification of neoplastic nuclei. Weyn *et al.* [219] used a combination of texture, morphometry and histometry to diagnose malignant mesothelioma. Multispectral images were analyzed in [173] where the authors derived statistical and spatial features useful in the quantification of chromatin structures.

2.3.2.1 Feature selection

Given the variety of recommended feature detectors, we have to select from the set of possible filters described above (Section 2.3.1) an "optimal" subset, as discriminating as possible, while keeping the number of selected descriptors to a reasonable number. This selection task (also called feature reduction) is a fundamental problem of statistical pattern recognition[4]. Indeed, reducing the number of detectors not only saves computing resources by discarding irrelevant or redundant features but it also alleviates the effects of the so-called *curse of dimensionality*, which links the ratio: sample size (in the learning set) / dimensionality of the feature vector to the classification performances [104]. Namely, when the dimensionality of the feature space increases, more parameters must be estimated, which enhances the risk of overfitting the model: adding new descriptors then usually first increases the classification performance, which attains a peak value, before decreasing as more descriptors are added (overfitting phenomena).

Given an objective function, which evaluates the performances of the classifier on an *a priori* set of measurements, the feature selection problem then boils down to a search problem in the combinatorial space defined by the feature descriptors. As trying out in a brute force manner every possible combination of features can be prohibitively costly when the number of descriptors is high (although, as argued by Cover [49], traversing the entire search space is the necessary condition to an optimal selection), specific sub-optimal selection strategies have been suggested in the literature. They either rely on *a priori* knowledge about the classification problem at hand to discard features, or use generic optimization heuristics when no domain-specific information is available (or when it is too difficult to exploit). Algorithms as diverse as stochastic annealing, genetic algorithms, max-min pruning, principal component analysis or neural node pruning have been introduced (see [106] for a taxonomy of feature selection algorithms).

[4]Described by many authors as at least as difficult as the design of the pattern recognition system itself.

2.3.2.2 Sequential forward floating selection

A requirement shared by most sequential feature selection methods is that the objective function (the criterion of feature set effectiveness) should not decrease when new features are added. This is the so-called set inclusion monotonicity property. As pointed out by Pudil *et al.* [165], most of the reported feature selection algorithms make use of separability measures satisfying this monotonicity property. However, the preference order induced over an arbitrary set of features by such measures is not necessarily coherent with that obtained from the classifier recognition rate [184]. Consequently, optimal feature sets for the given measures may yield poor classifier performances *in fine*. Pudil *et al.* then strongly advocate to consider the latter as the only effective objective function, which requires having to deal with non-monotonicity. They therefore introduced a family of suboptimal search algorithms (the Floating Search methods), particularly amenable on high-dimensional problems with non-monotonic functions. In view of the efficiency of that approach, which was also recommended in [106], we selected it for our texture classification problem.

The sequential forward floating feature selection (SFFS) scheme works as follows:

SFFS algorithm

---

step 0 (*a priori*): For each target anatomical structure to be segmented, we first compile a series of images (MR or histology) along with their associated manually segmented structures to form a training database. We then pick at random N voxels $p_i = I(x_i, y_i, z_i)$ which do and do not belong to the target structure. Intensity information collected in windows centered around these voxels ($5 \times 5 \times 5$ usually) is then fed to the entire set of texture descriptors (D descriptors in total). These texture measures, together with the input voxels with which they are associated, form the learning set $L = \left\{ S_i = \left\{ s_1^i, \ldots, s_D^i \right\}, c_i \right\}_{i=1}^{N}$ where $S_i \in \mathbb{R}^D$ is the set of outputs from the D texture descriptors for pixel p_i, and c_i is the label associated to p_i: $c_i = +1$ if it belongs to the target structure, $c_i = -1$ otherwise.

Let $Y = \{y_1, \ldots, y_D\}$ be the set of available descriptors. Let $X_k = (x_j \in Y,\ j = 1, \ldots, k)$, $k \leq D$ be the set of selected features at iteration k. Let J be the objective function (a measure, to be set for each classifier, of the performances of the classifier evaluated on the learning set of labelled samples).

step 1 (init): $X_0 = \emptyset$, $k = 0$

step 2 (inclusion):

$x^+ = \arg\max_{x \in Y - X_k} J(X_k + x)$

$X_{k+1} = X_k + x^+,\ k = k + 1$

step 3 (conditional exclusion):

$x^- = \arg\max_{x \in X_k} J(X_k - x)$

if $J(X_k - x^-) > J(X_{k-1})$ then

$X_{k-1} = X_k - x^-,\ k = k - 1$

if ($k \neq$ desired number of features) then goto step 3

else

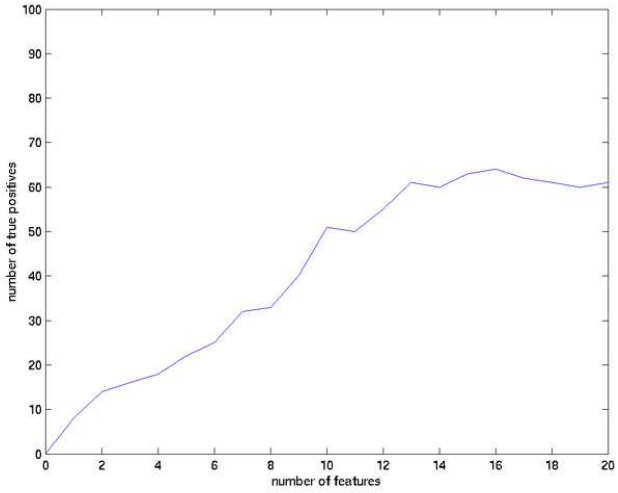


Figure 2.5: Performances of the LDA classifier as a function of the number of features, for the classification of the caudate nucleus in a typical T1-weighted MRI.

if ($k \neq$ desired number of features) then goto step 2

The efficiency of this feature selection approach then depends on a number of parameters:

- the desired number of features k: Figure 2.5 shows a typical recognition rate[5] versus number of features graph (corresponding to the feature selection with linear discriminant analysis for the caudate nucleus in a T1-weighted MRI). Obviously, the performance of the classifier first increases when new features are added until a plateau is reached (around $k = 15$, in this case). We can even notice a slight drop in performance as more features are added (overfitting phenomena). The choice of the number of features is then a matter of compromises: the more features, the longer the training of the classifier and the more computationally expensive the filtering.

- the structure of the learning set: As with any supervised learning approach, the performances of this feature selection algorithm (and also those of the classifier itself) depend on the quality and "representativity" of the learning set of texture samples. A sufficient number of voxels from the target structure and from decoy structure should be considered. However, even though choosing voxels from the target structure is easy (since its manual delineation is available), it is not clear how to choose a representative set of non-target voxels. This is a difficult and still open problem. Here, we assume that the most difficult voxels to classify are those on each side of the target structure's boundary and select an equal number of in-target and off-target voxels (typically, 50)[6].

[5] We record the number of true positives, that is, the number of target voxels that are correctly classified as belonging to the target structure.

[6] For our neural classifier, we have devised a neural learning strategy particularly suited to this voxel selection issue (see Section 2.4.3).

- the size of the floating window: This parameter affects both the feature selection and the classifier performance, as it controls the quality of the estimations delivered by the texture descriptors. Again, a compromise has to be found between too large a window over which estimation is more precise but also riskier as more than just one texture type can be present, and too small a window over which noise becomes preponderant. In our case, $5 \times 5 \times 5$ windows seemed to yield the best results.

The sequential forward floating selection (SFFS) algorithm was applied to the various texture classification algorithms we have used. Selected features are detailed and commented in the following Section 2.4.

2.4 Texture classification

In view of the difficulty of modeling the class distributions of the target textures, we opted for non-parametric classifiers. Various types of classifiers were evaluated, ranging from linear discriminant functions to hybrid neural networks, with an increase in both complexity and performance. *A priori* medical information was introduced in the form of a learning set of textures (one learning set per target structure as described in Section 2.3.2.2 above).

Modus operandis: For each target structure, we first identify a region of interest (ROI) around that structure in the input MRI or histological section. For efficiency reasons, the texture filters are applied only inside that area: this decreases both the learning and processing time, and enhances the performances of the classifier as less decoy structures or similar-looking off-target tissues must be discarded. Use of an atlas registration allows the identification of the ROI in a fully automated fashion with good precision (see Section 4.2.2 for details). We also compulse a series of already segmented structure instances and select voxels in them to form the learning set L (see Section 2.3.2.2 above). The SFFS algorithm is then applied to find the best combination of texture descriptors for each target anatomical structure where the objective function J depends on the selected classifier (LDA, SVM or neural network).

All input MR images were T1-weighted MRI with 1 mm^3 resolution. Note that our texture filters must be applied prior to any spatial transformation of the input images (before any registration between two acquisitions of the same patient, or registration to a standard stereotactic frame, for instance) to prevent brain texture distortions. Indeed, studies by Kovalev *et al.* [121] report major changes of textural properties for brain tissues even with reasonable spatial changes (between 85% and 115% in scale).

Evaluation of results: The performances of the various classifiers and texture descriptors were evaluated on a test set of MR images and histological sections (disjoint from those of the learning sets), whose associated manual segmentations were available. For each pair (target structure, texture classification system), we obtain a classification map (a [-1.0,1.0] image to the voxels of which is associated a label: between 0.0 and 1.0 when the used classification algorithm decided that the voxel belonged to the target structure and between -1.0 and 0.0 otherwise) and a binary classification map (thresholded classification map with label $+1$ for all voxels with associated value above 0.0 and -1 for all others) of the defined region of interest. We report the number of misclassified voxels in terms of:

- ratio of false positives: ratio of the number of voxels which do not belong to the target structure and are incorrectly classified by the algorithm as target voxel, over the overall number of pixels in the target structure.
- ratio of false negatives: ratio of the number of voxels which actually belong to the target structure and are incorrectly classified by the algorithm as non-target voxel, over the overall number of pixels in the target structure.

We give, for both ratios, the mean± standard deviation and the maximum over the entire test set, in percentage.

Delineation protocol: For each target structure, a particular delineation protocol was devised by expert neuroscientists and used to build the training set of manual delineations (see Appendix B). Protocols are always designed towards a specific *a priori* goal: for instance, the comparison of diseased and normal individuals, or the longitudinal study of a pathology, etc. They also have to ensure that the delineations can be carried out with reasonable accuracy by some trained operators. This might at times impose that the least visible parts of a structure be discarded (lest the manual delineation should introduce spurious edges and yield artificially high variability).

Consequently, the manual delineations from the training sets, which we consider our gold-standard, may not always conform with standard anatomical expectations about the shape of the target structures (for instance, our gold-standard caudate nuclei have a very short tail). We then have to take these delineation protocols into account when computing the error ratios of our classification approaches (and later, in Chapter 4, of our segmentation approach). For each target structure, we then devised a semi-automated means (which often relied on semi-automated masking) to correct the computation of the misclassified voxels. Namely, for each structure, its associated manual delineation protocol was applied to the output of the automated algorithm to discard from the error computation voxels outside of the range defined by the protocol. Typically, if the tail of a caudate nucleus were correctly recovered by our classifier, we cut it all the same, in a fashion similar to that prescribed in the caudate manual delineation protocol, so as not to artificially increase the false positive ratio.

This correction step, although necessary, unfortunately introduces artificial imprecisions. One should then consider the classification results with caution. In particular, it seems wise to nuance the performances of an algorithm with respect to the measured variability of the delineating human operators[7].

Note that while the obtained classification maps can be seen as segmentation results in and out of themselves, we rather use them as constraints imposed on the deformable templates presented in Chapter 4. Clearly, performances could be greatly enhanced by taking into account larger neighborhoods or non-iconic information, for instance. However we chose not to further complexify the texture classifiers as most of this additional information is more easily introduced in the framework of the deformable models.

Also note that the texture classification algorithms we present here are only a component of a more complex segmentation system. As such, their performances should be evaluated in the context of the global segmentation task: the best texture classification algorithm then becomes that which maximizes the segmentation performances, not that which provides the best classification maps. These maps nonetheless enable the qualitative understanding of the behaviour of

[7] Zou *et al.* describe in [228] a means to compare the results of automated algorithms with those of experts when the latter exhibit substantial variability.

the various classifiers at both a coarse and fine scale (systematic biases or regional difficulties in particular are more easily spotted on the maps than in the classification performances tables).

We first introduce in the following two sections the linear discriminant function and support vector machine classifiers: we briefly expose the underlying theory (see Duda *et al.* [62] for details) and report some experiments that demonstrate the need for a better and more specific classifier. We then introduce our answer to that need: the hybrid neural network architecture (Section 2.4.3).

2.4.1 Linear discriminant analysis (LDA)

Instead of hypothesizing the form of the probability densities of our texture classes and estimate from a learning set the values of the density parameters, we here make a linear assumption about the form of the decision boundary and use the samples from the learning set to estimate the parameters of that linear boundary.

2.4.1.1 Principles of linear discriminant analysis

With the notations introduced above, a linear discriminant function (LDF) g can be written:

$$g : \left| \begin{array}{ccc} \mathbb{R}^k & \rightarrow & \mathbb{R} \\ x & \mapsto & g(x) = w^t x + w_0 \end{array} \right. \tag{2.11}$$

where k is the number of selected feature descriptors, x is a k-dimensional vector of texture measurements (one measure per selected texture descriptor), $w = [w_1, \ldots, w_k]^t$ is the weight vector, and $w_0 \in \mathbb{R}$ the bias.

For a two-category classifier (our case here), $g(x) = 0$ defines a hyperplane that separates points assigned to the first class (target pixels: $g(x) > 0$), from points assigned to the second class (non-target pixels: $g(x) < 0$). From equation **??**, we deduce that a LDF separates the feature space in two halves with a hyperplane decision surface whose normal is given by the weight vector w, and whose location is determined by the bias w_0.

Given a learning set[8] of labeled samples $L = \left\{ S_i = \left\{ s_1^i, \ldots, s_k^i \right\}, c_i \right\}_{i=1}^N$, we want to find the weight vector w and bias w_0 that best classifies samples from L and, hopefully, generalizes well. More formally, a sample S_i is correctly classified if

$$c_i \left(a^t \hat{S}_i \right) > 0 \tag{2.12}$$

where $\hat{S}_i = \left[1, s_1^i, \ldots, s_k^i\right]^t$ and $a = [w_0, w_1, \ldots, w_k]^t$. A usual "normalization" consists of replacing the samples from class $c_i = -1$ (the non-target samples, here) by their negatives, to get rid of the labels altogether: a sample S_i is then correctly classified if

$$a^t \hat{S}_i > 0 \tag{2.13}$$

When the classification problem is linearly separable, we are then looking for a vector a such that $a^t \hat{S}_i > 0$ for all samples in L. To do so, we define a criterion J, a minimum of

[8] Note that we have already performed the feature selection at this point: each sample S_i is a k-dimensional vector.

which corresponds to a solution for a. J can be the number of misclassified samples, or, more pertinently, the so-called Perceptron criterion function, which is easier to minimize:

$$J(a) = \Sigma_{S \in \wp} \left(-a^t \hat{S} \right) \tag{2.14}$$

where $\wp$ is the set of samples from L misclassified by a. A basic gradient or Newton descent optimization then suffices to find the vector a :

gradient descent

step 0 (init): init a to an arbitrary value, $t = 0$

step 1:

$t = t + 1$
$a = a - \eta(t) \nabla J(a)$

step 2: if $|\eta(t) \nabla J(a)| > \theta$ then goto step 1

where θ is a termination threshold and $\eta(t)$ is the learning step at iteration t. θ and η are specified by the user.

However, when the task is not linearly separable (which is the case for our texture classification application), those optimization procedures do not converge[9]. We then opted to use the Ho-Kashyap [96] procedure which guarantees an answer in both the separable and non-separable cases. The trick is to look for a vector a and a vector $b = [b_1, \ldots, b_k]$ such that $a^t \hat{S}_i = b_i$, where the b_i are arbitrary positive constants (to each values of b corresponds a different solution). By minimizing the minimum squared error (MSE)

$$J(a) = \|Sa - b\|^2 \tag{2.15}$$

where S is the N-by-k matrix whose i^{th} row is the vector S_i^T, we get a good estimate of the optimal a and b but cannot guarantee that a is a separating vector in the separable case. The Ho-Kashyap procedure is a modification of the MSE algorithm that guarantees the separability in the separable case, while minimizing the MSE in any case:

Ho-Kashyap

step 0 (init): init a and b to an arbitrary value, $t = 0$

step 1:

$t = (t + 1) \ \mathbf{mod}\ N$
$e = Sa - b$
$e^+ = \frac{e + |e|}{2}$

[9] Since there does not exist any weight vector that can correctly classify all the samples from the learning set in the non-linearly separable case, gradient descent algorithms tend to oscillate between sub-optimal values. A number of heuristics have been presented in the literature to nonetheless extract a meaningful answer from them (see [62] for details).

$b = a + 2\eta(t)e^+$

$a = S^t b$

step 2: if $|e| > b_{min}$ or $t < t_{max}$ then goto step 1

---

where b_{min} is the lower bound on the margin error, and t_{max} is the maximal number of iterations (if it is reached, then no solution has been found). Please refer to Duda *et al.* [62] for convergence and separability proofs.

Note that since $a(t)$ is completely determined by $b(t)$, the margin vector at iteration t, the Ho-Kashyap algorithm essentially produces a sequence of margin vectors.

2.4.1.2 Application to MRI texture classification

corpus callosum: A first experiment tackled the LDA classification of the corpus callosum in 1 mm^3 T1-weighted MRI. Figure 2.6(a) shows a few typical mid-sagittal sections of the considered 3-D MR data. The learning set consisted of 3 corpora callosa. In each callosum, 50 voxels were selected both inside and outside the manual defined callosal boundaries. We used a $5 \times 5 \times 5$ window for the texture descriptors. The SFFS algorithm with $k = 3$ desired features selected the Haralick mean with $u = (2, 2, 2)$, Haralick energy with $u = (2, 0, 0)$ and Hölder family with "adaptive iso" (empirical experiments showed that using more descriptors only marginally improved the performances). Figure 2.6(b,c,d) shows the descriptor maps of the displayed corpora callosa. Note that the Hölder exponent, although not particularly discriminating inside the structures, proved helpful along the boundaries and helped achieve good precision.

We applied the Ho-Kashyap LDA to a test database of 16 corpora callosa, with $\eta(t) = 0.1$, $\forall t$, $b_{\text{init}} = [1, \ldots, 1]^t$ and $t_{\max} = 1000$. Figure 2.6(e,f) displays a few classification maps computed over the entire images, instead of just the region of interest, for illustration purposes. Quantitative results are reported in Table 2.1: the first column gives the standard classification errors; the second column reports errors calculated without considering the fornix in the numbering of the misclassified voxels. Overall, classification results were good: the corpus callosum boundaries were correctly recovered in most cases (see right-most column), most especially since its mean intensity is clearly above most of the surrounding tissues. Note that almost all of the true callosum voxels were recovered (the false negative rate is very low). However, as could be expected, the LDA algorithm could not fully discard the fornix.

ratio of misclassified voxels false positive (%) false negative (%)	*corpus callosum*	*corpus callosum without fornix*
mean	12.4 ± 2.1 5 ± 3	10.3 ± 1.2 4.1 ± 2.5
maximum	15.6 9	13 7.5

Table 2.1: Texture classification of the corpus callosum in T1-weighted MRI with the Ho-Kashyap LDA classifier.

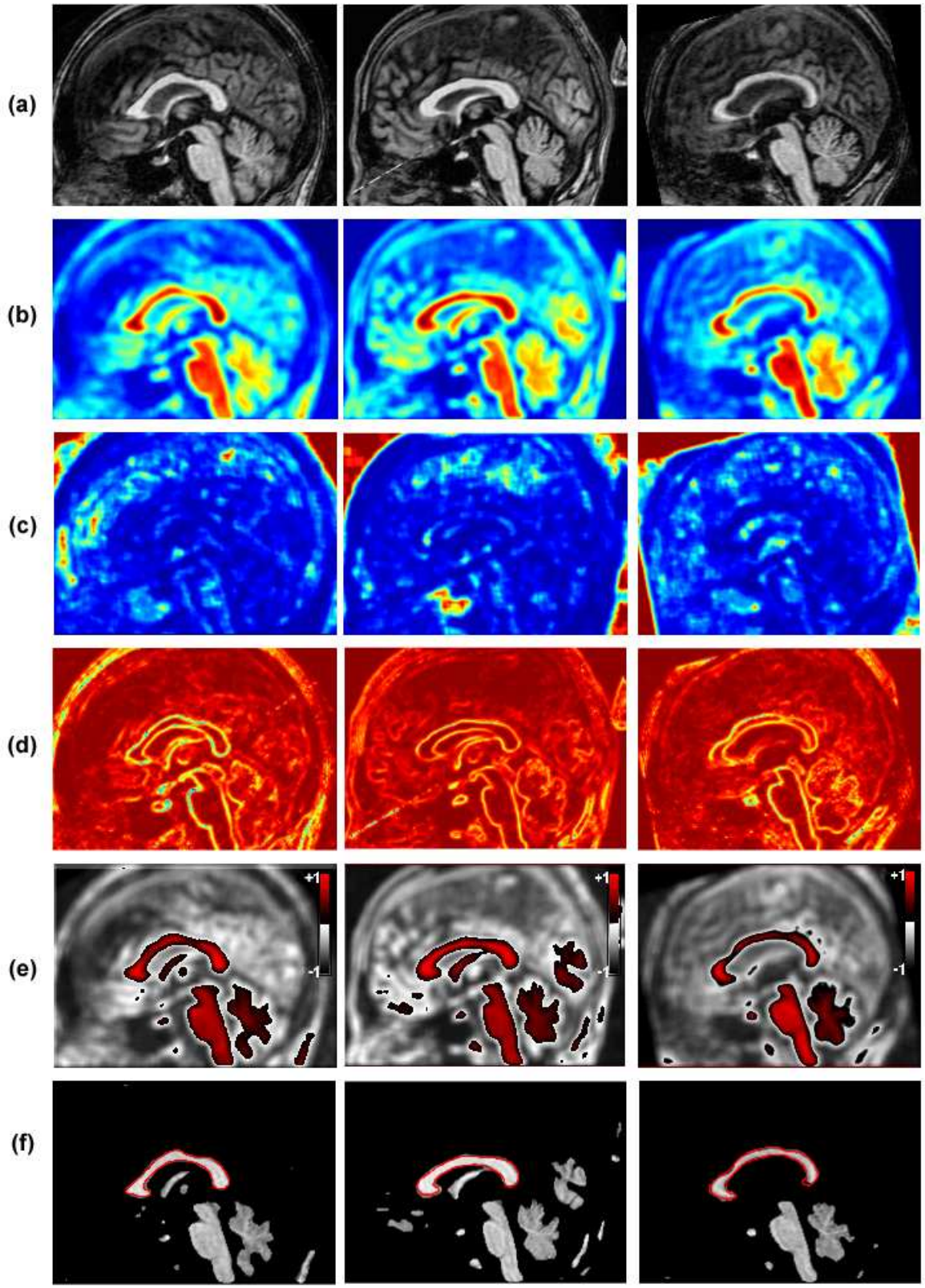


Figure 2.6: LDA classification of a few corpus callosum: (a) input MRI; (b) Haralick mean map; (c) Haralick energy map; (d) Hölder exponent map; (e) LDA classification map; (f) LDA extracted callosum voxels with true callosum outlines superimposed.

caudate nucleus: Similar settings were used to classify the caudate nucleus in T1-weighted MRI. We experimented with a variety of SFFS-selected feature sets (between 2 and 20 descriptors): the best results were achieved with 15 descriptors. The performances of the LDA classifier were however fairly poor: more than 35% of the target pixels and nearly 40% of the non-target pixels were incorrectly classified on average. This is not surprising as both the average intensity and the texture of the caudates are "intermediate". As illustrated in Figure 2.2, their intensity distribution falls between that of the ventricles and that of the surrounding tissues: even though texture could clearly discriminate between the ventricles and the caudate, no *linear* boundary in the texture space was found that could extract the caudates from their surrounding tissues. Clearly, this is too much of a non-linearly separable case for our LDA classifier to behave adequately: more degrees of freedom are required.

2.4.2 Support Vector Machine (SVM)

We have seen in Section 2.4.1 how linear discriminant analysis could provide gradient descent methods capable of reducing the classification error, even for non linearly separable cases. However, the use of a hyperplane decision boundary was not sufficient to enhance the contrast of the caudate nucleus. By projecting the data to a higher-dimensional space via use of some non-linear mapping functions, better separating hyperplanes can be found[10].

2.4.2.1 Principles of support vector machines

As in the Ho-Kashyap approach, an SVM classifier searches for the hyperplane which maximizes the separating margins between the two classes of the learning set. The first step consists of choosing the SVM non-linear mapping functions. In the general case, we can choose a different mapping function for each coordinate of the samples (one mapping function per texture measure in our case). We however restricted ourselves to a single non-linear operator (ϕ) for simplicity and parsimony reasons. The choice of ϕ depends on information about the specific problem domain. When none is available (or when it is too difficult to exploit), the classifier designer may use polynomials, Gaussian or radial basis functions.

With notations from Section 2.4.1, we are looking for the vector a which minimizes

$$J(a) = \frac{1}{2} a^t a = \frac{1}{2} \|a\|^2 \tag{2.16}$$

subject to the constraints

$$c_i \left(a^t \phi \left(\hat{S}_i \right) \right) \geq 1, \forall i \tag{2.17}$$

where $c_i = +1$ for samples that belongs to the target structure, and -1 for the others. The associated SVM classifier is then written:

$$g : \left| \begin{array}{ccc} \mathbb{R}^k & \rightarrow & \mathbb{R} \\ x & \mapsto & g(x) = a^t \phi(x) \end{array} \right. \tag{2.18}$$

[10]Clearly, for any input data, there always exists a choice of non-linear mapping functions that define a non-linear space of sufficient dimension to guarantee that a perfectly separating hyperplane in this high-dimensional space can be found. The true issue then becomes that of parsimony and generalizability.

Again, in the non-separable case, we relax the separability constraints by introducing arbitrary positive slack variables ξ_i. We get:

$$c_i \left(a^t \phi \left(\hat{S}_i \right) \right) \geq 1 - \xi_i, \forall i \tag{2.19}$$

The objective function (the so-called structural risk) becomes:

$$J(a, \xi) = \frac{1}{2} \|a\|^2 + \lambda \sum_{i=1}^{n} \xi_i \tag{2.20}$$

where $\lambda \in [0, 1]$ is a regularization coefficient, specified by the user, which balances the empirical risk $\Sigma_{i=1}^{n} \xi_i$ with the model complexity $\frac{1}{2} \|a\|^2$. The purpose of balancing these two terms is to avoid overfitting [180].

Lagrange multipliers enable us to reformulate the constrained minimization problem into an unconstrained one and derive a necessary condition for minimizing $J(a, \xi)$:

$$a = \sum_{i=1}^{k} \alpha_i c_i \phi \left(\hat{S}_i \right) \tag{2.21}$$

where $\alpha_i \geq 0$ are the Lagrange multipliers associated with the constraints 2.19.

We then substitute 2.21 into 2.18:

$$\begin{array}{rcl} g(x) & = & \sum_{i=1}^{k} \alpha_i c_i \phi \left(\hat{S}_i \right)^t \phi(x) \\ & = & \sum_{i=1}^{k} \alpha_i c_i K \left(\hat{S}_i, x \right) \end{array} \tag{2.22}$$

where $\forall x, z \; K(x, z) = \phi(x)^t \phi(z)$.

The Lagrange multipliers are solved using the dual form:

$$W(\alpha_1, \ldots, \alpha_k) = \Sigma_{i=1}^{k} \alpha_i - \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \alpha_i \alpha_j c_i c_j K \left(\hat{S}_i, \hat{S}_j \right) \tag{2.23}$$

subject to the following constraints:

$$\begin{array}{l} 0 \leq \alpha_i \leq \lambda, \forall i \\ \sum_{i=1}^{k} \alpha_i c_i = 0 \end{array} \tag{2.24}$$

As W is convex and quadratic in α, a numerical approximation of the solution can be obtained through quadratic programming. Classically, the α_i are partitioned in three classes, depending on their final values:

- $\alpha_i = 0$: The corresponding sample then satisfies $c_i g \left(\hat{S}_i \right) > 1$, is thus outside the decision margin and, consequently, correctly classified.
- $0 < \alpha_i < \lambda$: The corresponding sample satisfies $c_i g \left(\hat{S}_i \right) = 1$, is located exactly on the decision margin and is called a *margin support vector* of g.
- $\alpha_i = \lambda$: The corresponding sample satisfies $c_i g \left(\hat{S}_i \right) < 1$, is thus inside the decision margin[11] and is called an error support vector of g.

[11] Note that it may still be correctly classified.

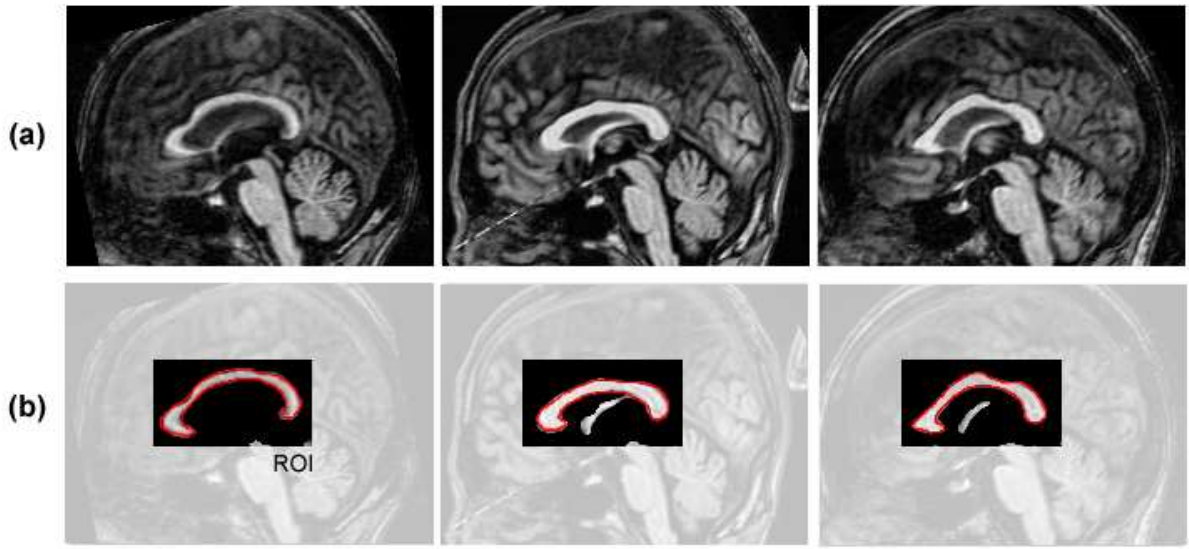


Figure 2.7: SVM classification of a few corpus callosum: (a) input MRI; (b) SVM extracted callosum voxels with true callosum outlines superimposed.

In most cases, only few samples from the learning set become support vectors. The decision function g is then entirely defined by the only support vectors and their associated mapping functions.

2.4.2.2 Application to MRI texture classification

corpus callosum: We first compared the performances of the SVM classifier with those of the LDA classifier when applied to the classification of the corpus callosum in T1-weighted MRI. The learning set consisted of the same 3 corpora callosa and the same 50+50 inside/outside voxels as above. The SFFS algorithm with $k = 3$ desired features too selected the Haralick mean with $u = (2, 2, 2)$, Haralick energy with $u = (2, 0, 2)$ and Hölder family with "adaptive iso".

We applied the SVM classifier to the same test database of 16 corpora callosa, with Gaussian mapping functions (standard deviation: 1.0). Figure 2.7 displays a few classification maps computed over the callosal regions of interest (ROI). Quantitative results are reported in Table

ratio of misclassified pixels false positive (%) false negative (%)	*corpus callosum*	*corpus callosum without fornix*
mean	10.2 ± 2.0 4.1 ± 2.5	8.5 ± 1.7 3.4 ± 2.1
maximum	13.2 7.1	11 5.8

Table 2.2: Texture classification of the corpus callosum in T1-weighted MRI with the SVM classifier.

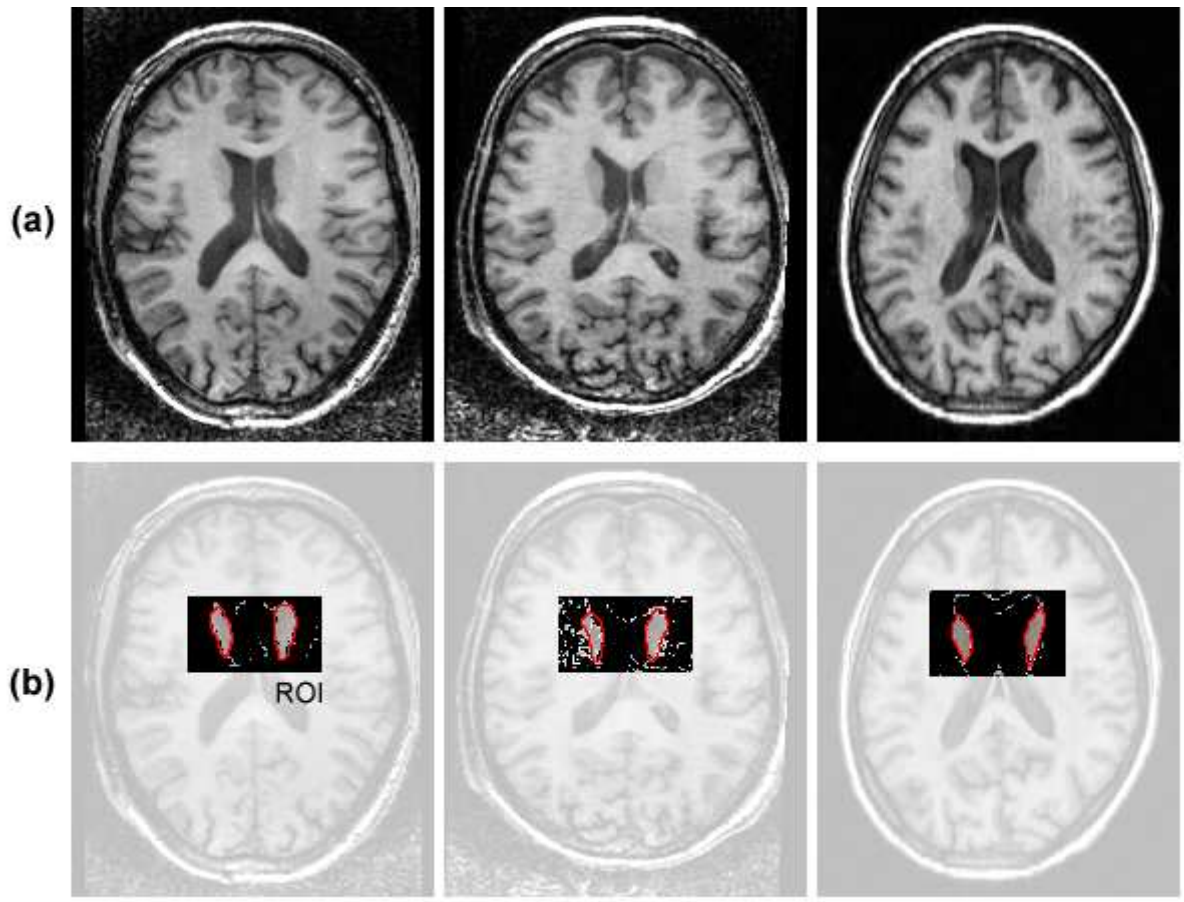


Figure 2.8: SVM classification of a few caudate nucleus: (a) input MRI; (b) SVM extracted caudate voxels with true caudate outlines superimposed.

2.2, again with and without taking the fornix into account. We can notice a slight improvement of the overall performances. In particular, the SVM classifier could discard a greater proportion of fornix voxels.

ratio of misclassified pixels / false positive (%) / false negative (%)	*caudate nucleus*
mean	21.5 ± 4.1 9.6 ± 2.5
maximum	37 14.2

Table 2.3: Texture classification of the caudate nucleus in T1-weighted MRI with the SVM classifier.

caudate nucleus: A more dramatic improvement of the performances was observed for the caudate nucleus. $k = 5$ SFFS selected features (Haralick mean with $u = (2, 2, 2)$, Haralick mean with $u = (2, 0, 0)$, Haralick contrast with $u = (2, 2, 2)$, Gabor filter at scale $n = 1$ and Hölder family with "adaptive iso") were sufficient to achieve good results (see Table 2.3 and Figure 2.8). Obviously, projecting to a non-linear space helped the optimization algorithm find an adequate linear classification boundary.

2.4.3 Hybrid neural network approach

Section 2.4.2 demonstrated how the careful selection of particular non-linear functions enabled the exploration of entirely new sets of solutions, and the definition of arbitrary shaped decision boundaries. The difficulty of designing a good classifier then becomes that of choosing an adequate set of non-linear functions. A number of techniques have been proposed in the literature (see [62] for a review). Yet, they often more closely resemble *ad-hoc* heuristics than actual optimal solutions. In the particular case where *a priori* information about the structure of the classification problem is available, the design of a good set of basis functions can be facilitated. This is unfortunately not our case, here. Consequently, we are looking for a method which could learn the non-linearity, that is, which could adapt to set of basis functions to the problem at hand, while optimising the discriminating function. This is the approach of neural networks.

2.4.3.1 Artificial neural networks

Artificial neural networks (ANN) are mathematical tools that mimic the densely interconnected and parallel structure of the brain, and the adaptive biological processes of the nervous system, both in terms of learning and of information processing. They are composed of a large number of processing elements (so-called neurons) linked together by weighted connections (analogous to synapses). Each neuron receives activation signals from other neurons and outputs a non-linear function of the weighted sum of these activations. This nonlinear mapping function (ϕ) is called the activation function. With notations from Section 2.4.1, output from a neuron $neuron_i$ is written:

$$\begin{array}{ccc} neuron_i : \mathbb{R}^{d_i} & \rightarrow & \mathbb{R} \\ x & \mapsto & neuron_i(x) = \phi\left(w_i^t x\right) + w_{i,0} \end{array} \tag{2.25}$$

where x is the d_i dimensional vector of input signals, w_i is the weight vector (or vector of synaptic weights), and $w_{i,0}$ is a constant bias.

A neural network is then a graph of interconnected neurons, whose connectivity matrix defines the connection pattern. We restricted our study to multiple-input/single-output feed-forward multi-layer networks, where this graph is a tree. The leafs of the tree form the input layer to which the data to be classified is fed, and the root of the tree is the output neuron, which, in our case, produces a number between -1 and +1, that is, a classification decision between non-target and target pixel, respectively.

As always, the shape of the activation functions should match the *a priori* knowledge about the classification problem at hand, whenever available. They can be any functions, provided they verify the following properties:

- ϕ must be non-linear to ensure maximal expressive power: given the proper activation function and a sufficient number of neurons, feed-forward multi-layer neural networks become universal approximizers: they can implement any arbitrary decision boundary (see [94] for a proof);
- it must saturate: in our case ϕ belongs to $[-1, +1]$;
- it must be continuous and smooth if efficient back-propagation algorithms are to be used;
- monotonicity also provides faster and easier convergence.

A commonly used activation function, the sigmoid function (hyperbolic tangent for instance) satisfies the above requirements.

Given a learning set of labeled samples, training an ANN then essentially consists of modifying the weights of its neurons so as to minimize the overall difference between the output values of the network and the target values from the learning set. With our notation, the functional to be minimized is written:

$$J(w) = \frac{1}{2} \sum_{j=1}^{n} |neuralnet(S_i) - c_i|^2 \tag{2.26}$$

where $neuralnet(S_i)$ is the output delivered by the network when S_i is fed to its input layer. The most popular learning technique is the so-called back-propagation algorithm, which is based on a straight-forward gradient descent technique, with:

$$\triangle w = -\eta \frac{\partial J}{\partial w} \tag{2.27}$$

Various more sophisticated learning techniques have been developed, please refer to [218] for a broad overview.

2.4.3.2 Texture classification with neural networks

A number of neural applications to texture classification have been reported in the literature, that rely either on the sole use of texture, or on the combined use of texture and local shape. A variety of texture descriptors and neural architectures have been investigated (see [63] for a review). Among others, Worth *et al.* [221] use analogue constraint satisfaction networks to segment MRI. Van Hulle *et al.* [205] process textures with a modular network. Hall *et al.* [89] segment tumors, edema and normal tissues. Wang *et al.* [213] use probabilistic neural networks, etc.

Most, if not all, of these approaches however partition the input image voxels into tissue type classes (gray matter, white matter, CSF, fat, ...) independently of the target structure. Yet, as argued above, the particular set of textures that a given anatomical structure exhibits is most useful in guiding the segmentation. In addition to *a priori* information about the texture of the target structure, experts also rely on intensity and textural information gathered from its vicinity to guide the segmentation process. Clearly, in Figure 2.2, the typical darker and homogeneous appearance of the ventricle is a useful clue in finding the boundary of the caudate nucleus.

We have developed a neural network architecture that draws on this anatomical experience.

Our neural classifier has two stages:

Stage I. A hybrid master classifier/mixtures of experts neural architecture is used in conjunction with adaptive spline neurons (ASN) to classify the textures of the input image and implement two internal feature selection procedures. We also use a dynamic scheme to evolve a learning set concentrated along the classification boundary between the target texture set and the others.

Stage II. We then use a multi-scale neural network, subject to node pruning, to further contrast the target structure and correct the possible mistakes of the first stage by introducing spatial awareness in the fashion of Markov random fields.

2.4.3.2.1 Hybrid master classifier / mixtures of experts (stage I)

Feature combination: Once computed, the output of the texture descriptors can be combined in a number of ways.

Merging: They can be concatenated into a single feature vector to be classified all together by the network. This poses the problem of normalization when some descriptor outputs have more dimensions than others.

Voting: Each feature can be classified separately and the results subsequently combined with a voter (with simple or absolute majority). The main constraint of this technique comes from the necessity for the classifiers to be statistically independent, lest the decisions of the correlated classifiers should always prevail to the detriment of the overall performances.

Master/slave: A master classifier [99] can be used to classify the results of several slave classifiers. The main problem is that each slave has to learn all the patterns in the training set, which might decrease the classifier capabilities as a feature from the learning set may not be separable in all the descriptor spaces.

Associative switch: In this approach, an independent classifier (the associative switch [222]) decides which feature filter is considered is to provide the final result. This technique is particularly computationally efficient as it does not require all the descriptors to be evaluated.

In stage I, feature combination is achieved by a neural architecture which blends together a master classifier (in the form of a 2-layered feed-forward perceptron [172]) and a specialized mixtures of experts [212] architecture (see Figure 2.9).

A specialized mixture consists of a number of experts (k here), each being a multi-layer perceptron (MLP) taking input from a number of texture descriptors, and a decision module, the gate, which receives input directly from the MR or histological image. With one output per expert, this gate controls the extent to which each expert contributes to the final classification result: the i^{th} output of the gate represents the confidence in the classification result of the i^{th} expert. This enables the specialization of the experts in the classification of the features they separate best.

In the original specialized mixture of expert design (see Figure 2.10(a)), Walter *et al.* [212] accumulated the decisions of the experts weighted by the output of the gate, starting with the experts with highest gate confidence. This scheme enabled them to evaluate the expert outputs

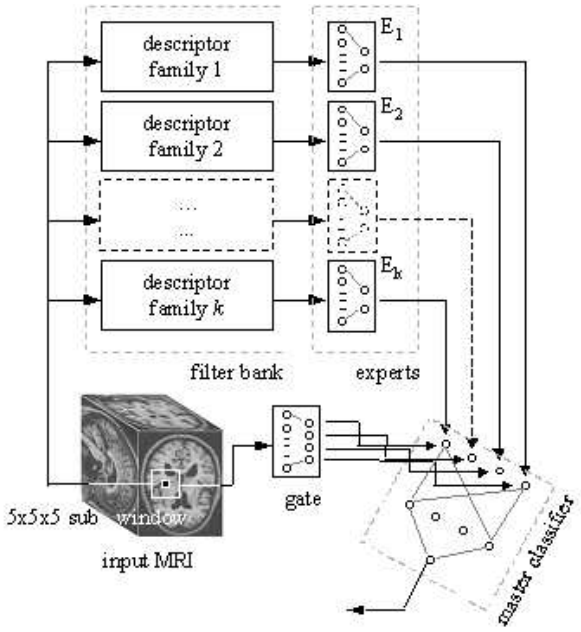


Figure 2.9: Stage I of our neural classifier.

on an as-needed basis, as determined by the confidence computations of the gate, which greatly penalized any weak arbitration by the gate. In spite of the authors' opinion that the feature fed to the gate did not have to contain "sufficient information to permit a complete classification" as the experts too contributed to the final decision, we believe it is too great a responsibility to give a single sub-network, linked to a single descriptor. In our design, we thus replaced the accumulator (and the threshold algorithm that the authors had to add to it to stop the computation when it became sufficiently "trustworthy") by an MLP which acts as a master classifier. The MLP receives input from the gate-weighted experts. This enables the entire pool of experts to contribute to the determination of the best expert for each feature, which releases the burden of the gate, even though the latter remains the main decision making component. As opposed to Walter's approach [212], this however requires a finer selection of the input descriptors.

Coarse/fine feature selection: Given the high learning cost of our network (or of most network in general), the SFFS algorithm, which requires a new evaluation of the performances of the classifier, that is, a new network training phase, with each new selection of feature descriptors, would lead to computationally intractable selection processes. Fine feature selection is here performed by the expert networks, which are fed a significant number of feature values from families of texture descriptors (one family per expert) and return a single output value. We then use the SFFS algorithm only to select the feature *families*. This implements a coarse feature selection process. Note that the MLPs, though their number of input units may vary, have a standard number of hidden neurons (10) which helps "normalize" the overall feature vector.

To take full advantage of the combination of both architectures (the mixtures and the MLP), we use adaptive spline neurons instead of standard sigmoid activated ones for the input layer of the master classifier network. The activation function of an ASN consists of a parameterized

Catmull-Rom spline, whose coefficients are adjusted during the training process (see [85] for a detailed presentation). By adapting the activation function of each ASN together with its weights, we in fact implement a second intrinsic feature selection scheme: when a neuron's activation function comes close to the horizontal axis, for instance, the corresponding feature expert plays only a minor role in the classification process. The initial activation functions take the form of sigmoid functions: $F = \{F_i : x \mapsto F_i(x) = \tanh(x)\}$ and consist of 11 control points.

Learning process: The learning process consists of 3 consecutive steps:

learning process stage I

---

step 1. We replace the master classifier by a two-level hierarchical mixture of experts, as described by Jordan and Jacobs in [109]. We then train the network using Walter *et al.*'s modified version [212] of their EM algorithm (Figure 2.10(a)).

This step actually implements two parallel learning processes: the gate learns which expert is best suited for each pattern from the learning set, and each expert specializes in the pattern subset to which it is most appropriate. The EM algorithm works as follows:

The goal of the training process is to maximize the likelihood H of the neural network for the learning set L, where H is written:

$$H = \prod_{i=1}^{N} \sum_{j=1}^{k} P\left(E_j | g_j^i, V\right) P\left(c_i | s_j^i, W_j, E_j\right) \tag{2.28}$$

k is the number of selected descriptor families, V is the weight matrix of the gate, W_j are the weight matrices of the k experts, g_j^i is the output of the gate for the i^{th} training pattern.

$P\left(E_j | g^i, V\right)$ is the so-called *a priori* probability, that is, the probability of expert E_j to deliver the correct classification based only on the output of the gate g_j^i for the i^{th} training sample. $P\left(c_i | s_j^i, W_j, E_j\right)$ is the probability of expert E_j, which we can compute as a function of the error with respect to the desired output c_i.

H is minimized using a classical Expectation Maximization algorithm with the following *a posteriori* probabilities:

$$\zeta_j^i = \frac{P\left(E_j | s_j^i, V\right) P\left(c_i | g_j^i, W_j, E_j\right)}{\sum_k P\left(E_k | g_j^i, V\right) P\left(c_i | s_j^i, W_k, E_k\right)} \tag{2.29}$$

As argued by Walter *et al.*, these are high for experts with high *a priori* probability and small error on the learning set, and low for experts with small *a priori* probability and large error. They can thus be seen as an improvement on the *a priori* probability induced by the gate: they then become the desired outputs of the gate. Alternatively, they represent the suitability of the expert to which they are associated for the patterns from the learning set, and serve to specialize the expert on the pattern subset to which it is most appropriate. We used the iterative EM minimization algorithms presented in [212] to maximize H.

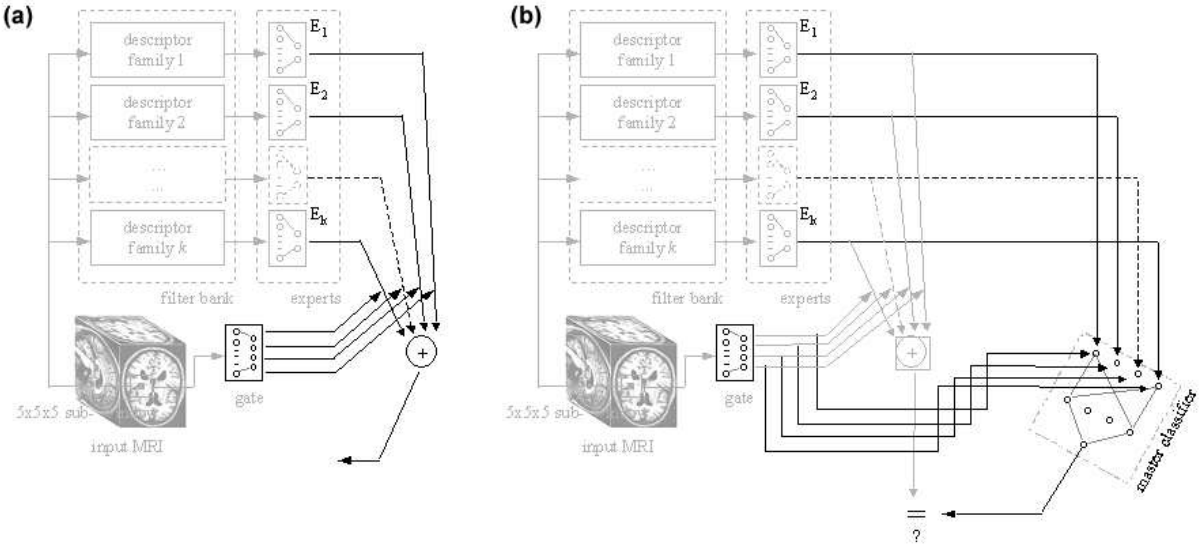


Figure 2.10: Learning steps for stage I.

step 2. The master classifier with ASNs is trained to mimic the behavior of the two-level mixture of experts when receiving the inputs from the expert networks and the gate. We use the gradient-based learning rules detailed in [85] (Figure 2.10(b)).

step 3. The entire network is re-trained by a gradient descent algorithm based on that of Guarnieri *et al.*[85]: we use an adaptive weight learning rate where the weight is small for the experts and the gate and high for the master classifier.

Step 3 enables us to specialize the master classifier with the ASN (and thus increase the classification performance) while still improving on the mixture of experts. Experimental results showed that this learning method yielded better results than those obtained when using a modified version of Jordan's gradient descent algorithm adapted to handle the ASN.

Note that modifying the activation function while training the network not only triggers faster convergence, but also enables the learning process to better escape local minima. Figure 2.11 shows the characteristic shapes of a few ASNs upon completion of the learning process, for the classification of the hippocampus (same learning set as above: 3 MRI with 50 voxels per MRI). Note the flat shape of the activation function of the neuron linked to the Gabor family at scale 1, which reflects its poor classification capabilities.

Dynamic learning set. A major problem of neural networks when used as classifiers lies in their lack of good rejection capabilities: a neural network has to assign every input feature vector to a class even though some vectors may not belong to any of the learnt classes. During the learning phase, classification boundaries are constructed that may acutely depend on the distribution of samples in the learning set. In our case, it is easy to select learning patterns belonging to the set of textures of the target anatomical structure and much more difficult to give a "representative" set of those which do not belong to it. We have implemented a dynamic

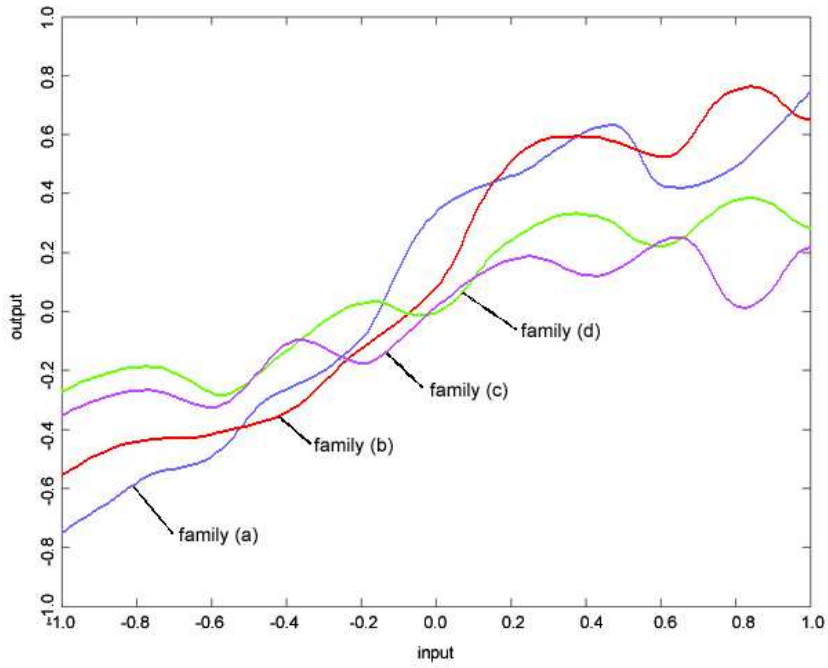


Figure 2.11: Typical activation function after the learning process of stage I for the classification of the hippocampus in a T1-weighted MRI: (a) Haralick family with u=(2,2,2); (b) Haralick family with u=(2,0,0); (c) Holder family with "adaptive iso"; (d) Gabor family at scale 1.

learning approach adapted from [175] and [112], which helps design a learning set that lies along the classification boundary. This both increases the classification performances and decreases the number of samples to include in the learning set. It works as follows:

dynamic learning approach

- The three-stepped learning process described above is run with the standard initial learning set (50 voxels drawn at random on each side of the structure boundaries).
- Then, we modify the learning set by inserting in it 90% of the misclassified samples and an equal number of correctly classified target voxels.
- Finally, we repeat the three-stepped learning process.

Additional performance improvements can be obtained by repeating this process, to the detriment of the learning speed, though.

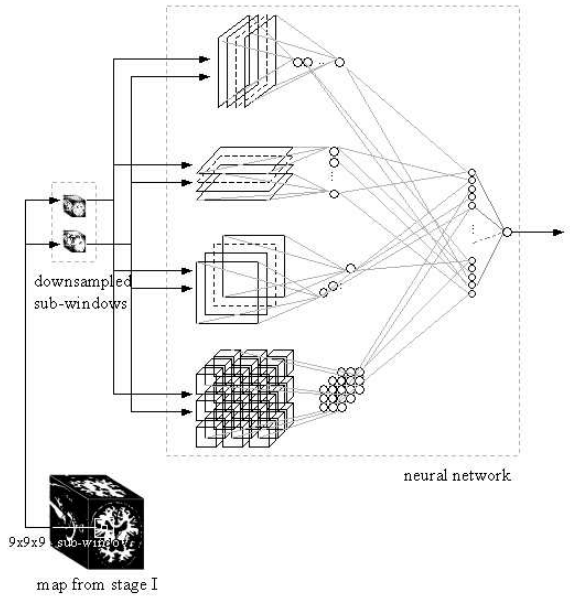


Figure 2.12: Stage II of our neural classifier.

2.4.3.2.2 Multi-scale neural network (stage II)

The first classification stage employs textural information in the input images in a very local fashion: a somewhat limited $5 \times 5 \times 5$ sub-window is used to compute the texture features and the voxels are considered independently from each other. Hence the need for a second classification stage whose task is to take into account a larger neighborhood for each voxel and make use of the nearby structures to help the classification task and to regularize the shape of the final classification map.

Architecture: We use a multi-scale neural network with standard sigmoid neurons (see Figure 2.12). A larger sub-window ($9 \times 9 \times 9$ pixels) and a downsampled version of it (scale 2) are fed into a layer of hidden units. We adapted an input connection pattern commonly used in the speech and character recognition literature. It consists of 4 distinct blocks. The first block looks at *X=constant* slices in the 3D data (each of the 9 neurons in this hidden layer block is fully connected to the 9×9 input neurons of the associated slice the input layer); the second and third block look in a similar fashion at *Y=constant* and *Z=constant* slices respectively; finally the fourth block looks at $3 \times 3 \times 3$ input sub-regions. These specialized groups of neurons are responsible for detecting local features and relative positions of nearby structures. The blocks are fully connected to a second hidden layer (30 neurons) connected to the output.

As the overall number of connections is quite large (~80000 weights), we use node pruning [135] to further adapt the architecture of the neural network to the classification problem at hand. This regularization issue, which involves controlling the complexity of the network by reducing the overall number of connections, is a pivotal network design element as there are still no general prescription for tuning a network architecture to a particular problem. Namely, if too many weights are used, then overfitting phenomena may decrease the generalization capabilities of the network, while, conversely, a network with too few weights may not be able to learn

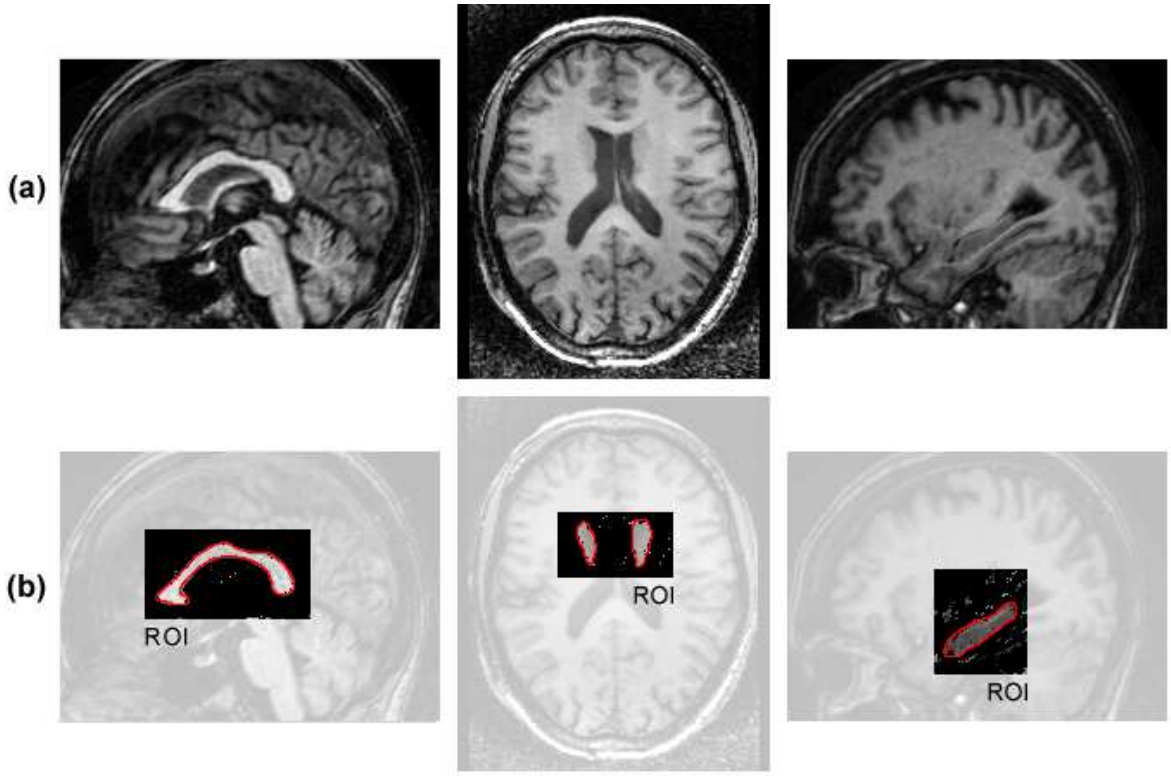


Figure 2.13: Neural classification of corpus callosum, caudate nucleus and hippocampus: (a) input MRI; (b) neural extracted structures with true outlines superimposed.

adequately. *In fine*, the Mao algorithm [135] usually leaves the network with between 2000 and 3000 weights.

Learning process: A classical back-propagation algorithm with momentum (see [62] for details) is then used to train the network on the final learning set from stage I.

ratio of misclassified pixels / false positive / false negative	*corpus callosum*	*caudate nucleus*	*hippocampus*
mean	5.3 ± 1.3 2.7 ± 2.3	11.2 ± 1.5 4.2 ± 1.6	20.5 ± 3.7 7.8 ± 3.1
maximum	8.3 5.3	13.5 6.3	25.4 12.3

Table 2.4: Texture classification of three target structures in T1-weighted MRI with our hybrid neural network.

2.4.3.3 Application to biomedical texture classification

MRI: We applied our neural classifier to three target structures: the corpus callosum, the caudate nucleus and the hippocampus. The settings were similar to those presented above: 3 instances of the target structure in each learning set with about 50 voxels selected both inside and outside the boundaries. Again, we used a $5 \times 5 \times 5$ window for the texture descriptors. The SFFS algorithm selected $k = 2$ feature families for the corpus callosum (Haralick descriptors with translation $u = (2, 2, 2)$ and Hölder family with "adaptive iso"), $k = 3$ families for the caudate nucleus (Haralick descriptors with translation $u = (2, 2, 2)$ and $u = (2, 0, 0)$ and Hölder family with "adaptive iso") and $k = 5$ families for the hippocampus (Haralick descriptors with translation $u = (2, 2, 2)$ and $u = (2, 0, 0)$, Hölder family with "adaptive iso" and Gabor family at scale 1 and 2).

Figure 2.13 displays a few classification results for 3 structures out of the 16 test corpora callosa, 20 caudate nuclei and 20 hippocampi. Quantitative results are reported in Table 2.4.

Comparison with LDA and SVM results demonstrates the superiority of the learnt non-linear approach. Overall, classification results were good: excellent for the corpus callosum with almost no errors at all, very good for the caudate nucleus (with a significant improvement over the SVM results) and somewhat good with the hippocampus, a difficult structure to segment in MRIs (the large standard deviation indicates that a few hippocampi were a lot more difficult to classify than others, a classical behavior when facing poorly contrasted structures).

Note that these results do not have to be perfect as the classification maps are used as *constraints* for the deformable templates, not as deformation target images (a number of post-processing steps -mathematical morphology cleaning for instance- are also applied to the classification maps prior to their use as constraints).

Cryo- and histological sections: In addition to MRI filtering (our main focus here), we have also applied the developed neural classifier to the segmentation of mouse brain tissues in cryosections and in Nissl-stained histological sections, within the framework of the Mouse Atlas Project[12].

A 2-D version of our hybrid neural network approach was first trained on a database of respectively 10 tissue/background labeled cryosections and 5 tissue/background labeled Nissl-stains sections. We used the SFFS algorithm to select $k = 1$ descriptor family for Nissl sections (Haralick descriptors with translation $u = (2, 2)$) and $k = 3$ descriptor families for cryosections (Haralick descriptors with translation $u = (2, 2)$, Gabor family at scale 1 and scale 2).

Figure 2.14 displays a few segmented Nissl-stained and cryosection images. Overall, the neural network had no difficulties segmenting the mouse brain tissues in the Nissl sections. Its performances for cryosections were less spectacular: the classifier was sometimes fooled by the tissues seen through the embedded parafin by transparency. The quality of the processed images was however judged qualitatively sufficient for them to be included in the atlas[13].

[12] The MAP (Mouse Atlas Project) project aims to produce a powerful imaging framework to house and correlate gene expression with anatomic and molecular information drawn from traditional and novel imaging technologies. Within the framework, genetic data acquired *in situ* or from tissue samples will be directly comparable with structural, metabolic, physiologic and genetic imaging data from well-characterized populations (see http://www.loni.ucla.edu/MAP/index.html for details).

[13] Note that a number of post-processing steps (mathematical morphological cleaning, contrast equalization, ...) were conducted prior to the insertion of the segmented images into the atlas.

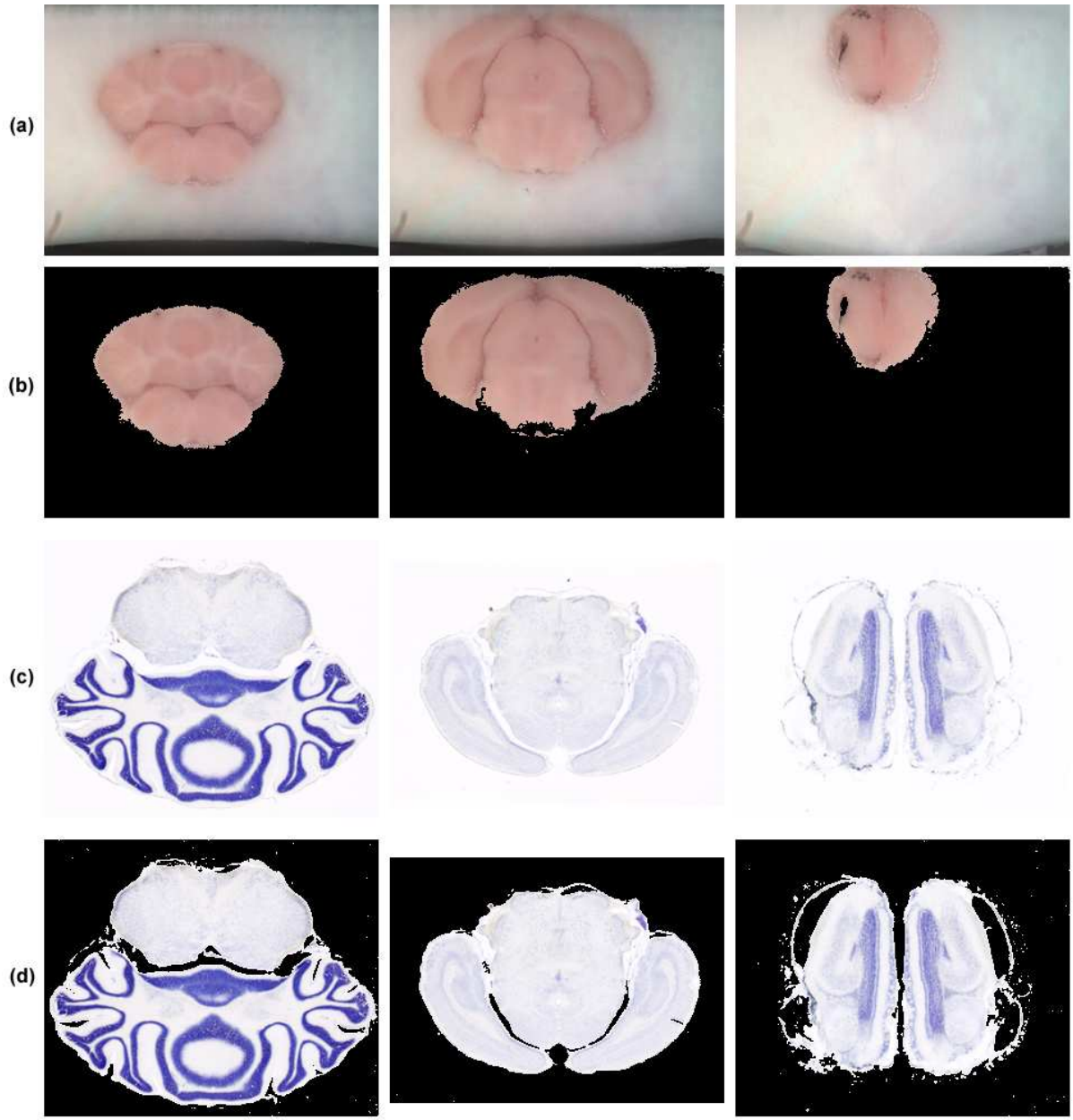


Figure 2.14: Neural classification of mouse brain tissues: (a) input cryosections; (b) extracted brain tissues; (c) input Nissl-stained histological sections; (d) extracted tissues.

2.4.3.4 Sensitivity to imaging characteristics

The sensitivity of the neural filtering approach to imaging parameters was evaluated on a series of MRIs acquired on different scanners (from three different medical hospitals), with different sequences for various individuals. 12 images were available with 3 MRIs per acquisition protocol.

We applied our neural approach to the classification of the corpus callosum, caudate nucleus and hippocampus. We tested the algorithm with the standard learning set whose labelled samples belonged to the homogeneous batch of MRIs we have used so far, with a learning set adapted

to each acquisition protocol (with labelled samples coming from similar acquisitions in terms of parameters and scanners), and with a mixed learning set with labelled samples coming from the entire set of heterogeneous MRIs. In all three cases, the performances were evaluated on a test subset different from that in the learning set.

Table 2.5 compares the performances of all three approaches on the 10 MRIs. As could be expected, even though the neural classifier with a standard learning set behaves correctly (with only a 10 to 15% drop in accuracy), performances were improved when an adapted learning set was used. With a mixed learning set, we observed adequate (even though inferior) performances. Again, this comes as no suprise as increasing the domain of application of a classifer cannot but decrease its performances locally.

Please refer to [1] for a more systematic approach of the dependence of texture analysis on the acquisition parameters.

2.5 Conclusion

We have presented in this chapter a selection of texture filtering approaches, which produce classification maps from input 3-D MR data or 2-D histological sections. These classification maps will serve as constraints on the evolution process of our deformable templates, as described in Chapter 4. We have proposed three classes of classifiers: linear (LDA), linear in a non-linear projective space (SVM) and adaptive non-linear (neural network), with an increase both in terms of performances and of computing resources.

A priori information about the classification task was introduced in the form of a learning set of *a priori* segmented target structures. These filtering methods are therefore *semi-explicit* as a learning algorithm automatically fits the parameters of a classification function from a manually-built learning set.

The use of neural networks helped design a better classifier, owing to the ability of neural approaches to adapt the structure of the decision boundary search space to the classification problem at hand, along with the search of the best fitting parameters (most especially thanks to the dynamic learning set). This adaptability compensates for the absence of truly exploitable *a priori* information about the shape of the classification boundary.

Note that the use of other modalities in addition to the sole T1 MRI could most certainly enhance the performances of our classifiers. Owing to its ability to reveal fiber orientations, diffusion tensor imaging for instance could be of precious help, provided suitable techniques are used to correct the associated noise and artifacts.

Finally, as stated in the No Free Lunch Theorem[14] and the Ugly Duckling Theorem[15], there is neither a "best" problem-independent learning system nor a best feature representation. To perform adequately, a classification system requires *a priori* information as to the characteristics of the input data to be classified (its category structure for instance or the density of distribution of its features, etc.). Designing a good classifier is then intimately linked to the ability of the designer to understand the nature of the classification problem at hand, an art rather than a science, really.

[14] The No Free Lunch Theorem states that in the absence of *a priori* information about the classification problem at hand, there are no reasons to prefer one learning classifier to another.

[15] The Ugly Duckling Theorem states that in the absence of *a priori* information about the classification problem at hand, there is no "best" discriminating feature. That is to say, the number of discriminating predicates shared by any two different patterns is constant and does not depend upon the choice of these two patterns.

ratio of misclassified voxels / false positive (%) / false negative (%)	standard	learning	set
	c.c.	*caudate*	*hipp.*
mean	5.9 ± 1.4 3.0 ± 2.2	12.6 ± 1.6 4.5 ± 2.0	23.1 ± 3.8 8.8 ± 3.4
maximum	7.5 6.0	14.5 7.1	27.3 13.1

ratio of misclassified voxels / false positive (%) / false negative (%)	adapted	learning	set
	c.c.	*caudate*	*hipp.*
mean	5.2 ± 1.4 2.8 ± 2.1	11.5 ± 1.8 4.4 ± 1.8	20.9 ± 3.5 7.6 ± 3.0
maximum	8.5 5.5	14.1 7.0	26.1 13.0

ratio of misclassified voxels / false positive (%) / false negative (%)	mixed	learning	set
	c.c.	*caudate*	*hipp.*
mean	5.5 ± 1.2 2.8 ± 2.2	12.2 ± 1.5 4.5 ± 1.8	22.3 ± 3.6 8.2 ± 3.2
maximum	7.0 5.3	14.2 6.5	26.0 12.5

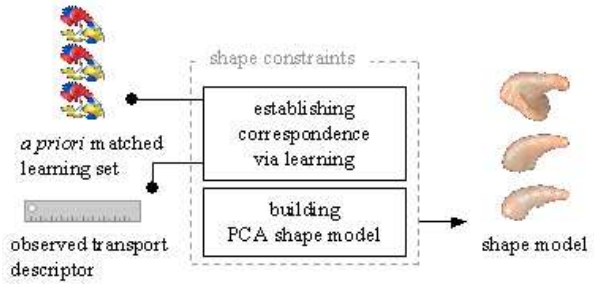
Table 2.5: Neural classification of the corpus callosum in MRI with different imaging characteristics and different learning sets.

Chapter 3

Statistical Shape Modeling
Modèles de Forme

Hamlet: Do you see yonder cloud that's almost in shape of a camel?
Polonius: By the mass, and 'tis like a camel, indeed.
Hamlet: Methinks it is like a weasel.
Polonius: It is backed like a weasel.
Hamlet: Or like a whale?
Polonius: Very like a whale.
William Shakespeare, *Hamlet*, Act 3-scene 2

Nous nous intéressons dans ce chapitre à l'étude statistique de la variabilité anatomique des structures corticales. Nous décrivons comment un modèle statistique de forme peut être établi, pour chacune des structures cibles, au moyen de l'analyse en composantes principales. Ces modèles serviront au chapitre suivante en tant que contraintes de forme imposées aux patrons déformables. Cette technique de modélisation requière cependant que les différentes instances de chaque structure cible soient reparamétrées, c'est à dire, qu'un champ de correspondence dense doit être établi a priori *entre elles. Nous introduisons en conséquence, ici aussi, de l'expertise médicale* a priori *par le biais de deux nouvelles techniques de reparamétrisation, en l'occurence deux algorithmes de reparamétrisation par apprentissage. Ces derniers établissent entre les instances un champ dense aux caractéristiques similaires à celles extraites d'un ensemble d'apprentissage de champs de correspondence* a priori. *Nous développons également un nouveau descripteur de forme, l' "observed transport descriptor", suivant en cela une liste de spécifications établie pour faciliter cette tache de reparamétrisation. En particulier, ce descripteur présente une plus grande robustesse au bruit et un pouvoir de discrémination accru.*



Overview of shape model building.

Even though, as illustrated above, a given brain structure can present a wide variety of forms, it seems that the notion of biological shape is reasonably well explained by a statistical description over a large population. Consequently, model-based approaches have attracted considerable attention [83, 202, 188, 46]. From a statistical description of the shape of a target structure (the so-called shape model), we can indeed derive a constraint to impose on the evolution process of our deformable templates. The most commonly employed techniques for building shape models use principal component analysis as a means to extract from a learning set of structure instances its most representative modes of variation and a mean shape. This process however requires that correspondences between the various instances be established *a priori*, a difficult problem in and out of itself, to which much attention must be devoted, for the quality of the shape model intrinsically depends on that of the established correspondences. Opening with a glance at the notion of shape, this chapter concerns itself with the building of the statistical shape model of a structure, via the controlled matching of its instances.

3.1 Introduction

As demonstrated in Chapter 2, anatomical structures are rarely clearly defined in MRI or histological slices. While noise, partial volume effects and quantization artifacts often significantly decrease the precision with which a structure's boundaries can be outlined, they can sometimes completely obliterate them. There is then no sufficient information in the image alone to allow a correct delineation. In this case, human operators rely on prior information about the shape of the target structure to interpolate the missing boundary pieces. Similarly, the use of a shape model allows the introduction of *a priori* medical knowledge about the shape of the target structure into the automated search for its borders.

A popular technique in the literature, principal component analysis (PCA) is often chosen to build, from a set of matched instances of the target structure, this shape model. The latter can then be used as a shape constraint on deformable templates. The templates are consequently not only constrained by the number of degrees of freedom dictated by their geometric representation, but also in that they must be linear combinations of their associated shape modes of variation. A pre-requisite to the building of the model via PCA, the matching of the structure's instances is a non-trivial and still open problem, all the more pivotal since the quality and representativity of the shape model crucially depends on the care with which the instances are matched.

After a brief overview of standard psycho-cognitive and mathematical approaches to shape (Section 3.2) and the underlying correspondence issue (Section 3.2.3), we introduce a standard non-learning approach to the object matching problem in Section 3.3 and a few learning approaches in Section 3.3.5 before presenting the classical principal component analysis approach to shape statistics (Section 3.4).

3.2 Shape

As is the case with texture, the concept of shape, although very common, is still in need of a satisfying definition, even centuries after the first inquiries into its nature. In Meno [161] for instance, Plato, through the voice of Socrates, first defines a shape as "the only thing which always follows color", a description which does not satisfy Meno. Asked to be more precise about what he means, Socrates then gives a definition "as for example in geometry": "I define figure to be that in which the solid ends; or, more concisely, the limit of solid" [1].

As pointed out by Veltkamp [208], shape is thus classically considered as "something geometrical", a definition that leaves a large part to psychovisual analysis.

3.2.1 Shape and the human visual system

One of the most advanced cognitive architectures of the mind, the human shape analysis system has been the subject of a large number of studies (see [47, 231, 132, 163] for instance). Their long term goal consists in developing an artificial equivalent in terms of visual perception and cognition capabilities. This is an ambitious endeavour which draws on the experience of a wide range of domains, from psychology and cognition to mathematics and engineering. As an illustration of this variety, we introduce in this section a short selection of classical and modern theories of visual perception (please refer to the works of Loncaric [131], Zusne [231] and Hake [87] for broader overviews).

3.2.1.1 Psycho-cognitive approaches

Gestalt: One of the most revolutionary psychological approach to the perception problem, the Gestalt theory [120] deals with shape in a remarkably consistent and systematic fashion. At a glance, it consists of a series of laws for visual shapes, which unfortunately also makes it a non-computational approach. Among these (many) laws, we quote from [231]:

- All visual forms possess at least two aspects, a figured portion called figure and a background called ground.
- Visual forms are transposable (with respect to translation, size, orientation, and color) without loss of identity.
- Forms may fuse to produce new ones.
- Space is anisotropic, it has different properties in different directions.
- etc.

[1] See Appendix A for an excerpt of the dialogue.

Reductionist approaches: A neuro-psychological theory of behavior was proposed by Hebb [93] where the pivotal role of neural structures (in particular cell assemblies: groups of mutually excitable neurons) and perception learning was demonstrated. In contrast to the holistic view of the Gestalt theory which treats form as a whole, Hebb submits that it is perceived as a collection of simple elements (lines and angles), recognized thanks to eye movements.

A similar stance was voiced in the work of Goldmeier [77], where he analyzes which clues people use in matching shapes. In particular, he studied which constraints (symmetry, area, verticality, etc.) were favoured by the human visual system. These properties, which he calls singularities, represent the "perceived geometrical essence" of the shape. They are special cases of more general representations. For instance, people's attention is attracted to vertical or horizontal lines, which are special cases of arbitrarily-oriented lines. Parallel lines are another of these singularities. The combination and ranking of these singularities enables the description of a shape with only a few fuzzy qualitative terms which, in turn, allow for flexible comparisons between shapes.

3-D cognition: Central to Gibson's theory of visual perception [76] is the notion of space as a real 3-D visual entity (as opposed to the inherently 2-D view of the world of the Gestalt approach, which only considers 2-D projections of 3-D objects) characterized by the forms it contains. Equally important is the concept of real world stimuli, which are properties of surfaces and are the basis for visual perception. Stimuli are physical gradients such as change in size of texture element (notion of depth), degree of convergence of parallel edges (notion of perspective), hue, shading, etc. This framework also offers a comprehensive taxonomy of forms, which can be solid, surface, outline, pictorial, perspective, etc.

3.2.1.2 Engineering approaches

In contrast to psychological or cognitive approaches which often remain qualitative, a number of more computational theories have been proposed.

Primal sketch: Marr *et al.* [136, 137, 140] have investigated in detail the various modules that shape the human visual system. Aside from the impressive body of work on the so-called "shape from x" approaches (shape from shading [98], from contour [203], etc.), this lead to the concept of "primal sketch", a paradigm of early vision: discontinuities in the first and second derivatives of an input image (the Marr-Hildreth edge detector [138] uses the zero-crossing of the Laplacian of the 2-D Gaussian of the image) are analyzed to produce this edge-based sketch.

Dynamic hierarchical approach: Contrary to classical theories which consider all pictorial details as a global shape feature, Koenderink et al. [119] proposed a dynamic multi-scale shape model, which exploited the well-known psycho-visual concepts of order and relatedness (at various resolutions). They developed a morphogenesis technique where partial differential equations helped describe the "morphogenetic sequence" of an input shape, by describing its evolution through a variety of resolutions.

Symmetry-Curvature evolution: The Symmetry-Curvature theorem, which links two fundamental concepts of visual perception, was demonstrated by Leyton in [125]. It states that any curve section that has only one curvature extremum has exactly one symmetric axis which terminates at this extremum. Leyton used it as a means to characterize the

deformation process of shapes [126]. Namely, his theory contends that all shapes are circles which have been subjected to various transformations (pushing, pulling, stretching, etc.). Symmetry-curvature can then be used both to infer the shape history of a given object and to infer the shape evolution between two objects.

3.2.2 Two computational definitions of shape

From a more mathematical, but not necessarily more objective, point of view, a shape can be defined as an equivalence class under a group of transformations. That is to say, given a similarity measure (or measure of resemblance), which must be invariant under this transformation group, the shape of a pattern is the pattern modulo the action of the group [82]. However, this particular framework is only interesting for establishing when two shapes are the *same*. For our applications this is insufficient: we need a measure of the *distance* between two shapes. The competing statistical theories of Bookstein [20] and Kendall [116] are two attempts at establishing a mathematical framework within which the notion of shape distance is more rigorously defined.

Kendall's approach: Let R be the space of all shapes, and G be a group of transformations acting on R. Let R_G be an equivalence relation:

$$R_G : \forall\ (r_1, r_2) \in R,\ \ r_1\, R_G\, r_2 \Leftrightarrow \exists g \in G\ s.t.\ r_1 = g.r_2 \tag{3.1}$$

Then, the shape space S is the quotient space $R\,|R_G$.

We are looking for a distance d on the shape space, derived from the Euclidean distance, which must be independent from orientation and scale.

In the particular case where $R = \mathbb{R}^{kn}$(statistics on k landmarks of $\mathbb{R}^n$), let us consider 2 icons (we call an icon a representative of a given orbit): $X = (x_1, \ldots, x_k)$ and $Y = (y_1, \ldots, y_k)$, their Euclidean distance is $d(X, Y) = \|X - Y\|_2^2$.

Kendall defines the full Procrustes distance as the minimum of $d(g.X, h.Y)$ where g and h travel G, provided a constraint on the norm of either X or Y is added, to prevent the distance from being always 0. This effectively defines a metric on the shape space.

More formally, we first remove the influence of translations from the set of icons:

$$F^{kn-n} = \left\{ X = (x_1, \ldots, x_k) \in \mathbb{R}^{kn}\ s.t.\ \Sigma_i x_i = 0 \right\} \tag{3.2}$$

To remove the influence of scale, we define the set of pre-shapes (which is isometric to the sphere $\mathbb{S}^{kn-n-1}$):

$$\mathbb{S}_*^{kn-n-1} = F^{kn-n} \bigcap \mathbb{S}^{kn-1} = \left\{ X \in \mathbb{R}^{kn}\ s.t.\ \Sigma_i x_i = 0 \text{ and } \|X\|_2 = 1 \right\} \tag{3.3}$$

Let Σ_k^n be the set of the orbits of $\mathbb{S}_*^{kn-n-1}$ under the action of rotations. In 1-D, $\mathbb{S}_*^{kn-n-1}$ is isomorphic to the sphere $\mathbb{S}^{k-2}$ and in 2-D, to the complex projective space $P^{k-2}\mathbb{C}$. The ambient metric structure of R (naturally defined on $\mathbb{S}_*^{kn-n-1}$ since it is a differential manifold) can then be entirely propagated to Σ_k^n. In 3-D, a differential structure can be

defined only almost everywhere as singularities appear, the defined space is however still usable.

The *full* Procrustes distance is then written:

$$\begin{aligned} \forall X, Y \in \mathbb{S}_*^{kn-n-1},\ d_{full}^2(X,Y) &= min_{r>0,\Gamma\in SO(n)} \left\| Y - rX\Gamma^T \right\|^2 \\ &= 1 - \left(\text{Trace}\left(\Lambda\right)\right)^2 \end{aligned} \tag{3.4}$$

where Λ is the diagonal matrix from the singular value decomposition of $Y^T X$, and $SO(n)$ the special orthogonal group.

Although this formulation is recommended for computational reasons, it is not derived *directly* from the Euclidean distance on the shape space, unlike the following Procrustes distance:

$$\begin{aligned} \forall X, Y \in S_*^{kn-n-1},\ d^2(X,Y) &= min_{\Gamma\in SO(n)} \arccos\left(Y, X\Gamma^T\right) \\ &= \arccos\left(\text{Trace}\left(\Lambda\right)\right) \end{aligned} \tag{3.5}$$

which shares the same topology on the shape space as both distances are equivalent: $d_{full} \leq d \leq \frac{\pi}{2} d_{full}$.

Boosktein's approach: While Kendall's approach is global, Bookstein proposes a more local method and defines a local metric based on the estimation of the costs associated with infinitesimal changes in a shape.

The influence of translation, rotation and scale is first discarded by the use of a special coordinate system (Bookstein-coordinates). Then linear transformations are used to represent infinitesimal variations of the shape in a bi-univocal fashion (n points are kept fixed while the others move). The SVD decomposition of the square norm of an infinitesimal variation of shape dS serves to compute the anisotropy of its eigenvalues and defines the new distance:

$$\|dS\|^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\mathbb{E}(\lambda) - \lambda_i\right)^2 \tag{3.6}$$

where $\mathbb{E}(\lambda)$ is the expected value of the eigenvalues λ_i . $\|dS\|^2$ is then nothing but the variance of the set of eigenvalues, which gives a zero cost for any similarity, and increases with the degree of anisotropy.

3.2.3 The need for correspondence

Both approaches however require that correspondences between shapes be available *a priori*, a constraint shared by most methods in the literature. This correspondence problem permeates a broad range of image related fields, from signal processing to pattern recognition. In computer vision for instance, the search for target patterns often requires matching a given template to pictorial elements in an input image [159]. In computer graphics, matched objects may be used to derive a series of intermediate shapes to "morph" one object into the other [113], etc. In our application, the objects to be matched are instances of a given anatomical structure, for which a statistical model is desired.

3.3 Object Matching

We approach the issue of object matching as a process of computing a dense correspondence field between two objects.

3.3.1 Correspondence field between two objects

At a glance, defining a correspondence between two objects entails finding in them pairs of corresponding elements that share particular similarities, in terms of shape, position, or both. More formally, given two objects $\mathcal{O}_1$ and $\mathcal{O}_2$ with any *a priori* parameterizations represented by two functions O_1 and O_2:

$$O_1 : \left| \begin{array}{ccc} I_1 \subset \mathbb{R}^m & \rightarrow & \mathbb{R}^n \\ x & \mapsto & O_1(x) \end{array} \right. , \; O_2 : \left| \begin{array}{ccc} I_2 \subset \mathbb{R}^m & \rightarrow & \mathbb{R}^n \\ x & \mapsto & O_2(x) \end{array} \right. \quad (m \leq n) \tag{3.7}$$

we are looking for a reparameterization of O_1 and O_2, that is, for two diffeomorphisms f_1 and f_2 , such that $O_1^* = O_1 \circ f_1$ and $O_2^* = O_2 \circ f_2$ and

$$\forall x_1 \in I_1, \forall x_2 \in I_2, \, x_1 \text{“close to”} \; x_2 \Rightarrow O_1^*(x_1) \text{ “very similar to” } O_2^*(x_2) \tag{3.8}$$

where “very similar to” is defined with respect to a given similarity metric.

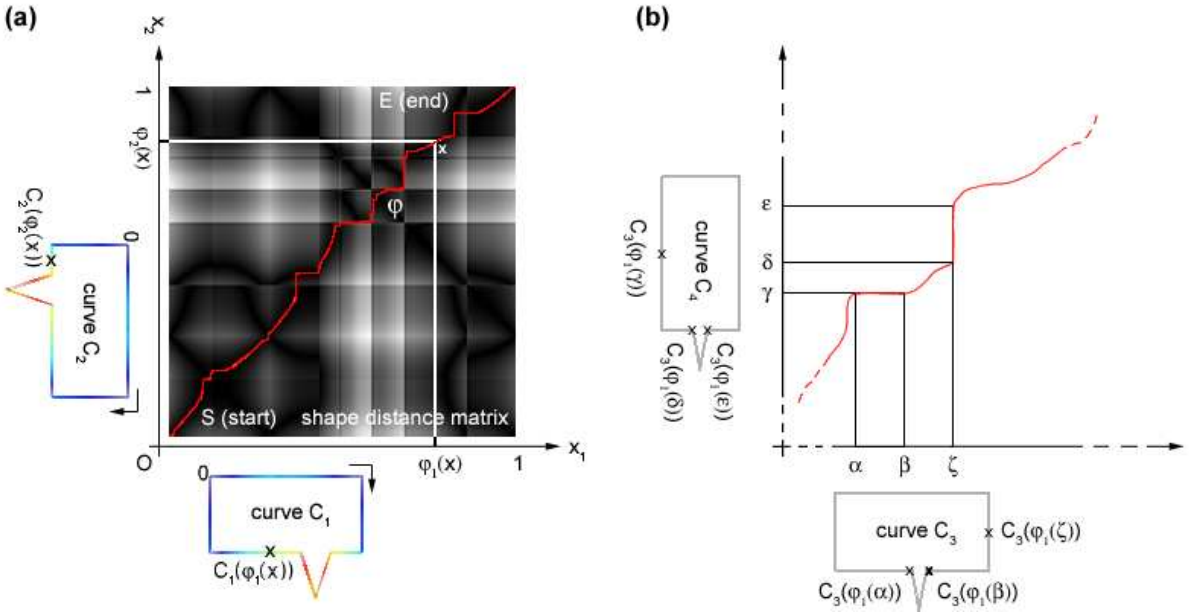


Figure 3.1: Illustration of the proposed matching framework in the case of two 2-D u-parameterized curves. Given a corresponding pair of matched points in both curves (with a point in each curve), a reparameterization corresponds to the red path ϕ between the starting point S and the ending point E. As x travels along ϕ, $\phi_1(x)$ in C_1 is matched to $\phi_2(x)$ in C_2.

Following [201], to allow multiple points in I_1 to be matched to a single point in I_2 and conversely, we restate our problem as that of finding a monotonically increasing and continuous function $\varphi : \left| \begin{array}{lcl} I \subset \mathbb{R}^m & \rightarrow & I_1 \times I_2 \\ x & \mapsto & (\varphi_1(x), \varphi_2(x)) \end{array} \right.$ such that:

$$\forall x \in I, \;\; O_1(\varphi_1(x)) \text{ “very similar to" } O_2(\varphi_2(x)) \tag{3.9}$$

Figure 3.1(a) illustrates that approach when those objects are two 2-D curves. It also demonstrates the inherent symmetry of this framework in (b) where an entire subcurve in C_3 is matched to a single point in C_4, and the other way around.

In addition, we would like the matching to satisfy the following properties:

symmetry: the diffeomorphism $\varphi_{O_1 \rightarrow O_2}$ associated with the matching of O_1 to O_2 should be the inverse of $\varphi_{O_2 \rightarrow O_1}$, the diffeomorphism associated with the matching of O_2 to O_1.

consistent self-matching: when trying to match an object to itself, the optimal reparameterization should be the identity: $\forall \, object \; O, \; \varphi_{O \rightarrow O} = (Id, \, Id)$. In the general case, we would like φ to be not too dissimilar from the identity.

3.3.2 Object matching approaches

A number of automated methods for curve/surface matching has been presented in the literature, that tackle some or all of the above issues.

Rather than giving a full-scale review of these approaches, which is already available elsewhere (in [209] or [131] for instance), we present in the following section a selection of techniques as varied as possible and comment in 3.3.2.2 on their unfortunate lack of explicit control, an issue we propose to solve in this chapter via a learning paradigm.

3.3.2.1 A few shape matching methods

Trouvé and Younes detailed in [201] an axiomatic formulation for 1-D matching: they introduced, among others, the concepts of symmetry ($\varphi_{O_1 \rightarrow O_2}$ should be the inverse of $\varphi_{O_2 \rightarrow O_1}$) and consistent self-matching($\forall \, object \; O, \; \varphi_{O \rightarrow O} = (Id, \, Id)$; in the general case, φ should not be not too dissimilar from the identity) and proposed a matching framework for 2-D piecewise lines that satisfies their axioms. In [37], Cohen *et al.* compared the bending and stretching energies of one curve (O_1) and a reparameterization of the other (O_2^*), in a PDE framework, to find the best match. Fleuté *et al.* [68] minimized the Euclidean distance between an input shape and a registered template, which assumed smooth transition paths in between them. Wang *et al.* [214] used geodesic interpolation to compute the dense correspondence field between two surfaces once an initial sparse set of corresponding points had been obtained with an automated shape-based matching algorithm. In [115], the first elliptical harmonics of the expansion of the input objects (which must have spheroidal shapes) served to establish a correspondence. In [181], Sebastian et al. used a dynamic programming approach similar to [201] to find the best match between two 2-D curves, using a similarity measure based on “alignment” between segments of the curves. Elastic registration and warping approaches have also been investigated. In [198], Thompson *et al.* mapped the input surfaces to two spheres whose coordinates were then warped under anatomical feature curve constraints. Davatzikos et al. [53] also identified corresponding points on object boundaries in two images before aligning them using elastic warping. Along different

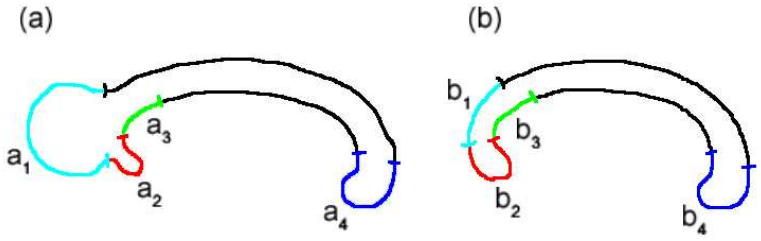


Figure 3.2: Matching two corpus callosum outlines.

lines, Davies *et al.* [56] presented a curve matching method, in the context of the search for the most compact statistical shape model. An information theoretic criterion was designed and controlled the correspondence between objects.

These approaches often rely on the computation of a shape descriptor, either in a local or in a global fashion. If objects to be matched are considered as filled contours (or regions), spatial moments (as described in Section 2.3.1.4) can be used for matching. They are a global measure, that is, they measure the shape of the object taken as a whole. A number of moment invariants can be derived from them: area, circularity, compactness, eccentricity, Euler number, etc. (see [164] for details). Often, object boundaries are used in place of their entire area. In the case of curves or surfaces, a number of analytical functions can be computed at any point: tangent, acceleration, tangent angle or curvature [206].

3.3.2.2 Common caveat

Despite their diversity, these matching approaches share the same drawback: their lack of control over the similarity measure introduced in equation 3.8, which is often defined *a priori*, once and for all, and uses only limited domain-based information (or information learned from examples). Typically, these matching processes can be reduced to optimizing a functional whose minimum corresponds to a "good" correspondence field. The difficulty of designing an adequate functional comes from the difficulty of characterizing an adequate correspondence field. In [37] for instance, the authors assume that points with similar curvature should be matched. This may suit some applications, but is not always desirable. Figure 3.2 illustrates such a situation: here two corpora callosa have been delineated and we wish to compute their average shape:

- Suppose that part of the fornix (a_1) has been improperly delineated together with the corpus callosum. Then we would like segments $\{a_1, (b_1, b_2)\}$, $\{(a_2, a_3), b_3\}$ and $\{a_4, b_4\}$ to be matched, in spite of the fact that the curvature signature of segment a_2 more closely resembles that of b_2 than that of b_3.
- On the other hand, we may decide to trust the delineation and assume that a lesion is the cause of the odd looking bulge (a_1) in the corpus callosum in 3.2.a. Then, we would like a match: $\{a_1, b_1\}$, $\{a_2, b_2\}$, $\{a_3, b_3\}$ and $\{a_4, b_4\}$.

Clearly, choosing between these 2 scenarios requires the introduction of *explicit knowledge* into the matching algorithm.

3.3.2.3 Learning the correspondence field

To overcome this issue, we have developed a learning approach where an *a priori* learning set of correspondence fields helps the matching algorithm compute between the input objects a correspondence field, whose characteristics bear close resemblance to those in the learning set. Our method relies on the use of a shape distance matrix, the matrix of the differences between the values of a local shape descriptor computed on every pair of points of the objects to be matched (see Figure 3.1 for an illustration and Section **??** for a mathematical description). We argue that this shape distance matrix embeds the matching problem in a highly expressive and redundant construct which is more easily manipulated. This matrix is both visually interesting (as it allows for visual inspection of the specific reparameterization problem at hand) and enables us to recast the matching problem as the search for a geodesic in another metrizable space: the space of reparameterizations (which is a group).

We introduce in the following Section 3.3.3 a new shape descriptor, the "observed transport" descriptor, and discuss the properties that make it particularly amenable to our matching problem. We then present the various non-learning (Section 3.3.4) and learning matching techniques (Section 3.3.5) that we have developed.

3.3.3 The observed transport shape descriptor

Before computing a shape distance matrix, we first have to select a shape descriptor. Evaluating the quality of a given descriptor is a difficult task which depends both on the characteristics of the shapes to be described and on the nature of the application at hand. No consistent evaluation criteria have been proposed yet. However several lists of requirements have been suggested.

For instance, Brady [22] proposes 3 criteria: rich local support (descriptors must preserve information and be computed locally), smooth extension and subsumption (to ensure a continuity of representation between local and global descriptions) and propagation (hierarchical property of representation). Although desirable, these properties are difficult to map to a mathematical representation, which renders comparison between descriptors at best suspicious. Here, we have adopted the Marr and Nishihara's criteria [139] which, although not directly numerically expressible either, seem more intuitive:

- accessibility, which measures the ease with which a given descriptor can be computed, in terms of both memory requirement and processing power;
- scope, which covers the class of shapes that can be described with the descriptor;
- uniqueness, which tests the existence of a one-to-one mapping between a shape descriptor signature and the described shape;
- stability and sensitivity, which describe how sensitive to noise and small changes the shape descriptor is.

In view of the corpus callosum example detailed above, we would like the selected descriptor to additionally exhibit the following application-dependent characteristics:

- it should resemble curvature, in that it should be able to discriminate high curvature points since the latter are especially significant for visual perception (see the psychological experiments of Attneave and Arnoult [8, 9]);

- it should be as discriminating as possible, at least more discriminating than curvature since we want a highly discriminating shape distance matrix;
- ideally, its sensitivity to variations could be adjusted from local to global.

As we could not find in the literature a suitable measure that satisfied those specifications, we have derived our own. We describe here the observed transport (OT) descriptor in a variety of n-D cases before detailing its properties.

3.3.3.1 1-D case (sub-manifolds of dimension 1)

Let $C : \left| \begin{array}{ccc} I \subset \mathbb{R} & \rightarrow & \mathbb{R}^n \\ u & \mapsto & x(u) \end{array} \right.$ be a curve in n-D (open or closed), parameterized with respect to a scalar u. We define the observed transport measure ρ_r as follows:

$$\forall u \in I,\ \rho_r\left(C(u)\right) = \int_{V_r(C(u))} \left\| C(u) - C(v) \right\| . \left| C'(v) \right| .dv \tag{3.10}$$

Visibility: $V_r\left(C(u)\right)$ spans the arc of C "visible" within range $r \in \mathbb{R}^+$ from $C(u)$:

$$V_r\left(C(u)\right) = \{v\ s.t.\ [C(u)\,C(v)] \cap C = \{C(u)\,;C(v)\}\ and\ \|C(u)\,C(v)\| \leq r\}$$

with $[C(u)\,C(v)]$ the line segment between points $C(u)$ and $C(v)$.

By considering only the points *visible* from $C(u)$, we obtain a measure whose value changes drastically with the sudden apparition of a vastly different surrounding (compare Figure 3.3 (a),(b) and (c)), within the range r, which helps achieve greater discrimination as those points are visually remarkable (Section 3.3.3.4 demonstrates the increased discriminating power of the OT descriptor over a measure with a similar formulation that takes into account all the points instead of only the "visible" ones).

Intrinsincness: $\left|C'(v)\right|$ insures the independence from reparameterization of the descriptor, as $\left|C'(v)\right| .dv = ds$, the arc length (see proof in Section 3.3.3.4 below);

Locality: r controls the locality of the descriptor. When r is small, we only consider a limited neighborhood around the point at which the descriptor is evaluated. When r increases, larger neighborhoods are considered, consequently a greater proportion of the object influences the value of the descriptor (see Section 3.3.3.4 for an investigation of the influence of r).

$\rho_r\left(C(u)\right)$ can be regarded as the minimal[2] total amount of work it takes to transport the elementary elements dv with mass $\left|C'(v)\right| \cdot dv$ that are visible within range r from point $C(u)$, from their location $C(v)$, to $C(u)$ in the fashion of a Monge-Kantorovich transport problem [88] (hence the name we gave to this descriptor).

[2]It is minimal because, since only the visible points are considered, the shortest overall distance is achieved when all visible points travel along straight lines.

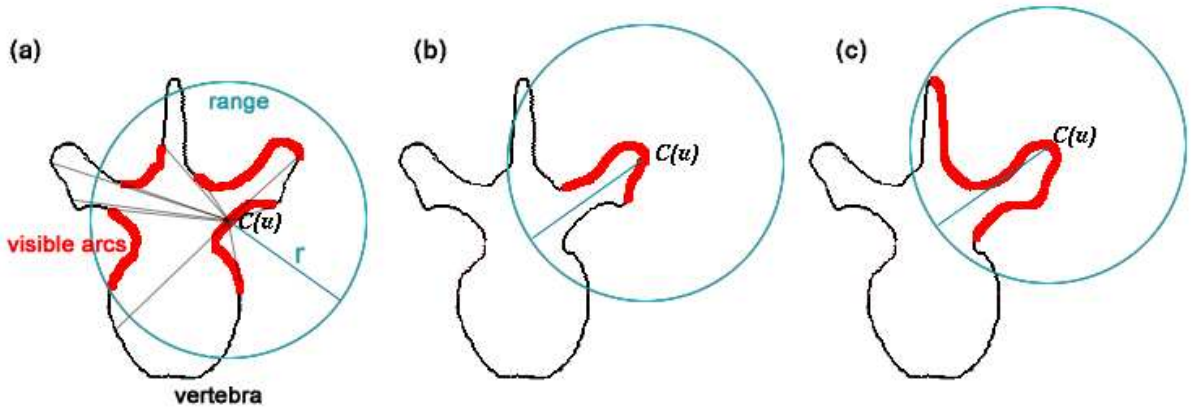


Figure 3.3: Observed Transport descriptor principle demonstrated on a human vertebra outline: the visible arcs (thick red lines) are defined for three locations of C(u).

3.3.3.2 2-D case (sub-manifolds of dimension 2)

The 1-D definition can be readily extended to the 2-D case.

Let $S : \left| \begin{array}{lll} \Omega \subset \mathbb{R}^2 & \rightarrow & \mathbb{R}^n \\ (u,v) & \mapsto & S(u,v) = x(u,v) \end{array} \right.$ be a n-D surface, parameterized with respect to scalars u and v. ρ_r becomes:

$$\forall (u,v) \in \Omega,\ \rho_r(S(u,v)) = \int\int_{V_r(S(u,v))} \|S(u,v) - S(a,b)\| \cdot \left\| \frac{\partial S}{\partial a} \wedge \frac{\partial S}{\partial b} \right\| .da\, db \qquad (3.11)$$

where $\left\| \frac{\partial S}{\partial a} \wedge \frac{\partial S}{\partial b} \right\| .dadb = dS$ is the elementary surface element, and V_r is defined analogously to the 1-D case.

Figure 3.4 shows in red the sub-surface visible from point P for a 2-D (u,v)-parameterized ventricle.

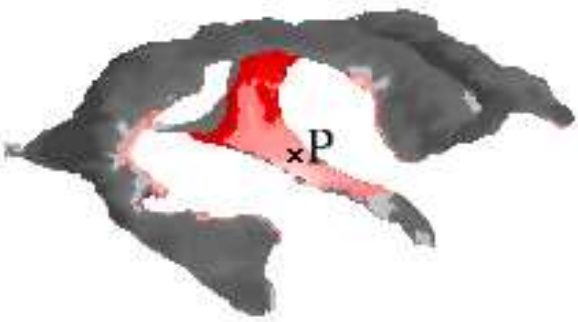


Figure 3.4: A 2-D (u,v)-parameterized ventricle (several patches) with, in red, the sub-surface visible from point P.

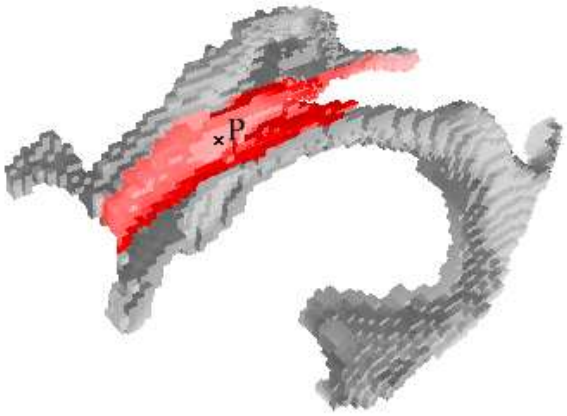


Figure 3.5: A set of 3-D isotropic voxels with, in red, the voxels visible from point P.

3.3.3.3 Discrete approximation

We define a discrete version of object $\mathcal{O}$ as an unsorted collection of n-D isotropic voxels of size s^n: $\mathcal{O} = \{O_i \in \mathbb{R}^n\}_{i=1}^N$. In this case, we do not assume any *a priori* parameterization.

We then derive a discrete version of ρ_r in the n-D discrete case:

$$\forall i \in 1 \dots N,\ \rho_r(O_i) = \sum_{j=1, O_j \in V_r(O_i)}^{N} \|O_i - O_j\| \, .dV_j(O_i) \tag{3.12}$$

where $dV_j(O_i)$ is the surface of voxel O_j visible from voxel O_i which we approximate by ns^{n-1}.

Figure 3.5 shows in red the voxels visible from point P in the 3-D discrete case.

3.3.3.4 Properties

We present in this section the notable properties of our descriptor. We also evaluate its adequacy to our specifications and compare it with other similar shape descriptors with respect to Marr's criteria.

Discriminating power

Figure 3.6 shows how ρ behaves on a few n-D objects. All values were recorded with a range $r = +\infty$. The descriptor values for the hand outline in (b) were saturated at a fifth of the maximal value to better visually illustrate the discriminating power of the observed transport approach.

As illustrated in curve (a), even though ρ evidently depends on the curvature at the point at which it is computed, it also takes into account a much larger neighborhood: $\rho(A) \neq \rho(B)$ and $\rho(C) \neq \rho(D)$, which correctly reflects the differences in the shape landscape surrounding those points (even though A & B and C & D have, respectively, the same curvature). As such, the observed transport descriptor is both a local measure of shape and an indicator of *context* (for large values of r): for instance, it adequately discriminates between the belly and the back of the corpus callosum (Figure 3.6(d)).

Curves (b) and (a) also demonstrate that high curvature points (black circles in Figure 3.6) are adequately discriminated (the intersection points of the circle and the straight line in

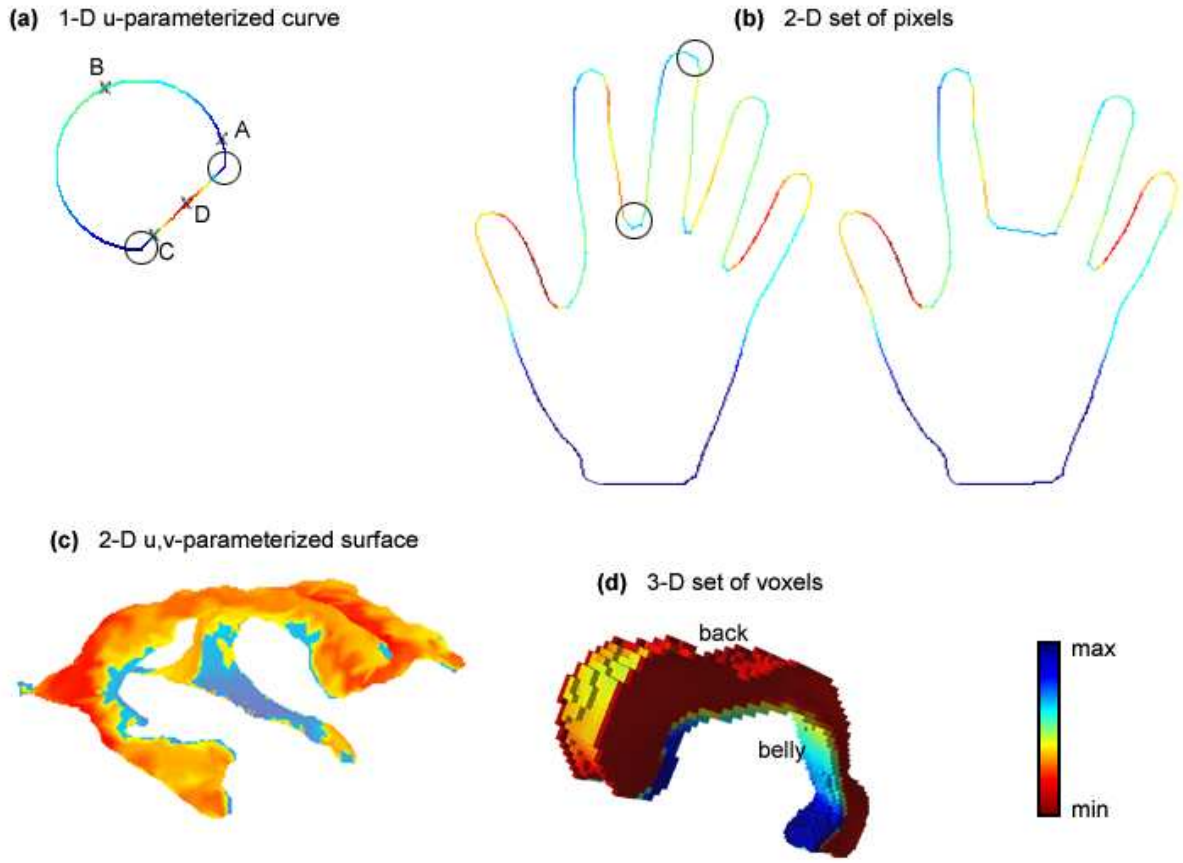


Figure 3.6: Observed transport descriptor computed over: (a) a u-parameterized 2-D curve (truncated circle), (b) a set of 2-D points (two hands), (c) a (u,v)-parameterized 2-D surface (ventricle) and (d) a set of 3-D points (corpus callosum). Black circles indicate a few high curvature areas.

curve (a) have pre-eminent values, the tips and junctions of the fingers in curve (b) are clearly distinguished, etc.).

Finally, a local shape modification only affects the immediate neighborhood of its geographic position, as illustrated with curve (b): the cut middle finger mostly modifies the descriptor values of the immediately surrounding fingers, slightly alters its values for points directly opposite on the other side of the hand (since the cut finger is visible from them), and, evidently, does not affect the left-most and right-most fingers.

Invariances

Parameterization invariance: The observed transport shape descriptor is independent of (*i.e.* invariant to) reparameterization.

Proof: For the sake of simplicity, we give here the proof in the 1-D case, from which a n-D case proof could be derived.

Let $C : I \subset \mathbb{R} \to \mathbb{R}^n$ be a curve, and let $f : I \to I$ be a diffeomorphism.
$C^* = C \circ f$ is a reparameterization of C.

$\forall u \in I$, let $t = f(u)$. We have:

$$\begin{aligned}
\rho_r(C^*(u)) &= \int_{V_r(C^*(u))} \|C^*(u)\, C^*(v)\| \cdot \left|C^{*\prime}(v)\right| .dv \\
&= \int_{V_r(C^*(u))} \|C(f(u))\, C(f(v))\| \cdot \left|C'(f(v)) \cdot \tfrac{df}{dv}\right| .dv \\
&\text{let } w = f(v),\ dw = \tfrac{df}{dv}.dv \\
&= \pm \int_{V_r(C(t))} \|C(t)\, C(w)\| \cdot \left|C'(w)\right| .dw \\
&= \pm \rho_r(C(t))
\end{aligned}$$

If f is decreasing $\left(\frac{df}{dv} < 0\right)$, then the sign of the change of variable is negative, while if f is increasing then the sign is positive. If f is decreasing, then we are integrating in the opposite direction of the integral determined by C and so the negatives cancel:

$\forall u \in I$, let $t = f(u)$, $\rho_r(C^*(u)) = \rho_r(C(t))$ $\square$

Invariance with respect to rigid transformations: Our descriptor is invariant with respect to rigid transformations since both the visible sets and the distances between visible points are invariant with respect to rigid transformations.

However, it is not scale invariant as we believe the scale of an object is an important shape characteristic when trying to match anatomical structures. We could easily introduce scale invariance by normalizing it to the largest observed transport value across the entire object, or by using a scale parameter in subsequent optimizations.

Influence of the locality parameter r

A first series of experiments was conducted to assess the influence of the locality parameter r over the behavior of the OT descriptor. The left-most columns of Figures 3.7 and 3.8 ($\sigma = 0.0$) demonstrate the behavior of the OT descriptor on two 2-D u-parameterized curves (a rectangle with a triangular indentation and a corpus callosum, respectively) for a variety of scopes r (we use a *relative* color scheme: for each callosum and rectangle, red encodes the minimal OT value for the considered curve and blue the maximal one).

When r increases, a larger neighborhood is taken into account; consequently a larger proportion of the curve influences the value of the OT descriptor. With r between 100 and 200, we can easily distinguish between the belly and the back of the corpus callosum, or between the flat and the indented side of the rectangle, as they exhibit different color ranges. With smaller r however, these differences tend to disappear, most especially when r goes below a remarkable distance in the curve. With $r = 10$, $\sigma = 0.0$ in Figure 3.7 for instance, we notice a large number of red segments surrounded by yellow parts, that we cannot differentiate. Similarly in Figure 3.8, when r goes from 50 to 25, it becomes smaller than the height of the rectangle and the curve turns almost entirely red.

Sensitivity to noise

A second series of experiments aimed at measuring the sensitivity of the OT descriptor to noise. The same two 2-D u-parameterized curves were selected and corrupted with a uniform noise $U[-\sigma, \sigma]$. Figures 3.8 and 3.7 demonstrate the behavior of the OT descriptor for a variety of σ and r.

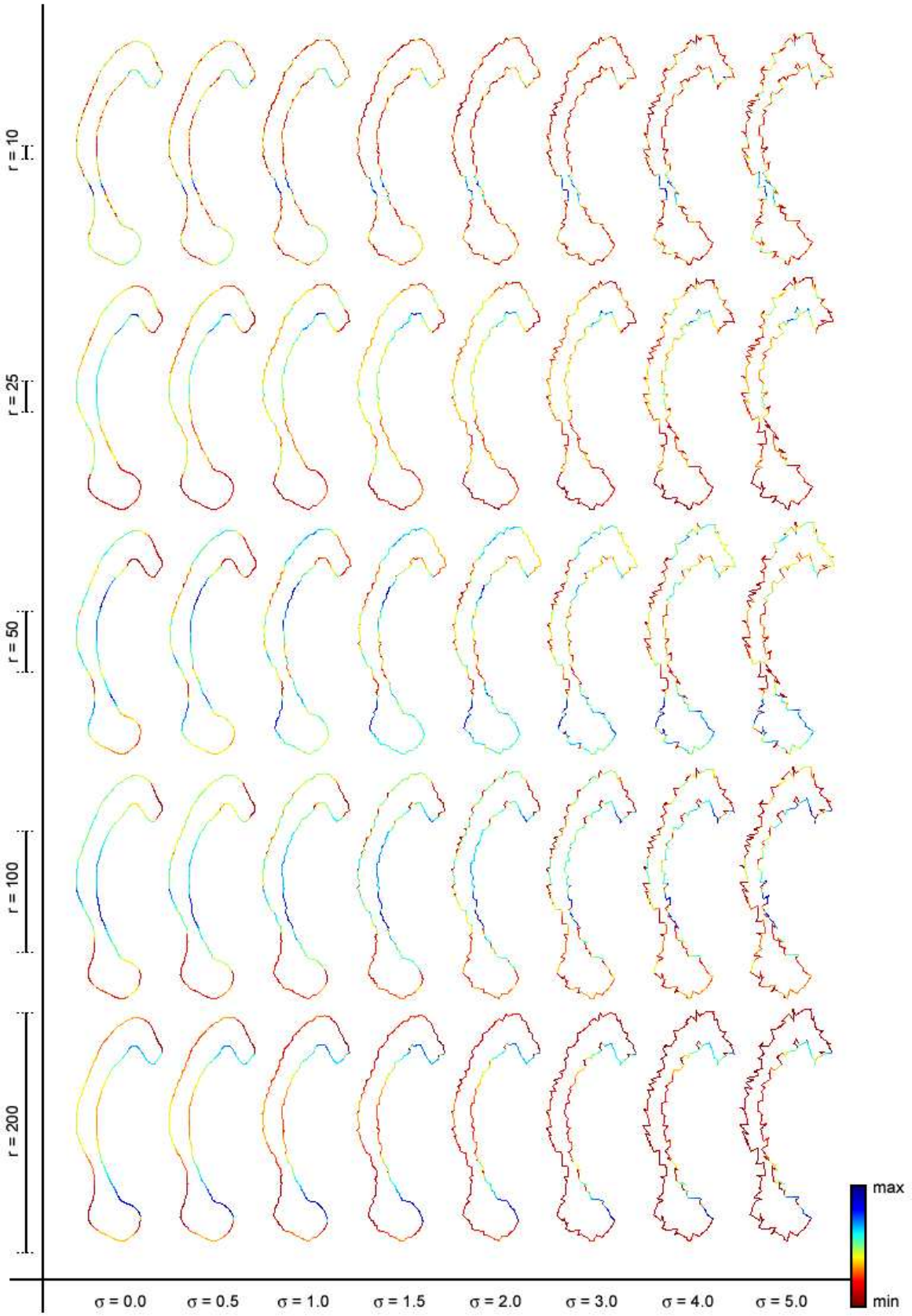


Figure 3.7: Observed Transport maps of a 2-D u-parameterized corpus callosum outline for a range of noise (sigma) and scope (r) values.

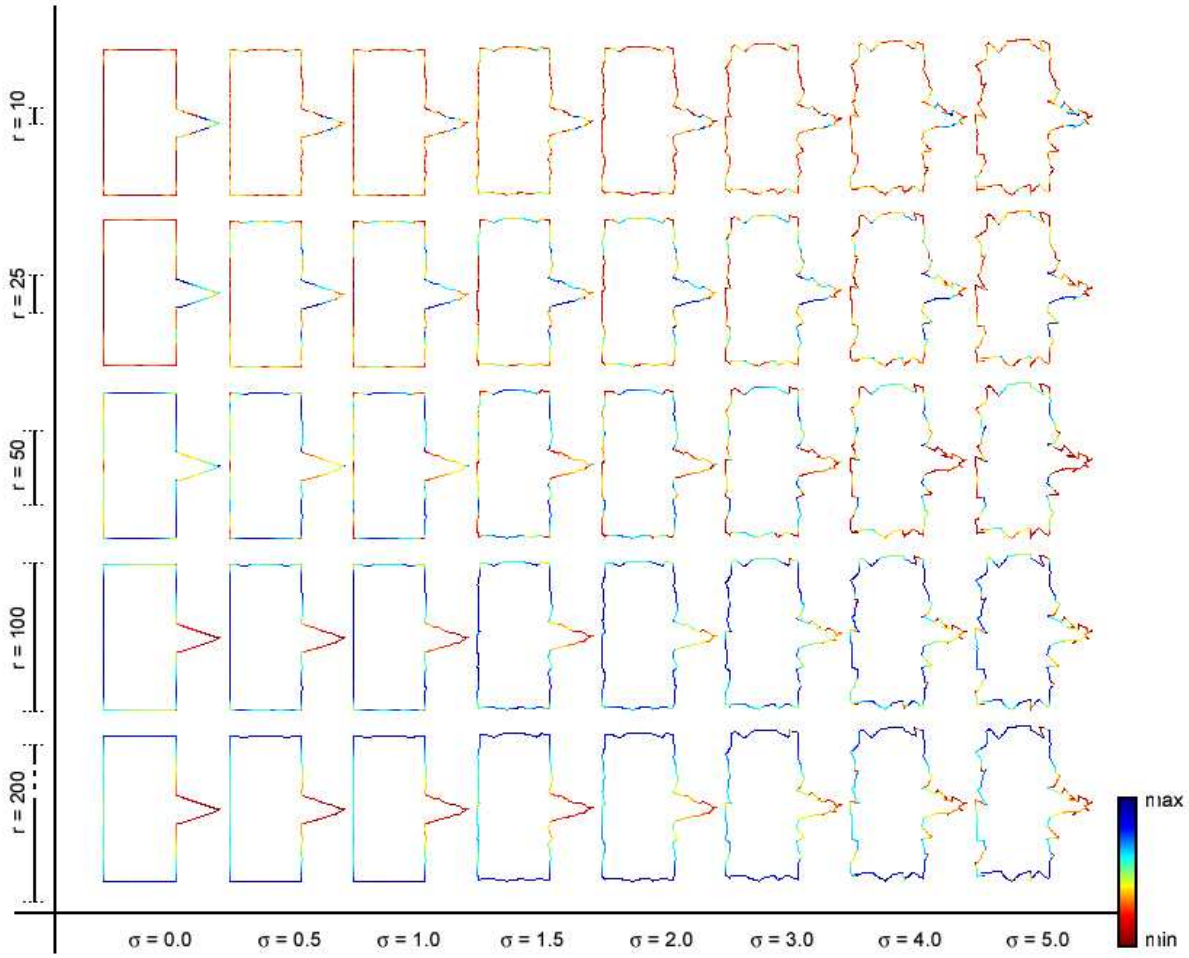


Figure 3.8: Observed Transport maps of a 2-D u-parameterized rectangle with triangular indentation for a range of noise (sigma) and scope (r) values.

The fact that the distances to the entire set of visible points are taken into account explain the remarkable stability of the OT descriptor, even over large ranges of σ: clearly, we can follow color patterns from the left-most column to the right-most one (for instance small blue segments in Figure 3.7 or the triangular indentation in red/yellow in 3.8). The OT descriptor is, not surprisingly, all the more stable that larger r are used. This stability is then maximal for circular shapes (where most of the points are visible from any point), and decreases for narrow or convoluted objects.

Note that the sensitivity of the OT measure to small variations for small values of r (compare for instance columns $\sigma = 0.0$ and $\sigma = 0.5$ for $r = 25$ in Figure 3.8) could be further decreased by introducing fuzziness into the computation of the visible arcs: the contribution of points which are "almost" visible (that is, that could be visible if their position was changed by a few degrees, for instance) could be taken into account once weighted adequately.

Comparison against other descriptors

curvature: If curvature was computed over curve (a) in Figure 3.6, we would get: *curvature(A)* = *curvature(B)* = 1 / radius of the circle, and *curvature(C)* = *curvature(D)* = 0 ,which exemplifies its intrinsically local nature.

Also, note that, again because of its locality, curvature is much more sensitive to noise than OT.[3]

OT without visibility: Figure 3.10 and 3.9 demonstrate the behavior of the OT descriptor on the same two pairs of curves for a variety of noise values σ, when all the points are taken into account instead of just the visible ones.

Even though taking all points into account significantly reduces the sensitivity to noise, it also yields a much less discriminating descriptor: in the rectangle case (Figure 3.10) in particular, the triangular indentation presents the same range of values than the opposite side in the rectangle, whereas its OT values (with visibility) are clearly distinct from that of all the other points.

Belongie's shape context: Our descriptor bears some resemblance with Belongie's "shape context" [13]. Given a shape represented by a set of 2-D points, the shape context of a point is the histogram of the positions of all the other points expressed in a log-polar coordinate system (centered on the given point). In addition to being a fully global descriptor (although the use of log-polar coordinates helps decrease the influence of points far away), the shape context approach delivers for each point a full array of measures, and not just a single scalar value. Given two shapes with their associated shape context signatures, finding a correspondence between them then requires matching the histograms of each point, which necessitates the additional definition of a histogram similarity measure (the χ^2 test in this case)[4]. Invariance to rigid transformation is at best awkward (translation-invariance is guaranteed but rotation-invariance is difficult to achieve). Finally, the shape context approach was devised for discrete sets of points and it is not clear if and how it could handle n-D manifolds. Indeed, it seems that computing the shape context signature of a u-parameterized curve or of a (u,v)-parameterized surface would require an *a priori* discretization, on the characteristics of which the shape context values would depend.

Hagedoorn's reflection metric: The concept of "visibility" is also present in Hagedoorn's "reflexion distance": in [86], Hagedoorn *et al.* first define a "visibility star" as the union of the line segments that connect the point at which the measure is evaluated to every other point from the curve, visible from it. The "reflexion star" is defined by intersecting the visibility star with its reflexion about this point. An affine invariant shape metric is then derived by integrating the area of the reflexion star over all points in the plane. However, in addition to being a global shape distance (and not a local shape descriptor[5]), its affine invariance makes it less discriminating than the OT descriptor.

[3] it could however be averaged over a neighborhood to increase robustness to noise.

[4] Obtaining a single scalar value is also particularly important to compute the shape distance matrix, in our application.

[5] The area of the reflexion star could not be used either as a local shape descriptor as its value is zero for all points on the curve.

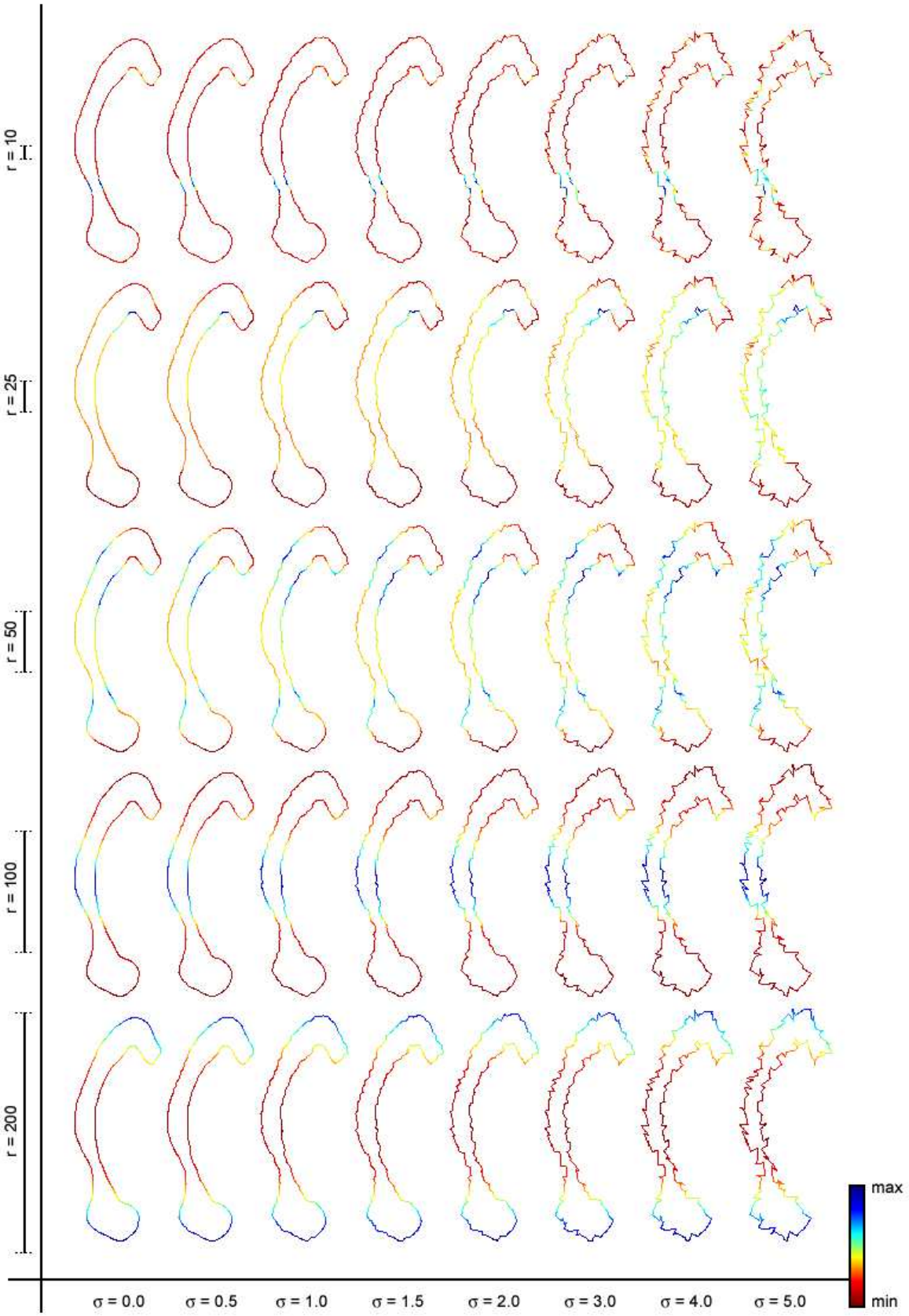


Figure 3.9: Observed Transport maps *without visibility* of a 2-D u-parameterized corpus callosum outline for a range of noise (sigma) and scope (r) values.

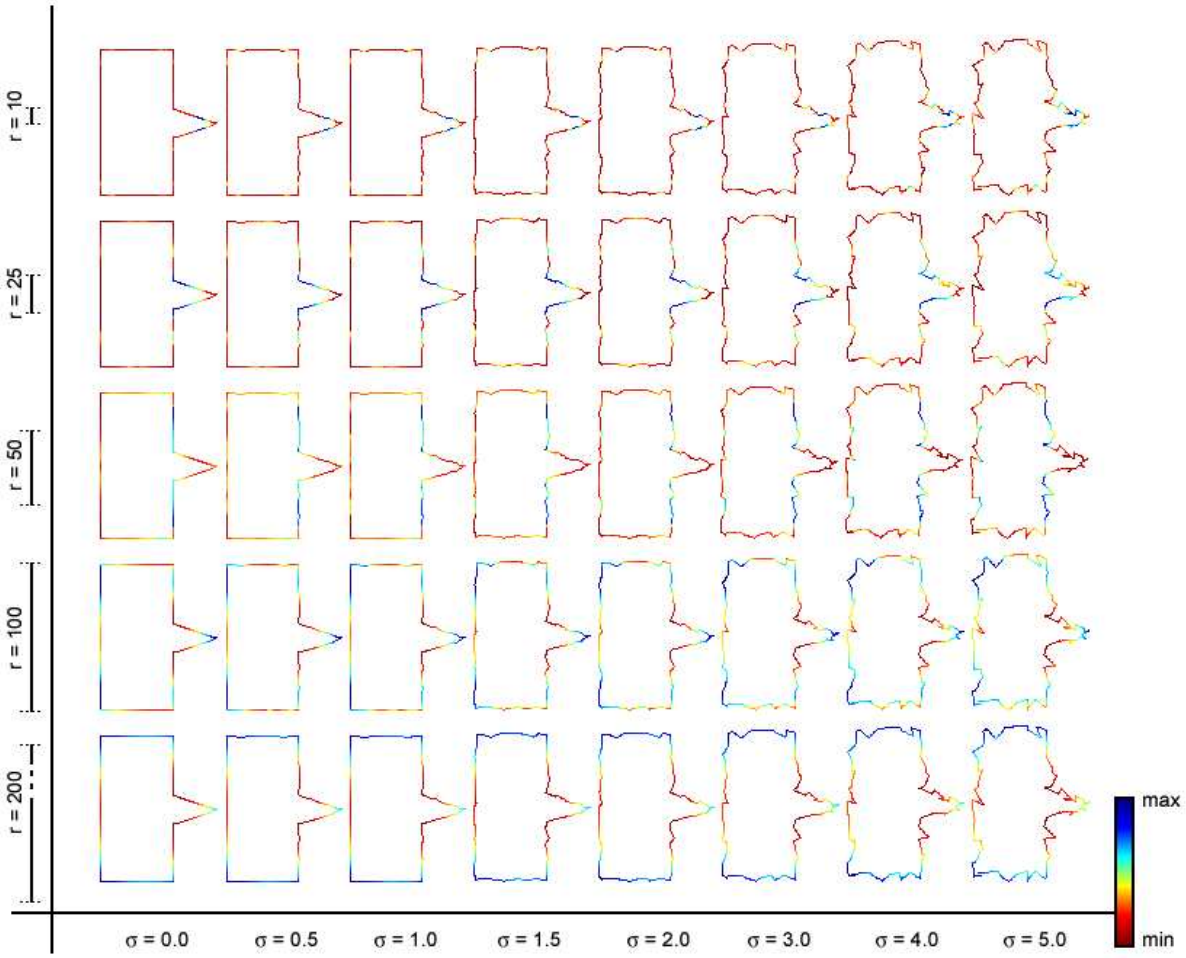


Figure 3.10: Observed Transport maps *without visibility* of a 2-D u-parameterized rectangle with triangular indentation for a range of noise (sigma) and scope (r) values.

3.3.3.4.1 Marr's criteria and application-dependent requirements. Table 3.1 compares four descriptors (OT, OT without visibility, curvature and shape context with respect to the Marr's criteria described above (accessibility, scope, uniqueness, stability/sensitivity).

accessibility: The OT descriptor is slightly more computationally expansive than the shape context as intersection tests have to be carried out to check whether a point is visible or not. When visibility is discarded, the computational requirements become similar. Curvature is usually easier to compute as it is only a local measure. Approximating the curvature of a discrete set of voxels is however not trivial.

scope: Clearly, the OT descriptor has the widest scope of application since it can be computed on both connected and disconnected[6] n-D manifolds and n-D sets of voxels/points.

[6]Note that for disconnected manifold, a parameterized representation of the object might be difficult to obtain.

	OT	*OT without visibility*	*curvature*	*shape context*
accessibility	4	2	1	2
scope	1	1	4	3
uniqueness	1	1	1	?
stability/sensitivity	2	1	4	2

Table 3.1: Comparison of four shape descriptors with respect to Marr's criteria: for each criterion, the best descriptor is labeled 1 and the worst one 4.

uniqueness: The fundamental theorem of space curves states that given a curvature and a torsion signature, there is only one curve that can be reconstructed from it. It seems to us that the same holds for the OT descriptor. We were however unable to find a rigorous proof at this point. We do not know if a similar theorem applies to the shape context descriptor.

stability/sensitivity: The OT descriptor without visibility exhibits the lowest sensitivity to noise, as illustrated by Figure 3.9 and 3.10. On the other side of the sensitivity scale, we find the curvature descriptor, again because of its intrinsic locality[7].

Finally, the OT descriptors covers the application-dependent requirements discussed above:

- it indeed discriminates high curvature points, as illustrated in Section 3.3.3.4;
- it is more discriminating than curvature;
- its sensitivity to variations can be adjusted from local to global.

Note that the transport and visibility "philosophies" behind the OT descriptor could be applied to other measures, such as the curvature for instance, to increase their discriminating power. In the 1-D case, equation 3.10 would become:

$$\forall u \in I,\ \rho_r^{curvature}\left(C(u)\right) = \int_{V_r(C(u))} \left\|C\left(u\right) - C\left(v\right)\right\| .curvature\left(C(v)\right) . \left|C'(v)\right| dv \qquad (3.13)$$

Figure 3.11 compares the behavior of the curvature with the "OT-ed" curvature and the OT descriptor on a truncated circle (u-parameterized curve) and demonstrates the increased discriminating power introduced by the "transport" framework.

3.3.4 Non-learning reparameterization approach

Equipped with a suitable shape descriptor, we can tackle the correspondence problem[8]. We describe here a standard non-learning algorithm. We will detail in the following Section 3.3.5.1 how learning strategies can be derived from this first approach.

[7] A radius-dependent curvature measure could also be computed.

[8] Note that the methods we introduce in this section could equally be applied to a different shape descriptor (the curvature, for instance). The enhanced discriminating power of the Observed Transport descriptor however makes it a better candidate.

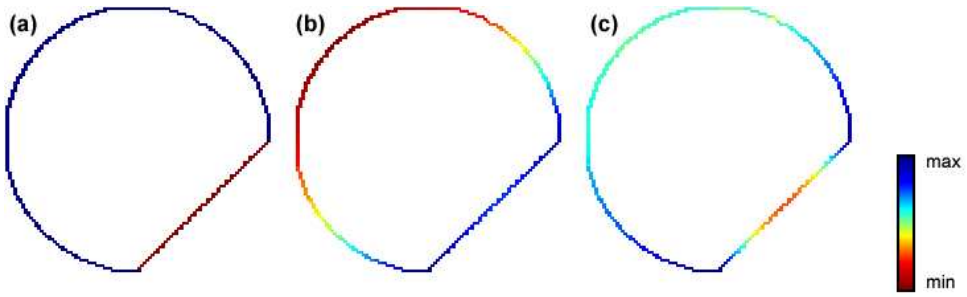


Figure 3.11: Comparison of the curvature map (a), transported curvature map (b) and Observed Transport map (c) of a truncated circle.

We restrict ourselves to the 1-D case in this section:

1-D case: 2-D and 3-D u-parameterized curves: we consider objects defined on an interval of $\mathbb{R}$, taking values in $\mathbb{R}^2$ or $\mathbb{R}^3$ respectively; $m = 1$, $n = 2$ or 3 with notations of Section 3.3.1.

We then define the best reparameterization $\varphi^*_{C_1 \to C_2}$ between curves C_1 and C_2 to be that which minimizes the overall cumulative distance between descriptor values computed for all pairs of matched points:

$$\varphi^*_{C_1 \to C_2} = \arg\min_{\varphi} \left\{ \int_I \left| \rho_{C_1} \left(\varphi_1 (s) \right) - \rho_{C_2} \left(\varphi_2 (s) \right) \right| .ds \right\} \tag{3.14}$$

where ds is the arc length.

In practice, we need to discretize both curves to compute $\varphi^*_{C_1 \to C_2}$. This sampling step must be performed carefully so as not to discard from the discrete set of pixels/voxels important curve features (in particular, one should pay attention to the sampling rate required to capture the curve details deemed important). Let D be the shape distance matrix associated with the discretized curves $C_1 = \left\{ C_i^1 \right\}_{i=1}^{N_1}$ and $C_2 = \left\{ C_i^2 \right\}_{i=1}^{N_2}$ where N_1 is the number of points in the discrete C_1 and N_2 the number of points in C_2:

$$D = [d_{ij}] \begin{array}{l} i = 1 \ldots N_1 \\ j = 1 \ldots N_2 \end{array}, \forall (i,j)\ d_{ij} = \left| \rho_r^{C_1} \left(C_i^1 \right) - \rho_r^{C_2} \left(C_j^2 \right) \right| \tag{3.15}$$

When the discrete C_1 and C_2 are very different in size (for instance, if C_1 was delineated in a high-resolution histological section and C_2 in a lower-resolution MRI and if their respective associated resolutions were not taken into account during the sampling step) then the OT descriptor may take vastly different ranges of values (since we decided not to make it scale invariant). This may significantly hamper the computation of an adequate correspondence field. Namely, if $\forall (i,j)\ \rho_r^{C_1} \left(C_i^1 \right) \gg \rho_r^{C_2} \left(C_j^2 \right)_{ij}$ then $\forall (i,j)\ d_{ij} = \left| \rho_r^{C_1} \left(C_i^1 \right) - \rho_r^{C_2} \left(C_j^2 \right) \right| \cong \left| \rho_r^{C_1} \left(C_i^1 \right) \right|$. To alleviate this issue, we can replace D by its "normalized" version, computed as follows:

$$\forall (i,j)\ d_{ij} = \left| \frac{\rho_r^{C_1}\left(C_i^1\right) - \min_k \rho_r^{C_1}\left(C_k^1\right)}{\max_k \rho_r^{C_1}\left(C_k^1\right) - \min_k \rho_r^{C_1}\left(C_k^1\right)} - \frac{\rho_r^{C_2}\left(C_j^2\right) - \min_k \rho_r^{C_2}\left(C_k^2\right)}{\max_k \rho_r^{C_2}\left(C_k^2\right) - \min_k \rho_r^{C_2}\left(C_k^2\right)} \right| \tag{3.16}$$

Finding the best reparameterization then boils down to finding in D (see Figure 3.1) the minimal cost path between points S (start) and E (end), which requires that a single matching pair of points ($M_1 \in C_1$, $M_2 \in C_2$) be given (this condition can be relaxed if circular shifts are included in the optimization as well).

Note that for all $X \in [S, E]$, we have:

$$\begin{aligned}
\varphi^*_{C_1 \to C_2} &= \arg\min_\varphi \left\{ \int_I \left| \rho_{C_1}\left(\varphi_1(s)\right) - \rho_{C_2}\left(\varphi_2(s)\right) \right| .ds \right\} \\
&= \arg\min_\varphi \left\{ \int_S^E \left| \rho_{C_1}\left(\varphi_1(s)\right) - \rho_{C_2}\left(\varphi_2(s)\right) \right| .ds \right\} \\
&= \arg\min_\varphi \left\{ \int_S^X \left| \rho_{C_1}\left(\varphi_1(s)\right) - \rho_{C_2}\left(\varphi_2(s)\right) \right| .ds \right. \\
&\quad \left. + \int_X^E \left| \rho_{C_1}\left(\varphi_1(s)\right) - \rho_{C_2}\left(\varphi_2(s)\right) \right| .ds \right\} \\
&= \arg\min_{\varphi[S,X]} \left\{ \int_S^X \left| \rho_{C_1}\left(\varphi_1(s)\right) - \rho_{C_2}\left(\varphi_2(s)\right) \right| .ds \right\} \\
&\quad + \arg\min_{\varphi[X,E]} \left\{ \int_X^E \left| \rho_{C_1}\left(\varphi_1(s)\right) - \rho_{C_2}\left(\varphi_2(s)\right) \right| .ds \right\} \\
&= \varphi^*_{C_1[S,X] \to C_2[S,X]} + \varphi^*_{C_1[X,E] \to C_2[X,E]}
\end{aligned} \tag{3.17}$$

where $C_i[A, B]$ is the restriction of C_i on $[A, B]$.

We can then use a dynamic programming approach to find the optimal path:

dynamic programming

```
// init the cost array:
cost[0][0] = 0.0; ∀(i, j) ≠ (0, 0), cost[i][j]=+∞

// init the direction array (an array of LEFT and DOWN which describes the optimal path
// in the shape distance matrix)
∀(i, j), direction[i][j]=NONE

// find the optimal path:
find_optimal_path(direction, cost, N1 − 1, N2 − 1, N1, N2);

////////////////////////////////

// recursive dynamic programming function:
float find_optimal_path(direction[][], cost[][], i, j, n, m) {

  // if cost[i][j] has not been computed:
  if (cost[i][j] = +∞) {
    p1 = +∞; p2 = +∞
    if (i > 0)
      p1 = d_{i,j} + find_optimal_path(direction, cost, i-1, j, n, m)
```

```
    if (j > 0)
      p2 = d_ij + find_optimal_path(direction, cost, i, j-1, n, m)
    if (p1 < p2) {
      direction[i][j] = LEFT; cost[i][j] = p1;
    } else {
      direction[i][j] = DOWN; cost[i][j] = p2;
    }
  } else
    // cost has already been computed:
    return cost[i][j];
}
```

3.3.4.1 Choosing S and E

The starting and ending points must then be chosen *a priori*:

- For open curves, S and E can be the extremities of the curve. The dynamic programming approach then yields an $\mathcal{O}(N_1.N_2)$ complexity.
- For closed curves, remarkable points, such as extrema of curvatures, or the left-most/right-most points could be selected, yielding the same complexity.
- In the general case where no pair of matching points is given, the complexity becomes $O\left(N_1.N_2^2\right)$ as we apply the $O(N_1.N_2)$ dynamic programming algorithm to all possible pairs $(M_1 \in C_1,\, M)\,,\, \forall M \in C_2$.
- The learning approach we introduce below in Section 3.3.5.1 can also provide better control over the automated choice of S and E.

3.3.4.2 Consistent self-matching

When a number of consecutive points have the same shape descriptor (in a circle for instance), there is not a unique best path with respect to the above criterion. To bias the search towards "natural" reparameterizations (the "consistent self-matching" axiom), we introduce in equation 3.14 a constraint to prevent the path from deviating too much from the diagonal of D, *i.e.* for some $\alpha \in [0,1]$:

$$\begin{aligned}\varphi^*_{C_1 \to C_2} \;=\; \arg\min_{\varphi} \Big\{ & \alpha . \int_I \left|\rho_{C_1}\left(\varphi_1(s)\right) - \rho_{C_2}\left(\varphi_2(s)\right)\right| .ds \;+ \\ & (1-\alpha) . \int_I \left|\varphi_1(s) . C_2'(s) - \varphi_2(s) . C_1'(s)\right| .ds) \Big\} \end{aligned} \tag{3.18}$$

3.3.4.3 A few optimal path examples

Figure 3.12 displays four pairs of reparameterized curves (a pair per column) along with the point-by-point Euclidean averages derived from them. We used $\alpha = 0$. Corresponding points are rendered with the same color in (c) and (d). Note in particular how the discriminating power of our shape descriptor enabled the triangular indentations to be correctly matched together in the first column, and against the corresponding points in the rectangle in the second

column. Also, the peculiar thinning towards the back of the mean callosum was *a posteriori* dimmed "reasonable" by an in-house medical doctor who confirmed the adequacy of the callosal reparameterization.

Note that our ambition at this point is not to demonstrate that the OT descriptor can produce "perfect" reparameterizations within this shape distance geodesic minimization framework since, as mentioned earlier, we submit that the quality of a reparameterization depends on *a priori* knowledge about the matching scenario. The adequate behavior of this matching scheme in combination with the OT descriptor however confirms that the latter is a good candidate as a matching similarity measure.

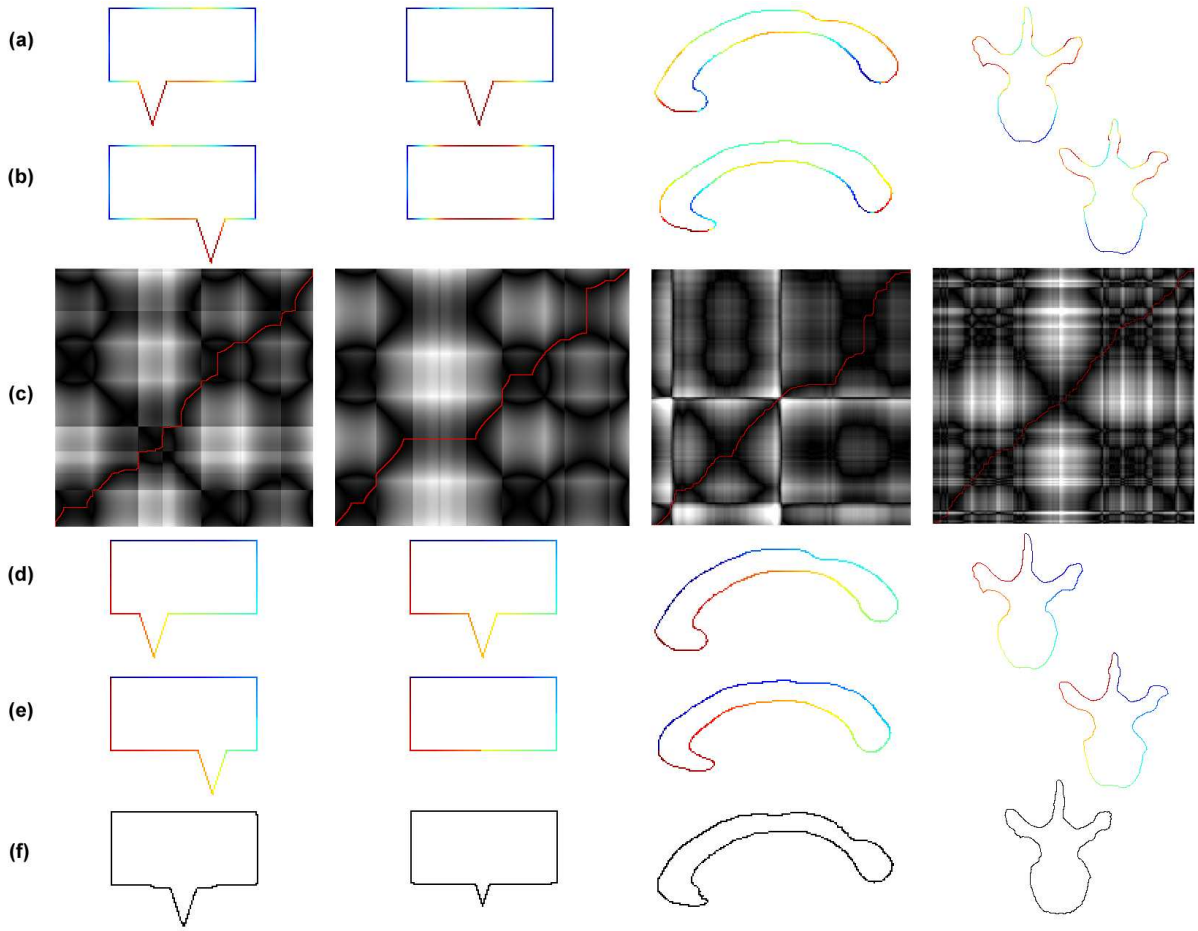


Figure 3.12: Non-learning reparameterization: (a,b) Observed Transport maps; (c) shape distance matrices and optimal paths (in red); (d,e) reparameterized curves; (f) point by point Euclidean average curves.

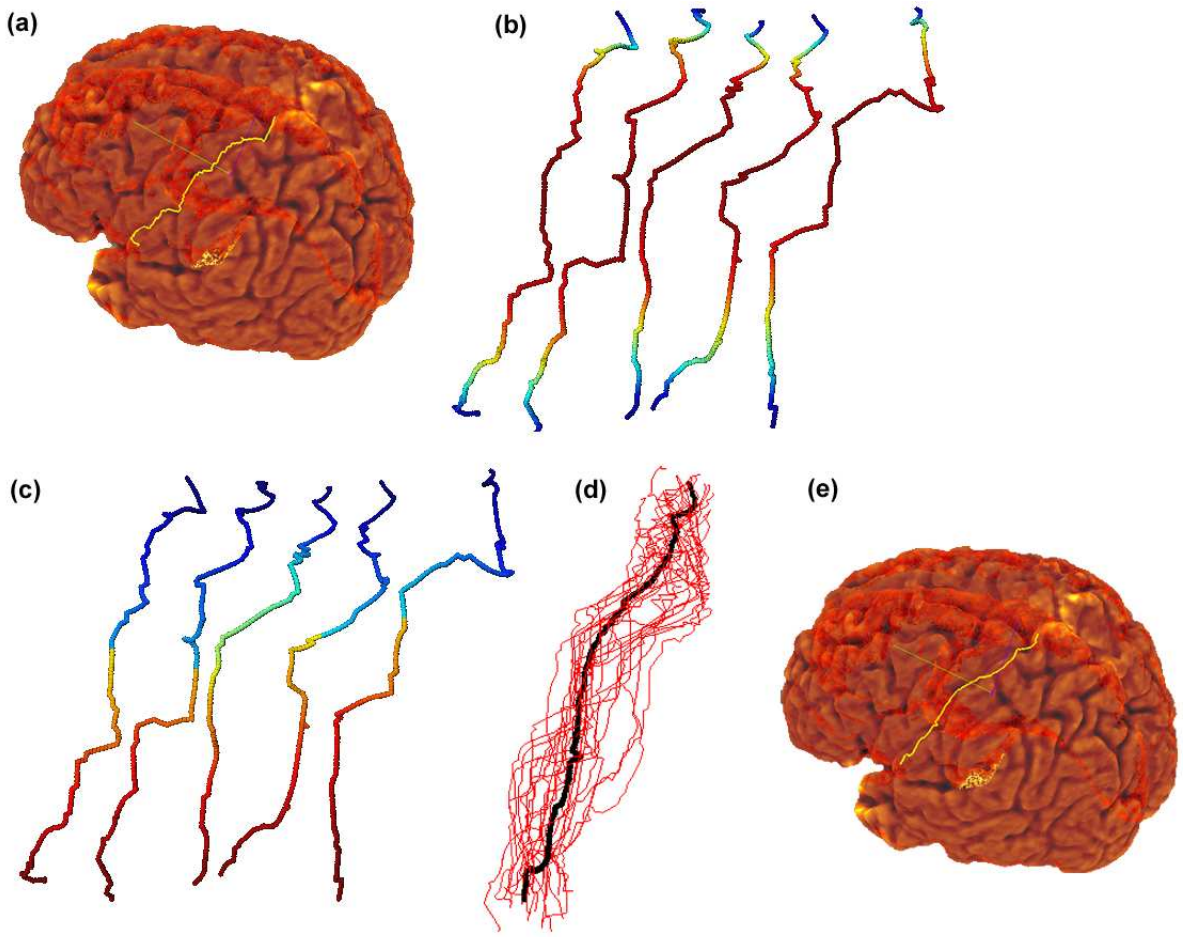


Figure 3.13: Averaging a series of left central sulcus delineations: (a) the left central sulcus (yellow) delineated on a human cerebral cortex (transparent red); (b) a few left central sulci OT maps; (c) the color-coded reparameterized sulci; (d) the point-by-point Euclidean left central sulcus average (black) with all 25 sulci; (e) average left central sulcus (yellow) on the same human cerebral cortex.

3.3.4.4 Genu of the central sulcus

We have applied this reparameterization technique to a series of 25 left central sulcus delineations (Figure 3.13). Each delineation consists of a u-parameterized 3-D curve ($m = 1$, $n = 3$). We used the observed transport descriptor with $r = +\infty$ and the matching framework with $\alpha = 0$. Figure 3.13 displays some OT maps (b), the associated color-coded reparameterized sulci (c) and the overall point-by-point average (d)[9]. Here also, an in-house medical doctor deemed the average sulcus "reasonably good looking".

[9]The overall average is the Euclidean point-by-point mean shape, once all the sulci have been reparameterized with respect to a selected reference sulcus.

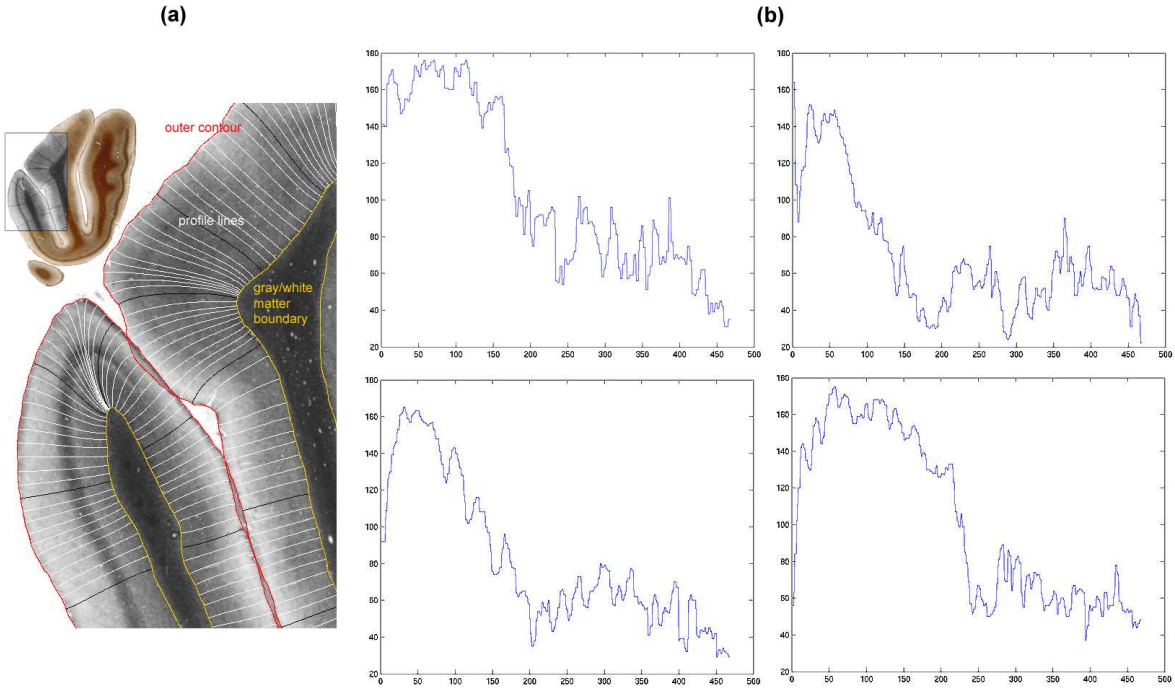


Figure 3.14: Cortical profile extraction: (a) a few regularly spaced cortical profiles computed between the pial surface (in red) and the white matter/gray matter border (in yellow) in a myelin-stained histological section of the human brain; (b) four staining density versus geodesic position graphs corresponding to four profile lines in the same myelo-architectonic area.

3.3.4.5 Average intensity profile in the cortical mantle

Our reparameterization approach also proved useful in the analysis of the myelo-architectecture of the human cerebral cortex [3]. In a nutshell, by solving a partial differential equation (which models the laminar patterns of the cortex) between the gray/white matter border and the pial surface in myelin-stained histological sections of the human brain, we generate cortical profile lines, orthogonal to the general direction of the cortical ribbon, and record along them the density of staining (see Figure 3.14(a)). A wavelet-based classifier then clusters these intensity profiles into a number of anatomically-meaningful classes (similar to Brodmann visual areas, in our case).

For each area, we then want to compute an average intensity profile to produce an area-specific template of cortical myelo-architecture. To do so, we assimilate each intensity profile to a 2-D u-parameterized curve, by considering its graph in a 2-D orthonormal coordinate system where the x-axis is the geodesic distance along the profile line starting from the pial surface, and the y-axis is the intensity value recorded along the profile (see 3.14.b). In view of the varying width of the constitutive layers of the cortical ribbon, no meaningful average could be obtained by simply averaging the staining densities with respect to their geodesic position across the set of profiles. We can however compute the OT descriptor along each profile and take advantage of its discriminating power to iteratively reparameterized the profiles within each class, and obtain a more meaningful area average where the various constitutive layers are matched adequately.

Figure 3.15 displays in (a) the result of the wavelet-based profile classification for a myelin-stained histological sections of the human visual cortex (to each color is associated a different

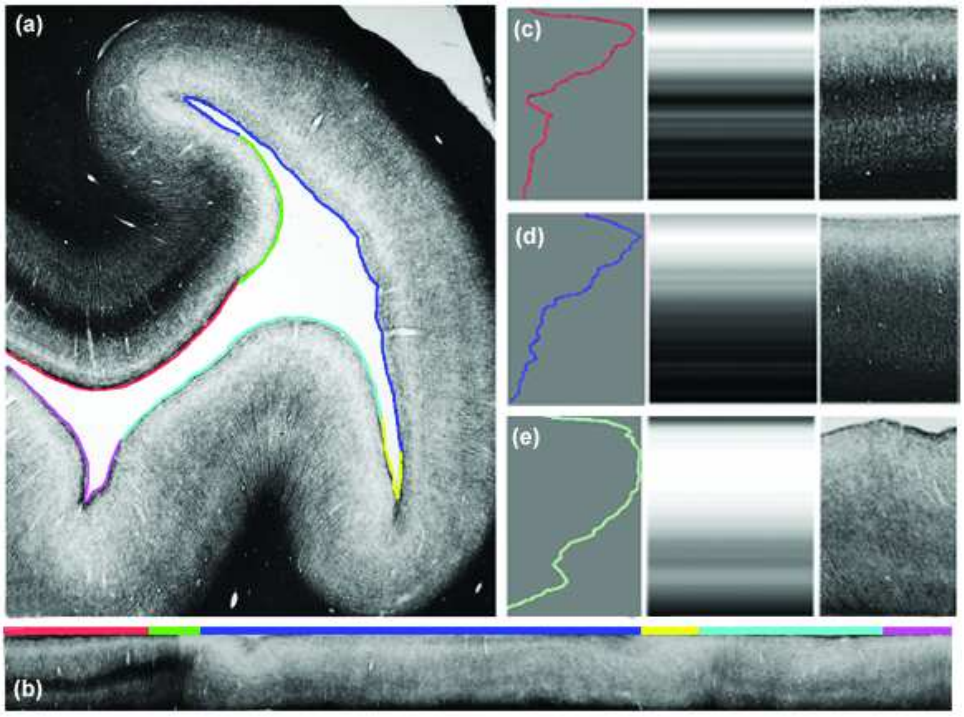


Figure 3.15: A few average cortical intensity profiles computed from the result of the wavelet-based classification method: (a) input myelin-stained histological section with color-coded classified areas; (b) straightened ribbon with color-coded areas; (c,d,e) : area profile average, synthetic area average and area sample.

profile class). The cortical ribbon and its associated color-coded classes are also shown in a straightened form in (b). The efficiency of our OT averaging technique is demonstrated in (c, d, e) where the averages of the red, blue and green areas, respectively, are computed. For each area, we also show the average intensity pattern generated from the average profile and compare it to a sample pattern extracted from the corresponding area in (a). The visual similarity between the synthetic averages and the actual cortical samples is evident in all three cases: our OT averaging technique satisfactorily matched the cortical layers of the ribbon (please refer to [3] for a detailed analysis).

3.3.4.6 Matching properties

Even though this matching framework is independent of the choice of the shape descriptor (any other descriptor could be substituted in place of the Observed Transport), its performances are subordinated to the behavior of the chosen shape measure.

From the characteristics of the OT descriptor stem the following properties:

Invariance with respect to rigid transformations: since the OT descriptor is rigid-invariant, the shape distance matrix associated with a pair of rigidly transformed objects is exactly the same as that associated with the same pair of original objects.

Robustness to noise: We evaluated the noise robustness of the OT descriptor in Section 3.3.3.4. As argued above, the noise-robustness of the matching framework should be directly related to that of the OT descriptor. Figure 3.16 qualitatively confirms this intuition. It shows the colored reparameterized curves, shape distance matrix, optimal paths and point-by-point Euclidean averages associated to the matching of the now familiar pair of rectangles with triangular indentation and pair of corpora callosa, when subjected to uniform noise (with $r = +\infty$).

Clearly, performances are excellent, even for large values of σ.

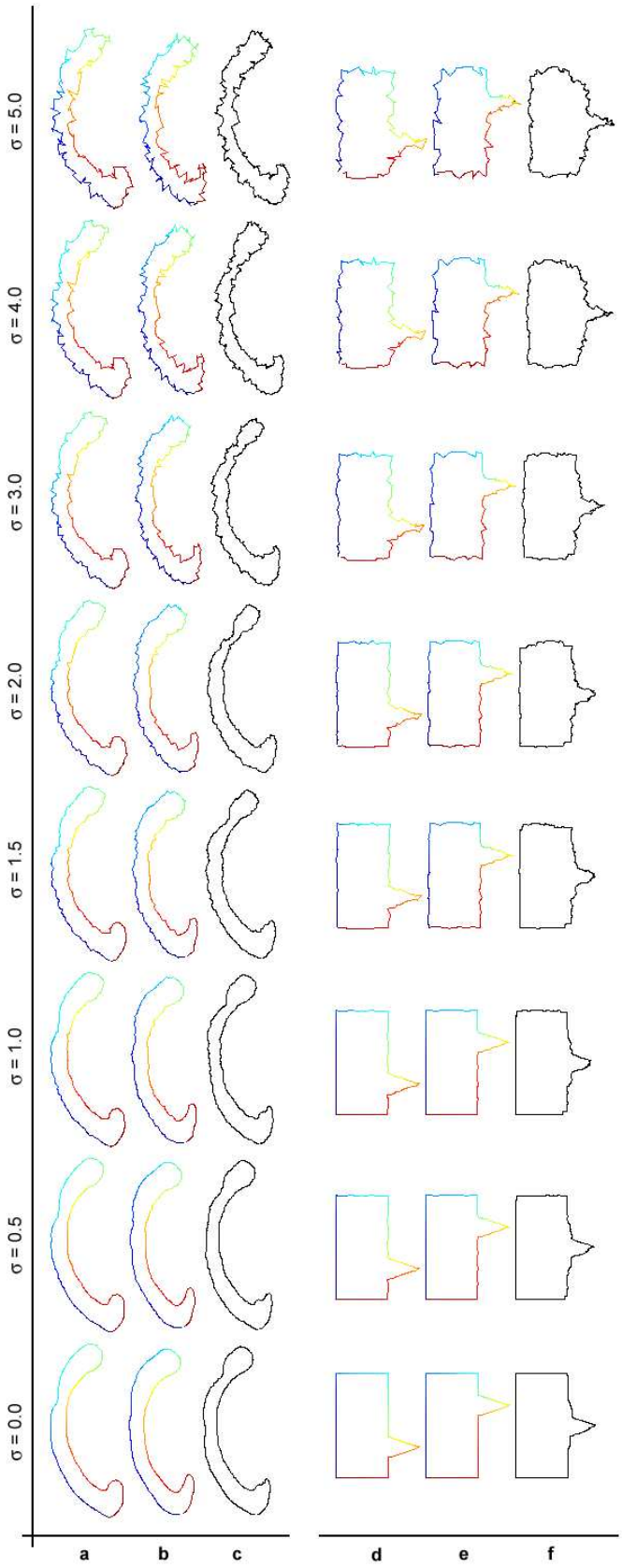


Figure 3.16: Noise-robustness of the non-learning reparameterization for two pairs of 2-D u-parameterized curves and a variety of noise intensities (sigma): (a,b) color-coded reparameterized pair of corpora callosa; (c) point-by-point Euclidean average callosa; (d,e) reparameterized pair of rectangles with triangular indentation; (f) point-by-point Euclidean average rectangles.

3.3.5 Learning reparameterization approach

We present in this section the learning algorithms we have developed to bias the search for a correspondence field between two objects towards instances that are admissible with respect to an *a priori* given learning set. It is through this learning set that we introduce *a priori* medical knowledge and expertise in the matching process.

We have tackled 3 distinct cases, to which all or only some of these methods can be applied:

1-D case: 2-D and 3-D u-parameterized curves: we consider objects defined on an interval of $\mathbb{R}$, taking values in $\mathbb{R}^2$ or $\mathbb{R}^3$ respectively; $m = 1$, $n = 2$ or 3.

2-D case: discrete 2-D point-set (unsorted collections of points of $\mathbb{R}^2$) and *(u,v)*-parameterized 2-D surfaces; $m = 2$, $n = 2$ or 3.

3-D case: discrete 3-D point-set (unsorted collections of points of $\mathbb{R}^3$); $m = 3$, $n = 3$.

3.3.5.1 A pattern matching approach to the 2-D reparameterization problem [1-D case: $m = 1$, $n = 2$ or 3]

An interesting feature of the shape distance matrix is that it embeds, in a highly redundant way, information about all possible reparameterizations between the two input objects. In Figure 3.17 for instance, we can notice clear patterns corresponding to the triangles. A local "matching scenario" (e.g. "discarding the fornix" in Figure 3.2, or "matching the triangles together" in Figure 3.12) then corresponds to a path in a sub-matrix extracted from the shape distance matrix of the objects. Note that even though our shape descriptor is independent of reparameterization, pairs of objects with different initial parameterizations will produce different looking shape distance matrices. Care should thus be taken to use the same (or similar) parameterization for the objects to be matched and the ones in the learning set.

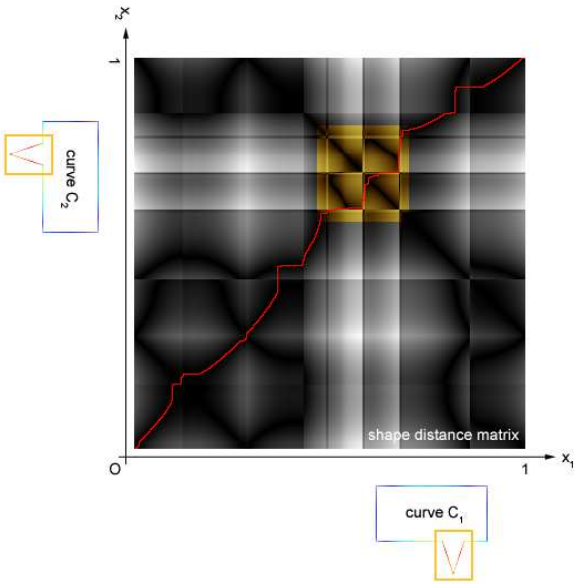


Figure 3.17: Shape distance matrix and matching scenarios. The submatrix highlighted in yellow embeds all the possible reparameterizations of the triangular indentations of curves C_1 and C_2.

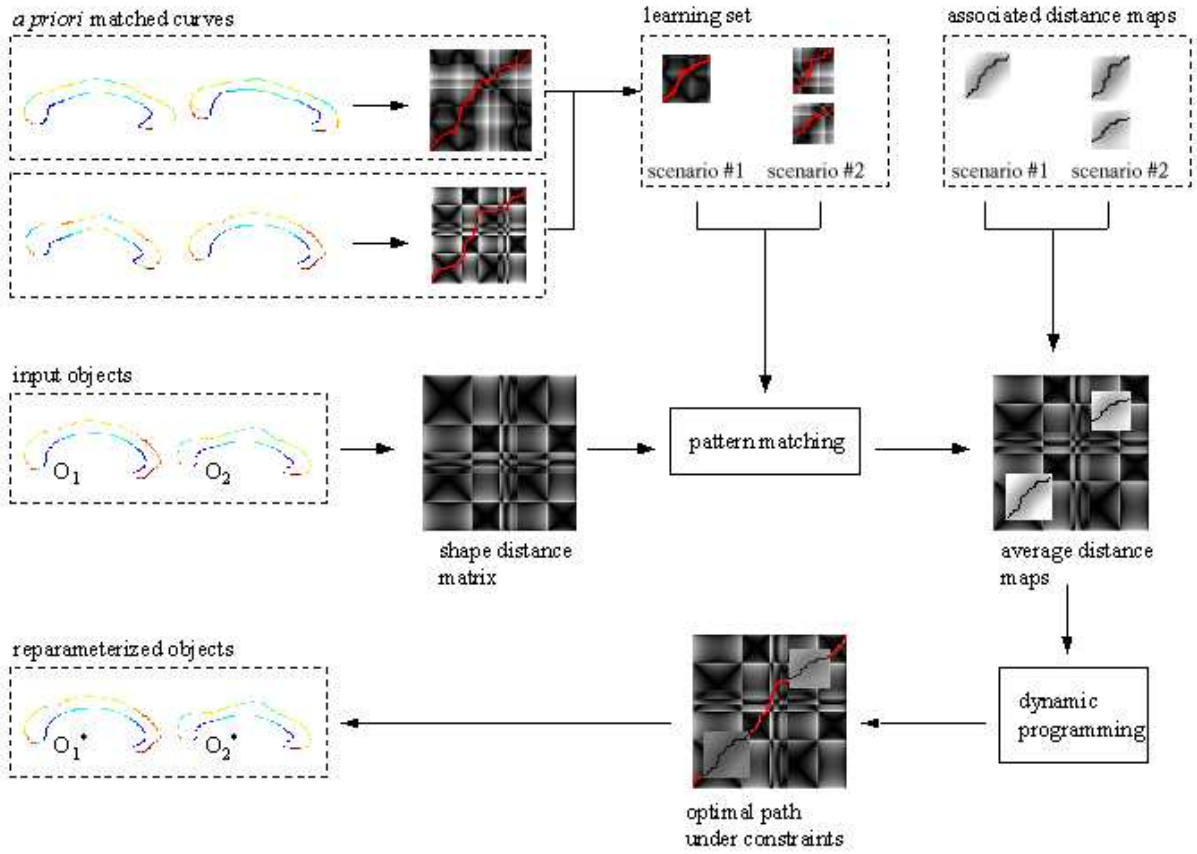


Figure 3.18: Pattern matching strategy.

We derive the corresponding algorithm (see Figure 3.18):

pattern matching algorithm

Step 1 (*a priori*). Given a number of desired local matching scenarios, a human operator first forms a learning set by selecting instances for each scenario (a careful process as the operator must ensure that the learning set adequately represents the desired matching characteristics). An instance consists of a 2-D sub-matrix $M_{i,j} \in \mathcal{M}_{m_{i,j} \times n_{i,j}}$, and its associated connected path $P_{i,j} = \left\{ \left(x_{i,j}^k, y_{i,j}^k \right) \right\}_{k=1}^{m_{i,j}+n_{i,j}}$. The sub-matrices are extracted from shape distance matrices computed from objects which should be "similar" to the ones the operator wants the algorithm to reparameterize. $P_{i,j}$ is the path in $M_{i,j}$ which represents a local matching scenario, in the same fashion that the optimal cost path in Section ?? represents the "optimal" global matching scenario. For each instance, we also compute the distance map of its path.
Let $S_1 = \{S_{1,1}, \ldots, S_{1,N_1}\}, \ldots, S_K = \{S_{K,1}, \ldots, S_{K,N_K}\}$ be the K scenarios, with their instances $S_{i,j} = (M_{i,j}, D_{i,j})$ where $M_{i,j}$ is the shape distance sub-matrix, and $D^{i,j}$ the associated distance map.

Step 2. Once we have computed the shape distance matrix $M \in \mathcal{M}_{m,n}$ from the two input objects $\mathcal{O}_1$ and $\mathcal{O}_2$, a pattern matching algorithm is used to find in M sub-matrices that

bear close resemblance to those of the learning set. We use a straightforward multi-scale registration approach where each sub-matrix $M_{i,j}$ in the learning set is matched against sub-matrices in M, at a number of positions and scales. For each $M_{i,j}$, we record the translation $t^*_{i,j}$ and scale $s^*_{i,j}$ for which the maximal similarity is achieved: $\left(t^*_{i,j}, s^*_{i,j}\right) = \arg\max_{t,s} \left(\text{similarity}\left(M_{i,j}, M\,|[t_x, t_x + s.m_{i,j}] \times [t_y, t_y + s.n_{i,j}]\right)\right)$ where $M\,|[t_x, t_x + s.m_{i,j}] \times [t_y, t_y + s.n_{i,j}]$ is the sub-matrix of M of size $s.m_{i,j} \times s.n_{i,j}$ which starts at index t_x, t_y (with $t = [t_x, t_y]^T$). We also discard instances for which the associated similarity measure is too low.

Step 3. For each scenario in the learning set, we then average the distance maps of the paths associated with their instances (once we have applied the proper translation and scale from step #2). The averaging process is done pixel by pixel. In Figure 3.18, we average the maps of the two instances of scenario #2; no averaging is required for scenario #1 since it only has 1 instance. The zero-level set of the average maps then gives the average path [124].

Note that the quality of the match between the sub-matrices from the learning set and the matrix M could be used to compute a weighted average distance map instead of an equal-weight one.

Step 4. We then insert the average paths inside the shape distance matrix M and use dynamic programming to find the optimal sub-paths in between those imposed by the scenarios.

3.3.5.1.1 Synthetic examples Figure 3.19 illustrates this approach on two geometric examples. In the first case (first row), we make sure to match the triangular indentations together, whereas in the second case (second row), we discard them as noise, and match them against the directly corresponding rectangle pieces. The learning set sub-matrices were taken from the matrices of Figure 3.12. The range of scales explored by the matching algorithm was {0.5, 1.0, 1.5}. The neighborhood explored consisted of a rectangular window twice the size of the considered sub-matrix centered around its position in the shape distance matrix where it comes from. Every position within the rectangular window was considered. We used the correlation coefficient as a similarity measure and discarded matches below 0.2.

Incidentally, the same method can be used to rule out certain sub-matches. In this case, we give to all points in the sub-matrix found by the pattern matching technique of step 2 an arbitrarily high value: the dynamic programming algorithm will then avoid this area in search for the overall minimum cost path (this could be helpful to absolutely prevent the matching of the corpora callosa with parts of the fornix when we know that they are mistakes in the delineations to be processed, for instance).

3.3.5.1.2 Biomedical examples. Figure 3.20 demonstrates the behavior of our matching approach on a series of corpus callosum delineations. Two pairs of corpus callosum were *a priori* manually reparameterized (a): a number of pair of corresponding points were manually selected, and a continuous corresponding field was interpolated between them. We computed their associated shape distance matrix and selected in it two "scenarios" (control of the matching of the head and of the tail of the callosum) to serve as a learning set (b). We applied our learning approach to a test set of 20 callosum. Here also, the range of scales explored by the

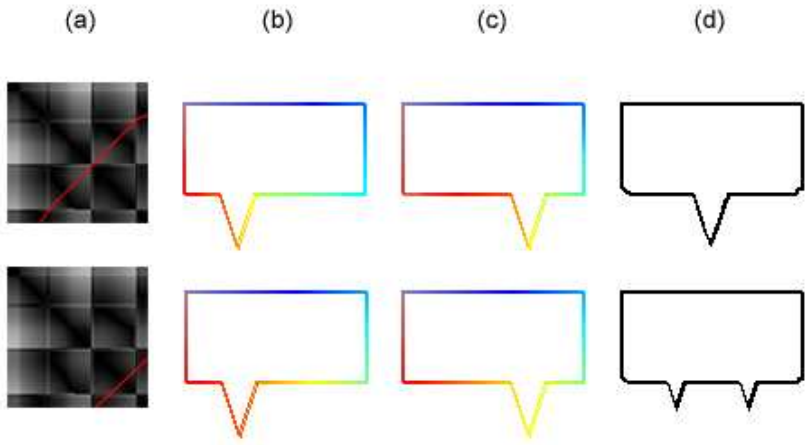


Figure 3.19: Pattern matching examples: (a) learning set, (b and c) reparameterized curves, (d) the resulting point-by-point Euclidean average curve.

matching algorithm was {0.5, 1.0, 1.5} and we considered the same exploration neighborhood (a rectangular window twice the size of the considered sub-matrix centered around it). We used the correlation coefficient as a similarity measure and discarded matches below 0.2. Figure 3.20(c) shows 5 out of the 20 reparameterized corpora callosa, and the overall callosum average in (d), computed as the point-by-point Euclidean mean of the reparameterized callosum. Clearly, both the head and the tail of the callosum were correctly matched across the set of instances.

3.3.5.1.3 Robustness of the matching approach The robustness of this approach mostly depends on that of the pattern matching algorithm used in step 2, that is, on the characteristics of the used similarity measure and of the explored neighborhood.

Similarity measure: Since we are matching shape distance sub-matrices against each other, a similarity measured designed for mono-modal registration should be used. Here, we selected the correlation coefficient, which assumes a linear relationships between both sub-matrices. When a greater dissimilarity is expected (that is, when the samples from the learning set might be fairly different from the objects to be matched), more general purpose measures could be used, the correlation ratio or mutual information for instance.

Exploration neighborhood: The size of the exploration neighborhood depends on the differences between the *global* shapes of the objects to be matched and the *a priori* matched objects in the learning set:

- when they differ substantially, the exploration neighborhood must be relatively large (it could even be extended to the entire shape distance matrix) as the sub-matrix in the shape distance matrix of the objects to be matched might occupy a completely different position than that of the corresponding sub-matrix in the learning set.
- when the object are globally similar in shape, then the shape distance matrices must also look alike: the exploration neighborhood can consequently be reduced as the corresponding sub-matrices should occupy close locations.

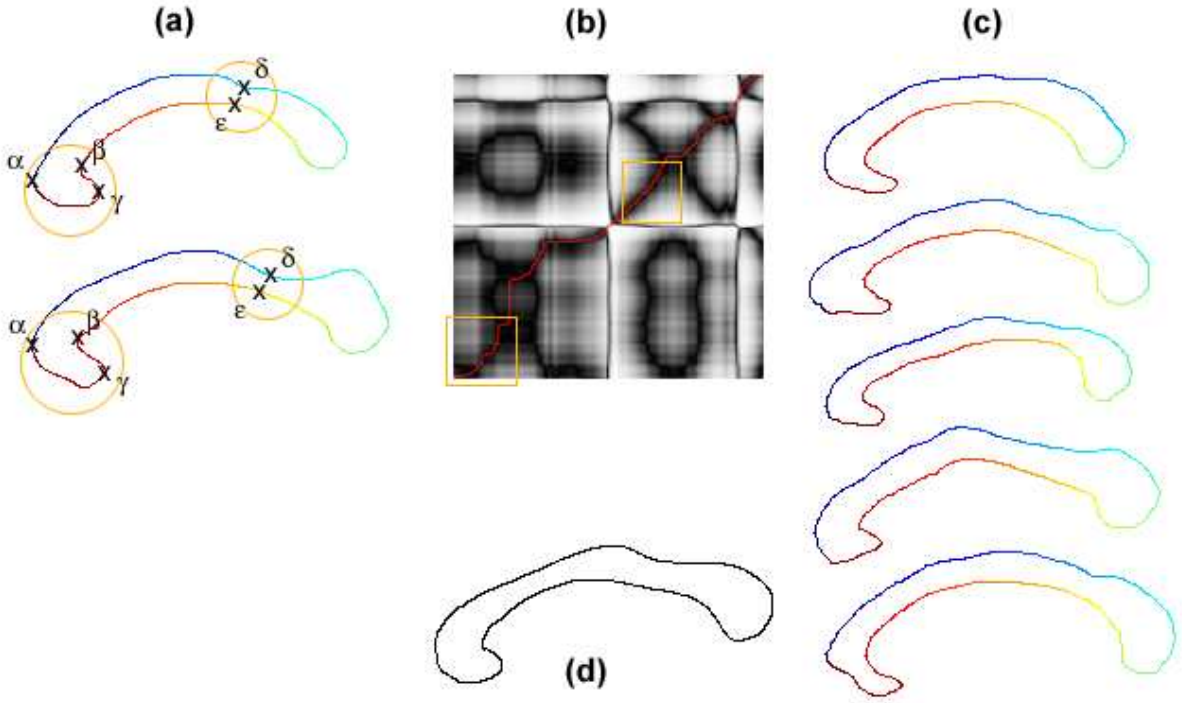


Figure 3.20: Pattern matching reparameterization of a few corpus callosum: (a) Two pairs of *a priori* parameterized callosum (greek letters indicate matching pairs of points); (b) their associated shape distance matrix with the corresponding path in red and the selected scenario/shape distance sub-matrices in orange; (c) a few reparameterized corpus callosum; (d) the overall point-by-point Euclidean average callosum.

Scale: The range of explored scale also depends on the expected differences between the learning set and the objects to be matched:

- if the sizes between these are expected to be about the same (for anatomical structures for instance) then a small range of scale is to be used, it should only be sufficiently large to allow for variations caused by noise;
- when the sizes are expected to be more different, then obviously a larger range should be used.

3.3.5.1.4 Choosing the starting and ending points S and E. Recall that the use of the dynamic programming framework requires that the starting and ending point of the curve ϕ be specified *a priori*. Our matching approach provides an alternative means to select those points, provided that they exhibit particular characteristics. Namely, we can extract from the shape distance matrix a scenario for the starting and for the ending point and use shape matching to find them in the shape distance matrix associated to a pair of input curves to be reparameterized. In view of the redundant nature of the shape distance matrix, this provides us with a selection technique more robust than those using maxima of curvature or position, for instance.

3.3.5.2 Towards a registration approach to the n-D reparameterization problem [m-D case: $m \in \mathbb{N}_*, n \in \mathbb{N}_*, m \leq n$]

Even though noticeable patterns are still present in higher dimensional distance matrices, the lack of a single-scalar parameterization for n-D objects prevents us from using the dynamic programming approach. However, we can still capitalize on the advantageous aspects of the shape distance matrix by considering the problem of reparameterization between two objects to be that of deforming and adapting a hyper-surface given *a priori* (associated to an *a priori* shape distance matrix) to the shape distance matrix of the input objects. In doing so, we avoid the issue of the parameterization of the input objects (and can thus consider collections of points).

The resulting algorithm (see Figure 3.21) is very similar to that of Section 3.3.5.1:

- Given a number of 2m-D shape distance matrices computed from pairs of already matched objects (and their associated matching hyper-surfaces), we non-linearly register them to the shape distance matrix computed from the two input objects.
- The resulting non-linear transforms are then applied to the distance maps of the hyper-surfaces associated to the learning items.
- These transformed distance maps are then averaged and the zero-level set of the average map becomes the new reparameterization. Note that a matching criterion (the integral of the deformation field for instance) could be used to compute a weighted average.

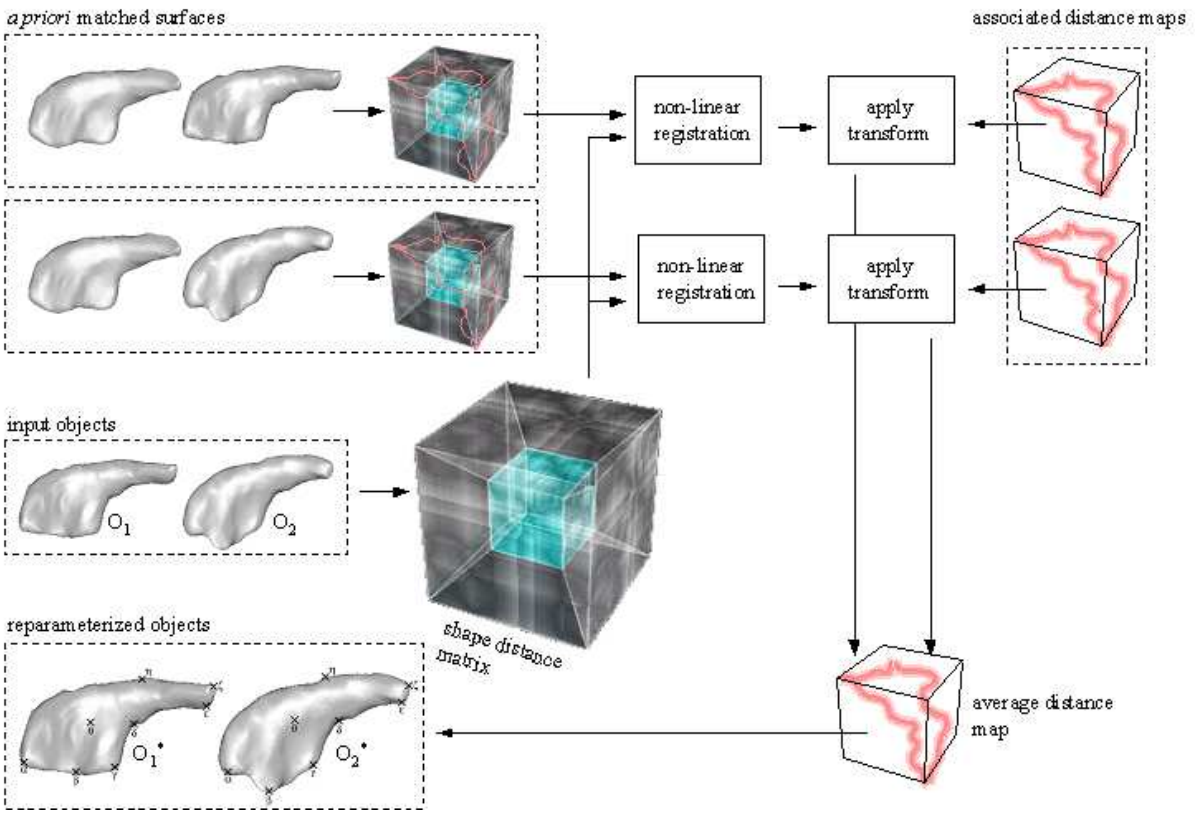


Figure 3.21: Overview of the registration approach to reparameterization.

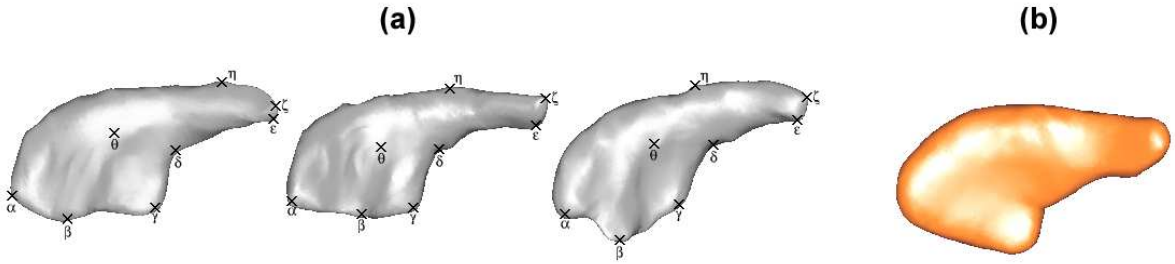


Figure 3.22: Illustation of the registration approach: (a) three reparameterized sample caudate nuclei out of a database of 20 instances; (b) overall Euclidean point-by-point average of the 20 reparameterized nuclei.

With this approach, we transform an m-D matching issue into a 2m-D registration problem[10]. Despite the curse of dimensionality, we are left with a somewhat simpler problem given the high expressivity of the distance matrices (see [102] for a similar dimension increase for surface matching). Consequently, the performance of our method depends on the robustness and accuracy of the non-linear registration algorithm. In the 1-D case (2-D shape distance matrix) we use the PASHA method [31] where the amount of regularization depends on the estimated discrepancy between the instances in the learning set and the objects to be reparameterized. We have adapted it in 4-D to treat the 2-D case (4-D shape distance matrix). Even though extending it to 6-D (3-D case) is not theoretically impossible, the size of the search space makes the registration intractable.

Note that what we propose here is very much work in progress and should be considered more as a series of guidelines towards a learning approach to the m-D reparameterization problem than an actual m-D extension. In particular, the problem of the inital parameterizations of the objects (which must be "similar") is a non-trivial issue in an out of itself.

Figure 3.22 shows how our registration method behaved on a series of 20 caudate nuclei (*(u,v)*-parameterized surfaces obtained from sets of 2-D slice-by-slice delineations in MRI). One caudate was selected as a target and the remaining 19 others were resampled together with it, using a 2-caudate learning set built by an expert neuroanatomist. We show in (a) 3 reparameterized sample caudates (out of the 20) with some corresponding points (Greek letters) and the resulting mean caudate in (b), obtained by averaging homologous points across the resampled caudate set. Visual inspection confirmed the agreement between the parameterization of the structures in the learning set and those in the test set.

Although these preliminary results are promising, the outrageously long processing time required to achieve the reparameterization of a 20-instance learning set (about 50 hours on a SGI Octane workstation) confirms the technical difficulties linked to the curse of dimensionality. To prove useful in practice, sparse space management approaches and faster registration processes should be implemented.

[10]Non-linear registration was also used in [39] by Cohen-Or *et al.* to find the correspondence field between 3-D voxel sets.

3.3.5.3 Building the learning set

Our learning approaches require that the correspondences between the objects of the learning set be established *a priori*. This may not be a trivial task for 3-D objects with complex shapes. However, it only has to be specified once and for a small number of instances. Also a sparse subset of the correspondence field could be specified by the user to generate a learning set. Most of the fully automated techniques presented in the introduction could produce a meaningful set that could then be manually corrected if need be.

Note that using a learning set implies that the objects we want to reparameterize should not be too different from those in the learning set. In fact, similarities between objects do not matter so much as similarities between the pairs of objects to be reparameterized and the pairs of objects in the learning set. Of course, the former is a sufficient condition for the latter. However, a unique advantage of our approach lies in its ability to learn a matching strategy for even very dissimilar objects, provided that we apply it to the same dissimilar matching situations.

As such, this technique could prove particularly useful to put into correspondence the "odd looking" instances of an object set as they might require more attention than the "ordinary" ones which could be treated with direct non-learning algorithms.

3.4 Building the shape models

Once the instances of the target structures have been reparameterized, we can build the associated statistical shape models, which will serve as constraints on the deformable templates, as detailed in Chapter 4.

We present here an overview of the classical principal component analysis (PCA) approach to shape statistics and a few results on one target structure.

3.4.1 PCA shape statistics

Given a set $S = \{S_1, \ldots, S_N\}$ of N instances of a target structure, we want to build the associated shape model. That is, we want to extract from S the average structure shape and its main modes of variations. This shape model, which synthesises the variability of the target structure, can then serve both to analyze a given instance of the structure and to generate new instances which are "similar" to the target, thereby providing a means to constrain our deformable models [46, 44, 202].

3.4.1.1 Reparameterizing the instances

As argued above, we first need to put the structure instances in correspondence and compute a correspondence field between them. In the previous sections, we described a number of reparameterization techniques and introduced a few new learning approaches. As the segmentation system introduced in Chapter 4 mostly uses 3-D simplex meshes, we used the approach of Fleuté *et al.* [68] to reparameterize the instances. For the 2-D applications, we used our learning pattern matching techniques.

3.4.1.2 Aligning the set of instances

We first need to align the structure's instances into a common coordinate frame. Namely, we want to apply a transformation t to each instance to bring them all in a comparable configuration

with respect to a set of axes. A number of techniques have been detailed in the literature (please refer to [45] for a general overview), the most popular approach being Procrustes Analysis [81].

Specifically, if $\mathbb{T}$ is the group of allowed transformations and S_i are vectors of $\mathbb{R}^k$ (our case here since S_i are 3-D simplex meshes, that is, lists of 3-D vertices), then the mean shape $\hat{S}$ is given by:

$$\hat{S} = \inf_{S \in \mathbb{R}^k} \inf_{t_1,\ldots,t_N \in \mathbb{T}} \sum_{i=1}^{N} \|S - t_i S_i\|^2 \tag{3.19}$$

In [158], Pennec showed that $\hat{S}$ corresponded to the arithmetical mean of the S_i once they had been transformed to an optimal position. That is, once we have found $t_1, \ldots, t_N \in \mathbb{T}$ such that:

$$\sum_{1 \leq i < j \leq N} \|t_i S_i - t_j S_j\|_2^2 = \inf_{t_1,\ldots,t_N \in \mathbb{T}} \sum_{1 \leq i < j \leq N} \|t_i S_i - t_j S_j\|_2^2 \tag{3.20}$$

then

$$\hat{S} = \frac{1}{N} \sum_{i=1}^{N} t_i S_i \tag{3.21}$$

We restrict $\mathbb{T}$ to the group of similarities (rotations, translations, changes of scale) and used the technique described in Raynaud's approach [168] (closely derived from Fleuté's [68] mesh deformation technique) to find the corresponding set of transformations $\{t_1, \ldots, t_N\}$. This consists of an iterative procedure where the various instances at iteration l are registered to their mean at iteration $l-1$. The process starts by selecting one of those instances as a target for the first round of registrations and at the end of each such round. We eventually obtain a mean structure with the desired property that there does not exist any transformation that could further decrease the distances between the structure's instances and the mean.

Note that such approach is cruder than the 1-D techniques we introduced earlier. It however works well in practice and is very fast.

3.4.1.3 Capturing the statistics

Given the set of N aligned reparameterized instances $\{S_i^* = t_i S_i\}_{i=1}^N$ and the mean shape $\hat{S}$, we compute the centered covariance matrix K:

$$K = \frac{1}{N-1} \sum_{i=1}^{N} \left(S_i^* - \hat{S}\right) \left(S_i^* - \hat{S}\right)^T \tag{3.22}$$

The eigenvectors[11] of K describe the modes of variation, and the vectors corresponding to the largest eigenvalues describe the most significant ones. Any shape S_i^* can then be approximated:

$$S_i^* = \hat{S} + Q a^i \tag{3.23}$$

where $Q = (q_1 | \ldots | q_m)$ is the matrix of the first m eigenvectors of K, and $a^i = \left(a_1^i, \ldots, a_m^i\right)^T \in \mathbb{R}^m$ is the so-called shape vector. By varying a^i, we can generate new instances of the shape model.

[11] We use the classical technique of Turk *et al.* [202] to apply principal component analysis to our N k-dimensional vectors S_i^*, with $N < k$.

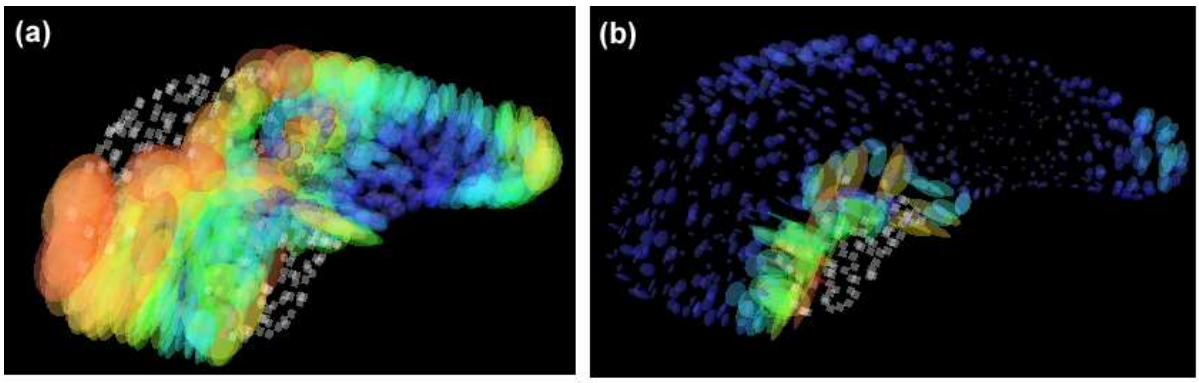


Figure 3.23: Variability of the caudate nucleus before similarity registration (a) and after (b).

3.4.1.4 Choosing the number of modes

There are no general prescriptions for choosing the number of modes. The simplest way to do it consist of making sure that the shape model represents a given proportion of the total variance exhibited by the structure's instances. Since the eigenvalues λ_i of K represent the variances of the data about the mean shape $\hat{S}$ in the direction of their associated eigenvectors q_i, we can choose m such that:

$$m = \arg\min_n \left| \sum_{i=1}^{n} \lambda_i - \gamma \sum_{i=1}^{N} \lambda_i \right| \tag{3.24}$$

where $\gamma \in [0, 1]$ is the desired proportion of the total variance $\sum_{i=1}^{N} \lambda_i$.

Alternatively, we could choose the number of modes which yields a model that best approximates all the instances of the target structure. Also, the number of mode can depend on the noise associated to the vertices of our instances.

3.4.2 PCA analysis of the corpus callosum

This PCA approach was applied to a series of 2-D corpus callosum delineations (the callosa were manually outlined in the mid-sagittal section of MRIs for 20 patients). We used our learning reparameterization approach to compute the dense correspondence fields between all 20 delineations. The learning set we used was that described in Section 3.3.5.1.2. Figure 3.24 shows the first three modes of variation. For each mode (#1, #2, #3), we present the mean shape (middle), and mean shape plus (right) or minus (left) three standard deviation (that is, plus or minus three times the square root of the associated eigenvalue).

From an analysis of the figure, it seems that the first mode of variation corresponds to the amount of bending of the callosum, while the second mode may be linked to the width of its body and the thrid mode to the shape of the head and tail.

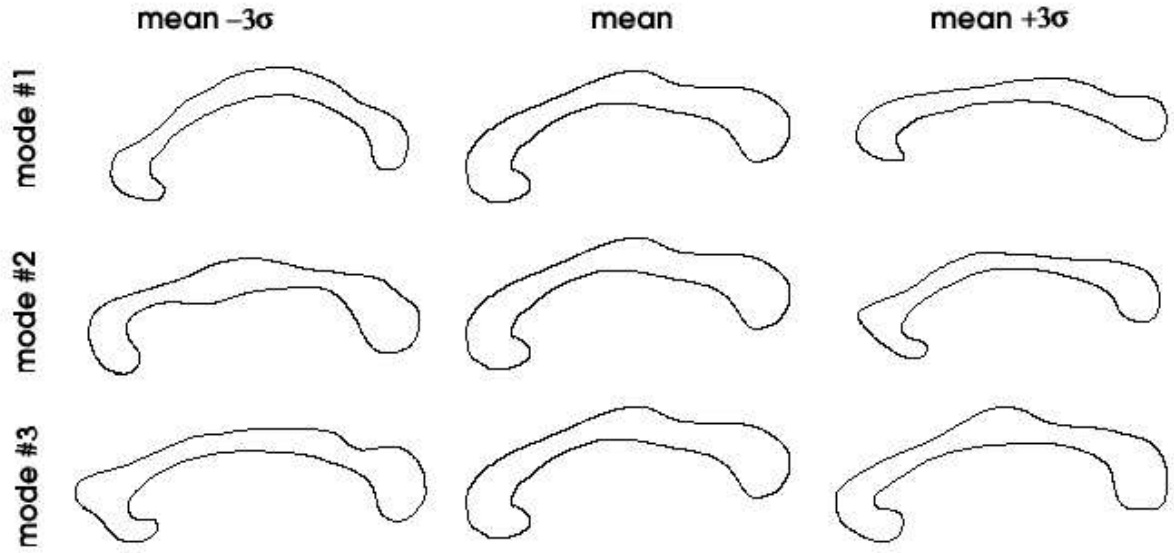


Figure 3.24: Principal modes of variation of the corpus callosum in the mid-sagittal section of the human brain.

3.4.3 PCA analysis of the caudate nucleus

For each target structure (corpus callosum, caudate nucleus, ventricles and hippocampus), a number of instances are available in the form of 3-D simplex meshes (see Chapter 4 for details). We can then apply the PCA approach to the set of instances, where each instance is considered as lists of 3-D points in $\mathbb{R}^{3n}$ with n is the number of vertices in the associated mesh.

Figure 3.23 demonstrates the necessity to align a structure's instances before computing its modes of variations. It shows the variability of the caudate nucleus before (a) and after (b) the similarity registration of the caudate instances. We display, for each vertex of the PCA computed mean mesh, an ellipsoid whose characteristics depend on the eigenvalues and eigenvectors of the coordinate covariance matrix associated to that vertex, across the set of instances (the eigenvectors give the 3-D orientation of the ellipsoid, whose lengths are proportional to the eigenvalues). Each ellipsoid is also colored according to the norm of the eigenvectors, from red (lowest) to blue (highest). Gray cubes are displayed instead of ellipsoids when the variability is too high. In addition to illustrating the need for prior registration, this figure also indicates that the greatest amount of variability should be expected in the ventral part of the caudate.

The first three modes of variation (which together represent 80% of the total variation) are shown in Figure 3.25. Again, we present for each mode the mean shape (middle), and mean shape plus (right) or minus (left) three standard deviation (that is, plus or minus three times the square root of the associated eigenvalue). Clearly the first mode mostly explains the ventral variability of the head of the caudate (as expected by the analysis of the variability map, Figure 3.23). The second mode seems to control the thickness of the caudate and its antero-external variability. The third mode is harder to interpret.

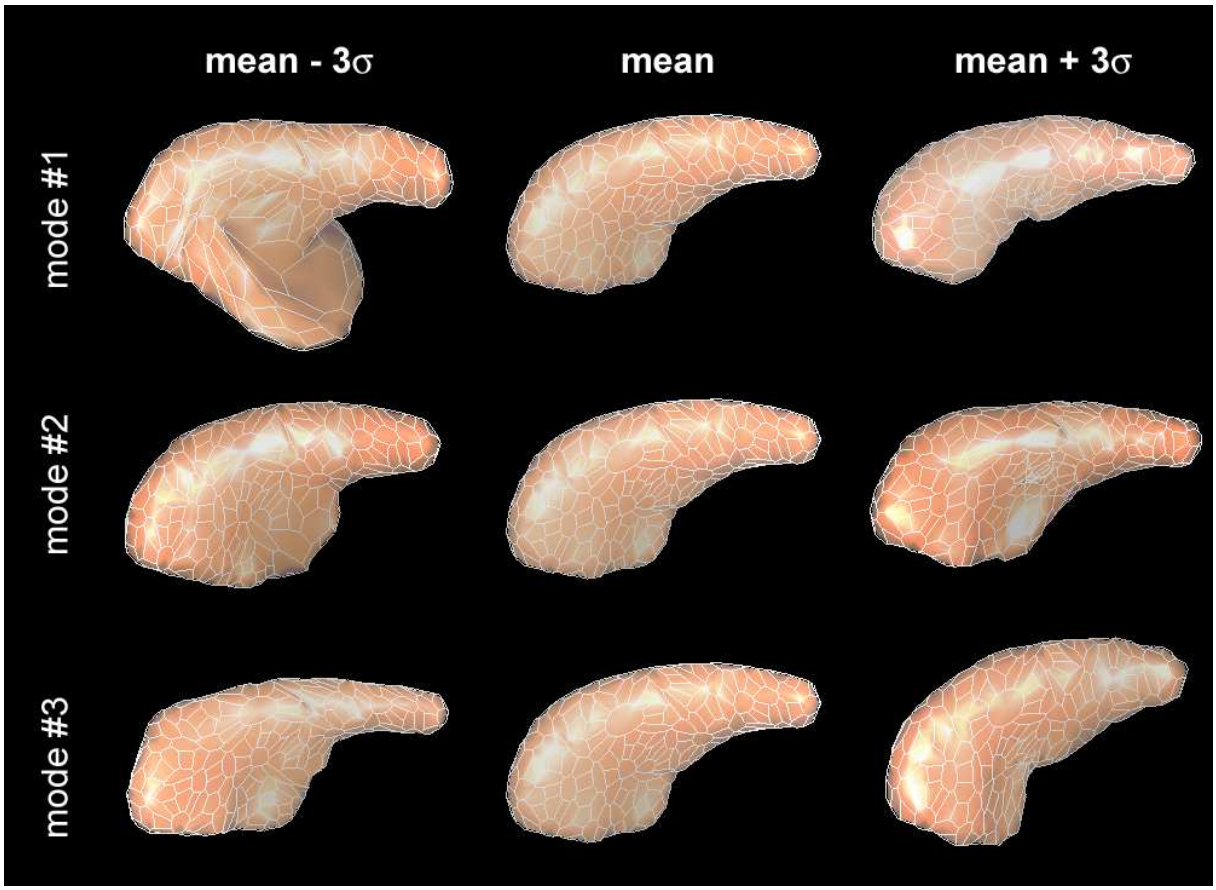


Figure 3.25: Principal modes of variation of the caudate nucleus.

3.5 Conclusion

This chapter dealt with the processes involved in the construction of meaningful shape models for anatomical structures. Built from a set of delineated instances, each model will serve as a constraint imposed on the deformable templates of our segmentation system, as detailed in Chapter 4, to bias the search for a structure's boundaries towards the most anatomically probable ones.

A careful analysis of the object matching problem in the context of statistical shape modeling motivated the need for *explicit control* over the reparameterization process that the structure's instances must undergo. We have consequently developed two learning approaches to the matching problem (based upon a classical shape difference geodesic search technique), where a learning set of *a priori* corresponding instances drive the matching algorithm towards a correspondence field whose characteristics bear close resemblance to those in the learning set.

These matching approaches required that a suitable shape descriptor be selected. As those encountered in the literature did not suit our set of requirements, we developed our own: the observed transport descriptor, whose remarkable discriminating power and robustness to noise make it particularly amenable to our matching framework.

We have applied both the OT descriptor and our matching algorithm to a variety of applications and demonstrated their adequate behaviour and flexibility. These also illustrated the

inherent simplicity of our approach: given a matching scenario, the visual qualities of the shape distance matrix make it a lot easier to build a learning set by selecting sub-matrices in it than having to design the suitable similarity measure and set of associated constraints that a classical approach would require to properly handle the scenario in the same satisfactory fashion.

Finally, a classical principal component analysis approach was used to build the models out of the reparameterized instances.

Chapter 4

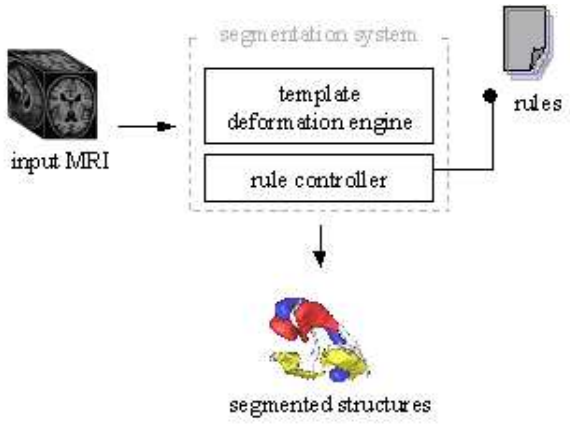
Knowledge-driven Segmentation
Segmentation Guidée par l'Expertise

[On the remote pages of the Celestial Emporium of Benevolent Knowledge,]
it is written that animals are divided into:

1. those that belong to the Emperor
2. embalmed ones,
3. those that are trained,
4. suckling pigs,
5. mermaids,
6. fabulous ones,
7. stray dogs,
8. those included in the present classification,
9. those that tremble as if they were mad,
10. innumerable ones,
11. those drawn with a very fine camelhair brush
12. others,
13. those that have just broken a flower vase,
14. those that from a long way off look like flies.

Jorge Luis Borges, *The Analytical Language of John Wilkins*

Nous détaillons dans ce chapitre le système de segmentation en lui-même. Après une revue des modèles déformables (dont nous aurons auparavant justifié l'usage) organisée suivant une taxonomie implicite/explicite, nous décrivons le cadre général de déformation des patrons. Celui-ci s'appuie sur une combinaison de contraintes à la fois implicites/semi-implicites (texture, forme, distance) et une séries de règles explicites. Les résultats obtenus démontrent l'efficacité de cette approche et la capacité de l'architecture règles/méta-règles à réagir de façon adéquate en présence de difficultés particulières.



Segmentation system.

Developed as a replacement for manual delineation processes plagued by tediously long completion times and too great an intra-/inter- operator variability, automated segmentation systems for biomedical images are a first step towards the quantitative analyses required to better understand the correlations between the functions of the brain and the morphology of its structures. As argued above, the effective segmentation of anatomical structures in MR images proves all the more challenging that the required precision is high, since the quality of the subsequent studies often depends on how accurately the various structures in the image can be delineated. In view of the characteristics of the input images, deformable templates seem the adequate paradigm within which developing an automated segmentation system, as they allow the introduction of the *a priori* medical expertise necessary to warrant an accurate delineation. After an overview of expert-guided segmentation methods, we detail in this chapter how the various constraints that we have studied previously can be incorporated into a knowledge-guided segmentation system to bias the evolution of a group of deformable templates under the overall control of a series of segmentation rules.

4.1 Introduction

A key component to the comprehension of the inter-relationships between brain structures and functions, the segmentation of anatomical structures is a major objective of neuro-informatics.

After a rapid definition of image segmentation (Section 4.1.1), we propose in Section 4.1.3 a taxonomy of deformable models, sorted with respect to the amount of explicit or implicit knowledge they rely on. Our expert-knowledge guided segmentation system is then detailed in Section 4.2. Finally, segmentation results and considerations about sensitivity to noise and parameters are discussed in Section 4.3.

4.1.1 Image segmentation

From a mathematical point of view, image segmentation is an injective process which maps sets of pixels (low-level numerical bits of information) to decorated regions (high-level semantic information). Note that these decorations can be actual labels (in reference to some dictionary of labels) or simply ordinal numbers to differentiate the associated regions.

More formally, we get:

Definition 1.2.1: Let I be an input image, defined by its intensity value in each point of its domain $I : \Omega \rightarrow \mathbb{R}$. Let ω be a non-empty subset of Ω. Then let Λ be a predicate which assigns the value true or false to ω.

Definition 1.2.2: A segmentation of image I for predicate Λ is a partition of Ω into n disjoint non-empty subsets $\{\omega_i\}_{i=1}^{n}$ such that:

- $\forall i,\ 1 \leq i \leq n,\ \omega_i \subset \Omega$;
- $\forall i,\ 1 \leq i \leq n,\ \omega_i \neq \emptyset$;
- $\forall i, j,\ 1 \leq i, j \leq n,\ i \neq j, \omega_i \cap \omega_j = \emptyset$;
- $\bigcup_{i=1}^{n} \omega_i = \Omega$;
- $\forall i,\ 1 \leq i \leq n,\ \Lambda(\omega_i) = \text{true}$; and
- $\forall i, j,\ 1 \leq i, j \leq n,\ i \neq j,\ \Lambda(\omega_i \cup \omega_j) = \text{false}$.

For instance, if regions in I were defined by a constant intensity, then a choice for Λ could be:

$$\begin{array}{rcl} \Lambda : \Omega & \rightarrow & \{\text{true, false}\} \\ \omega & \mapsto & \begin{cases} \text{true} & \text{if } \exists c \in \mathbb{R},\ \forall x \in \omega,\ I(x) = c \\ \text{false} & \text{otherwise} \end{cases} \end{array}$$

Here, Λ depends both on static and dynamic information, static in the form of the intensity in I and of prior knowledge about the target structure, and dynamic as Λ varies along with the evolving templates. It should also be noted that Λ often relies on a neighborhood of ω_i to assign it a boolean value (in that, it is not a uniformity predicate).

In this chapter, we are concerned with the segmentation of the input MR image into as many classes as there are target structures plus a "miscellaneous" class which contains all the voxels with do not belong to any target.

4.1.2 Deformable models

In view of the many shortcomings of manual delineation (see Section 1.2.1), a variety of automated segmentation methods have been discussed in the literature to extract anatomical structures from biomedical images, using a no less varied range of feature descriptors and shape models. The choice of a segmentation paradigm is pivotal as it conditions the ability of the segmentation system to extract anatomically meaningful delineations. A careful analysis of the features of the input images (MRI, in our case) lead us to favor deformable templates as a support for our segmentation system, for the following reasons.

First, they can adequately handle the various discontinuities and irregularities that sampling artifacts or noise may induce along the boundaries of the target structures. Then, their compactness enables the description of a wide variety of shapes while minimizing the overall number of parameters or masking these behind a small and easily manageable set of physical principles. They often provide a local, if not global, analytical model of the structure once segmented, which facilitates its subsequent analysis. Finally, *a priori* knowledge about the shape, the location, or the appearance of the target structure can be used to guide the deformation process: deformable templates are then the framework of choice within which mixing bottom-up constraints computed from the input MRI with *a priori* top-down medical knowledge[1].

[1] as opposed to model-free pixel-level techniques, which work at a too local a level to avoid generating an annoying amount of spurious contours.

4.1.3 Taxonomy of deformable models

Reviews of various deformable template-oriented techniques can be found in [144] and [148]. We propose here a taxonomy of model-based approaches partitioned in two categories: implicit knowledge and explicit knowledge techniques.

Given a learning set of *a priori* segmented instances of an anatomical structure, implicit knowledge algorithms have to automatically discover the relationships and functional dependencies of the various parameters of the model. However, *explicit* information about the target structures is often available, in the form of medical expertise. For instance, the relative positions of most of the deep gray nuclei is fairly constant across individuals, anatomical structures are not to intersect, etc. From these observations, a series of rules can be derived to better drive the segmentation process. Broadly speaking, *explicit knowledge* approaches can be seen as a special case of *implicit knowledge* algorithms where the additional medical expertise provides short cuts in the search for the target structure.

We further note that most of the used models encode the shape of the target object, its intensity or a combination of both.

4.1.3.1 Implicit knowledge methods

4.1.3.1.1 Implicit shape models

Since the seminal work on spring loaded templates by Fischler *et al.* [66], many deformable shape template approaches have been proposed in the literature to incorporate prior knowledge about the shape of a structure. Often times, *a priori* knowledge about both the expected shape of a target structure and its modes of variation is projected onto a series of global shape parameters which may serve as a regularizing constraint.

Hand-crafted models: Early work frequently relied on highly specific models. Yuille *et al.* [224] chose to use circles and parabola to retrieve eye and mouth patterns in face pictures. The associated parameters were the radius of the circle and the characteristics of the parabolic curves, respectively. Noticing the elliptical shape of the vertebra in axial cross section images of the spine, Lipson *et al.* [129] used deformable ellipsoidal templates to extract their contours. Both methods present the advantage of providing a very economic description of the shape in terms of the number of required parameters but lack flexibility in that a new model with new parameters has to be developed with each new object.

Active shape models: Developed by Cootes and Taylor [46], active shape models are represented by both a set of boundary/landmark points and a series of relationships established between these points from the different instances of the training set. New shapes are modeled by combining in a linear fashion the eigenvectors of the variations from the mean shape. These eigenvectors encode the modes of variation of the shape, and define the characteristic pattern of a shape class. The shape parameter space serves as a means to enforce limits and constraints on the admissible shapes, and insure that the final extracted shape presents some similarity with the shape class, as established from the training set. Many variants have been presented in the literature [214, 100]. They either introduce more constraints or decrease the control over admissible shapes. These models however make a Gaussian hypothesis on the distribution of the points in space and require that *a priori* correspondence between the shape instances be established, a problem that we tackled in Chapter 3.

Medial representations: Blum introduced the medial representation in [19] as a model of biological growth and a natural geometry for biological shape. Pizer *et al.* derived a sampled medial model in[160]. Joshi *et al.* [111] used it within a Bayesian framework to incorporate prior knowledge of the anatomical variations. A multi-scale medial representation was used to build the exemplary templates needed to obtain prior information about the geometry and shape of the target anatomical structure. Within this framework, the anatomical variability of a structure corresponds to the distribution of the admissible transformation of the shape model. Medial representations are however difficult to build for non symetrical shapes and are notoriously plagued by topological difficulties.

Fourier decomposition: Staib *et al.* [188] used a similar Bayesian scheme to control the coefficients of an elliptic Fourier decomposition of the boundary of a deformable template. They introduced a likelihood functional, which encoded the spatial probability distribution of each model, to be maximized under a Bayesian framework. The distribution of the model parameters was derived from a learning set of instances of the target object, and served to constrain the deformable template towards the most likely shapes. Szekely *et al.* [193] added an elastic property to a Fourier decomposition to create elastically deformable Fourier surface models. A mean shape and its associated modes of variation were extracted via statistical analysis of a learning set of Fourier decomposed instances of the target structure. The elastic fit of the mean model in the shape space was used as a regularization constraint.

Spherical harmonics: Styner *et al.* [191] combined a fine-scale spherical harmonics boundary description with a coarse-scale sampled medial description. The SPHARM description, introduced by Brechbühler [23] is a global, fine scale parameterized description which represents shapes of *spherical* topology. It uses spherical harmonics as a basis function. Styner's medial models were computed automatically from a predefined shape space using pruned 3-D Voronoï skeletons to determine the stable medial branching topology.

Superquadrics: Metaxas *et al.* [145] devised deformable superquadrics which combined the global shape parameters of a conventional superellipsoid with the local degrees of freedom of a membrane spline. The relatively small number of parameters of the superellipsoid captured the overall shape of the target structure while the local spline component allowed flexible shape deformation in a Lagrangian dynamics formulation. Vemuri *et al.* [210] used the properties of an orthonormal wavelet basis to formulate a deformable superquadric model with the ability to continuously transform from local to global shape deformations. Such model can continuously span a large range of possible deformation: from highly constrained with very few parameters, to underconstrained with a variable degree of freedom. Here again, a Bayesian framework biased the deformable models towards a range of admissible shapes.

Moment invariants: Poupon *et al.* [?] proposed the use of 3-D moment invariants as a way to embed shape distributions in deformable templates. They devised a framework capable of dealing with several simultaneously deforming templates, thanks to their fairly low updating cost, whose aim was the segmentation of deep grey nuclei in 3-D MRI. The remarkable stability of the invariant moments allowed them to study the anatomical variability of the segmented structures.

4.1.3.1.2 Implicit intensity models

Intensity models are used either alone, or in conjunction with a shape model. When used alone, they usually rely on region-growing algorithms where a seed positioned *a priori* grows voxel by voxel under some intensity distribution based constraints. The parameters of the distribution can be learned from a training set of instances. We will not elaborate on them here to focus on the approaches that use them in conjunction with a shape model.

Active appearance models: AAM [44] incorporate both a statistical model of the shape of the target, and a description of the statistical distribution of the gray-level intensities of the structure. Cootes *et al.* applied principal component analysis to both families of descriptors, thereby allowing for a robust search in a lower dimensional space. Their method used a learned correlation between model displacement and the resulting difference image.

Intensity profiles: A given instance of the target structure may not always exhibit homogeneous intensity distribution along its boundaries. Yet, the intensity may be locally characteristic. Intensity profiles, computed along the border of the structure models, then provide an efficient means to introduce *a priori* knowledge. Cootes *et al.* [43], for instance, modeled the statistical distribution of the intensity profile on each side of the structure surface. The mean profile, established from a structure learning set, was compared against the image data to determine the cost of a particular configuration of the model and guide the segmentation process. Brejl *et al.* [24] used a somewhat similar border appearance model to automatically design cost functions that served as a basis for the segmentation criteria of edge-based segmentation methods.

4.1.3.1.3 Indirect modeling

Indirect methods do not consider the target structure *per se* to build a shape model, but rather utilize the properties of related objects that interact with them, such as deformation fields. Davatzikos *et al.* [54] for instance suggested an analysis of shape morphometry via a spatially normalizing elastic transformation. Inter-subject comparisons were made by comparing the individual transformations. A similar approach in 3-D was developed by Joshi *et al.* [110] to compare hippocampi. Subject were compared by analyzing the transformation fields of a viscous fluid transformation. However, the analysis of fields in both methods has to cope with the high dimensionality of the transformation and the sensitivity to the initial position. Furthermore, the shape changes provided by the deformation fields are difficult to interpret. It may be that the use of principal component analysis or of other dimension reduction techniques could provide a more compact representation of the changes and allow for a convenient shape modeling.

4.1.3.2 Explicit knowledge methods

Most explicit methods combine shape and intensity descriptions within the same framework. Often time, explicit information is complemented or generalized by implicit information (for instance, a purely explicit position rule can be made more robust as a fuzzy condition, which however introduces non-explicit elements: the α parameter of the cut-off, the amount of diffusion, etc.).

Extended shape models: Noting that even though ASM can handle disconnected shapes it would be easier to partition a complex shape, that consists of several connected components, into a series of simpler and more manageable elements, Bernard *et al.* [14] devised a two-level hierarchical scheme to model both the shape and the topology of the resulting complex model. Each individual structure is controlled by its own ASM, and subject as well to an overall global ASM responsible for the relative positions and orientations of the set of components. The shape and appearance model, which describe plausible variations of shapes and appearances of individual structures, form the lower level, while the topological model, which describes the plausible topological variations of the articulated structure, forms the upper level.

Topological models: Amit [2] used a complex graph of landmarks, automatically chosen from the input images, to guide the registration process. A dynamic programming algorithm was used on decomposable subgraphs of the template graph to find the optimal match to a subset of the candidate points.

Fuzzy systems: In view of its ability to represent and merge uncertain or imprecise statements, fuzzy theory is particularly well suited to model shape.

Among others, Chang *et al.* [120] developed a fuzzy-controlled rule-based system capable of segmenting MR images of diseased human brains into physiologically and pathologically meaningful regions by incorporating expert knowledge about both brain structures and lesions. They introduced a novel technique to generate the memberships function of the fuzzy sets in a fully automated fashion. They used the distance between pixels and ventricle boundary as a fuzzy property of periventricular hyperintensity to help diagnose the studied disease.

Barra *et al.* [12] used information fusion to combine medical expertise with fuzzy maps of morphological, topological, and tissue constitution data to segment anatomical structure in brain MRIs. For instance, they encoded expert information about the relative position of 2 structures as a fuzzy distance map.

Wen *et al.* [217] used fuzzy-embedded human expert knowledge to evaluate the confidence level of two matching points using their multiple local image properties such as gradient direction and curvature.

Studholme *et al.* [190] merged region labeling information with classic iconic image registration algorithm via information fusion.

Expert systems: When the medical expertise can be captured by a series of simple positional, geometric or intensity rules, expert systems provide a convenient framework to assist the segmentation tasks. They can be used in conjunction with fuzzy logic fusion.

Ardizzone [4] developed a descriptive language to express the geometric features and relationships among areas of images belonging to the sections of interest. The descriptive language embedded first order logic, unification with backtracking and the sequential clause execution mechanism of Prolog. Complex processing could thus be performed by solving the relationships among the sets without any strict algorithmic description, thereby allowing greater flexibility.

Matesin *et al.* [142] too used a rule-based system to organize and classify features (such as brightness, area, neighborhood and relative position to the central symmetry axis) of

regions that had been automatically extracted via region growing. The IF-THEN rules were described in the language of the CLIPS expert system [75].

Brown *et al.* [28] segmented lung boundaries in chest X-ray images by matching an anatomical model to the image edges using parametric features under the guidance of a series of rules. They clearly distinguished between low-level processing that operated on raw image data, and high-level analysis which manipulated the symbolic objects of a knowledge-based domain.

Growe *et al.* [84] used a rule-based semantic net to extract and match control points for the registration of remotely sensed images. The appropriate control points and corresponding features are selected by an explicit knowledge-based scheme and the best correspondence is found by an A* algorithm.

Li *et al.* described in [127] a knowledge-based image interpretation system for the segmentation and labeling of a series of 2-D brain X-ray CT-scans. Their model, which contained both analogical and prepositional knowledge on the brain structures, helped interpret the image primitive information produced by different low-level vision techniques.

4.1.4 Our approach

We approach here the issue of boundary finding as a process of fitting a series of deformable templates to the contours of the anatomical structures. We chose simplex meshes [58] to model the templates, owing to their fairly simple geometry, which facilitates the incorporation of external and internal constraints.

Simplex meshes: they are discrete model representations (set of vertices and edges) with prescribed vertex connectivity. In a fashion similar to triangular meshes (of which they are the duals), simplex meshes can represent surfaces of all topologies.

To encode the structure surfaces, we use 2-simplex meshes: each vertex is then connected to exactly three neighbors. This inherent geometric simplicity greatly eases the imposition of constraints (both internal and external) to bias the segmentation process. Additionally, "zones" (subsets of vertices with their associated edges) can be defined on the simplex meshes to further specify the constraints.

More formally, we define a k-simplex as a union of p-cells, $0 \leq p \leq k$, where p-cells are defined recurrently as follows:

- a *0*-cell of $\mathbb{R}^d$ is a point (vertex) in $\mathbb{R}^d$,
- a *1*-cell of $\mathbb{R}^d$ is an edge of $\mathbb{R}^d$ (that is, 2 distinct vertices),
- a p-cell c of $\mathbb{R}^d$ is a union of (p-*1*)-cells such that
  - every vertex in c belongs to p distinct (p-*1*)-cells
  - the intersection of two (p-*1*)-cells is either empty of a (p-*2*)-cell.

A *2*-cell is therefore a set of edges that have one and only one vertex in common, that is, a closed polygonal line of $\mathbb{R}^d$ or face. A k-simplex mesh of $\mathbb{R}^d$ is then a (k+*1*)-cell of $\mathbb{R}^d$. It is therefore a (k+*1*)-connected mesh, where each vertex has (k+*1*) neighboring vertices.

Figure 4.1 shows a sample simplex mesh on which a "zone" (sub-set of vertices) has been defined.

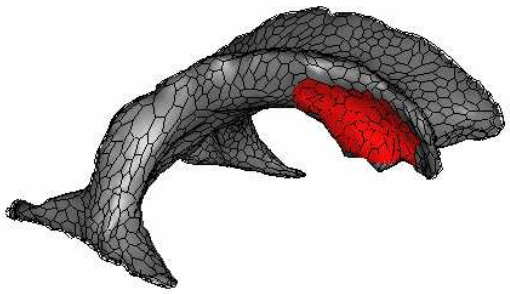


Figure 4.1: Gouraud rendering (gray) of the simplex mesh (black lines) associated to some lateral ventricles with a defined zone (in red).

We have focused on devising a segmentation system where maximum use is made of the available medical expertise, either in the form of implicit knowledge (the shape of the structures, their appearance, ...) or of explicit information (the relative distance between structures, the rules for the feedback loops, ...), implemented as constraints on the meshes. We then incorporate techniques from a variety of the approaches described in the above taxonomy.

Figure 4.2 illustrates the proposed framework. Briefly, non-linear registration of a hybrid iconic/surfacic atlas serves to initialize a series of simplex meshes (one mesh per target structure). Each simplex mesh is then iteratively modified to minimize a hybrid local/global energy which incorporates an internal regularization energy, an external term which couples the models to the underlying image features and a global shape-constrained term. Those meshes evolve in parallel, within a rule-controlled framework whose purpose is to maximize the achieved match over each structure while respecting some distance, position, etc. constraints, derived from medical expertise.

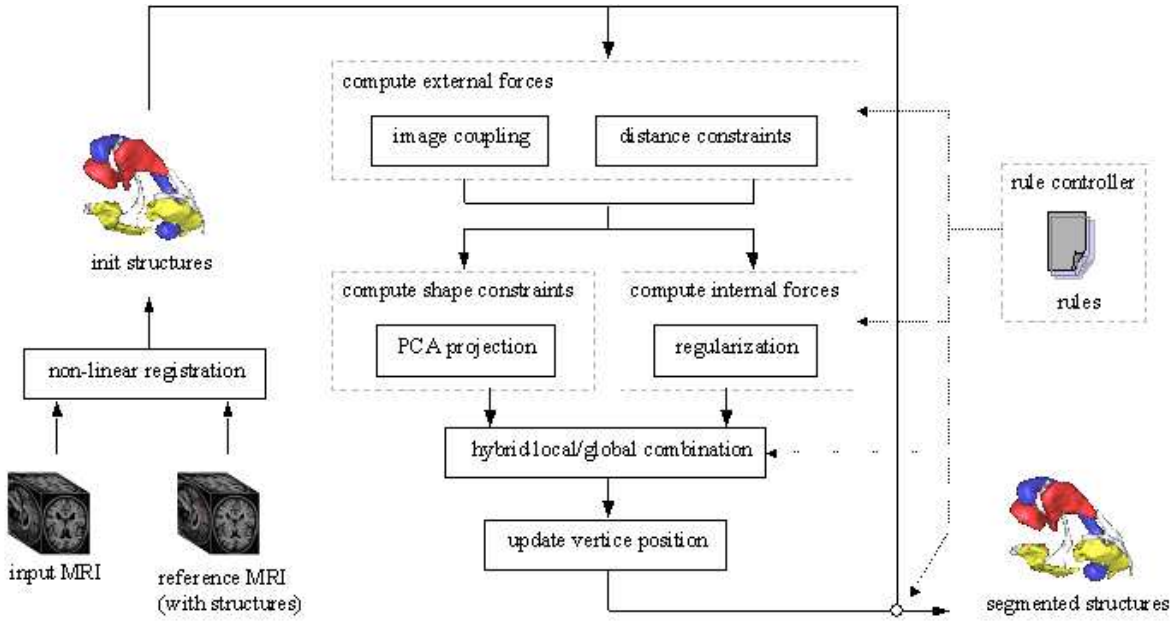


Figure 4.2: Overview of the proposed segmentation system.

4.2 Expert-knowledge guided segmentation system

We detail in this section the various components of our segmentation system and how they interact with each other under the supervision of segmentation rules.

4.2.1 Deformation model

We define the input MR image by its intensity at each point: $I : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}$, and a 2-simplex mesh as a set of points $\Pi_j = \left\{ P_j^i \in \mathbb{R}^3 \right\}_{i=1}^{N}$ along with its associated connectivity matrix (which is constant in our case since we do not allow topological changes). The algorithm's goal is then to find in I a pictorial object whose overall boundary fits that of Π_j. To guide the deformation process and drive the template towards the required object shape, we introduce a compound energy E whose minimum we aim to determine. Classically, E is made up of three terms:

- an internal (or regularization) energy $E_{internal}$ which characterizes the possible deformations of the template,
- an image coupling energy E_{image} which couples the template to the image, and
- a constraint energy $E_{constraint}$ which regroups the various constraints (shape, texture, etc.) we have mentioned in the previous chapters.

E_I is then written:

$$E_I = \alpha.E_{internal} + \beta.E_{image} + \gamma.E_{constraint} \tag{4.1}$$

with α, β, $\gamma \in \mathbb{R}$.

Within a Newtonian framework, we get the following iterative point updating procedure:

$$\begin{aligned} \Pi_j^{t+1} = \; & \Pi_j^t + (1-\delta)\left(\Pi_j^t - \Pi_j^{t-1}\right) \\ & +\alpha.f_{internal}\left(\Pi_j^t\right) + \beta.f_{image}\left(\Pi_j^t\right) \\ & +\gamma.f_{constraint}\left(\Pi_j^t, \left\{\Pi_k^t\right\}_k\right) \end{aligned} \tag{4.2}$$

where $\delta \in \mathbb{R}$ is a damping coefficient.

Image force: A number of image forces are available [148]. They can be based on the gradient of the input image, on a smoothed version of its associated edge-image, on intensity profiles, etc. We use here a force which depends on the distance to the closest strong gradient in the underlying image as it exhibits a good trade-off between precision and robustness [58]:

$$\forall P_j^i \in \Pi_j^t,\; f_{image}\left(P_{j,i}\right) = \left(\overrightarrow{\left(closest\left(P_j^i\right) - P_j^i\right)}.\overrightarrow{n_{\nabla\Pi_j}}\left(P_j^i\right)\right).\overrightarrow{n_{\nabla I}}\left(P_j^i\right)$$

where $\overrightarrow{n_{\nabla\Pi_j}}\left(P_j^i\right)$ is the normal of the simplex mesh Π_j at point P_j^i, and $closest\left(P_j^i\right)$ is computed by searching along the direction given by $\overrightarrow{n_{\nabla\Pi_j}}\left(P_j^i\right)$ for the strongest gradient in the image within a given exploration range (of size $size_{exploration}$), which depends on the expected distance between the point in the mesh and its final position in the target structure.

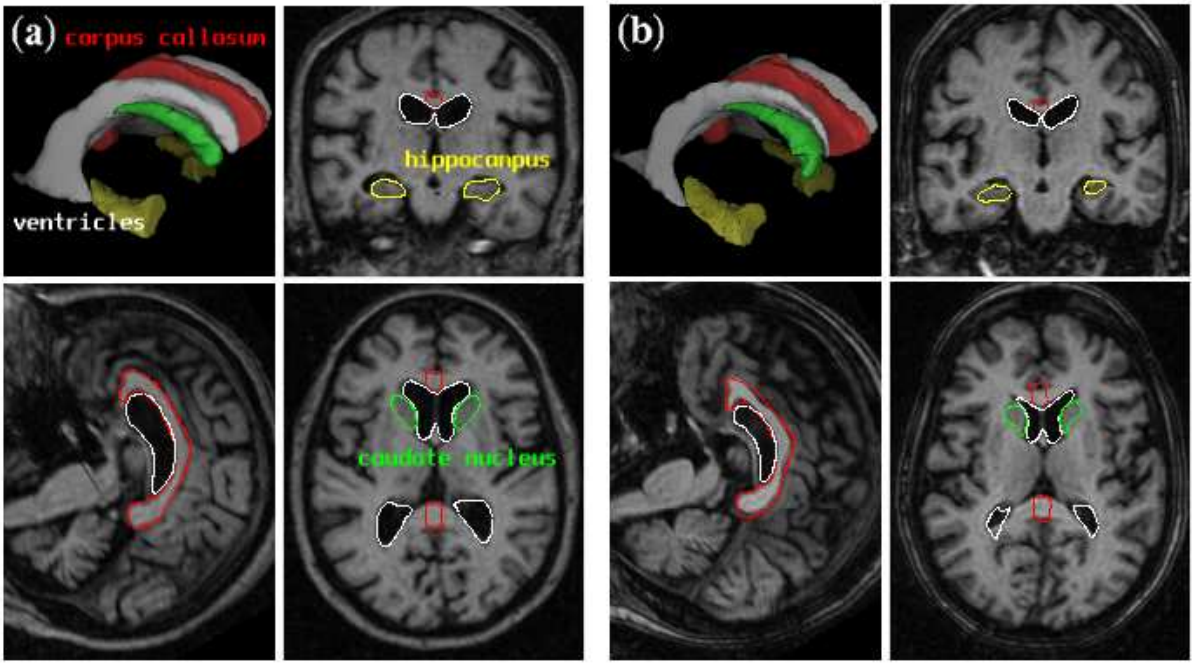


Figure 4.3: (a) reference MRI with superimposed manually delineated structures; (b) reference MRI registered to an input MRI and transformed structures;

Internal force: We implement an internal regularization by averaging the positions of simplex vertices over a spherical neighborhood of size $size_{regul}$.

4.2.2 Initialization

Once we have reduced the segmentation problem to an energy minimization task, we face a multi-modal, non-linear and possibly discontinuous function of many variables. Given the size and non-convexity of the solution space, most minimization technique would only lead to weak sub-optimal solutions (where the deformation model adapts to mere noise or decoys or maybe only follows parts of the desired boundaries) if the search space were not drastically reduced by assuming that a good approximation to the solution was available. This could be either in the form of a set of pose parameters (position, orientation, scale) or shape descriptors.

Various approaches have been presented in the literature to overcome this robustness issue. In [16] for instance, a coarse to fine strategy, the Graduated Non-Convexity Algorithm, is implemented, where a scalar parameter controls the amount of "local" convexity. Alternatively, the templates can be initialized at a number of locations and evolved in sequence: the deformed template with the best final match is then selected. In [159], a hybrid evolutionary algorithm controls a family of deformable templates that are evolved simultaneously and explore the search space in a robust fashion[2]. Here, we use nonlinear registration to initialize the templates reasonably close to their expected positions.

An MRI brain dataset was selected for its "standard" appearance (the reference MRI), and

[2]This same evolutionary programming approach also serves as an alternative way to introduce shape constraints into the deformable template paradigm and evade local minima, as detailed in Section 4.2.3.1.

we carefully segmented in it the target structures (see Figure 4.3(a)) following their established delineation protocols (see Appendix B). Given an input MRI to be processed, we first register the reference MRI to it with a non-linear registration algorithm with an elastic prior (the MAMAN algorithm [31]). The obtained transform is then applied to the meshes segmented in the reference MRI. Those transformed meshes serves as initial guesses for the segmentation of the target structures (Figure 4.3(b)).

The transformation is also applied to regions of interest (ROI) defined in the reference MRI around each target structure (see Section 2.4). The transformed ROIs then serve to speed up the computation of the texture maps and increase the performances of the texture classifiers (for efficiency reasons, we actually consider as a texture ROI the parallelepiped with edges aligned along the axis of the 3-D image closest to the transformed ROI).

Note that the non-linear registration of a true iconic atlas [42] (the average of a number of MRIs linearly or nonlinearly registered to themselves) used as reference MRI yielded inferior results since some of the strong features used by the MAMAN algorithm (edges, ridges, ...) were not as clearly defined in the true atlas than in the "average-looking" MRI.

Also, even though the affine registration of the reference MRI too gave good initializations, better results were achieved with a non-linear algorithm, most especially when the MRI to be segmented was substantially different from the reference MRI: in this case, a global affine transformation was less effective in aligning the internal structures.

4.2.3 Knowledge-based constraints

The evolution process of our deformable templates is biased by a number of constraints (either in the form of energy terms or of a series of rules) towards shapes that are dimmed more probable with respect to the *a priori* medical knowledge we have gathered on the target structures.

4.2.3.1 Statistical shape constraints

As detailed in Chapter 3, a shape model is available for each target structure. Namely, principal component analysis helps extract from a set of instances the average shape $\bar{S}$ and the principal modes of variations (given by the eigenmodes q_i of the covariance matrix K of the homologous vertices). The deformable templates must then be constrained accordingly.

In [44], the pose and shape parameters of the templates are adjusted by projecting the local deformation induced by the external energy onto the shape space as follows:

> Let $d\Pi^t_{image} = f_{image}\left(\Pi^t_j\right) - \Pi^t_j$ be the deformation induced by the image coupling forces. Let $\bar{S}$ be the mean shape computed for the target structure, and $Q = \{q_1, \ldots, q_m\}$ its m first eigenmodes. The shape-constrained deformation is written:

$$d\Pi^t_{shape} = \sum_{i=1}^{m} \left\langle \Pi^t_j + d\Pi^t_{image} - \bar{S}, q_i \right\rangle . q_i \tag{4.3}$$

This however limits the range of possible shapes to only the projections onto the shape space. Alternatively, hybrid deformation models can be crafted where the shape constraint bias the

deformation process in a less restricted manner. We present two such approaches here: a hybrid evolutionary algorithm[3] and a hybrid local/global deformation scheme.

Hybrid evolutionary programming

As opposed to sequential methods which progress step-by-step from an initial solution to the optimization problem (here, the initial estimate of the target structure) to a neighbor solution by performing some elementary modification (parameter updates), evolutionary techniques deal with a population of solutions [146]. The pivotal idea is to use the collective properties of a group of solutions to search the global solution space. Once an initial population of solutions has been generated, it is improved by a cyclic two-phase evolution process: an exploration phase (cooperation) and an exploitation phase (self-adaptation). In the exploration phase, solutions of the current population exchange information (recombining in a probabilistic fashion) with the aim of producing new solutions which inherit good attributes. In the exploitation phase, solutions change their internal structure without any interaction with the other members. Optimal solutions are thus evolved by iteratively producing new generations of solutions and selectively favoring the most successful ones.

A promising innovation in the field of general heuristics, hybrid evolutionary systems embed a sequential local search within the framework of a population-based strategy. The combination of a local search and a global control heuristic yields an algorithm with significantly better performance than that of both search methods running separately [48]. In a hybrid algorithm, the role of the sequential search is to explore thoroughly disjoint areas of the solution space, whereas the evolutionary procedure provides global guidance through the space. Hybrid evolutionary techniques have been successfully applied to many combinatorial optimization problems: the traveling salesman problem [149], the quadratic assignment problem [70], the bin-packing problem [65] or the graph coloring problem [67] among others. Our approach proceeds along these lines by integrating a local elastic template matching method within a global evolutionary scheme.

The shape model described in Chapter 3 can then be used within an evolutionary heuristic to constrain the dynamics of a *population* of deformable templates. Our evolutionary approach is similar to the guided evolutionary simulated annealing (GESA) method detailed in [223]. We have, however, modified it to incorporate the local deformation scheme from Section 4.2.1, thus turning it into a hybrid evolutionary heuristic which explores the solution space by interspersing periods of local optimization with phases of interaction and diversification.

Initialization: We consider a population of individuals $\{I^i\}_i$, each of which consists of a pose and a shape parameter: $I^i = \{s^i, \theta^i, T^i, a^i_1, \ldots, a^i_m\}$ with s^i the scale, θ^i the rotation, T^i the translation and $a^i = (a^i_1, \ldots, a^i_m)$ the shape vector.

The population is made up of families. Each family $family^i$ consists of a parent $parent^i$ and c_i children $child^{i,j}$. The pose parameter of each of the parents is selected at random to cover the entire input image. Their shape is that of the mean structure. The children in each family are generated by randomly modifying the pose parameters of the parents and choosing valid instances of the shape model:

$$\begin{aligned} family^i: \quad parent^i &= \left(s^i, \theta^i, T^i, 0, \ldots, 0\right) \\ child^{i,j} &= \left(\delta_s^{i,j}.s^i, \delta_\theta^{i,j}.\theta^i, \delta_T^{i,j}.T^i, \epsilon_1^{i,j}.\sqrt{\lambda_1}, \ldots, \epsilon_m^{i,j}.\sqrt{\lambda_m}\right) \end{aligned} \tag{4.4}$$

[3]historically developed for the 2-D segmentation of anatomical structure in brain MRI.

with $\delta_s^{i,j}$, $\delta_\theta^{i,j}$,$\delta_T^{i,j} \in \mathbb{R}$, and $\epsilon_1^{i,j}, \ldots, \epsilon_m^{i,j} \in \mathbb{R}$ drawn at random from a uniform distribution whose range decreases with each generation.

Algorithm: The hybrid evolutionary algorithm works as follows (stars in brackets indicate points further detailed below):

---

hybrid EP

- initialize the f families (select the parents, generate the children)

local search + local competition:

- for each family $family^i$
  - evolve the parent and the children under the local deformation template matching scheme for N_l iterations.
  - select the child $child^*$ with best fitness (e.g. minimum global energy)
  - if fitness $(child^*)$ < fitness $\left(parent^i\right)$ or
    if fitness $(child^*)$ > fitness $\left(parent^i\right)$ with probability $e^{\frac{\text{fitness}\left(parent^i\right)-\text{fitness}(child^*)}{T}}$
    then $child^*$ replaces $parent^i$ for next generation
  - project the new parent onto shape space [*]

2nd level competition [***]

- for each family family^i
  - $count = 0$
  - for each child $child^{i,j}$
    - if fitness $\left(child^{i,j}\right)$ < fitness $\left(parent^i\right)$ or
      if fitness $\left(child^{i,j}\right)$ > fitness $\left(parent^i\right)$ with probability $e^{\frac{\text{fitness}\left(parent^i\right)-\text{fitness}\left(child^{i,j}\right)}{T_{mode}}}$
      then $count = count + 1$ [**]
  - $acceptance\, count_i = count$

- for each family $family^i$
  - $c_i = acceptance\, count_i * c_i$/total number of children
  - generate the children

- decrease the temperature T

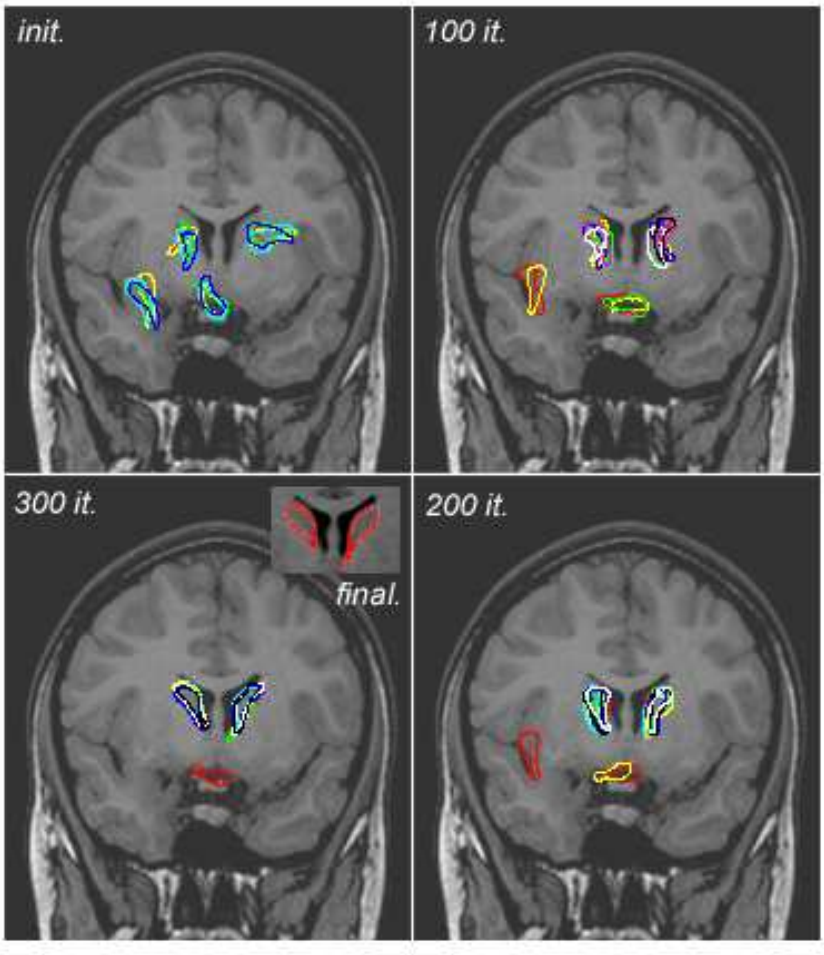


Figure 4.4: Dynamics of a population of templates under the supervision of the hybrid evolutionary heuristics.

[*] Projection onto the shape space: We use an approach similar to that described at the beginning of the Section.

Following the analysis of [22] and [48], we first register the template to the shape model. We then restrict the projection of onto the shape space so that the distance between the deformed template and the corresponding instance of the shape model is lower than three times the empirical standard deviation for each mode. We get:

$$d\Pi^t_{shape} = \sum_{i=1}^{m} \begin{array}{l} \text{sign} \left\langle \Pi^t_j + d\Pi^t_{image} - \bar{S}, q_i \right\rangle . \\ \min \left(\left\langle \Pi^t_j + d\Pi^t_{image} - \bar{S}, q_i \right\rangle , 3\sqrt{\lambda_i} \right) . q_i \end{array} \tag{4.5}$$

[**] To favor the children whose shapes are closer to the first eigenmodes, we vary the temperature of the Boltzman distribution that controls the acceptance number of each family as follows:

$$T_{mode} = \rho[mode] * T \tag{4.6}$$

where T is the simulated annealing (SA) temperature, ρ is an array of decreasing weights, and $mode = \arg\max_i |\langle d\Pi, q_i \rangle|$.

[***] As in [50], the approach implements two levels of competition. At a local level, the children of the same family compete with each other to generate the parent for the next generation. At a second level, there is a competition between the families themselves in that the number of children allocated to each family depends on the combined fitness of all the children, and is biased toward the first eigenmodes to favor the most likely shapes.

The number of children actually reflects the interest in a given area of the search space: the more good candidate solutions in a given area, the more attention we devote to it. The entire algorithm can be viewed as parallel SA with competition: each family is a multiple-trial-parallel simulated annealing machine with the children contributing the trials in parallel: it is, therefore, a very efficient global search technique as argued in [50]. An illustration of this evolutionary programming approach in the 2-D case is presented in Figure 4.4. We show the dynamics of a population of 2-D templates (four families, five children per family initially) applied to the segmentation of the caudate nuclei in a 2-D MR slice. Note how the families that are too far from the target structure quickly disappear as children cannot achieve a good match.

Note that by projecting the new parents into the shape space, we bias the search for a solution toward the set of valid instances of the shape model. Yet, incorporating an independent local search allows a better traversal of the search space. As illustrated in Figure 4.5, the considered individual could not leave its current location to reach the target minimum if it were restricted to move only in the lower dimensional shape space (as it would have to go up-hill) whereas the local deformable matching scheme can find a path (in white) in the solution space.

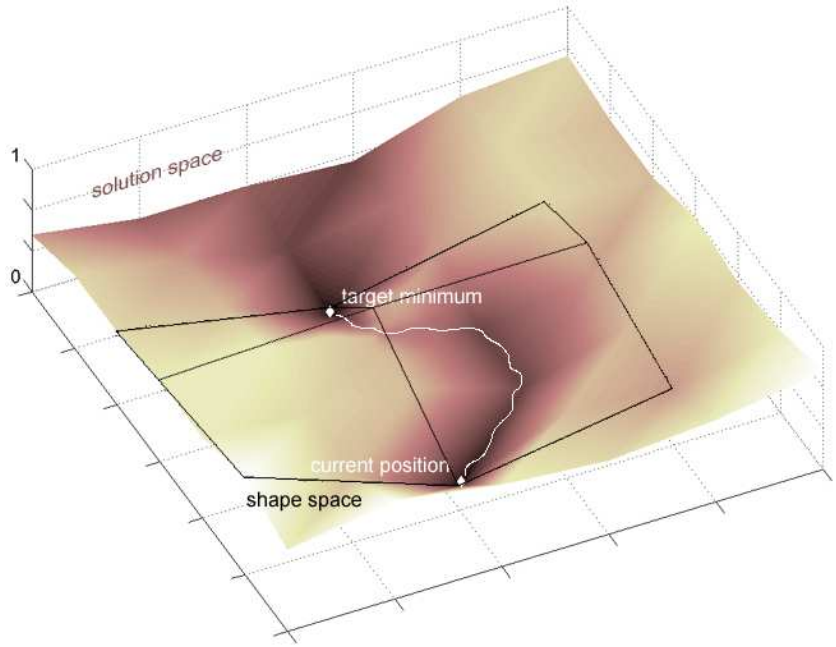


Figure 4.5: Shape space versus local deformation matching scheme.

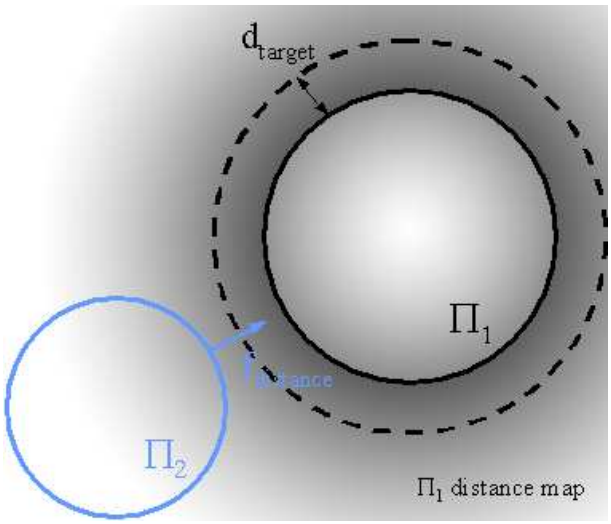


Figure 4.6: Illustration of the distance constraint mechanism.

Hybrid local/global deformation scheme

In spite of the remarkable robustness of the evolutionary programming approach (see [159] for details), we chose to use for our 3-D segmentation system a simpler and much less computationally intensive technique in the form of a hybrid local/global scheme [147]. Here, deformations are regularized through a combination of global (shape constrained) and local (external) forces.

The point updating rule becomes:

$$\begin{aligned}\Pi_j^{t+1} = \; & \Pi_j^t + (1-\delta)\left(\Pi_j^t - \Pi_j^{t-1}\right) \\ & +\lambda\left\{\alpha.f_{internal}\left(\Pi_j^t\right) + \beta.f_{image}\left(\Pi_j^t\right) + \gamma.f_{constraint}\left(\Pi_j^t, \{\Pi_k^t\}_k\right)\right\} \\ & +(1-\lambda)\,.\left\{d\Pi_{shape}^t\right\}\end{aligned} \tag{4.7}$$

where λ is the "locality" parameter which controls the contribution of the global shape-model constraint.

4.2.3.2 Distance constraint

The positions (and shapes) of nearby anatomical structures are not independent from each other. For instance, as the caudate nuclei are juxtaposed to the lateral ventricles, any change in the shape or position of the latter will affect those of the former. Information about the respective position of structures can then help the segmentation process.

In [12], fuzzy logic was used to express distance and positional relationships between structures. Here we choose distance maps (see Figure 4.6), owing to their ability to model more precise constraints (to guarantee non intersection for instance). Given a deformable template Π_0^t, we wish to impose on it a distance constraint with respect to template Π_1^t. We first compute the distance map D_1^t associated with a discrete sampling of Π_1^t. We use a classical Chamfer map

[21] algorithm[4] to compute a signed distance map, positive outside the discrete sampling of Π_1^t and negative inside. At each vertex $P_{0,i}^t$ of Π_0^t, we can then compute a "distance force" $f_{distance}$ whose magnitude depends on the value of the distance map at the considered vertex.

Two types of constraints can be applied:

- We can either wish the force to attract the vertex, along the direction of the gradient of the distance map, up to an exact distance d_{target} of the target mesh:

$$f_{distance}\left(P_{0,i}^t\right) = -\frac{\nabla D_1^t\left(P_{0,i}^t\right)}{\left\|\nabla D_1^t\left(P_{0,i}^t\right)\right\|}.\left(D_1^t\left(P_{0,i}^t\right) - d_{target}\right) \tag{4.8}$$

- Alternatively, we may want to only enforce that this vertex should remain at distance inferior or superior to d_{target} (to prevent intersections between structures for instance). We get:

  if $D_1^t\left(P_{0,i}^t\right) < d_{target}$ then

  $$f_{distance}\left(P_{0,i}^t\right) = -\frac{\nabla D_1^t\left(P_{0,i}^t\right)}{\left\|\nabla D_1^t\left(P_{0,i}^t\right)\right\|}.\left(D_1^t\left(P_{0,i}^t\right) - d_{target}\right)$$

  else

  $$f_{distance}\left(P_{0,i}^t\right) = 0$$

  to ensure a minimum distance ($d_{target} = 0$ for non-penetration) or

  if $D_1^t\left(P_{0,i}^t\right) > d_{target}$ then

  $$f_{distance}\left(P_{0,i}^t\right) = \frac{\nabla D_1^t\left(P_{0,i}^t\right)}{\left\|\nabla D_1^t\left(P_{0,i}^t\right)\right\|}.\left(D_1^t\left(P_{0,i}^t\right) - d_{target}\right)$$

  else

  $$f_{distance}\left(P_{0,i}^t\right) = 0$$

  to ensure a maximum distance.

Note that these forces can also be applied to a subset of the mesh vertices (so-called "zones") to enforce more local constraints. This however assumes that the simplex zones, which initially correspond to actual anatomical areas between which a distance constraint is to be established, will still correspond to these areas during the deformation process, a not unreasonable assumption when only limited displacements are considered (our case here, since the initialization step delivers remarkable performances).

[4] Note that a more precise Euclidean distance map computation could be carried out [143] if enhanced precision truly were necessary.

4.2.3.3 Texture constraint

A similar distance map scheme serves to build texture constraints from the texture classification maps produced by the classifiers described in Chapter 2. Namely, we compute the distance map D_i^t of the thresholded classification map associated to each target structure and derive a "texture force" $f_{texture}$ as follows:

$$\forall \Pi_j^t,\ \forall P_{j,i}^t,\ f_{texture}\left(P_{j,i}^t\right) = -\frac{\nabla D_j^t\left(P_{j,i}^t\right)}{\left\|\nabla D_j^t\left(P_{j,i}^t\right)\right\|}.D_j^t\left(P_{0,i}^t\right) \tag{4.9}$$

Note that since the texture maps are only computed inside regions of interest, the texture forces are also only available there.

4.2.3.4 Rule-controlled framework

With all the above mentioned contributions taken into account, the complete point updating rule is written:

$$\begin{aligned}\Pi_j^{t+1} = \ & \Pi_j^t + (1-\delta)\left(\Pi_j^t - \Pi_j^{t-1}\right) \\ & +\lambda\left\{\alpha.f_{internal}\left(\Pi_j^t\right) + \beta.f_{image}\left(\Pi_j^t\right)\right. \\ & \left. + \gamma_{distance}.f_{distance}\left(\Pi_j^t, \left\{\Pi_k^t\right\}_k\right) + \gamma_{texture}.f_{texture}\left(\Pi_j^t\right)\right\} \\ & +(1-\lambda).\left\{d\Pi_{shape}^t\right\}\end{aligned} \tag{4.10}$$

with δ, λ, α, β, $\gamma_{distance}$ and $\gamma_{texture} \in \mathbb{R}$, with $\alpha + \beta + \gamma_{distance} + \gamma_{texture} = 1$.

In view of the complexity of the segmentation task, choosing a value for the various scalar parameters that control the contributions of these constraints is not a trivial matter. Similarly, there is not a single prescription for the amount of regularization to apply to the templates. Instead of setting *a priori* sub-optimal values, those parameters could evolve dynamically along with the deformation process. Additionally, rather than segmenting the structures independently and running the risk of them intersecting one another, better segmentation results could be obtained by evolving the templates in parallel while controlling their inter-relationships.

We have consequently build a catalog of rules to control the dynamic properties of our deformable templates. For each target structure or pair of structures, a set of rules was developed that took into account recommendations from clinicians as well as low-level image observations (see Section 4.2.3.4.1). Note that in contrast to the expert system approached presented in Section 4.1.3.2, we did not implement a particular inference engine or logical framework for these rules. The latter were established in plain English and implemented using ad-hoc tools.

A feedback rule (Section 4.2.3.4.2) was also developed to monitor the possible mistakes during the deformation process and react adequately.

4.2.3.4.1 Rules

lateral ventricles: As the ventricles are fairly highly contrasted in T1-weighted MRIs with respect to the immediately surrounding tissues, the non-rigid transformation obtained via registration of the reference MRI to the input MRI usually gives an excellent estimate of the true boundaries. The texture filter also delivers excellent maps and we set $\gamma_{texture} = 0.6$.

With that in mind, and in view of the large variability of the ventricles, no shape constraint was used for their segmentation (Table 4.1 confirms that adding a shape constraint decreases the segmentation performances). For the same reason, only a small internal regularization energy is used: $\alpha = 0.1$.

$\delta = 0.1 \quad \gamma_{distance} = 0.0$
$\alpha = 0.1 \quad \gamma_{texture} = 0.6$
$\lambda = 1.0 \quad \beta = 0.3$

caudate nuclei: With the exception of the caudate tail, which the delineation protocol discards, the caudate nuclei from our training set were not that variable. We consequently use a reasonable shape weight: $\lambda = 0.3$. To prevent intersections with the lateral ventricles, a distance constraint is added. We define on each caudate (left and right) a zone corresponding to the contact area with the near-by lateral ventricle. A distance constraint with $d_{target} = 1$ ensures a good juxtaposition and prevents inter-penetrations.

$\delta = 0.1 \quad \gamma_{distance} = 0.3$
$\alpha = 0.1 \quad \gamma_{texture} = 0.3$
$\lambda = 0.3 \quad \beta = 0.3$

corpus callosum: A fairly variable structure (at least based on the analysis of the callosal instance in our training set), we did not use any shape constraint for the corpus callosum (here also, Table 4.1 agrees with this choice). A distance constraint with $d_{target} = 2$ ensures the non-intersection with the lateral ventricles.

$\delta = 0.1 \quad \gamma_{distance} = 0.2$
$\alpha = 0.1 \quad \gamma_{texture} = 0.6$
$\lambda = 1.0 \quad \beta = 0.1$

hippocampi: In view of the lack of contrast of the hippocampus with respect to some of its neighboring structures, the use of a shape constraint proved necessary ($\lambda = 0.3$) to interpolate the missing information. Since the performances of the texture classifier were not particularly high, we give the texture constraint a moderate weight: $\gamma_{texture} = 0.3$.

$\delta = 0.1 \quad \gamma_{distance} = 0.0$
$\alpha = 0.1 \quad \gamma_{texture} = 0.3$
$\lambda = 0.3 \quad \beta = 0.6$

parameter dynamics:

- We progressively refine the image gradient used in computing the external forces to guarantee deformation at early stages and later ensure a precise delineation (dynamic coarse-to-fine approach): the standard deviation of the Gaussian used to compute the gradient of I is initialized at 3.0 and decreased by 0.2 every 10 iterations.

- The locality parameter λ is slowly increased by 0.02 every 10 iterations as the deforming templates approach the borders of their target structures to allow a better adaptation to these borders.

4.2.3.4.2 Feedback rule

While classical rules control the behavior of the deformable templates, feedback rules controls the applicability of the rules themselves to ensure that no mistake is being made during the deformation process. As such, they can be considered as meta-rules. We have only developed and implemented one feedback rule for the time be.

leak prevention: To ensure that the deformable templates do not "leak" outside of the correct boundaries, we check at each iteration that their distances (mean distances averaged over all vertices) to their associated shape-constrained projections stay reasonable (below a structure-dependent threshold). Each time the threshold is reached, we increase the amount of regularization (α is increased) and the shape constraint if used (λ is decreased). As the deformation process goes along, the structure-dependent thresholds are increased to allow for finer deformations.

4.3 Segmentation results

We present here some qualitative and quantitative segmentation results for our four target structures: corpus callosum, caudate nucleus, lateral ventricles and hippocampus. For each of them, 20 manually segmented instances were available and served as ground truth.

4.3.1 A few segmented target structures

Figures 4.7, 4.8 and 4.9 display a few 3-D rendering of the target structures segmented with our system, along with the associated 2-D synchronized views, for a previously unseen T1-weighted MRI, with imaging characteristics similar to those from the learning set.

- Figure 4.7 illustrates the successful delineation of all four structures with the complete segmentation system (all applicable constraints and rules). Note that, as explained earlier in Chapter 2, the overall shapes of the targets might look different from standard anatomical expectations: the corpus callosum is too narrow, only the superior horns of the ventricles are kept (even though the segmentation system, in the absence of shape model for the ventricle, also segmented part of the third and fourth ventricles), the tail of the caudate nucleus is too short. However, those outlines conform with the established delineation protocols for our study (see Appendix B) and are thus satisfactory.

- Figure 4.8 demonstrates the interest of the distance constraint: the lateral ventricles, caudate nuclei and corpora callosa of the same MRI were segmented by our system without non-penetration distance constraints (for the corpus callosum, ventricle and caudate nucleus). As could be expected, these structures intersect.

- In Figure 4.9, no shape constraints were used to segment the structures from the same MRI. Not surprisingly, comparison with Figure 4.7 confirms the interest of shape models to control the deformation of templates when little or spurious intensity information is available: the segmentation of the hippocampus greatly suffered from this lack of a priori shape knowledge (we could even observe changes of topology). Conversely, in the absence of shape constraint, we obtained on this MRI a segmentation of the caudate nuclei in better agreement with anatomical expectations, in that they both present longer tails (which are still within the guidelines of the delineation protocol).

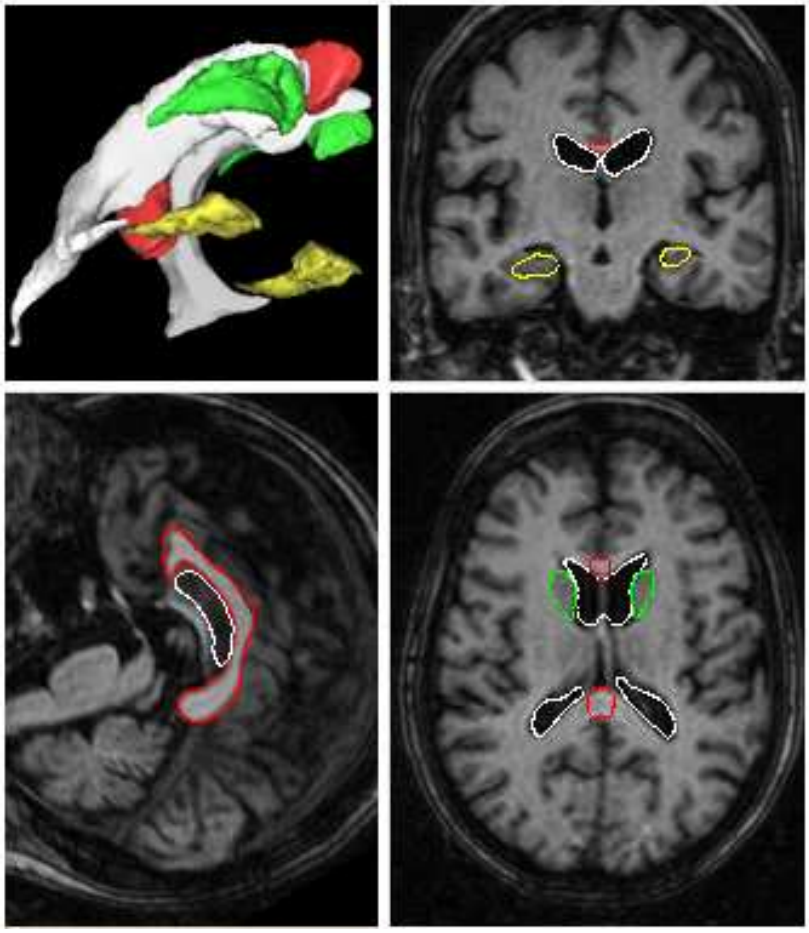


Figure 4.7: Segmented target structures (in color) in a typical T1-weighted MRI with the complete segmentation system (all rules, all constraints).

4.3.2 Segmentation accuracy

The accuracy of our segmentation system was evaluated following the methodology presented in [74]. We used as error metrics the partial Hausdorff distance and the mean absolute surface distance[5]. As argued above (Section 2.4), all segmentations were adjusted to take into account the delineation protocols.

Hausdorff distance: Given a deformed simplex mesh Π_j^t, its Hausdorff distance to a gold standard segmentation GS_j (represented by a set of 3-D voxels) is the largest distance between them both, computed in an asymmetric way, as the maximum over all voxels v of a discretized version of Π_j^t of the minimum Euclidean distance between v and its closest

[5]We favoured this error methodology over the computation of the false positive and false negative ratios that we used in Chapter 2 as it better illustrates the global behavior of the segmentation system (as argued in [74]). In particular, it is less sensitive to small delineations errors: for instance, if a deformed template lies only 1 pixel away from the true boundary everywhere along the target structure, then both ratios are maximal in spite of the segmentation being fairly good.

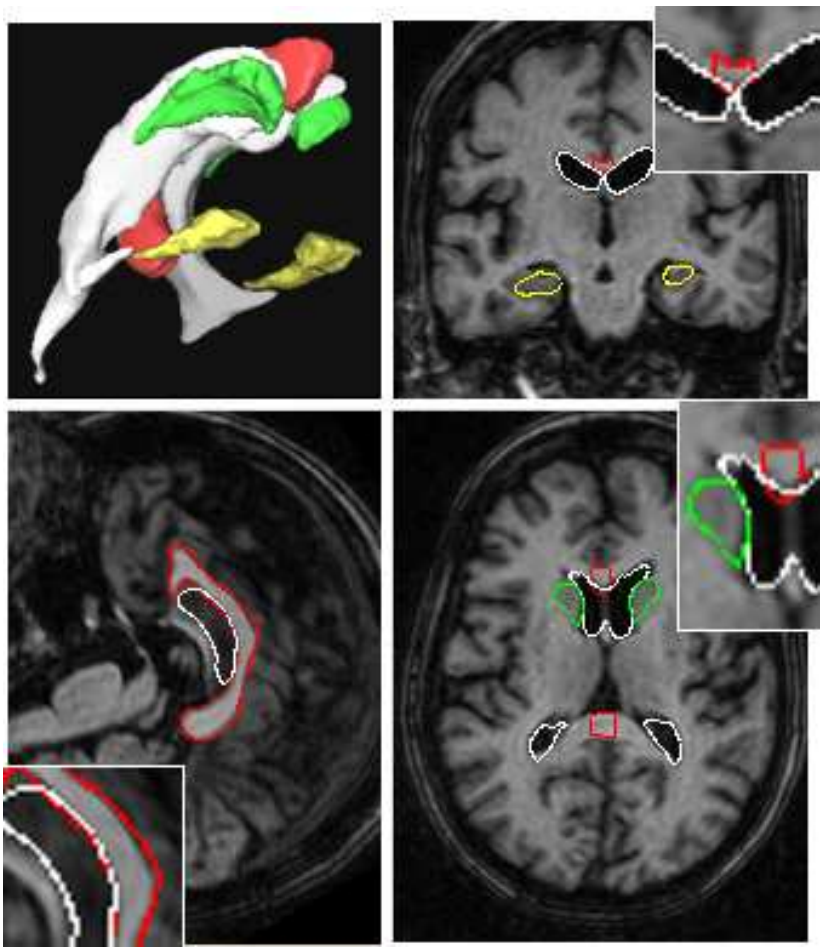


Figure 4.8: Segmented target structures (in color) in a typical T1-weighted MRI (same as previous) with our segmentation system without distance constraints.

voxel w on G_j:

$$H_{asym}\left(\Pi_j^t,\, GS_j\right) = \max_{v \in \Pi_j^t} \left(\min_{w \in GS_j} d_{euclidean}\left(v, w\right) \right) \tag{4.11}$$

It can be symmetrized by taking the maximum of both asymmetric measures:

$$H_{sym}\left(\Pi_j^t,\, GS_j\right) = \max\left(H_{asym}\left(\Pi_j^t,\, GS_j\right),\, H_{asym}\left(GS_j,\, \Pi_j^t\right)\right) \tag{4.12}$$

For efficiency reasons, we evaluated the Hausdorff measure by integrating the values of the Euclidean distance map of one surface along the contour of the other one, as described in [74]. Finally, in view of its high sensitivity to outliers, we rather considered the 95% quantile Hausdorff distance.

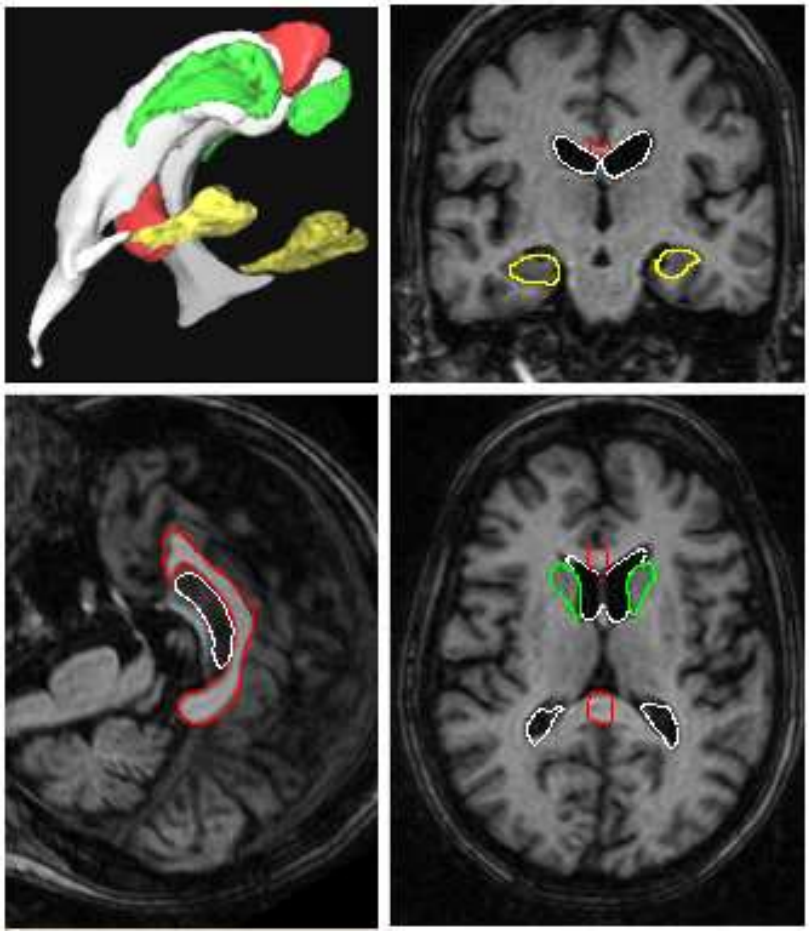


Figure 4.9: Segmented target structures (in color) in a typical T1-weighted MRI (same as previous) with our segmentation system without shape constraints.

Mean absolute surface distance: A similar distance map strategy enables the computation of a mean distance between the deformed template and its target gold standard. Namely, the mean distance from a deformed simplex mesh Π_j^t to its gold standard segmentation GS_j is computed by first computing the distance map of GS_j and then integrating the distance to GS_j over the voxels of a discretized version of Π_j^t . A symmetric version is obtained by taking the maximum of both asymmetric measures, as described above.

Table 4.1 reports both measures for all 4 structures, averaged over the 20 test instances (different from training instances). All MRI data set were T1-weighted 1mm^3 resolution MRI. We present the accuracy of several versions of our segmentation system to demonstrate the influence of its various components.

On average, segmentation were performed in approximately 5mn on a standard Pentium III, 1GHz PC, for all four selected target structures. Overall, the segmentation performances compared favorably with those reported in the literature [74, 160].

system	*distance (mm)*	*corpus callosum*	*ventricles*	*caudate nucleus*	*hippocampus*
basic	mean	1.3	4.4	4.2	3.5
framework	95% sym.	2.2	5.4	4.7	8.2
with shape	mean	1.4	4.8	3.9	3.2
constraint	95% sym.	2.4	5.5	5.5	7.8
with distance	mean	1.2	4.3	3.8	N/A
constraint	95% sym.	2.2	5.1	4.2	N/A
with texture	mean	0.2	2.2	2.1	2.5
constraint	95% sym.	0.4	3.3	3.0	3.5
with shape	mean	1.3	3.5	2.2	N/A
and distance	95% sym.	2.2	4.7	4.9	N/A
with all	mean	0.2	1.9	1.8	2.3
constraints	95% sym.	0.4	2.8	2.3	3.5
with feedback	mean	0.2	1.8	1.6	2.1
rule	95% sym.	0.4	2.6	2.0	3.0

Table 4.1: Performance of our segmentation system over the target structures for a variety of phantom images. No distance constraints were used for the hippocampus.

- Segmentation of caudate and callosum were good and further improved with the use of shape, distance and texture constraints. A few odd-looking caudates (far from the mean shape) worsened the performances when a shape constraint was added (the 95% symmetric Hausdorff distance is worse than that computed with the basic framework while the mean error is better). This illustrates the difficulty of designing a learning set representative enough for the shape model to adequately cover all the encountered shapes. Clearly, a compromise must be found between too exhaustive a learning set which would induce only poor shape constraints overall, and too generic a learning set which might hamper the segmentation system on a few cases while improving performances in most cases. Incidentally, principal component analysis may not be the optimal tool to build a shape model representative of the true anatomical variability. Here also, *a priori* information could be used to build a better shape model.

- The less accurate segmentation of the ventricles is explained by the inability of our deformable templates to reach as far as the apex of the inferior horns as they would have to go through partial volume effect voxels. However, even though these voxels were included in the delineation protocol, manual delineations exhibit a large variability in this area which nuances the relatively low performances of our automated approach. Furthermore, the model-to-manual maximum Hausdorff distances $H_{asym}\left(\Pi_j^t,\, GS_j\right)$ were good (2.2 with all constraints), since our approach correctly segmented the parts of the ventricles that were "reachable".

  As an attempt to improve the ventricle segmentation performances, we defined on the ventricle simplex mesh a zone which covered the apex area, and locally decreased for the vertices of the zone the amount of regularisation ($\alpha = 0.01$) and increased the influence of the image force ($\beta = 0.5$, $\gamma_{texture} = 0.49$). We obtained a better mean error: 1.5 mm with a 2.2 mm symmetric Hausdorff measure.

- The feedback rule was overall particularly effective, most especially in reducing the maximal errors. However, poor contrast and noise hampered the hippocampus deformable templates. The interest of the texture constraint was particularly evident for this structure.

It should be noted that those segmentation results are to be considered in the light of the intra/inter-operator variabilities associated to the delineation of the structure's boundaries. While variability values reported in the literature are fairly small for the corpus callosum (2.5% in a study of the choice of the mid-sagital section around which the callosa slices are delineated [167], 1mm RMS error for inter-operator variability in [151]), the average inter-operator error can be as high as 13% volume difference for the hippocampus in [152], a difficult structure to outline in an MRI. Those variabilities affect both the *a priori* delineated samples from the learning sets use to build the shape and texture constraints, and the gold-standards that we use to evaluate the performances of the automated segmentation algorithm. In this regard, our segmentation system seems to be doing as well as a manual operator could do.

4.3.2.1 Noise robustness.

The noise sensitivity of the overall segmentation system mostly depends on that of its various components. The sensitivity of the texture filter was studied in Chapter 2 (Section 2.4.3.4). We use here the same set of T1-weighted MRIs with various acquisition protocols to evaluate the noise and image characteristic robustness of the segmentation system as a whole. Table 4.2 reports the distance measures for all four structures with the standard learning set, with an adapted learning set (that is, a learning set of MRIs with characteristics similar to those to be segmented) and with a mixed learning set (with MRIs from various scanners and protocols).

Performances were good for all structures with a significant decrease for the hippocampi however, mostly due to the decreased performances of the texture filter for that structure. Here also, the segmentation system performed better when an adapted learning set was used.

Comparable results were obtained with a mixed learning set. We however observe a slight decrease in performances with respect to those obtained with the adapted learning set. Mixed learning sets also seem to induce a greater variability in the quality of the segmentation, which is probably explained by the lack of representativity of the learning set. Clearly, the latter is much harder to make representative when a large variety of MR characteristics must be represented. Additional experiments with a greater number of samples in the learning set confirmed that intuition (again, this is mostly due to the performances of the texture filter).

4.4 Conclusion

We have presented in this chapter a general framework for the automated segmentation of anatomical structures in brain MRI. A hybrid combination of external and internal energies, modeling a variety of aspects of prior medical knowledge, drives a series of deformable template towards the boundaries of these target structures. A number of explicit rules, also derived from medical expertise, further increases the overall accuracy and robustness of the method.

The validity of this approach was demonstrated on the four selected target structures. The developed framework could of course readily be extended to tackle the segmentation of additional structures. A more in-depth study of the multivariate relations between the various parameters of the deformation scheme and how they affect the accuracy of the match should also be conducted.

A number of additional rules could be implemented to increase the overall performances: in particular, additional feedback loops could be devised to tackle the segmentation of difficult

ratio of misclassified voxels / false positive (%) / false negative (%)	standard	learning	set	
	c.c.	ventricles	caudate	hippocampus
mean	0.3	2.3	1.8	2.7
95% sym.	0.6	2.9	2.5	4.2

ratio of misclassified voxels / false positive (%) / false negative (%)	adapted	learning	set	
	c.c.	ventricles	caudate	hippocampus
mean	0.3	2.0	1.6	2.3
95% sym.	0.5	2.7	2.2	3.8

ratio of misclassified voxels / false positive (%) / false negative (%)	mixed	learning	set	
	c.c.	ventricles	caudate	hippocampus
mean	0.3	2.2	1.7	2.5
95% sym.	0.9	2.6	2.6	4.3

Table 4.2: Performance of our segmentation system over MRIs with different images characteristics and different learning sets.

images where robustness becomes a more pressing issue, when lesions are apparent for instance. We could also incorporate meta-meta rules to monitor the number of times the error-checking rules (leak prevention, for instance) have been triggered and either interact with the human operator (to alert him about a particularly difficult segmentation, or require assistance in an area of the image, etc.) or select an entirely different set of parameters and shape/texture constraints.

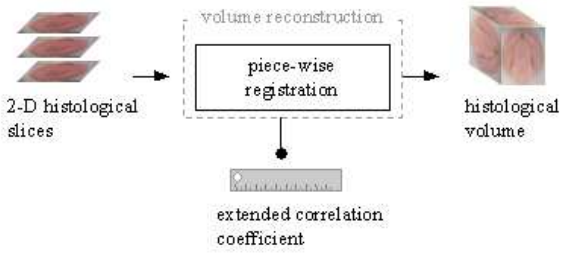
The unavailability of source code for alternative segmentation systems unfortunately prevented us from performing a detailed comparison of our system against others. As it is, the comparison of segmentation performances is a tricky issue. Clearly, in view of the complexity of the segmentation problem at hand, there are no general prescriptions for selecting a "good" segmentation algorithm. This choice must not only be driven by the characteristics of the input image (type of noise and noise-signal ratio, texture characteristics, contrast of the target object with respect to surrounding pictorial elements, bias fields, etc.) but also by the possible usage constraints (algorithmic complexity with respect to available memory/CPU resources, time limits if real-time applications are envisioned, etc.) and the downstream treatments that follow this segmentation step (diagnosis, morphometric analysis, shape recognition, etc.). Consequently, assessing the true performances of a given segmentation tool in and out of itself is a difficult, if not irrelevant, task. It seems more sound to compare segmentation algorithms by measuring the overall quality of the complete chain of processes of which they usually are part: the best segmentation technique then becomes that which maximizes the overall system performance.

Chapter 5

Piecewise Affine Registration
Recalage Affine par Morceau

« Non mais, t'as déjà vu ça ?... En pleine paix, y chante
et pis crac ! un bourre-pif ! Mais il est complètement fou ce mec ...
Mais moi, les dingues, j'les soigne, je m'en vais lui faire
une ordonnance, et une sévère ! Je vais lui montrer qui c'est Raoul.
Aux 4 coins de Paris, qu'on va le retrouver éparpillé par petits bouts,
façon puzzle ... Moi, quand on m'en fait trop, je correctionne plus,
je disperse ... je ventile ... »
Michel Audiard, *Les Tontons Flingueurs*, 1963.

Nous introduisons dans ce chapitre un modèle plus fidèle des transformations induites par les processus de traitement histologique, l'approche affine par morceaux, qui permet la conception d'un algorithme de recalage biomédical mieux adapté à la reconstruction de volumes histologiques 3-D. Plus précisément, notre modèle consiste en un ensemble de transformations rigides ou affines appliquées à des régions soigneusement délimitées dans les images à recaler et encapsulées dans une transformation non-linéaire. Nous développons une approche en plusieurs étapes (calcul d'une carte de similarité, clustering hiérarchique, recalages indépendants robustes, érosion et composition hybride affine/non-linéaire) où chaque étape bénéficie de l'expertise médicale a priori, *sous forme soit explicite (suppression du fond, etc.) ou implicite (distance EMD de distribution, etc.). Nous détaillons également une nouvelle mesure de similarité, le coefficient de corrélation contraint, dont le compromis entre précision et robustesse en fait un bon candidat au recalage biomédical.*



Piecewise registration approach.

We saw in Chapter 4 how the registration of a hybrid MRI/structure atlas to the input MRI helped initialize the deformable templates closer to the location of the target structures. Clearly, additional information about the microscopic characteristics of the underlying tissues, unavailable to the MRI modality (restricted as it is to the macroscopic level), could further increase the performances of the segmentation process. *Post-mortem* histology then becomes a natural choice to build a more precise and information-rich combined MRI/histology atlas.

To obtain such a 3-D atlas, we first need to reconstruct a 3-D volume from the set of 2-D slices produced by the histological treatments. Care should be taken to perform this reconstruction in an *anatomically meaningful* way, lest the added-value of the atlas should be diminished. The use of as much *a priori* anatomical information as available is center to the philosophy and design of the registration methodology we detail in this chapter.

5.1 Introduction

As argued above, segmentation performances are increased when registration of an *a priori* built hybrid MRI/structure is used to initialize the deformable templates. While the interest of using such a mono-modal MRI-to-MRI registration is evident, the resolution of usual MR scanners restricts the atlas to a macroscopic level of description. Yet, information about the microscopic characteristics of the tissues underlying the brain structures could help build a better atlas, in terms of precision of the delineated structures, of their relative placements or of their mutual interactions. In turns, a better atlas is bound to increase the accuracy and robustness of the segmentation process: not only could it guarantee a better initialization, but it could also foster a better understanding of the inter-relationships between structures, thereby providing new and improved rules to guide the segmentation process. Post-mortem histology (see Section 2.2.2 for a description) then becomes the microscopic modality of choice: the variety of available stains can help discriminate between structures that might be blurred and/or blended in MRI (the amygdala for instance is difficult to segment out of the hippocampus in a 1 mm^3 resolution MRI), much greater resolutions can be achieved (of an order of magnitude 100 to 1000 times bigger), etc.

Histology however produces 2-D sections. Section to section registration is then needed to reconstruct a 3-D biological image from these series of 2-D images. Subsequent fusion with 3D data acquired from tomographic imaging modalities (*e.g.* MRI) then allows the tissue properties

to be studied in an adequate anatomic framework (using *in vivo* reference data) and the building of the desired hybrid atlas[1].

After a review of image registration in the context of 3-D reconstruction (Section 5.1.1.1) which highlights the need (and lack) for both problem-specific and anatomically-driven registration techniques (Section 5.1.1.2), we introduce the piecewise affine transformation model we have developed (Section 5.2) as an attempt to satisfy these requirements. We then discuss in Section 5.3 some registration results on a variety of biomedical images, and the sensitivity of our technique to noise conditions and parameters.

5.1.1 Image registration

A key component of medical image analysis, image registration essentially consists of bringing two images, acquired from the same or different modalities, into spatial alignment. This process is motivated by the hope that more information can be extracted from an adequate merging of these images than from analyzing them independently. For instance, mono-modal registration of a population's MRIs can be used to build anatomical atlases [42, 195], while mono- or multi-modal registration of the same patient's data can help determine the nature of an anomaly [108] or monitor the evolution of a tumor [90] or other disease processes [169].

More formally, given two input images, registering the floating (*i.e.*, movable) image to the reference (*i.e.*, fixed) one entails finding the transformation that minimizes the dissimilarity between the transformed floating image and the reference. As such, it can be decomposed into three elements:

- a transformation space, which describes the set of admissible transformations from which one is chosen to apply to the floating image;
- a similarity criterion, which measures the discrepancy between the images; and
- an optimization algorithm, which traverses the transformation space, in search of the transformation that will minimize the similarity criterion.

A large variety of transformation spaces have been discussed in the literature: among others, one finds linear transformations (rigid, affine) and non-linear transformations (polynomial [220], elastic [52, 73] or fluid [34]). Similarly, many similarity criteria have been presented: Studholme *et al.* [190] use normalized mutual information, Collins *et al.* [41] cross-correlation, Roche *et al.* [171] the correlation ratio, Ashburner *et al.* [6] the squared intensity difference, etc. Optimization algorithms range from the straightforward Powell method [40] to sophisticated multi-scale Levenberg-Marquardt techniques [194] or stochastic search [216].

In view of the plethora of approaches, mathematical frameworks and envisioned applications, a general review of image registration approaches, even restricted to medical imaging, is outside the scope of this chapter. We focus here on the reconstruction methods of a 3-D volume from a series of 2-D biomedical images. Please refer to Maintz [133], van den Elsen *et al.* [204] and Cachier [30, Chapitre 1] for thorough reviews of medical image registration.

[1] We will not tackle here this 3-D fusion, please refer to the work of Malandain and Bardinet [134] for some advanced fusion methodologies.

5.1.1.1 Registration algorithms for 3-D reconstruction

By registering each pair of consecutive slices in the stack of histological slices, we can recover a geometrically coherent 3-D alignment of the 2-D images. This problem essentially consists of the mono-modal registration of similar objects (even though non-coherent distortions may occur between consecutive slices). A number of techniques have been proposed in the literature:

manual registration: Classically, people have relied on their anatomical expertise to register images manually [61, 178], a time-consuming, operator-dependent and non-reproducible process.

fiducial markers: The use of fiducial markers helps increase the reliability of the registration process [69, 79, 78, 101]. Physical markers (straight rigid needles for instance) are inserted into the brain to be processed, prior to the slicing step. They are later tracked in the images. A least square minimization process then attempts to recover the original geometrical configuration.

While obviously a lot less operator-independent and more reproducible than manual registration, the biases introduced by the non-orthogonality between the needle's axes and the cutting planes and the geometrical deformations undergone by the needle holes during the histological preparations (the chemical baths for instance can collapse the holes) may significantly hamper the ability of the least square minimization to recover the correct transformation.

geometrical approaches: These require that feature elements be extracted *a priori*.: those elements are usually geometrically remarkable points such as lines, curvature extrema or contours. Hibbard *et al.* [95] for instance use principal axes to align digital autoradiograms. Better performances in terms of precision can be achieved by matching contours [36, 225], edges [114, 117] or points [166].

While it could be argued that these techniques enable a better control over the registration process, the segmentation of feature elements can sometimes be as difficult a problem as the overall registration process itself, with similar drawbacks: non-reproducibility if the feature are extracted manually, lack of precision for fully automated segmentation, etc.

iconic approaches: They rely on the comparison of the intensities associated to voxels in the input images. Usually, they can be casted into a similarity-minimization framework where the transformation space is traversed in search for the transformation which minimizes the intensity dissimilarity between the transformed floating image and the reference image [52, 34, 190, 216, and many others].

5.1.1.2 Application-tailored registration

A careful study of the above-mentioned registration methods and reviews highlights the unfortunate lack of specificity of most of the approaches introduced in the literature. All too often, these techniques are presented in a very general context and advertised as generic, even though they ought to be applied to specific registration problems. This is all the more surprising that specific approaches abound in closely related fields like medical image segmentation for instance, where the combined use of atlases and a variety of shape and appearance constraints (see previous chapter) helps tailor the segmentation methods to the segmentation problems at hand.

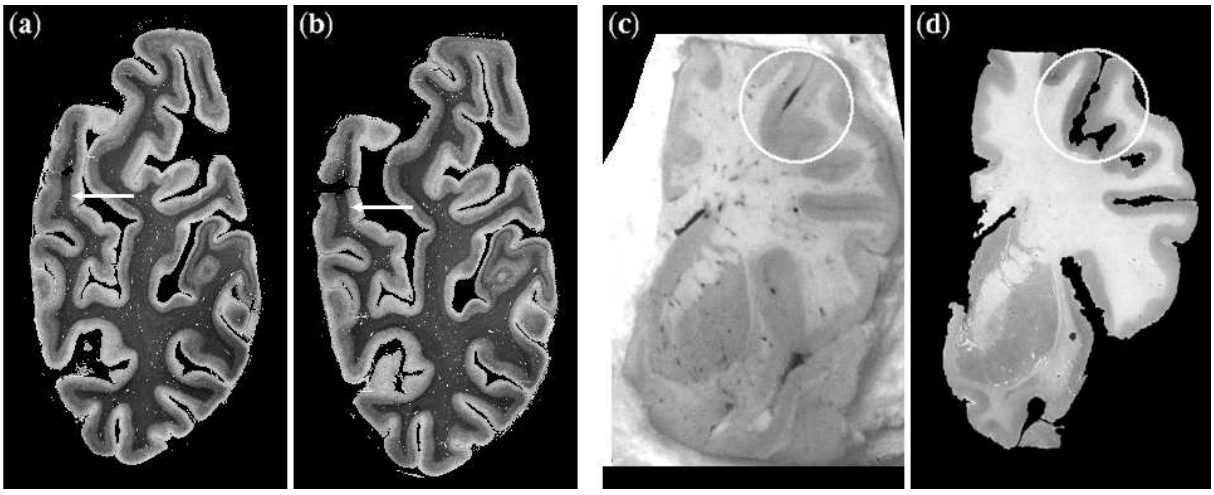


Figure 5.1: Two consecutive myelin-stained histological sections of the human brain (a, b); Human brain cryosection (c) and its associated Nissl-stained section (d). White arrows and circles indicate moving gyri.

Furthermore, again with few remarkable exceptions[2], very limited *a priori* medical expertise is used in the design of the registration algorithms.

Yet, as argued and demonstrated all through this manuscript, prior medical knowledge is pivotal to both the conception of methodologies and the analysis of their performances. This is especially true of medical image registration, given that it is an ill-posed problem in the sense of Hadamard. First, even though a given image transformation may adequately put the floating and reference image in correspondence, it may not correspond to the actual physical transformation that took place, if the latter only exists (for multi-patient registration for instance, it might be difficult to establish the existence of such physical transformation). And second, a number of transformations, and not only one, may give very similar result in terms of visual correspondence, most especially when the number of degrees of freedom of the allowed transformations is large. Evaluating the quality of a registration process then becomes particularly difficult when the images are considered independently from their medical context, that is, when the problem to be solved is not that of putting in correspondence two views of an underlying medical truth but that of registering two images, taken as mere sets of voxels with associated intensity values.

We submit that much benefit is to be gained from the use of medical expertise in the design of a registration methodology. We have consequently developed a new registration paradigm, the piecewise approach, adapted to the registration of 2-D biomedical images, which we motivate and detail in the following section.

[2]For iconic (*i.e.*, intensity-based) methods, Roche *et al.* [171] derived optimal similarity measures from an analysis of the expected relationships between the input images, with different hypotheses leading to different measures.

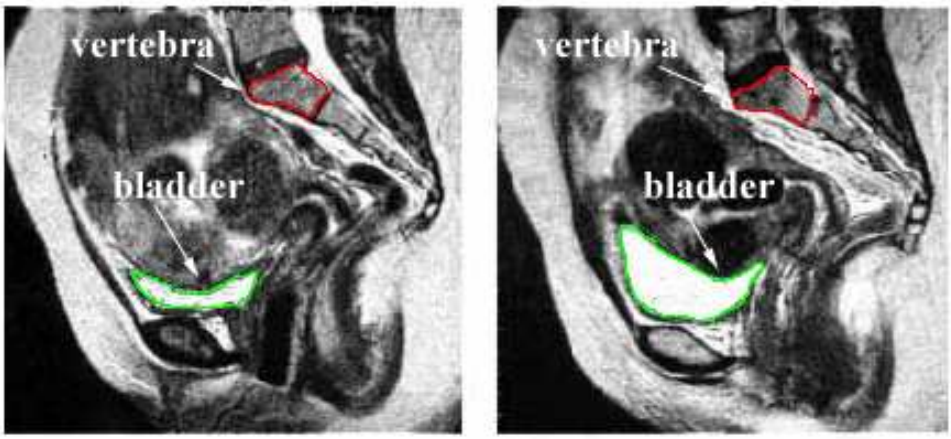


Figure 5.2: Two abdominal MRIs of the same patient, with corresponding vertebra (red) and bladder (green) outlined.

5.2 Piecewise Registration

As described above, a registration algorithm basically consists of three elements: a similarity measure, a transformation search space, and an optimization algorithm, which must all three conform to *a priori* medical expertise.

Adapted similarity measure. In a fashion similar to Roche's tailorisation of the similarity measure to the expected relationships between the input images, we have extended a classical similarity measure, the correlation coefficient, with the goal to allow a better modelling of the combined transformation and intensity relationships via the building of a more coherent similarity map of the input floating image (see Section 5.2.1.2).

Special care was also taken to distinguish the background from the actual pieces of tissues to be registered, as treatments are only applied to the latter (see Section 5.2.1.1).

Histology-tailored transformation space. *A priori* knowledge about the acquisition process for biological images should also allow the transformation space to be modeled more accurately. In our case, the cutting process, successive chemical treatments, and the glass mounting step that a slab of tissue undergoes during a histological preparation yield a fairly flexible global transformation that is however locally affine for some identifiable components of the section[3]. In brain sections for instance, each gyrus (compare white arrows in Figure 5.1(a) & (b), and white circles in (c) & (d)) undergoes an affine transformation (due to successive manipulations) relatively independent from those of other gyri: in (b) the operator introduced an artificial rotation to the marked sulcus when he positioned the tissue on the glass slide; holes and tears appeared in (d) during the slicing and chemical bath steps. Consequently, in spite of the large variety of available transformation spaces in the literature, their functional form may not reflect our specific needs.

Discussions with neuro-anatomists and histologists prompted us to consider the input biological images as a set of independent components, subject to linear transformations. We then

[3]Even though the chemical baths introduce non-linearities, an affine model remains a good approximation.

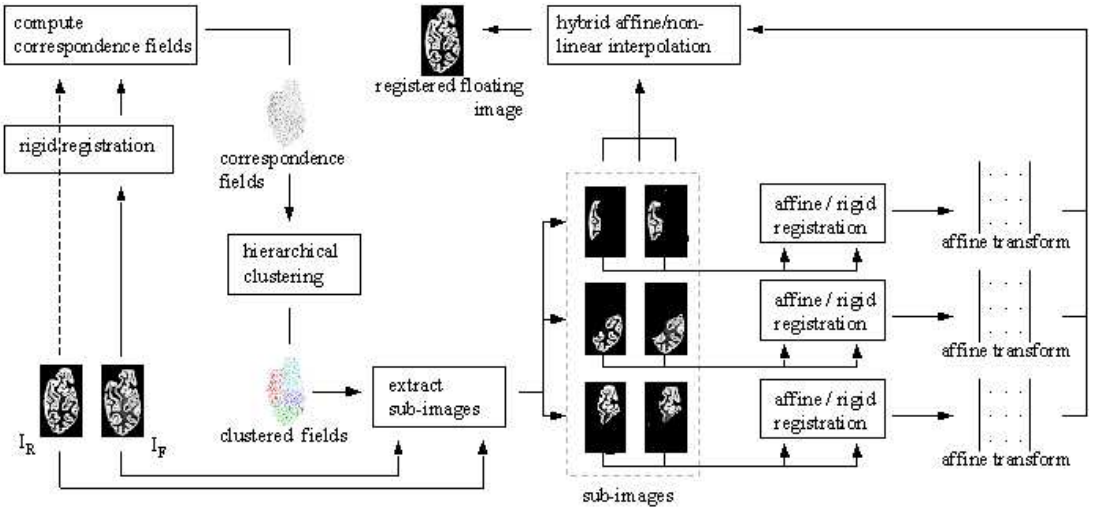


Figure 5.3: Detailed overview of the piecewise registration approach.

model the composite transformation yielded by the chain of physical histological processes as a number of affine or rigid transformations applied to carefully delimited areas, with non-linear transformations interpolated in between.

Note that the utility of this transformation model extends beyond biological images (our primary motivation here) to medical images as well. For instance, abdominal or torso MRIs (see Figure 5.2) often include rigid structures such as bones (ribs, vertebrae, etc.), deformable organs (liver, heart, etc.), and elastic tissues. Two abdominal MRIs of the same patient are then linked by a complex transformation which can be rigid in some regions (for bones) but potentially exhibits large local dilations (in deformable organs). Global rigid or affine transformations cannot adequately handle such a case. Also, a single rigid transformation could not correctly register all the vertebrae along the spinal column simultaneously. Furthermore, high degree of freedom (*e.g.*, fluid) transformations could correctly map one image onto the other, but they may not ensure that specific components (*e.g.*, bones) will be only rigidly transformed.

Optimization algorithm. We selected a block-matching approach (see Section 5.2.1.1). Here also, prior segmentation of the tissues helps discard background blocks from the chain of treatments. A number of tests on the blocks (intensity variance, number of active pixels, etc.) further adapt the search for an adequate transformation towards anatomically meaningful ones.

Figure 5.3 displays an overview of the piecewise approach that we detail in the following sections. Briefly, given the input floating and reference images, I_F and I_R respectively, we first rigidly register I_F to I_R, before computing a dense similarity map (or correspondence field) between them with a block matching algorithm (Section 5.2.1). A hierarchical clustering algorithm then partitions the correspondence field into a number of classes from which we extract independent pairs of sub-images (Section 5.2.2). Our clustering algorithm relies on a distribution metric (the Earth mover's distance) to agglomerate blocks, and uses the estimated transformations associated with each cluster to guide the grouping process. The pairs of sub-images are then, separately, rigidly or affinely registered. Finally, an hybrid affine/non-linear interpolation scheme is used to compose the registered floating image (Section 5.2.3).

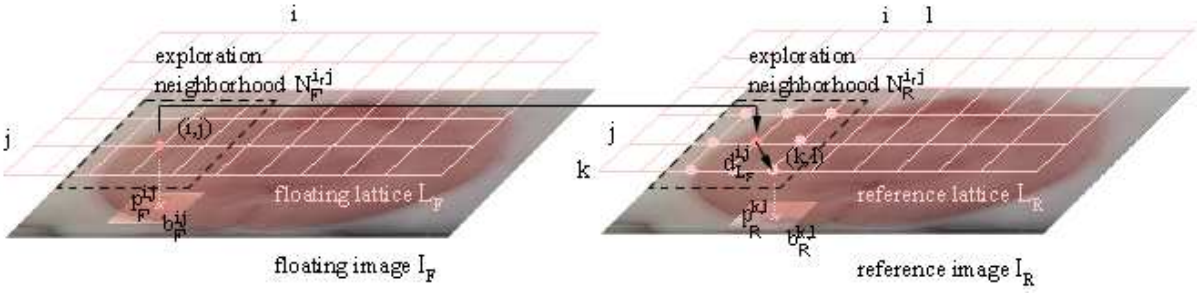


Figure 5.4: Block-matching algorithm.

5.2.1 Computing the initial correspondence field

The first step of our approach consists of *automatically* partitioning the input floating and reference images (I_F and I_R) into a number of pairs of corresponding sub-images, where each sub-image is associated with an independent (in terms of transformation) image component.

We approach this segmentation issue as a process of partitioning a correspondence field computed from I_F to I_R. Our method is motivated by the following observation. When both images are composed of pairs of independent components, where each component is subject to some linear transformation, the associated correspondence field should exhibit rather homogeneous characteristics within each component, and heterogeneous ones across them. Consequently, by clustering the fields with a criterion based on local characteristics, we hope to extract from them the desired independent components.

We use a block-matching algorithm [103] to compute the correspondence field.

5.2.1.1 Block-matching algorithm

We associate with I_F and I_R two rectangular lattices $L_F = \{(i,j) \in [1,\ldots,w_F] \times [1,\ldots,h_F]\}$ and $L_R = \{(i',j') \in [1,\ldots,w_R] \times [1,\ldots,h_R]\}$ respectively, whose sites correspond to pixels in the input images.

We may choose to associate a site to each pixel of the input images, in which case w_F, h_F and w_R, h_R are the width and height of I_F and I_R, respectively. We could also consider a sparser regular or non-regular site distribution. In our case, we use sparse regular lattices and discard, for histological sections, sites which lie on the background. A prior segmentation step is then required to identify these background blocks: depending on the modality and quality (in terms of signal to noise ratio, and contrast) of the input images, we either use simple intensity thresholding (the case for the myelin-stained human brain section of Figure 5.1), or a more sophisticated segmentation algorithm (the neural classifier introduced in Chapter 2 was used to discard the background of the cryo- and Nissl-stained mouse brain sections of Figure 2.14 prior to the reconstruction of the 3-D cryo- and Nissl- volumes).

The block-matching algorithm works as follows (see Figure 5.4): for each site (i,j) in L_F, we consider a neighborhood $b_F^{i,j}$ in I_F of the pixels associated with (i,j) (usually a square neighborhood of constant size b_{size} called a "block", whose centroid is denoted by $p_F^{i,j}$). We then compute the similarity measures (given a similarity metric sim) between block $b_{I_F}^{i,j}$ and every block $b_R^{k,l}$ in I_R associated with sites (k,l) in the corresponding neighborhood $N_R^{i,j}$ of (i,j) in

L_R (the "exploration neighborhood"). For every site (i,j) in L_F, we then get a 2-D spatial similarity distribution (the values$sim\left(b_F^{i,j}, b_R^{k,l}\right)$ defined in the neighborhood $N_R^{i,j}$ of (i,j)). We also record the "arg max" displacement $d^{i,j}$ defined by $d^{i,j} = p_R^{(k,l)_{max}} - p_F^{i,j}$ where $(k,l)_{max}$ is the site of L_R that is associated with the block $b_R^{k,l}$ in $N_R^{i,j}$ which is the most similar to $b_F^{i,j}$, *i.e.* $(k,l)_{max} = \arg\max_{k,l} sim(b_F^{i,j}, b_R^{k,l})$. This displacement field will serve to estimate transformations on clusters (see Section 5.2.2.1).

The quality of both the similarity map and the displacement field is essentially determined by three parameters: the size of the blocks, the similarity metric and the size of the exploration neighborhood in L_R.

- The similarity metric and the size of the blocks must reflect the expected relationship between the intensity distributions of blocks in the floating and reference images, and the scale of the features of interest within those blocks respectively (see [42] and [59] for details).
- The size of the exploration neighborhood is linked to the expected magnitude of the residual displacements after global alignment. It conditions the extent to which our registration algorithm can recover large deformations: the further apart corresponding components are, the larger the size of the neighborhood must be.

Section 5.3.3 provides an exhaustive list of the parameters of all the algorithms used in this study, along with comments on their impact on the overall registration quality and some standard values.

As a pre-processing step, we first rigidly register I_F to I_R to remove from the subsequently computed correspondence fields the global rigid transform that uniformly affects all components. We use the fully automated intensity-based registration algorithm presented in [156], where a *robust* multi-scale block-matching strategy was introduced. Not accounting for this would only degrade the quality of the field and affect the efficiency of the clustering.

5.2.1.2 Constrained correlation coefficient

A ubiquitous choice for image registration [171], the correlation coefficient represents, in the context of block-matching, a measure of the affine dependency between the block of interest $b_F^{i,j}$ in the floating images and every block $b_R^{k,l}$ in the corresponding exploration neighborhood in the reference image. It is written:

$$\mathrm{cor}\left(b_F^{i,j}, b_R^{k,l}\right)^2 = \frac{\mathrm{cov}\left(b_F^{i,j}, b_R^{k,l}\right)^2}{\mathrm{var}\left(b_F^{i,j}\right).\mathrm{var}\left(b_R^{k,l}\right)} = \frac{\sum_{u,v}\left(b_F^{i,j}(u,v) - \mu_F^{i,j}\right).\left(b_R^{k,l}(u,v) - \mu_R^{k,l}\right)}{\sum_{u,v}\left(b_F^{i,j}(u,v) - \mu_F^{i,j}\right)^2 . \sum_{u,v}\left(b_R^{k,l}(u,v) - \mu_R^{k,l}\right)^2} \tag{5.1}$$

where $\mu_F^{i,j}$ and $\mu_R^{k,l}$ are the mean intensities of $b_F^{i,j}$ and $b_R^{k,l}$ respectively. To make the affine dependency clearer, equation 5.1 can be rewritten:

$$1 - \mathrm{cor}\left(b_F^{i,j}, b_R^{k,l}\right)^2 = \frac{\min_{A,B}\mathbb{E}\left[b_F^{i,j} - A.b_R^{k,l} - B\right]^2}{\mathrm{var}\left(b_F^{i,j}\right)} \tag{5.2}$$

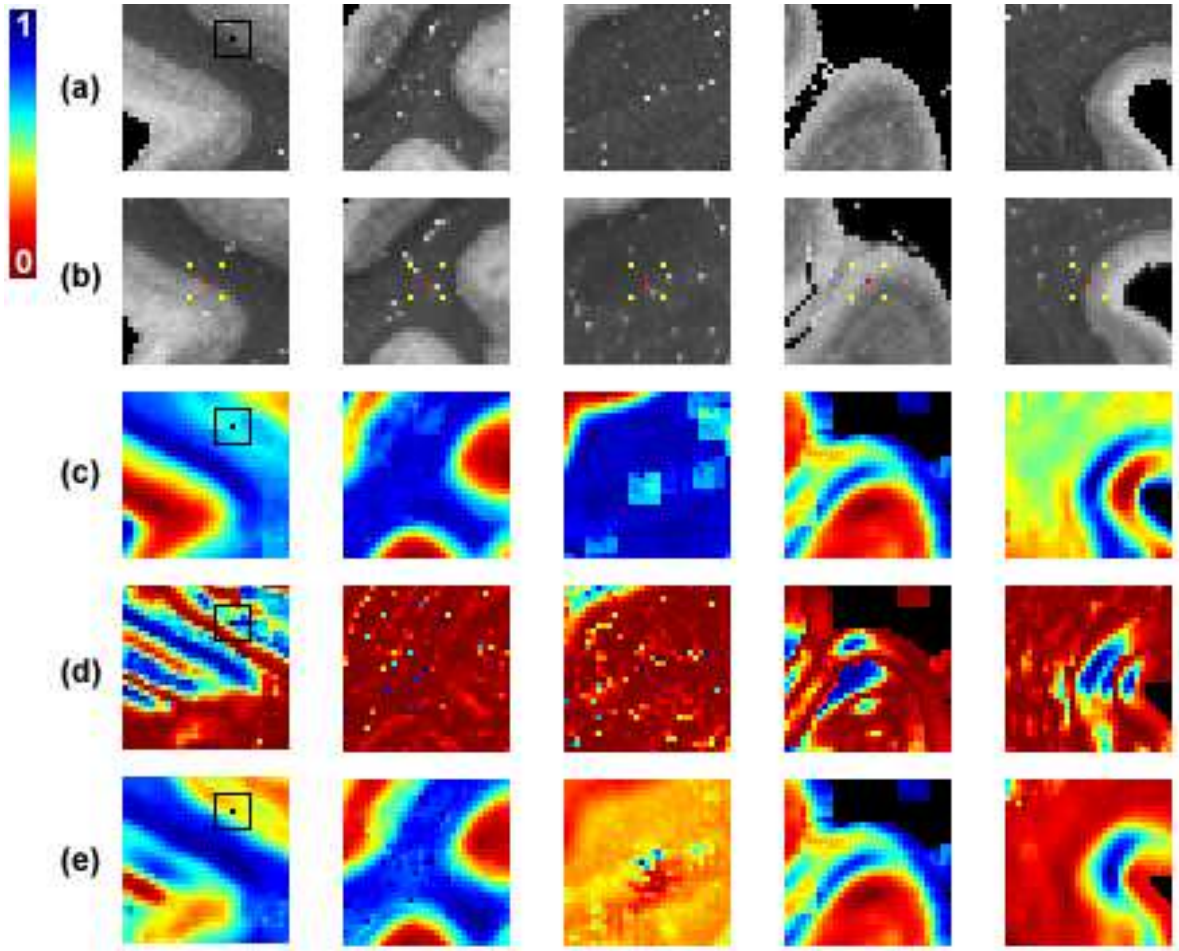


Figure 5.5: Similarity maps induced by three different similarity measures: (a) input reference sub-images; (b) input floating sub-images with considered floating blocks in red and yellow; (c) maps induced by the SSD measure; (d) maps induced by the classical correlation coefficient; (e) maps induced by the constrained correlation coefficient.

where $\mathbb{E}$ is the statistical expectation and A and B represent the affine coefficients of the intensity dependency function between $b_F^{i,j}$ and $b_R^{k,l}$.

This affine formulation stems from the assumption that each block contains at most two different tissues, a reasonable hypothesis when dealing with small image windows. A variety of studies have documented the effectiveness of the correlation coefficient in not only mono- but also multi-modal registration applications [171, 156].

Yet, when building the similarity map of $b_F^{i,j}$ (and thus, also when computing the "arg max" displacement field), different *implicit* A and B are used with every block in the reference exploration neighborhood. Comparing similarity values, both within the similarity distribution associated with a single floating block and across the distributions associated to different floating blocks, when obtained under those conditions then becomes questionable. A simple way to alleviate this issue consists of *explicitly* estimating A and B over the entire neighborhoods $N_R^{i,j}$

(the "exploration" neighborhood of (i,j) in L_R) and $N_F^{i,j}$ (the corresponding, same size, neighborhood in L_F), to keep them constant during the computation of the similarity distribution of a given floating block. Equation 5.1 then becomes:

$$\mathrm{ccor}\left(b_F^{i,j}, b_R^{k,l}\right)^2 = \frac{\sum_{u,v}\left(b_F^{i,j}(u,v) - M_F^{i,j}\right)\cdot\left(b_R^{k,l}(u,v) - M_R^{i,j}\right)}{\sum_{u,v}\left(b_F^{i,j}(u,v) - M_F^{i,j}\right)^2 \cdot \sum_{u,v}\left(b_R^{k,l}(u,v) - M_R^{i,j}\right)^2} \tag{5.3}$$

where $M_F^{i,j}$ and $M_R^{i,j}$ are the mean intensities of $N_F^{i,j}$ and $N_R^{i,j}$ in the floating and reference image respectively.

- By replacing $\mu_R^{k,l}$ by $M_R^{i,j}$ in equation 5.3, we validate the comparison of the similarity measures within the exploration neighborhood in I_R of the floating block $b_F^{i,j}$, and thus the choice of the "arg max" displacement vector $d^{i,j}$.
- By replacing $\mu_F^{k,l}$ by $M_F^{i,j}$, we homogenize by propagation (as the exploration neighborhoods $N_F^{i,j}$of adjacent floating blocks overlap in I_F) the computation of the similarity field (and of the displacement field) over the entire image. This also justifies the computation of distances between our similarity distributions (see Section 5.2.2.1).

Note that by estimating A and B on a larger neighborhood, we also make a stronger hypothesis. Whereas, with the classic correlation coefficient, we assume an affine relation between small blocks (equation 5.1), we here extend that assumption to larger areas: the exploration neighborhoods (equation 5.3). As it is, this extended hypothesis is still reasonable in our context (better experimental results were obtained with the constrained coefficient). It should however be carefully re-considered for multi-modal registration applications (for instance, we had to revert to the classic coefficient for the registration of autoradiographs to MRI data in the monkey brain, see Figure 5.11).

Figure 5.5 compares the behavior of the constrained correlation coefficient, its classical formulation and the sum of square difference (SSD) measure on a few sub-images extracted from the myelin-stained histological sections of Figure 5.1. We selected a number of sub-images from the histological sections, and computed the similarity map of each measure, where the exploration neighborhoods coincided with the entire sub-images. We used 7 by 7 pixel blocks in a 15 by 15 pixel neighborhood. For each pair (floating sub-image, reference sub-image), we show the associated similarity map in color. The comparison of the three maps demonstrates the good compromise achieved by the constrained coefficient between the regularity of the SSD and the good precision of the classical correlation coefficient. The benefits of the constrained formulation over the classical one are particularly obvious in the third and fourth column, where the correlation coefficient cannot manage a background composed of low variance blocks (high noise). Finally, comparison between maps (d) and (e) in the left-most column shows that while a high similarity value (dark blue) is obtained with the classical correlation coefficient for a pair of blocks with inverted tissues (white matter in the bottom-left corner/ gray matter in the top-right corner of the floating block and the other way around for the reference block, in black), a lower one (yellow) is obtained with the constrained coefficient.

Figure 5.6 displays the similarity distributions for both the constrained and the classical correlation coefficient and the "arg max" displacement field (for the constrained coefficient) for

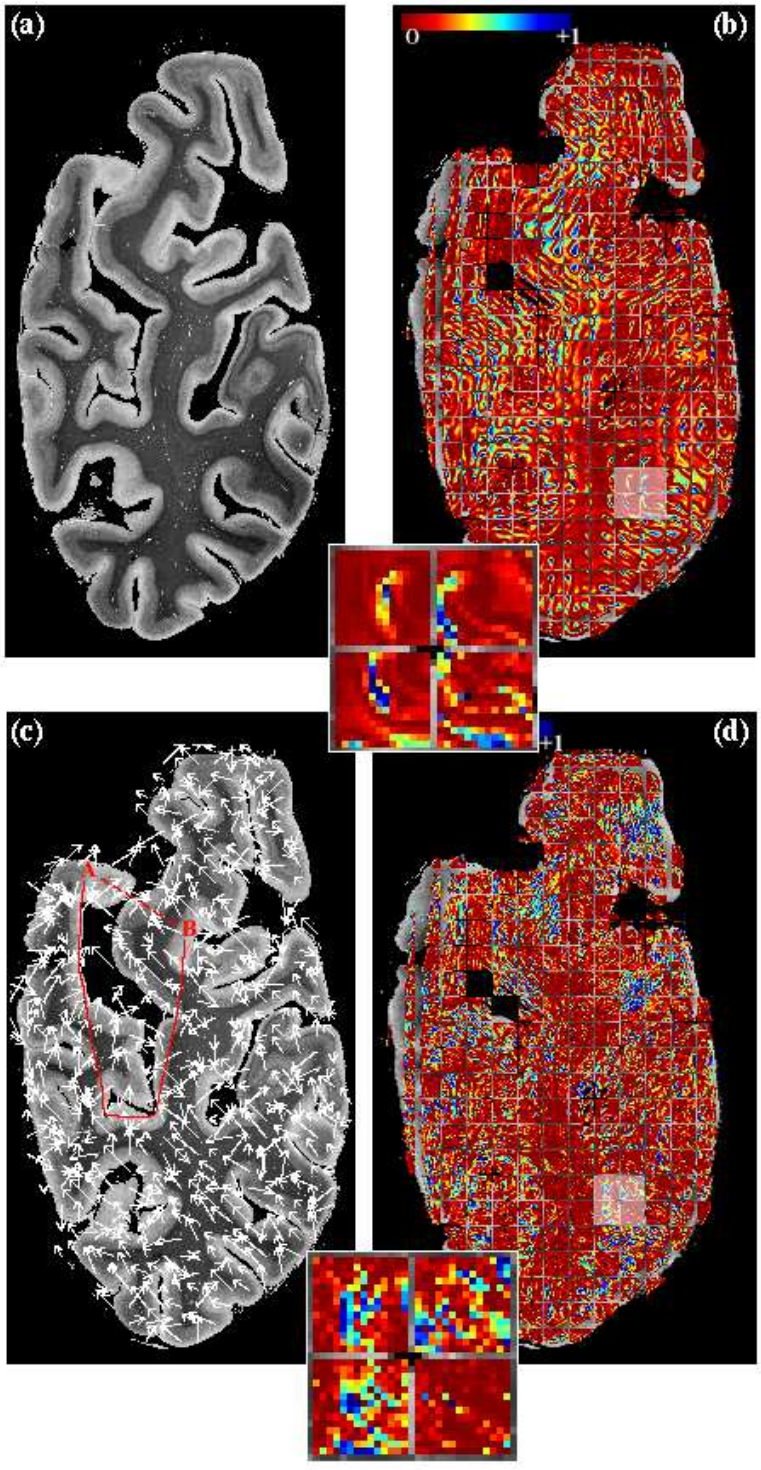


Figure 5.6: Dense correspondence field for two consecutive myelin stained histological sections: (a) input reference image; (b) input floating image with superimposed similarity distribution with the *extended* correlation coefficient; (c) floating image with superimposed arg max displacement field; (d) input floating image with superimposed similarity distribution with the *classical* correlation coefficient. The colorbars show the range of values of the similarity function, for each block. The red lines connecting A and B represent the geodesic (full) and Euclidean (dotted) paths.

the two consecutive histological sections of the brain of Figure 5.1 (60 μm myelin stained coronal sections through the occipital cortex). For every site of the floating lattice, a color square shows the similarity measures between the corresponding floating block and the reference blocks in its neighborhood (middle column). The optimal "arg max" displacement field is rendered with arrows whose length and direction are those of the optimal displacement vector associated to the lattice site at which the arrow originates. For visualization purposes, only half of the "similarity squares" are rendered in the similarity maps. Obviously, the similarity squares associated to the constrained coefficient present clear patterns (much clearer than those nonetheless exhibited by the classical coefficient), and, more importantly, conspicuous differences in patterns across the image, that will help the clustering algorithm segment the input images. Additionally, the "arg max" displacement field, although globally rather chaotic, tends to present more homogeneous patterns at a local scale, from which transformations can be estimated in a robust fashion. This will help both cluster the input images and register the extracted sub-images.

5.2.2 Extracting the image components

Equipped with a dense correspondence field, we can partition our input images. We describe in the following section the way the correspondence field is clustered, and how we extract sub-images from the clustered sites (Section 5.2.2.2). Those sub-images will be later used to estimate local transformations.

5.2.2.1 Clustering the correspondence field

We are looking for a hierarchical clustering of L_F, that is, a sequence of partitions in which each partition is nested into the next partition in the sequence [10]. Cluster analysis (unsupervised learning) essentially consists of sorting a series of multi-dimensional points into a number of groups (clusters) so as to maximize the intra-cluster degree of association and minimize the inter-cluster one. It is particularly well suited here as it behaves adequately even when very little is known about the category structure of the input set of points. That is, it does not require strong hypotheses to be formulated beforehand.

For simplicity's sake, we rewrite L_F as an ordered set of sites:
$L_F = \{s \, s.t. \, \exists!(i,j) \in L_F, s = (i,j)\}_{t=1}^{w_F.h_F}$. Our clustering method is adapted from the standard agglomerative hierarchical clustering algorithm described in [107]:

clustering algorithm

step 1: initialize a cluster list by placing each site of L_F in an individual cluster, and let the distance between any two of those clusters be the distance between the sites they contain (the more similar the clusters, the smaller the distance).

step 2: find the closest pair of clusters, remove them from the cluster list, merge them into a new single cluster and add the new cluster to the cluster list.

step 3: compute the distances between the newly formed cluster and the other ones in the cluster list.

step 4: repeat steps 2 and 3 until the desired number of clusters have been reached.

The number of clusters can either be specified by the user (our case here), or pre-indicators like the Davies-Bouldin index [55] or the cophenetic correlation coefficient [10] can assist this choice. The influence of this parameter over the registration performances is discussed in Section 5.3.3.1.

To store the distances between any two clusters in the cluster list at each iteration, we maintain a variable-size distance matrix M which summarizes their proximity (or similarity). At each iteration, M is therefore a square symmetric matrix whose size is the number of clusters in the cluster list at that iteration. The computation of similarity matrix M is the pivotal element of the clustering algorithm. The distance measure between clusters should be consistent with both the model we chose for the input images and the relationships we expect between them.

5.2.2.1.1 Distance between sites. To define a distance on clusters, we first need a distance on sites. This distance is defined as a linear combination of two distances, a distance between the centroids of the associated blocks and a distance between the associated similarity distributions:

$$D_{site} = \alpha \; D_{centroid} + (1 - \alpha) \; D_{distribution} \tag{5.4}$$

Distance between the centroids. To satisfy the model constraint, we have to ensure that close blocks are more likely to be clustered than blocks far apart. It appears that the Euclidean distance is not the most suitable here. Indeed, if the input images contain several pieces of tissues (*e.g.*, in histological images, they can easily be identified by thresholding) that are potentially non convex, a geodesic distance within each piece will be more convenient to define the proximity of two points from an anatomical point of view.

We recall that the geodesic distance between two points is the length of the shortest path that connects these points within a component that must contain them (the continuous red line in Figure 5.6 is the geodesic path between A and B, the dotted line the Euclidean path). By convention, when two sites cannot be connected (when they belong to disjoint components), we define the geodesic distance as the Euclidean distance between their associated centroids plus the radius of the input image. Computation of the geodesic distances was done using a variant of the circular propagation algorithm introduced in [50] which achieves a good trade-off of precision over speed.

Given two sites t and u, their centroid distance is written:

$$D_{centroid}\,(t, u) = D_{geodesic}\left(p_F^t, p_F^u\right) \tag{5.5}$$

Distance between similarity distributions. The high expressivity of the similarity distributions described above, which summarize the similarity landscapes associated with the neighborhoods of the blocks of interest, makes them remarkably well suited to capture the actual differences between those blocks, in spite of noise or decoys, and thus allows for a better discrimination. We use a normalized version ρ of these distributions to ensure that they all have the same overall unit mass (see [185] for a similar distributional approach in the context of image-flow computation).

Given a site t in L_F, the associated 2-D normalized distribution ρ^t is defined for sites u in the neighborhood N_t of t in L_R by $\rho^t\left(p_R^u - p_F^t\right) = \frac{sim\left(b_F^t, b_R^u\right)}{\sum_{v \in N_t} sim\left(b_F^t, b_R^v\right)}$. Such distributions are depicted in Figure 5.6(b) and (d).

As a distance between distributions, we chose the Earth mover's distance [177], a discrete solution to the discrete Monge-Kantorovich mass-transfer problem [88]. Given the so-called

"ground distance" (the distance between elements of the distribution, the Euclidean distance in our case), the Earth mover's distance (EMD) between two distributions becomes the minimal total amount of work (= mass × distance) it takes to transform one distribution into the other. As argued by Rubner *et al.* [177], this boils down to a bipartite network flow problem, which can be modeled with linear programming and solved by a simplex algorithm.

More formally, if $x = \{(x_1, s_1), \ldots, (x_m, s_m)\}$ and $y = \{(y_1, t_1), \ldots, (y_n, t_n)\}$ are the distributions between which we want to compute a distance, with $x_1, \ldots, x_m$ and $y_1, \ldots, y_m$ the centroids of the reference blocks, and $s_1, \ldots, s_m$, $t_1, \ldots, t_n$ the associated similarity values, then the EMD is defined by the following linear program:

$$EMD(x,y) = \frac{1}{\Lambda} . \min_{F=(f_{i,j}) \text{ in } F(x,y)} \sum_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} f_{i,j} d(x_i, y_j) \tag{5.6}$$

where the minimum is taken over the set of feasible flows $F(x,y)$ defined by the following constraints:

$$\begin{array}{ll} f_{i,j} \geq 0 & \forall i,j \\ \sum_{j=1}^{n} f_{i,j} \leq s_i & \forall i \\ \sum_{i=1}^{m} f_{i,j} \leq t_i & \forall j \\ \Lambda = \sum_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} f_{i,j} & \end{array} \tag{5.7}$$

where Λ can be regarded as the minimum of the total weight of x and of the total weight of y.

Among other advantages, the EMD is a true metric, is not impaired by quantization problems (as opposed to histogram-based approaches for instance) and can handle variable-size distributions (our case here). For sites t and u, we obtain:

$$D_{distribution}(t,u) = D_{EMD}\left(\rho^t, \rho^u\right) \tag{5.8}$$

To summarize, given two sites t and u, their site distance is written:

$$D_{site}(t,u) = \alpha D_{geodesic}\left(p_F^t, p_F^u\right) + (1-\alpha) D_{EMD}\left(\rho^t, \rho^u\right) \tag{5.9}$$

where α is a real-valued positive weight ($0 \leq \alpha \leq 1$).

5.2.2.1.2 Distance between clusters. Once we have a distance between sites, a cluster distance can be defined. We adapted the standard complete link distance [10] to additionally take into account the transformations that can be estimated on the already formed clusters.

Namely, when the size of a cluster reaches a given threshold (we usually take $\theta_{cluster}$=20, even though experiments showed that the value of that threshold does not really impact the quality of the clustering), a rigid or affine transformation can be estimated, in a robust fashion, from the associated set of "arg max" displacement vectors (via a least-square regression combined with an LTS (Least Trimmed Sum of Squares) estimator, see Section 5.2.3). The decision to merge

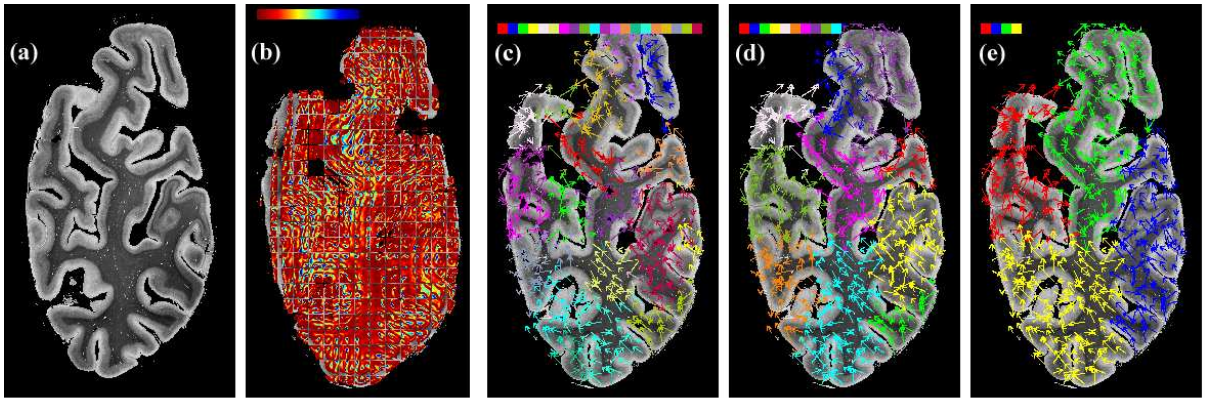


Figure 5.7: Clustering the correspondence field: (a) input reference image; (b) input floating image with super-imposed similarity map; (c,d,e) floating image with super-imposed clustered "arg max" field for respectively 20, 10 and 4 clusters.

two clusters can then be biased by the agreements between their associated estimated transformations, again as this might indicate that they belong to the same component. Incidentally, when the distance between a cluster with an associated transformation and another one without enough sites to have allowed an estimation must be computed, we choose to return 0. Although theoretically possible, such a case almost never occurs in practice as a hierarchical clustering algorithm tends to aggregate sites in small clusters at early stages before merging them into large ones in subsequent iterations, not leaving single sites un-aggregated very long (see [10] for details). This so-called "chaining effect" also motivates the use of transformation distances.

Given two transformations T^a and T^b, we use a standard symmetric distance:

$$D_{trsf}\left(T^a, T^b\right) = \begin{cases} \sum_{i,j}\left[T^a T^{b^{-1}} - Id\right]^2_{i,j} + \sum_{i,j}\left[T^{a^{-1}} T^b - Id\right]^2_{i,j} & \text{if both } T^a, T^b \text{ are defined} \\ 0 & \text{otherwise} \end{cases} \tag{5.10}$$

(where i, j are matrix indices).

Finally, given two clusters of sites $C^a = \{a_1, \ldots, a_{n_a}\}$ and $C^b = \{b_1, \ldots, b_{n_b}\}$, the cluster distance between them is the longest distance from any site in C^a to any site in C^b (complete-link) *plus* the "transformation distance" wherever it can be computed:

$$D_{cluster}(C^a, C^b) = \beta \max_{i,j} D_{site}(a_i, b_j) + (1 - \beta)\ D_{trsf}(T^a, T^b) \tag{5.11}$$

where β is a real-valued positive weight $(0 \leq \beta \leq 1)$.

Figure 5.7 shows the clustering process of the correspondence field for the two consecutive histological sections of Figure 5.1 from 20 down to 4 clusters. It illustrates the nesting property of the hierarchical clustering approach: the clusters in (c) are merged to form the clusters in (d).

Note that as an alternative, we could have used one of the numerous optical flow segmentation algorithms developed in the literature ([215] for instance) to segment the input images. However,

a number of modifications would need to be made to allow for the registration of multi-modal images as they violate the principle of intensity conservation. Additionally, taking into account geodesic distances might also prove difficult. Finally, we believe that better results can be obtained by considering the *complete* similarity map associated with a block instead of choosing *a priori* a single displacement to perform the classification.

5.2.2.2 Extracting the sub-images

We have described above how we cluster the floating lattice L_F. We detail here how to extract, from the input floating and reference images, pairs of sub-images that will later be registered independently.

Let N_C be the final number of clusters, $C = \left\{C^1, \ldots, C^{N_C}\right\}$ the cluster partition of L_F, and $\left\{c_1^i, \ldots, c_{n_i}^i\right\}$ the n_i sites of the i^{th} cluster C^i. We want to build a set of N_C sub-images $\left\{I_F^i\right\}_{i=1}^{N_C}$, each of them associated with a single cluster. Given the partition of L_F, a partition of I_F can be built in many ways. For instance, one could compute a Voronoï diagram of the sites c_i^j (or equivalently of their centroids) and draw a partition of the pixels (x, y) of I_F from it. However, our clustering method does not ensure that the borders between clusters are sufficiently precise to adequately represent the sub-images' borders. Moreover, as we are going to use these sub-images to find local transformations, it is often better to choose larger supports to avoid boundary effects.

Consequently, rather than build a partition of I_F from the partition of L_F, we build a covering of I_F, *i.e.*, a set of sub-images that could overlap. To do so, we aggregate in I_F^i the pixels of I_F in the vicinity of the sites of the cluster C^i. We get:

$$I_F^i = \{(x, y) \in I_F \text{ such that } D((x, y), c_j^i) \leq cover_{radius} \text{ for some } c_j^i \in C^i\} \tag{5.12}$$

In practice we use the L_∞ distance. Then, with blocks of size b_{size} associated to the sites, taking $cover_{radius} = \frac{b_{size}}{2}$ we get $I_F^i = \bigcup_j b_{I_F}^{c_j^i}$. In our experiments, to ensure a large support, we chose $cover_{radius} = \frac{3}{4}\, b_{size}$.

The corresponding reference sub-images I_R^i are built identically, but with the centroids $p_R^{(k,l)_{max}}$ of the most similar blocks:

$$I_R^i = \{(x, y) \in I_R \text{ such that } D((x, y), c_j^i + d^{c_j^i}) \leq cover_{radius}, \text{ for some } c_j^i \in C^i\} \tag{5.13}$$

Again, we use the L_∞ distance here, with $cover_{radius} = b_{size}$ (a larger extent than that of the floating sub-image) to ensure that I_F^i can be effectively registered against I_R^i (see Figure 5.9).

5.2.3 Composing the registered floating image

5.2.3.1 Registering the sub-images

Once we have extracted the reference and floating sub-images, we use the robust affine block-matching algorithm described in [156] to register them, independently, pair by pair. Briefly, this algorithm first estimates a sparse "arg max" displacement field, using a block matching approach (our block-matching algorithm is closely derived from this approach, and we feed both of them the same parameters and similarity measure). From this field, a least square regression extracts a rigid or an affine transformation. As an illustration, in the rigid case we are looking for R^* and t^* such that:

$$(R^*, t^*) = \arg\min_{R,t} \sum_{i,j} \left\| \left(p_F^{i,j} + d^{i,j}\right) - R.p_F^{i,j} - t \right\|^2 \tag{5.14}$$

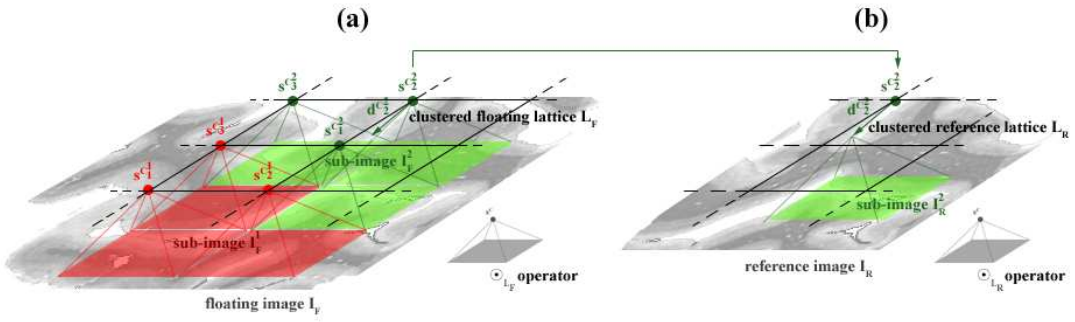


Figure 5.8: Extracting the sub-images from a clustered floating similarity map: (a) floating image with clustered set of sites; (b) reference image.

where $\left(p_F^{i,j} + d^{i,j}\right) - Rp_F^{i,j} - t$ is the residual error and $\|.\|$ the L_2 Euclidean norm.

However, given the rather noisy appearance of the displacement field, an LTS estimator (Least Trimmed Sum of Squares, see [174] for details) is used in place of the least square one to ensure a robust estimation of the transformation. At a glance, instead of minimizing the total sum of the squared residuals (equation 5.14), a LTS estimator will iteratively minimize the sum of the h smallest squared residuals (we take h at 50% of the number of residuals), to reduce the influence of outliers.

Finally, a better trade-off between robustness and registration precision is achieved via a multi-scale implementation. Note that even though this block-matching algorithm computes displacements (actually, translations) only *locally*, it is able to recover *global* rotations and translations, thanks to its iterative nature. For instance, a robustness study on rat brains sections presented in [156] demonstrated its ability to recover rotations up to 28 degrees.

Then, for each pair of sub-images $\left\{I_R^l, I_F^l\right\}$, $l \in 1 \ldots N_C$, we obtain a rigid or an affine transform T^l. Note that since these registrations are robust, the sub-images do not need to perfectly correspond to the anatomically separate components.

5.2.3.2 Composing the final images

We selected the Little *et al.* method [130] to compose the final registered floating image. In their approach, a user selects a number of pairs of corresponding rigid structures in the input images along with associated linear transformations (also given by the user). A number of pairs of landmarks further constrain a hybrid affine/non-linear interpolation scheme that acts as a local registration algorithm. This technique then essentially applies user-provided affine transforms to user-defined structures and ensures a smooth interpolation in between them.

In our application, the set of floating sub-images forms a covering of the input floating image, so we have to erode the sub-images to leave space for interpolation. Furthermore, the floating sub-images must be cut to ensure that they do not overlap, once transformed, as this may impair the interpolation scheme (note the overlap of the red and green sub-images, and the gap between the read and yellow ones in Figure 5.9). This erosion algorithm works as follows:

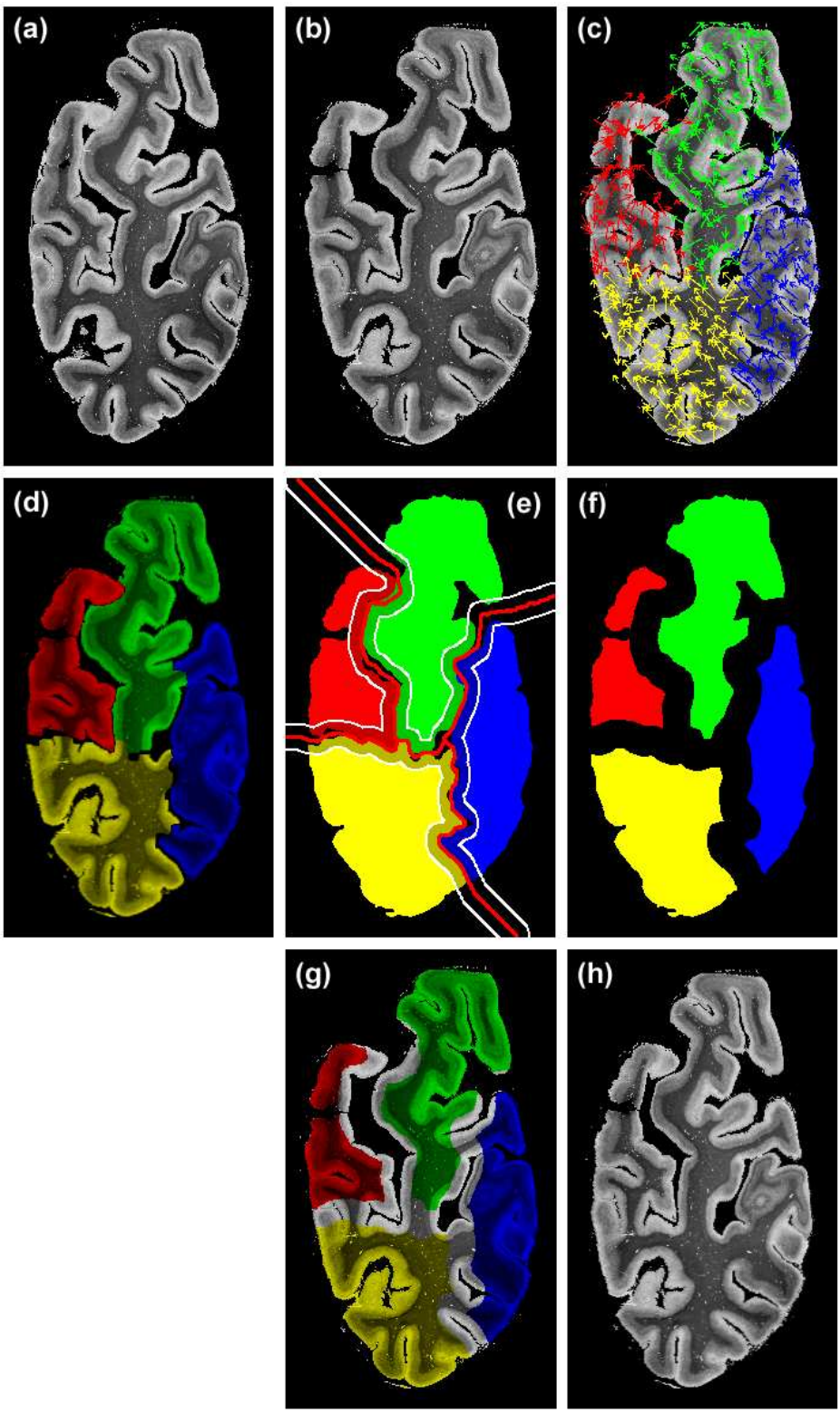


Figure 5.9: Composing the locally registered floating image: (a) input reference image; (b) floating image; (c) floating image with clustered "arg max" displacement field (d) registered floating sub-images; (e) binarized sub-images with darkened eroded pixels; (f) eroded registered floating sub-images; (g) input floating image with super-imposed colored eroded floating sub-images; (h) final composed locally registered floating image.

composition

- We first apply the transformations to the floating sub-images ($\forall l \in 1 \dots N_C$, $T^l\left(I_F^l\right)$ is the transformed floating sub-image), binarize them (zero for background, one for tissue) and fill in the holes.
- We superimpose the binarized transformed sub-images in a single image J and compute a distance map over the background of that image.
  - A series of morphological operations (erosion) first ensures (on a need for basis) that the $T^l\left(I_F^l\right)$ are disjoint.
  - A Euclidean distance map of the background of J is computed.
- A medial axis algorithm then extracts the skeleton of the background of J.
- We compute the distance map of this skeleton.
- We identify in J pixels whose corresponding distance to the skeleton is smaller than a given threshold ν. This ensures a minimum distance of 2ν between any two sub-images. Let N be the set of these pixels. We then remove from the floating sub-images their inverse transformed intersection with N: $\forall l \in 1 \dots N_C$, $\overline{I_F^l} = I_F^l - T^{l\,-1}\left(T^l\left(I_F^l\right) \cap N\right)$.

We choose as landmarks the corners of the original images, I_R and I_F (after the initial rigid registration), to further constrain the interpolation scheme, and use the modified Hardy multiquadric recommended in [130] as a basis function for interpolation, as this agrees with an affine transform at infinity. Figure 5.9 exemplifies our composition process on the myelin stained histological section of Figures 5.1. We show in (e) the transformed floating binarized sub-images in color, the skeleton of the background of J in red and the $2\nu = 20$ pixel wide band of eroded pixels in darkened colors with white borders.

Note that the entire registration process could easily be included within an iterative multi-scale framework to achieve a better trade-off between accuracy and complexity. Such a framework could also be useful for handling both large-scale and small-scale components.

5.3 Results

We present here the various experiments we have conducted to assess the performances of our piecewise registration approach. We first discuss the ability of the clustering algorithm to correctly classify phantom images (Section 5.3.1) before detailing some biomedical registration results (Section 5.3.2). Finally, we discuss in Section 5.3.3 the influence of the various components and parameters of our registration system on the quality of the registration.

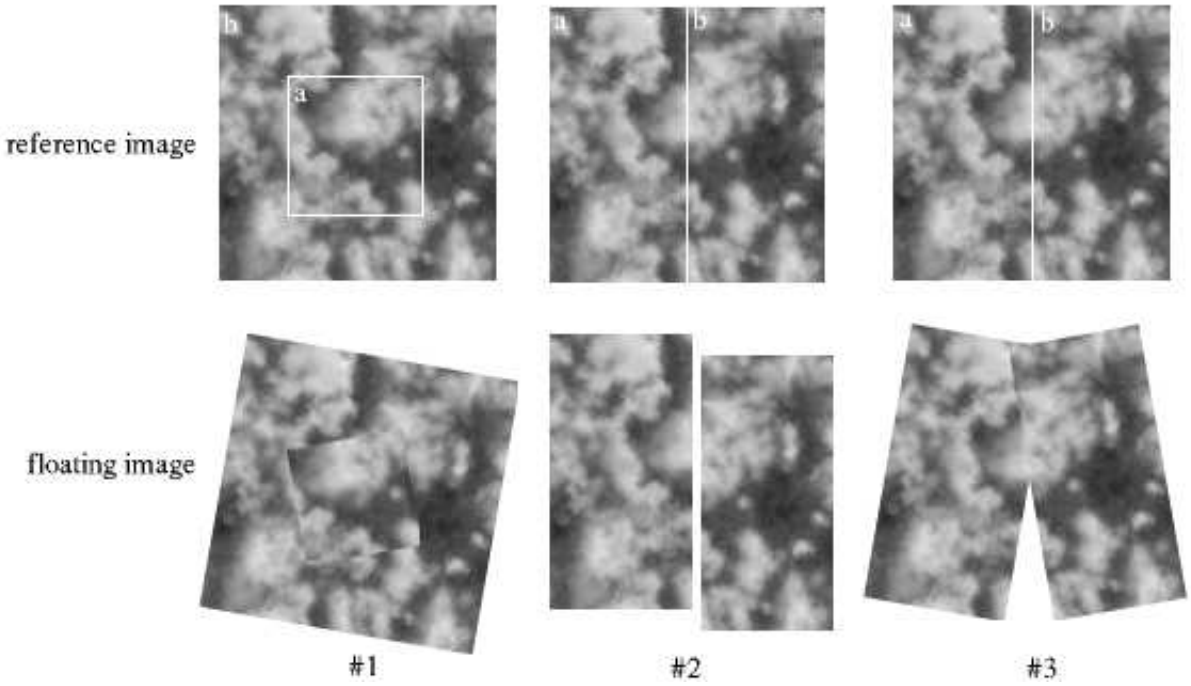


Figure 5.10: Three pairs of 2-component phantom structures.

5.3.1 Phantom images

Phantom images enable us to evaluate the quality of the clustering algorithm under controlled textural conditions. Namely, pairs of synthetic images were created with artificial textures (grayscale cloud pattern, see Figure 5.10).

We illustrate the behavior of our approach on images consisting of several connected components. We considered 3 structures with 2 components each: in #1 a small square inside a larger one, in #2 and #3 the two vertical halves of a large square. To each component, a different affine transformation was applied:

- component 1a: T^{1a} = counter-clockwise $\frac{\pi}{8}$ rotation;
- component 1b: T^{1b} = clockwise $\frac{\pi}{8}$ rotation;
- component 2a: $T^{2a} = (-10, -20)$ translation;
- component 2b: $T^{2b} = (-5, +15)$ translation;
- component 3a: T^{3a} = counter-clockwise $\frac{\pi}{8}$ rotation;
- component 3b: T^{3b} = clockwise $\frac{\pi}{8}$ rotation.

Table 5.1 reports the classification results with and without the transformation based distance, and with and without distribution distance (EMD). We set $\alpha = 0.5$, $\beta = 0.5$, $\theta_{cluster} = 20$ and $N_C = 2$, and used the constrained correlation coefficient for rigid registration. For each

components	with transformation distance		without transformation distance	
	with EMD	without EMD	with EMD	without EMD
# 1a	43 / 47	40.1 / 34	40 / 43.2	37.6 / 41
# 1b	53 / 57	66 / 59.9	56.8 / 60	59 / 72.4
# 2a	86.6 / 26.8	68.3 / 46.2	74.3 / 32.4	72 / 34.2
# 2b	73.2 / 13.4	53.8 / 31.7	67.6 / 25.7	65.8 / 28
# 3a	76.5 / 19	59.2 / 26.5	54.3 / 27.5	57.2 / 11.5
# 3b	81 / 23.5	73.5 / 40.8	72.5 / 45.7	88.5 / 42.8

Table 5.1: Performance of our local registration algorithm for a variety of phantom images.

structure, we show, for each component, both the number of pixels of that component that were correctly classified and the number of vectors of the complementary component that were wrongly classified as this one, as a percentage of their respective number of pixels.

We observed better performances when the EMD distribution distance was used than when it was not. The similarity distribution distance actually helps the clustering algorithm to form, at early stages, sensible clusters that are then adequately agglomerated with the aid of the robust estimation of the associated transformations (even though, due to the noisy nature of the "arg max" displacement field, those may sometimes hinder the clustering process). Finally, use of the geodesic distance was particularly beneficial for components 2a/b and 3a/b.

Note that a *perfect* clustering is not a necessary condition for our method to perform well, (1) since the subsequently extracted sub-images will be larger than the obtained clusters and will overlap, and (2) since the registration algorithm we use to register these sub-images is robust.

5.3.2 Biomedical images

5.3.2.1 Two detailed examples of mono- and multi-modal registration

We illustrate here the validity of the piecewise paradigm on two examples of mono- and multi-modal registration.

Figure 5.11 displays the results of our local registration for an autoradiographic monkey brain section and its associated MRI (obtained by co-registration of a series of contiguous autoradiographic sections with an MR volume of the same monkey [134]) and Figures 5.12 for the pair of myelin-stained histological sections introduced in Figure 5.1, and . In both cases, a gyrus (bottom left corner in Figure 5.11 and top left corner in Figure 5.12) was detached during the histological preparation and manually realigned in an unsatisfactory fashion.

In Figure 5.12, we show the reference image (a), the transformed binarized floating sub-images with darkened eroded pixels (b), the locally registered floating image (c), the colored superposition of the reference image and the globally affinely registered floating image (d), the image of a regular grid convected with the associated hybrid affine/non-linear transformation with the superimposed transformed eroded floating sub-images (e), and the colored superposition of the reference and the piecewise registered floating image (f). In Figure 5.11, the transformed sub-images composition is replaced by the floating image with superimposed clustered "arg max" displacement field (b), the colored superpositions are replaced by the reference image on which we superimposed the edges of the globally registered floating image (d) and of the locally registered floating image (f).

These pairs of images were piecewise *rigidly* registered by our approach, with $\alpha = 0.5$, $\beta = 0.5$, $\theta_{cluster} = 20$ and standard parameters for the block matching algorithm (see Table 5.2) with a constrained correlation coefficient and $N_C = 4$ for the myelin sections, and a classic coefficient with $N_C = 2$ for the autoradiography (tests with normalized mutual information as a similarity measure yielded less good results). Our clustering algorithm adequately isolated in a separate sub-image the moving part in the monkey case (green area in Figure 5.11(b)) and the floating gyrus of the myelin section pair (red area in Figure 5.12(b)) which were subsequently correctly registered to their counterpart in the reference image. An affine transform would of course further decrease the discrepancy between the pairs of sub-images. However, in the general case, when one suspects only a rigid transformation between sub-images, opting for an affine registration would only introduce unnecessary over-parameterization which, among other disadvantages, could substantially alter textures.

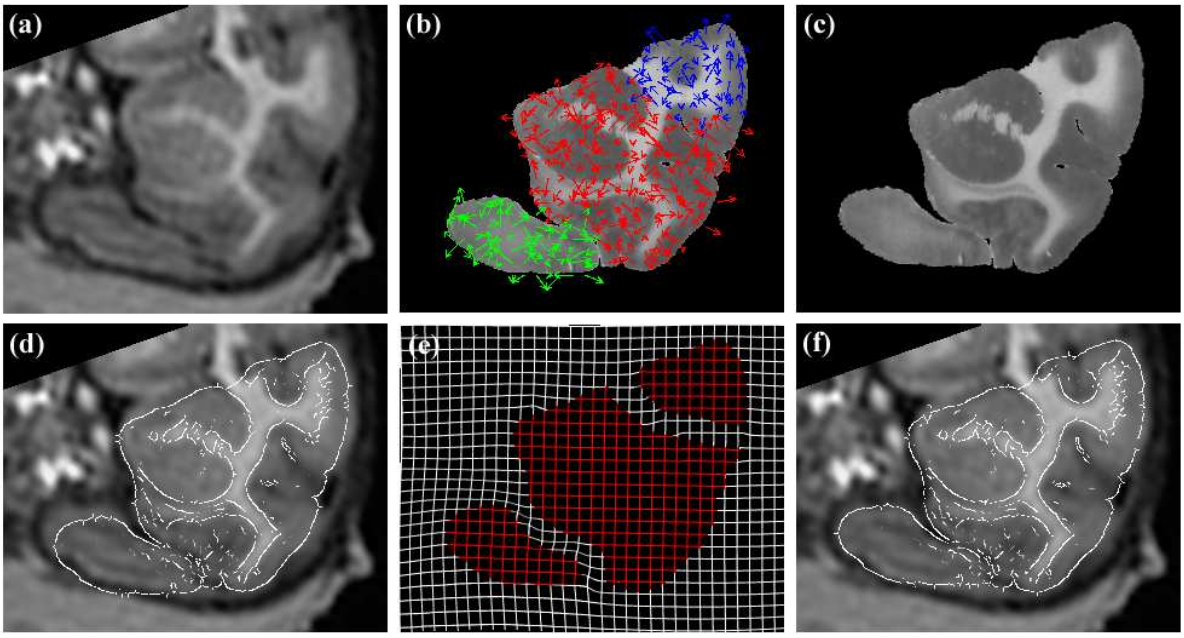


Figure 5.11: Registration of an autoradiographic monkey brain section and its associated MRI: (a) reference image, (b) floating image with clustered optimal arg max displacement field, (c) locally registered floating image, (d) reference image with superimposed edges of the globally affinely registered floating image, (e) image of a regular grid convected with the associated hybrid affine/non-linear transformation with superimposed transformed eroded floating sub-images (in red), (f) reference image with the superimposed edges of the locally registered floating image.

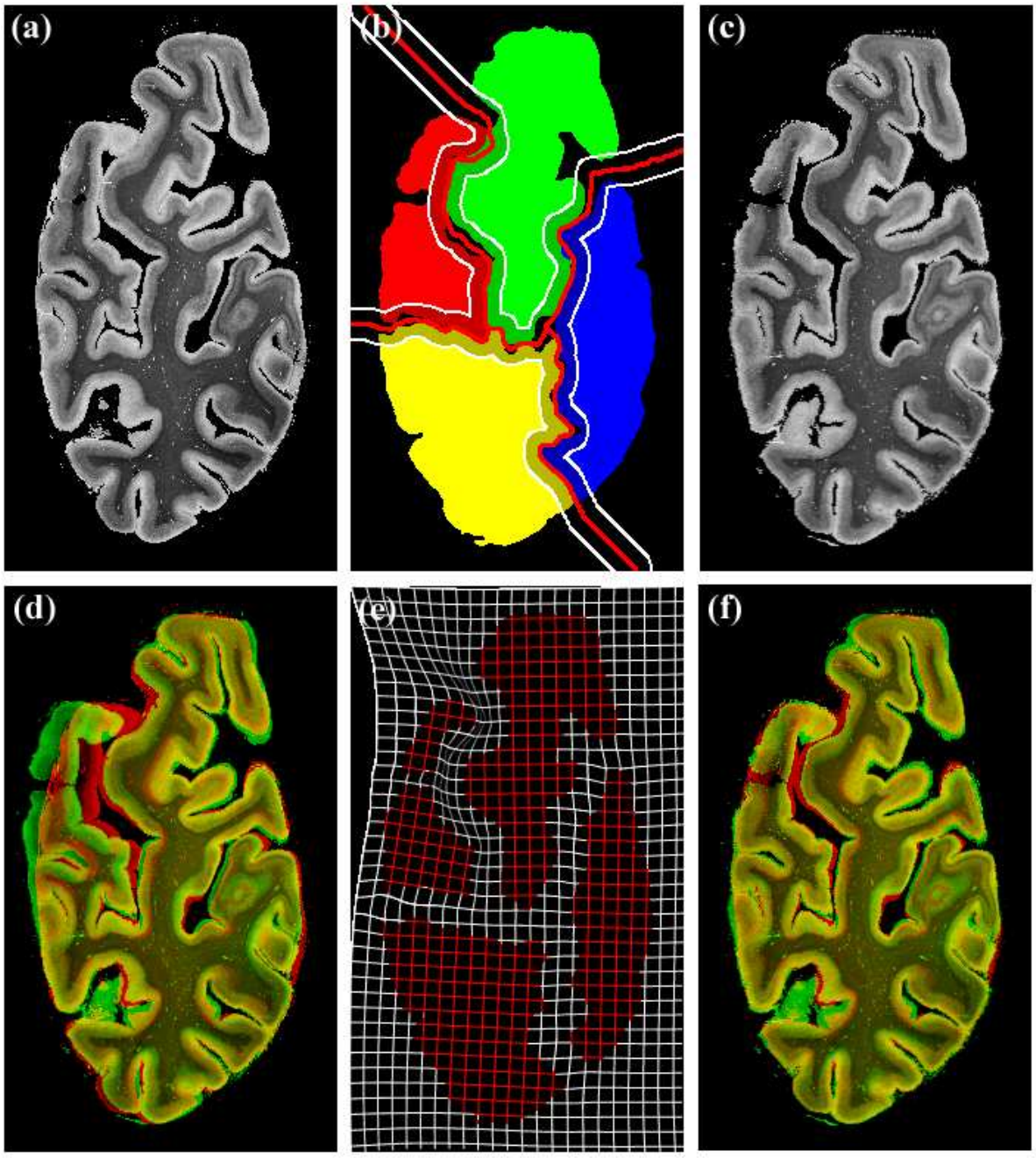


Figure 5.12: Registration of two consecutive myelin-stained histological sections of the human brain: (a) reference image, (b) transformed binarized floating sub-images with darkened eroded pixels, (c) locally registered floating image, (d) superposition of the reference image (red) and of the globally affinely registered floating image (green), (e) image of a regular grid convected by the associated hybrid affine/non-linear transformation with superimposed transformed eroded floating sub-images (in red), (f) superposition of the reference image (red) and of the locally registered floating image (green).

5.3.2.2 Reconstruction of a 3-D histological volume

Even though the deformations recovered by our registration method may sometimes be rather subtle, as exemplified by the registration of the two pairs of images presented above (Figures 5.12 and 5.11), they become a clear nuisance when entire stacks of sections must be aligned.

We aim to reconstruct a 3-D volume from a series of histological or autoradiographic images. As argued above, previous work [156, 134] showed that by registering (affinely or rigidly) each pair of consecutive slices in the stack we can recover a geometrically coherent 3-D alignment of the 2-D images and provide a satisfying 3-D reconstruction. However, local rigid/affine piecewise transformations, as described in Section 5.1.1, still impair this registration process and must be accounted for.

We have reconstructed of a 3-D histological volume from a series of 70 images. These were $50\mu m$ thick myelin-stained histological sections of the human brain cut roughly orthogonal to the calcarine fissure, including the primary (striate) visual cortex. Reconstruction was performed using the classic pair-wise approach described above. Note that this process require the choice of a reference section: if we let $Img(ref)$ be this reference section, with $1 < ref < 70$, the reconstruction algorithm is then as follows:

```
for i from ref+1 upto 70
  rigid piecewise register Img(i) to Img(i-1)
for i from ref-1 downto 1
  rigid piecewise register Img(i) to Img(i+1)
```

We used here the same parameters as for the registration of the myelin-stained sections of Figure 5.12): $\alpha = 0.5$, $\beta = 0.5$, $\theta_{cluster} = 20$ and standard parameters for the block matching algorithm with a constrained correlation coefficient and $N_C = 6$.

Figure 5.13 compares the volume reconstructed with our piecewise approach and that built with the robust rigid registration algorithm we use to register the sub-images. In both cases (rigid and piecewise rigid) we show in (b) a coronal view (middle) of the 3-D reconstructed volume corresponding to the 51^{st} image of the stack with the associated axial (top) and sagittal (left) views, in (a) the 50^{th} image (the immediately preceding section) with edges of the 51^{st} one superimposed in red, and in (c) the 52^{nd} image (the immediately following section) with edges of the 51^{st} image superimposed.

Note the greater regularity of the 3-D structures in both the sagittal and axial views of the piecewise reconstructed volume, with respect to the global rigid volume. A better registration of the separate gyri, illustrated by the better superposition between the red edges and the underlying images, explains this smoother aspect.

Visual inspection all through the piecewise 3-D volume confirmed the enhanced continuity of the 3-D structures.

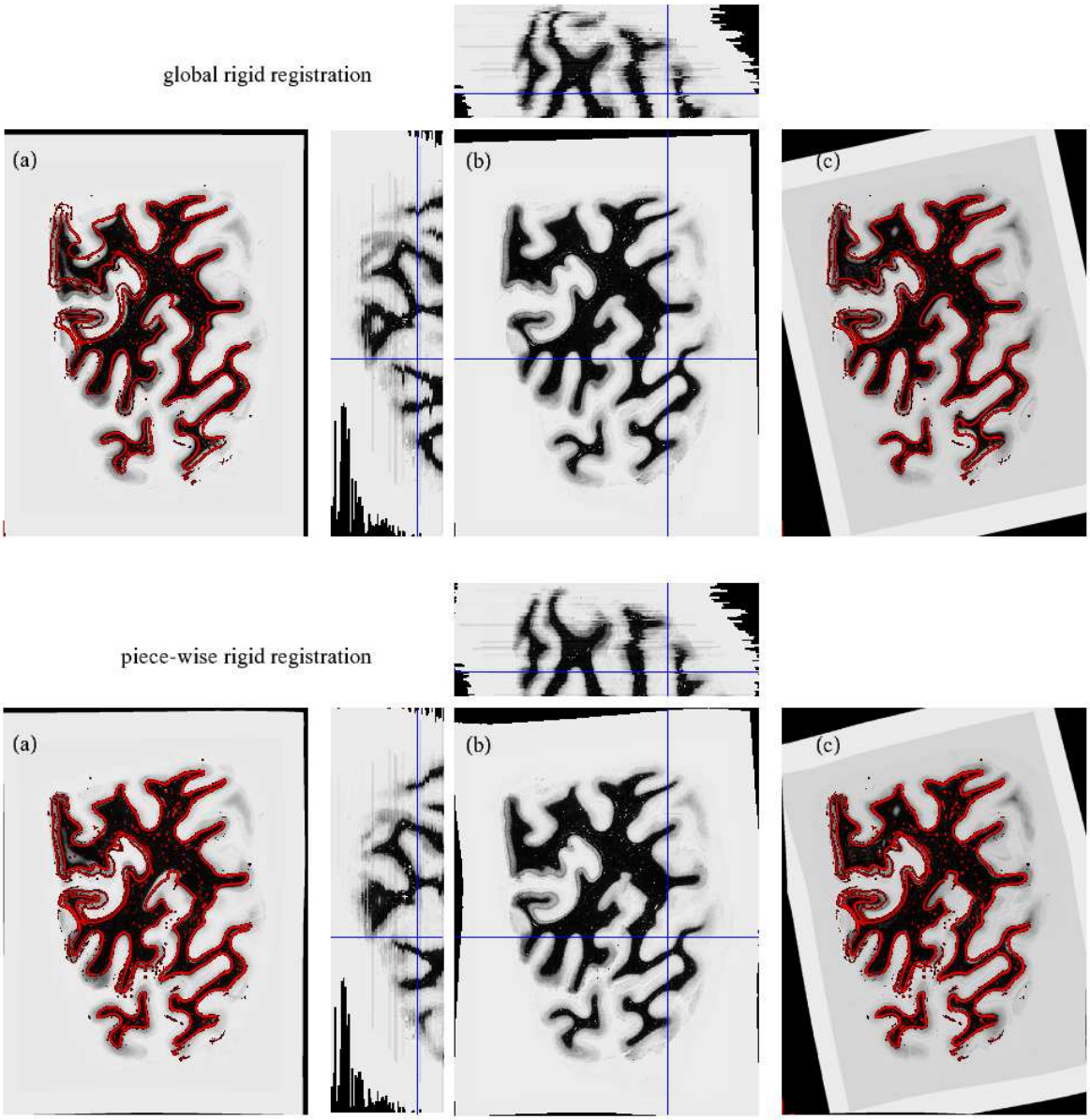


Figure 5.13: Reconstruction of a 3-D histological volume with a globally rigid (top) and our piecewise rigid (bottom) registration algorithm: (b) coronal view (middle) of the 3-D reconstructed volume corresponding to the 51st image of the stack with the associated axial (top) and sagittal (left) views, (a) 50th image (immediately preceding section) with edges of the 51st one superimposed in red, (c) 52nd image (immediately following section) with edges of the 51st image superimposed

5.3.3 Sensitivity study

We discuss in this section the dependence of our piecewise registration approach on its various parameters. Standard values for most of these parameters are also reported in Table 5.2.

components	*algorithm*	*name*	*typical value*
block size	block-matching	b_{size}	6×6
size of $N_R^{i,j}$	block-matching	$<none>$	20×20
lattices step size	block-matching	$<none>$	5×5
similarity measure	block-matching	*sim*	*cor* or *ecor*
LTS cut-off	block-matching	h	50%
centroid weight	clustering	α	0.5
transformation distance weight	clustering	β	0.5
transformation distance threshold	clustering	$\theta_{cluster}$	20
number of clusters	clustering	N_C	6
covering radius	extraction	$cover_{radius}$	$\frac{3}{4} b_{size} = 5$
space between eroded structures	sub-images erosion	ν	10

Table 5.2: Standard values for the piecewise registration algorithms.

5.3.3.1 Number of clusters

The number of clusters determine the number of degrees of freedom of the overall transformation, and consequently influences the quality of the final match. The middle column ($\nu = 10$) of Figure 5.14 qualitatively illustrates the behaviour of the piecewise registration system when the number of clusters varies (between 3 and 8).

A quantitative analysis of the influence of the specified number of clusters on the final registration quality was also carried on, using the synthetic images of Figure 5.10. In both the qualitative and quantitative cases, we observe that when the specified number of cluster increases above the number of actual components, we get sub-components that are correctly included in the components they come from. The associated transformations are also part of the transformation of the enclosing component (with minimal error, 2% on average).

This comes as no surprise. Indeed, in a hierarchical clustering, each partition is nested into the next partition in the sequence. Therefore, when the number of desired clusters increases above the actual number of components, the new sub-images (associated with the new clusters) are sub-parts of actual components. Since actual components are supposed to be rigid or affine by definition, affinely registering the new sub-images should produce transformations very similar to the transformations associated with the nesting sub-image. Conversely, when the specified number of cluster drops below the number of actual components, performances decrease and tend towards those of a robust global affine registration. In the limit where a large number of clusters are used, the piecewise registration tends to resemble a block-matching approach with the notable difference that complete affine (or rigid) transformations are estimated instead of simple translations at each block.

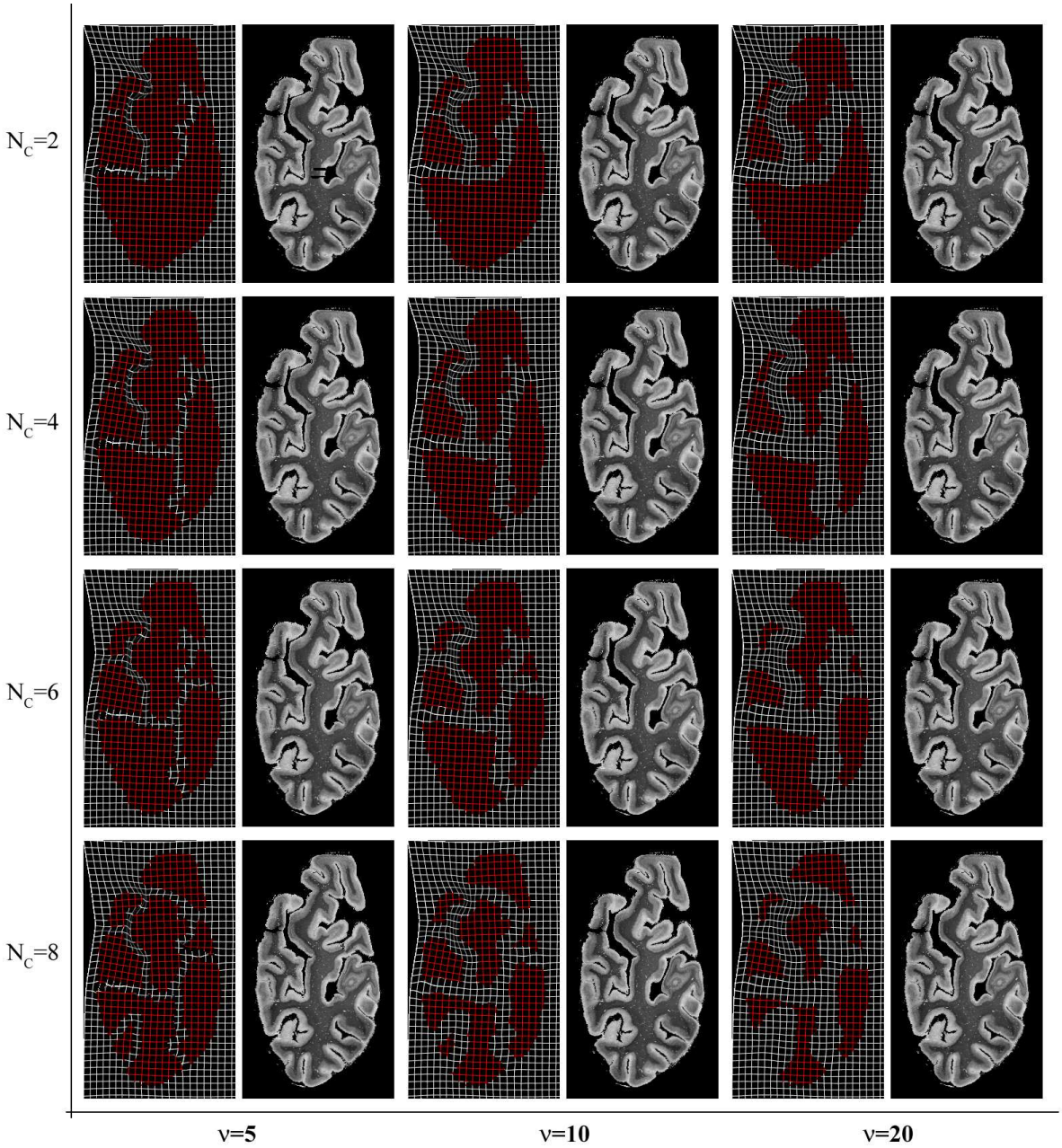


Figure 5.14: Sensitivity of the piecewise affine registration method to the number of clusters and to the inter sub-image space. For each choice of parameters, we show the regular grid convected with the composed transformation with the superimposed transformed eroded sub-images (left) and the final composed registered floating image (right).

Note, that even though obtaining a *perfect* clustering may not actually be necessary, our clustering algorithm suffers from its inability to modify clusters that have already been created: namely, clusters can only be aggregated to form larger clusters, they cannot be re-cut or broken

down. This is unfortunate as better cluster boundaries could probably be obtained based on the associated cluster transformations, which are computed only after the clusters have reached a sufficient size to ensure a correct estimation. The use of stochastic clustering approaches or the introduction of uncertainty in the clustering process should alleviate this issue.

5.3.3.2 Block-matching parameters

As argued above, the quality of the similarity map computed between the floating and the reference image depends on the block matching parameters: similarity measure, size of the exploration neighborhood, step in that neighborhood, size of blocks, etc. We already studied in Section 5.2.1.2 how the use of a constrained correlation coefficient helped increase both the homogeneity and the precision of the similarity map. Our block-matching algorithm is similar to that detailed [156] and [155] to which we report the reader for a detailed sensitivity investigation of the other parameters.

Note that the selection of the "arg max" displacement vector is clearly sub-optimal and somewhat arbitrary when many blocks in the reference exploration neighborhood have close associated similarity measures. Better estimation of the sub-image transformations might be obtained by taking into account the full spectrum of displacement vectors, together with their similarity measures.

5.3.3.3 Parameters of the registration algorithm for the sub-images

The robust affine block-matching algorithm used in Section 5.2.3.1 to register the sub-images too requires that a number of parameters be set. In addition to block matching parameters similar to ours, the cut-off of the robust estimator, the parameters controlling the multi-scale system within which it works, or the parameters of the various variance and intensity tests performed on the block to discard them from the robust estimation must be managed. Again, we report the reader to [155] for details about their influence on the registration performances.

5.3.3.4 Composition parameters

The choice of the amount of space to leave in between structures (2ν pixels) depends on the input images and should be set accordingly. However, there is no general prescription for selecting a good value for ν which would work well for all images and, within a single image, for all sub-images. Clearly, as the amount of space decreases, the band in between sub-images becomes more stretched (which might induce substantial textural changes).

Figure 5.14 illustrates the relationships between the selected number of clusters and the space to leave between sub-images. As a rule of thumb, the greater the number of clusters is, the smaller ν should be, as the size of the components tend to decrease. Note that even though we restricted ourselves to a constant value for ν across the image in this study, a variable ν should increase the overall registration performances. Namely, for each pair of sub-images, ν should depend on the differences between the transformations associated to the sub-images. If the difference is large, then a large ν is to be used to prevent large distortions of the textures of the underlying tissues. Conversely, when the two sub-images share very similar transformations (even though these transformation might be large with respect to the identity), ν can be much smaller (for instance, for components A and B in Figure 5.14).

The choice of ν could also depends on the nature of the underlying material: large deformations will not impair the quality of the registration if they occur over the black background.

Use of the tissue/background segmentation maps described in Section 2.4 could then also help choose locally an optimal value for ν.

5.4 Conclusion

We have presented in this chapter a fully automated local registration method, capable of dealing with a variety of 2-D images. It builds complex spatial transformations by elastically interpolating between rigid or affine transforms that are locally defined on pairs of sub-images. Consequently, we manage to minimize the overall number of degrees of freedom of the transformation, thereby agreeing with the guidelines of the parsimony principle (see [118] for a discussion of the problems inherent to high-dimensional transformations). These sub-images represent geometrically coherent components (in our biomedical applications, they are even anatomically coherent components). They are automatically extracted from an initial displacement field computed between the images to be registered. All user interaction is avoided, by contrast with other approaches [130].

The use of a hierarchical clustering approach and a similarity distribution distance proved very promising: while the distribution distance can effectively deal with noise and textural issues to discriminate between image blocks, our clustering algorithm manages to extract the expected sub-images.

Results on real data in both 2-D registration and 3-D reconstruction cases demonstrated the adequacy of the proposed method for several specific problems in biomedical imaging. In our atlas building case, once a 3-D histological volume has been constructed, the next step, which we did not tackle in this chapter, consists of registering it to an anatomical MRI to build a geometrically correct hybrid histology/MRI atlas. This will serve to enhance the performances of the segmentation system. The developed framework could also ideally benefit from the *in vivo* microscopic capabilities of the future MR scanners.

Finally, even though the presented registration method works in 2-D, it could readily be extended to 3-D with a close-to-linear increase in processing time.

Chapter 6

Conclusion
Conclusion

Mathematicians are like Frenchmen: whatever you say to them
they translate into their own language,
and forthwith it is something entirely different.
Johann W. von Goethe

Nous avons ouvert ce manuscript sur la notion d'image en tant que représentation. Nous espérons avoir démontré au fil de ses pages qu'une image ne peut en conséquence être considérée indépendamment des mécanismes qui ont présidé à son élaboration (ni donc des processus physiques qui ont effectivement eu lieu, ni encore des appareils de mesure qui auront capturé un état à un instant de leur déroulement). Nous revenons dans ce dernier chapitre sur les objectifs et hypothèses attenants à notre étude, et concluons sur le recalage en tant que paradigme unificateur pour l'ensemble des méthodes et algorithmes développés avant de donner quelques perspectives futures.

Jasper Johns, *0 through 9*, 1960

We stated in the introduction that images are first and foremost representations. We demonstrated, all through this manuscript, that they consequently cannot be considered independently from the mechanisms that produced them (neither from the physical processes that actually took place nor from the measuring tools that captured a snapshot of their evolution). Image analysis then consists in extracting meaning out of one or of a series of images.

In view of the plethora of data encompassed at a variety of scale in a given image, this analysis must be conducted following the guidelines provided by *a priori* given hypotheses and goals, a requirement all the more pivotal since most real world applications are unfortunately ill-posed.

Those hypotheses are set forth in the following section along with a summary of the five chapters of this manuscript. We then present a general registration framework for our algorithms (Section 6.2) before concluding with some future studies (Section 6.3).

6.1 Wrap-up

Chapter 1 (Introduction) motivated the objectives of our study. Confronted with the necessity to delineate with a controlled precision an ever growing number of cerebral structures in MRIs, to establish much needed diagnosis, monitor brain dysfunctionings or study anatomical variability, automated segmentation appeared as an obvious choice. The interest of using maximum *a priori* knowledge, both in an explicit form (established by a human operator) and in an implicit from (discovered by an automated, or semi-automated, algorithm), was submitted to handle the particularly variable shapes and appearances of anatomical structures. As argued in that chapter, the need for a convenient framework into which incorporating this *a priori* medical expertise led us to favour deformable templates as a

deformation paradigm. We approached the issue of boundary finding as a process of fitting a group of deformable templates to the contours of some selected target structures (corpus callosum, lateral ventricle, caudate nucleus and hippocampus). Those templates evolved in parallel under the supervision of a series of rules derived from the analysis of both the template's dynamics and medical experience. They were subjected to a variety of constraints also designed based on prior medical knowledge about the textural, shape and underlying histological properties of the target structures, which we detailed in the subsequent chapters.

Chapter 2 (Neural Texture Filtering) introduced the main two (complementary) modalities on which we focused, MRI and histology. Their characteristics prompted us to consider texture as a discriminating element for the target structures. From a textural classification of regions of interest defined in the input images, texture constraints could be built to bias the deformable templates towards more probable boundaries. We reviewed a large panel of texture descriptors and insisted on the necessity to select an optimal subset from them, the so-called feature selection problem. We presented three texture classifiers: two classical approaches in the form of linear discriminant analysis and linear support vector machines, and a non-linear classifier in the form of an original neural network architecture, with an increase in both complexity (from linear, to linear in a non-linear projection space, to learnt non-linear) and performances. The neural approach, in the form of a 2-stage hybrid network, implemented two levels of classification (the second correcting the mistakes of the first one and incorporating spatial awareness) and relied on a dynamic learning phase to deal with the issue of representativity of the texture learning set. *A priori* information was introduced in a semi-implicit form via this learning set.

Chapter 3 (Statistical Shape Modeling) tackled the statistical variability of anatomical structures. We described how a standard principal component analysis approach helped build a statistical model of each target structure, which then served as a shape constraint for the deformable templates. This technique however required that the various structure's instances be reparameterized. That is, a dense correspondence field must be established *a priori* between them. There again, we introduced *a priori* medical expertise in the form of two original learning approaches to the n-D matching problem. These computed between instances a dense field bearing close resemblance to those in an *a priori* given learning set. A new local shape descriptor, the observed transport descriptor, was also developed following specifications designed to facilitate this reparameterization task, namely enhanced noise robustness and greater discriminating power.

Chapter 4 (Knowledge-driven Segmentation) detailed the segmentation system itself. We described the general deformable template framework, which used a combination of both implicit/semi-implicit constraints (texture, shape, distance) and explicit rules to achieve better performances. Results demonstrated the adaptability of our approach and the ability of the rule/meta-rule architecture to adequately react when facing difficult segmentations.

Chapter 5 (Piecewise Affine Registration) introduced a more faithful model of the transformations induced by histological processes, the piecewise affine approach, which allowed the design of a biomedical registration algorithm better suited to the reconstruction of 3-D histological volumes. Namely, our model consisted of a number of affine or rigid transformations applied to carefully delineated areas with non-linear transformations interpolated

in between. We designed a multiple-stepped approach (similarity map computation, hierarchical clustering, independent robust registration, erosion and hybrid affine/non-linear composition) with explicit (background removal, etc.) and implicit (EMD distribution distance, etc.) constraints and guidelines cast into each step. We also detailed a novel similarity measure, the constrained correlation coefficient, whose compromise between precision and robustness made it a good candidate for biomedical registration.

6.2 Registration as a unifying framework

In addition to the use of *a priori* medical expertise and knowledge, we propose registration as a general framework for our algorithms. We first introduce a few formal definitions of the related concepts before describing how the methods and methodologies we developed in this study can be cast into that framework.

6.2.1 Definitions

Registration is a process by which two objects (A and B) are put in correspondence. That is to say, by registering A to B, we associate elements of A *at a given scale* to element of B at the *same* scale.

As an example, when objects are 2-D images, elements can be pixels at a given resolution (the scale is then the image resolution). Note that they could also be entire regions (sets of pixels) at the same resolution.

Type. Objects come in a variety of types: they can be n-D images, surfaces, lists of remarkable points, etc. Registration can then be performed between objects of the same type (between two 2-D images for instance), or across types (between a triangular mesh and a 3-D voxel-based image for instance).

Correspondence. Depending on the requirements of the application at hand, a correspondence can be:

- surjective: for each element in A, there are many corresponding elements in B;
- injective: for each element in A, there is at most one corresponding element in B; or
- bijective: for each element in A, there is exactly one corresponding element in B, and vice versa.

It can also be symmetric (when A and B are interchangeable) or asymmetric.

Entity. To each object we also associate an entity, of which it is a representation, as types can be versatile (the same type can be used to represent two different objects). For instance, registering two images can amount to either putting in correspondence the representation of a *region* and that of a *model* (that is, finding in the first image a region similar in appearance to the second image: this is a segmentation and labeling process), or putting in correspondence the representation of two regions (which would be a classic intensity-based image registration process).

6.2.2 Unification

With this formalism, we can project most of the developed algorithms into a registration framework:

- Texture classification is a surjective correspondence between pixel sets (regions) in the input image and the various textural representations associated to the target structures.
- Reparameterizing two instances of a structure is putting in correspondence points on one with points on the other. For instance, in the 2-D discrete case, objects are the lists of pixels corresponding to the two instances, elements are the pixels themselves and the entities are both models of the structure.
- Segmenting internal brain structures corresponds to the example detailed above: putting in correspondence the representation of a region (the structure in the image to be segmented) and that of a model (the deformable template).
- Of course, both the registration of a reference MRI to the input image, and the piecewise registration of 2-D biomedical images are registration processes between objects of the same type (images) and same entity (region).

6.3 Future prospects

We present in this section directions of research that could be further investigated, in addition to those already mentioned in the conclusion section of each chapter.

6.3.1 Robustness

Algorithms can arguably never be too robust. Even though we attempted to design approaches as insensitive to noise and imaging characteristics as possible, there is still much room for improvement. In particular, most of the texture descriptors we described in Chapter 2 depend to some extent on the orientation of the patient's head in the MR scanner, or on the quality of the used interpolation and registration schemes. Enhanced robustness could be obtained by either using orientation independent descriptors or taking those imaging parameters into account in the classification schemes.

Similarly, additional segmentation rules and meta-rules could be developed to better segment structures when the input data correspond to pathological cases. For instance, we could envision a two-step application where an automated diagnosis system first decides whether the input image is typical of a given pathology (from a bank of available pathologies) and then a pathology-dependent segmentation system delineates the boundaries of the target structures using an adapted set of rules/meta-rules.

The use of better acquisition models could also greatly improve the robustness of our registration approach. As it is, the piecewise paradigm is only a coarse approximation of the true transformations induced by histological processes. It could be perfected if acquisition noise were taken into account, or if a fine model of the non-linear shrinkings yielded by the successive chemical baths were introduced.

Robustness is also a pivotal issue for the learning reparameterization algorithms we introduced in Chapter 3 as their performances depend on the effectiveness of the pattern matching algorithm in 1-D and on the non-linear registration method in n-D.

6.3.2 Other registration paradigms

In addition to the piecewise affine model we developed for the registration of histological sections (which could also serve for the registration of mono-modal mono-patient MRIs as illustrated in Figure 5.2), more models could be designed to facilitate the registration of other types of images. From the sulci models given by the so-called primal sketches of Cachia *et al.* [29] for instance, precise anatomical correspondences could be establish to guide the registration of 3-D multi-patient MR data.

6.3.3 Going full circle

Finally, while our segmentation system relies on various registration steps (to initialize the templates, to build the shape models, etc.) a precise segmentation of cerebral structures could conversely be used within a registration algorithm to increase its performances in a controlled way. For instance, the fine registration of the brain sulci does not matter as much in a Parkinsonian atlas of the basal ganglia as the precise correspondence between the central anatomical structures. Use of an *a priori* segmentation could then help locally enhance the precision in that specific area. Clearly, segmentation and registration are but two faces of the same coin.

Bibliography

[1] EC COST B11 Action. In *Eur. Co-op Sci. Technical Res.* available online at: http://www.uib.no/costb11.

[2] Y. Amit and A. Kong. Graphical Templates for Model Registration. *Research Report*, 1994.

[3] J. Annese, A. Pitiot, I.D. Dinov, and A.W. Toga. A myelo-architectonic method for the structural classification of cortical areas. *NeuroImage*, 2003.

[4] E. Ardizzone, D. Peri, R. Pirrone, A. Palma, and G. Peri. Knowledge based Approach to Intelligent Data Analysis of Medical Images. In *Proceedings of Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP'02)*, 2001.

[5] J.B. Arnold, J-S Liow, K.A. Schaper, J.J. Stern, J.G. Sled, D.W. Shattuck, A.J. Worth, M.S. Cohen, R.M. Leahy, J.C. Mazziotta, and D.A. Rottenberg. Qualitative and quantitative evaluation of size algorithms for correcting intensity nonuniformity effect. *NeuroImage*, 13:931–943, 2001.

[6] J. Ashburner and K. J. Friston. Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4):254–266, 1999.

[7] D. Atkinson, D. Hill, P. Stoyle, P. Summers, and S. Keevil. Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. *IEEE Trans. on Medical Imaging*, 16(6):903–910, 1997.

[8] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193, 1954.

[9] F. Attneave and M.D. Arnoult. The quantitative study of shape and pattern perception. *L. Uhr, editor, Pattern Recognition*, pages 123–141, 1966.

[10] E. Backer. *Computer-assisted reasoning in cluster analysis.* Prentice Hall, 1995.

[11] J. Barbosa and V. Barroso. Multiple and Simultaneous Image Compression for Video Sequences. In *Proceedings of Conferencia de Telecomunicacoes (conftele'01)*, 2001.

[12] V. Barra and J.Y. Boire. Automatic Segmentation of Subcortical Brain Structures in MR Images Using Information Fusion. *IEEE transactions on Medical Imaging*, 20(7):549–558, 2001.

[13] S. Belongie, J. Malik, and J. Puzicha. Shape Matching and Object Recognition Using Shape Context. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522, April 2002.

[14] R. Bernard, B. Lika, and F. Pernus. Segmenting Articulated Structures by Hierarchical Statistical Modeling of Shape, Appearance, and Topology. In *Proceedings of MICCAI (MICCAI'01)*, 2001.

[15] J. Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society*, 36:344–348, 1974.

[16] A. Blake and A. Zisserman. *Visual Reconstruction.* MIT Press, 1987.

[17] R.E. Blanton, J.L. Levitt, P.M. Thompson, L.F. Capetillo-Cunliffe, T. Sadoun, T. Williams, J.T. McCracken, and A.W. Toga. Mapping Cortical Variability and Complexity Patterns in the Developing Human Brain. *Psychiatry Research*, 107(1):29–43, 2001.

[18] D. Blostein and N. Ahuja. Shape from Texture: Integrating Texture-Element Extraction and Surface Estimation. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 11(12):1233–1251, 1989.

[19] T.O. Blum. *Transformation for Extracting New Descriptors of Shape. Models for the Perception of Speech and Visual Form.* MIT Press, 1967.

[20] F.L. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge University Press, 1991.

[21] G. Borgefors. Distance transformations in arbitrary dimensions. *Computer Vision, Graphics, and Image Processing*, 27:321–345, 1984.

[22] M. Brady. *Smoothed local symmetries and their implementation.* Academic Press, 1984.

[23] C. Brechbühler. *Description and Analysis of 3-D Shapes by Parameterization of Closed Surfaces.* Ph.d. thesis, IKT/BIWI, ETH ETH Zürich, 1995.

[24] M. Brejl and M. Sonka. Object Localization and Border Detection Criteria Design in Edge-Based Image Segmentation: Automated Learning from Examples. *IEEE TMI*, 19(10):973–985, 2000.

[25] B.H. Brinkmann, A. Manduca, and R.A. Robb. Optimized homomorphic unsharp masking for MR grayscale inhomogeneity correction. *IEEE transactions on Medical Imaging*, 17:161–171, 1998.

[26] G. Brown, G. Michon, and J. Peyrière. On the multifractal analysis of measures. *Journal of Statistical Physics*, 66:775–790, 1992.

[27] M.A. Brown and R.C. Semelka. *MRI: Basic Principles and Applications.* Wiley-Liss, 2003.

[28] M.S. Brown, L.S. Wilson, B.D. Doust, R.W. Gill, and C. Sun. Knowledge-based Method for Segmentation and Analysis of Lung Boundaries in Chest X-ray Images. *Computerized Medical Imaging and Graphics*, 22:463–477, 1998.

[29] A. Cachia, J.-F. Mangin, D. Rivière, F. Kherif, N. Boddaert, A. Andrade, D. Papadopoulos-Orfanos, J.-B. Poline, I. Bloch, M. Zilbovicius, P. Sonigo, F. Brunelle, and J. Régis. A Primal Sketch of the Cortex Mean Curvature: a Morphogenesis Based Approach to Study the Variability of the Folding Patterns. *IEEE Transactions on Medical Imaging*, 22(6):754–765, 2003.

[30] P. Cachier. *Recalage non rigide d'images médicales volumiques - contribution aux approches iconiques et géométriques.* PhD thesis, Ecole Centrale des Arts et Manufactures, 2002.

[31] P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache. Iconic Feature Based Nonrigid Registration: The PASHA Algorithm. *CVIU — Special Issue on Nonrigid Registration*, 2003. In Press.

[32] J.M. Chassery and C. Garbay. An Iterative Segmentation Method Based on a Contextual Color and Shape Criterion. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 6(6):794–800, 1984.

[33] C.H. Chen, L.F. Pau, and P.S.P Wang (eds.). *The Handbook of Pattern Recognition and Computer Vision (Chapter 2.1: Texture Analysis).* World Scientific Publishing Co., 1998.

[34] Gary E. Christensen. Consistent linear-elastic transformations for image matching. In Attila Kuba, Martin Samal, and Andrew E. Todd-Pokropek, editors, *Proc. Information Processing in Medical Imaging (IPMI'99)*, volume 1613 of *LNCS*, pages 224–237, Visegrád, Hungary, 1999. Springer.

[35] J.M. Coggins. *A Framework for Texture Analysis Based on Spatial Filtering, Ph.D. Thesis.* 1982.

[36] F.S. Cohen, Z. Yang, Z. Huang, and J. Nissanov. Automatic Matching of Homologous Histological Sections. *IEEE Transaction on Biomedical Engineering*, 45(5):642–649, 1998.

[37] I. Cohen, N. Ayache, and P. Sulget. Tracking Points on Deformable Objects using Curvature Information. In *Proc. of ECCV*, pages 458–466, 1992.

[38] M.S. Cohen, R.M. Dubois, and M.M. Zeineh. Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. *Human Brain Mapping*, 10:204–211, 2000.

[39] D. Cohen-Or, D. Levin, and A. Solomovici. Three-Dimenstional Distance Field Metamorphosis. *ACM transactions on Graphics*, 17(2):116–141, 1998.

[40] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality image registration using information theory. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Proc. Information Processing in Medical Imaging (IPMI'95)*, Computational imaging and vision, pages 263–274. Kluwer, 1995.

[41] D. L. Collins, A. C. Evans, C. Holmes, and T. M. Peters. Automatic 3D segmentation of neuro-anatomical structures from MRI. In Y. Bizais, C. Barillot, and R. Di Paola, editors, *Proc. Information Processing in Medical Imaging (IPMI'95)*, Computational imaging and vision, pages 139–152. Kluwer, 1995.

[42] D. L. Collins, A. P. Zijdenbos, T. Paus, and A. C. Evans. Use of registration for cohort studies. In Joseph Hajnal, David Hawkes, and Derek Hill, editors, *Medical Image Registration.* 2003.

[43] T.F. Cootes, D.H. Cooper, C.J. Taylor, and J. Graham. Trainable method of parametric shape description. *Image Vision Comput*, 10(5):–, 1992.

[44] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active Appearance Models. In *Proc. of ECCV*, pages 484–498, 1998.

[45] T.F. Cootes and C.J. Taylor. *Statistical Models of Appearance for Computer Vision.* Tutorial at MICCAI 2001, MICCAI 2001, 2001.

[46] Timothy F. Cootes, A. Hill, Christopher J. Taylor, and J. Haslam. Use of Active Shape Models for Locating Structures in Medical Images. *Image and Vision Computing*, 12(6):355–366, 1994.

[47] T.N. Cornsweet. *Visual Perception.* Academic Press, 1970.

[48] D. Costa. An evolutionary tabu search algorithm and the NHL scheduling problem. *INFOR*, 33(3):161–178, 1995.

[49] T.M. Cover and J.M. Van Campenhout. On the possible orderings in the measurement selection problem. *IEEE transactions on Systems Man and Cybernetics*, 7:657–661, 1977.

[50] O. Cuisenaire. *Distance Transformations: Fast Algorithms and Applications to Medical Image Processing.* Ph.d. thesis, 1999.

[51] J.G. Daugman. Uncertainty relation for resolution in space, spatial-frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of Optical Society of America*, 2:1160–1169, 1985.

[52] C. Davatzikos. Spatial transformation and registration of brain images using elastically deformable models. *Computer Vision and Image Understanding*, 66(2):207–222, 1997.

[53] C. Davatzikos, J. Prince, and N. Bryan. Image Registration Based on Boundary Mapping. *IEEE Transacstions on Medical Imaging*, 15(1):212–215, 1996.

[54] D. Davatzikos, M. Vaillant, S. Resnick, J.L. Prince, S. Letovsky, and R.N. Bryan. Computerized Method for Morphological Analysis of the Corpus Callosum. *Computer Assisted Tomography*, 20:88–97, 1998.

[55] D. L Davies and D. W. Bouldin. A cluster separation measure. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1:224–227, 1979.

[56] R.H. Davies, C.J. Twining, T.F. Cootes, J.C. Waterton, and C.J. Taylor. A Minimum Description Length Approach to Statistical Shape Modelling. *IEEE Trans. on Medical Imaging*, 21(5):525–537, 2002.

[57] G. Van de Wouwer, B. Weyn, P. Scheunders, E. Van Marck, and D. Van Dyck. Wavelets as chromatin texture descriptors for the automated identification of neoplastic nuclei. *Journal of microscopy*, 197:25–35, 2000.

[58] H. Delingette. General object reconstruction based on simplex meshes. *International Journal of Computer Vision*, 32(2):111–146, 1999.

[59] J. Dengler. Estimation of discontinuous displacement vector fields with the minimum description length criterion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'91)*, pages 276–282, Lahaina, Maui, Hawaii, 1991. IEEE Computer Society Press.

[60] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 9:39–55, 1987.

[61] M.H. Deverell, J.R. Salisbury, M.J. Cookson, J.G. Holman, E. Dykes, and F. Whimster. Three-dimensional reconstruction: methods of improving image registration and interpretation. *Analytical Cellular Pathology*, 5(5):253–263, 1993.

[62] R.O. Duda, E. Hart, and D.G. Stork. *Pattern Classification (second edition).* Wiley-Interscience, 2000.

[63] M. Egmont-Petersen, D. de Ridder, and H. Handels. Image processing with neural networks - a review. *Pattern Recognition*, 35(10):2279–2301, 2002.

[64] Alexandre X. Falcao, Jayaram K. Udupa, Supun Samarasekera, Shoba Sharma, Bruce Elliot Hirsch, and Roberto de A. Lotufo. User-Steered Image Segmentation Paradigms: Live Wire and Live Lane. *Graphical Models and Image Processing*, 60(4):233–260, 1998.

[65] E. Falkenauer. A hybrid grouping genetic algorithm for bin packing. *Journal of Heuristics*, 2(1):5–30, 1996.

[66] M. Fischler and R. Elschlager. The Representation and Matching of Pictorial Structures. *IEEE Transasctions On Computers*, 22(1):67–92, 1973.

[67] C. Fleurent and J.A. Ferland. Object-oriented implementation of heuristic search method for graph coloring, maximum clique, and satisfiability. In *Proceedings of DIMACS Implementation Challenge*, pages 619–652, 1996.

[68] M. Fleuté, S. Lavallée, and R. Julliard. Incorporating a Statistically Based Shape Model into a System for Computer-Assisted Anterior Cruciate Ligament Surgery. *Medical Image Analysis*, 3(3):209–222, 1999.

[69] T.S. Ford-Holevinski, M.R. Castle, J.P. Herman, and S.J. Watson. Microcomputer-based three-dimensional reconstruction of in situ hybridization autoradiographs. *J Chem Neuroanat*, 4(5):373–385, Sep-Oct 1991.

[70] B. Freisleben and P. Merz. New genetic local search operators for the traveling salesman problem. *Lectures Notes in Computer Science 1141*, pages 890–899, 1996.

[71] K.S. Fu. *Syntactic Pattern Recognition and Applications.* Prentice-Hall, 1982.

[72] A. Gagalowicz. *Vers un modèle de textures.* PhD thesis, Université Pierre et Marie Curie, 1983.

[73] J. C. Gee, M. Reivich, and R. Bajcsy. Elastically deforming 3D atlas to match anatomical brain images. *Journal of Computer Assisted Tomography*, 17(2):225–236, March/April 1993.

[74] G. Gerig, M. Jomier, and M. Chakos. Valmet: a new validation tool for assessing and improving 3D object segmentation. In *Proc of MICCAI*, pages 516–528, 2001.

[75] J. Giarratano and G. Riley. *Expert Systems, Principles and Programming.* PWS Publishing Company, 1993.

[76] J.J. Gibson. *The perception of the visual world.* Houghton, 1950.

[77] E. Goldmeier. Similarity in visually perceived forms. *Psychological Issues*, 8(1), 1972.

[78] A.F. Goldszal, O.J. Tretiak, P.J. Hand, S. Bhasin, and D.L. McEachron. Three-dimensional reconstruction of activated columns from 2-[^{14}C]deoxy-D-glucose data. *Neuroimage*, 2(1):9–20, March 1995.

[79] A.F. Goldszal, O.J. Tretiak, D.D. Liu, and P.J. Hand. Multimodality multidimensional image analysis of cortical and subcortical plasticity in the rat brain. *Ann Biomed Eng*, 24(3):430–439, May-June 1996.

[80] M. A. González-Ballester, A. Zisserman, and M. Brady. Estimation of the partial volume effect in MRI. *Medical Image Analysis*, 6(4):389–405, December 2002.

[81] C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statisical Society*, 53(2):285–339, 1991.

[82] U. Grenander, Y. Chow, and D. Keenan. *HANDS: A Pattern Theoretic Study of Biological Shapes.* Springer, 1991.

[83] U. Grenander and M.I. Miller. Representations of knowledge in complex systems. *Journal of the Royal Statistical Society*, 56(3):549–603, 1994.

[84] S. Growe and R. Tonjes. A Knowledge Based Approach to Automatic Image Registration. In *Proceedings of International Conference on Image Processing (ICIP'97)*, pages 228–231, 1997.

[85] S. Guarnieri, F. Piazza, and A. Uncini. Multilayer Feedforward Networks with Adaptive Spline Activation Function. *IEEE transactions on Neural Networks*, 10(3):672–683, 1999.

[86] M. Hagedoorn and R. Veltkamp. Metric pattern spaces. Technical Report UUCS-1999-03, Utrecht University, 1999.

[87] H.W. Hake. *From discrimination and the invariance of form.* L. Uhr, editor, Pattern Recognition: Theory, Experiments, Computer Simulations, and Dynamic Models of Form, Perception, and Discovery, 1966.

[88] S. Haker, S. Angenent, and A. Tannenbaum. Minimizing Flows for the Monge-Kantorovich Problem. *SIAM Journal of Mathematical Analysis*, 2003. to appear.

[89] L.O. Hall, A.M. Bensaid, and L.P. Clarke. A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain. *Neural Networks*, 3(5):672–682, 1992.

[90] Sean M. Haney, Paul M. Thompson, Timothy F. Cloughesy, Jeffrey R. Alger, and Arthur W. Toga. Tracking tumor growth rates in patients with malignant gliomas: A test of two algorithms. *American Journal of Neuroradiology*, 22(1):73–82, January 2001.

[91] R. Haralick. Statistical and structural approaches to textures. *Processings of the IEEE*, 67(5):786–804, 1979.

[92] J.K. Hawkins. Textural properties for pattern recognition. In B. Lipkin and A. Rosenfeld, editors, *Picture Processing and Psychopictorics.* 1969.

[93] D.O. Hebb. *The organization of behavior.* John Wiley, 1949.

[94] R. Hecht-Nielsen. Theory of the backpropagation neural network. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN '89)*, pages 589–605, 1989.

[95] L.S. Hibbard and R.A. Hawkins. Objective image alignment for three-dimensional reconstruction of digital autoradiograms. *J Neurosci Methods*, 26(1):55–74, November 1988.

[96] Y.-C. Ho and R.L. Kashyap. Algorithm for Linear Inequalities and its Applications. *IEEE Transactions on Electronic Computers*, 14:683–688, 1965.

[97] L.T. Holly and K.T. Foley. Intraoperative spinal navigation. *Spine*, 28(15):554–561, 2003.

[98] B. Horn. Height and gradient from shading. *International Journal of Computer Vision*, 5(1):37–75, 1990.

[99] Y.S. Huang, K. Liu, and C.Y. Suen. The Combination of Multiple Classifiers by a Neural Network Approach. *International Journal of Pattern Recognition and Artificial Intelligence*, 9:579–597, 1995.

[100] J. Hug, C. Brechbuhler, and G. Székely. Model-based Intitialisation for Segmentation. In *Proceedings of European Conference on Computer Vision (ECCV'00)*, 2000.

[101] J.L. Humm, R.M. Macklis, X.Q. Lu, Y. Yang, K. Bump, B. Beresford, and L.M. Chin. The spatial accuracy of cellular dose estimates obtained from 3D reconstructed serial tissue autoradiographs. *Phys Med Biol*, 40(1):163–180, January 1995.

[102] E.G. Huot, H.M. Yahia, I. Cohen, and I. Herlin. Surface Matching with Large Deformations and Arbitrary Topology: A Geodesic Distance Evolution Scheme on a 3-Manifold. In *Proc. of ECCV*, pages 769–783, 2000.

[103] A. K. Jain. Image data compression: a review. *Proceedings of the IEEE*, 69(3):349–389, March 1981.

[104] A.K. Jain and B. Chandrasekaran. Dimensionality and sample size considerations. In P.R. Krishnaiah and L.N. Kanal, editors, *Pattern Recognition Practice*, volume 2, pages 835–855. 1982.

[105] A.K. Jain and F. Farrokhnia. Unsupervized Texture Segmentation Using Gabor Filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[106] A.K. Jain and D. Zongker. Feature Selection: Evaluation, Application, and Small Sample Performence. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, 1997.

[107] S. C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 2:241–254, 1967.

[108] F. A. Jolesz, A. Nabavi, and R. Kikinis. Integration of interventional MRI with computer-assisted surgery. *Journal of Magnetic Resonance Imaging*, 13(1):69–77, January 2001.

[109] M.I. Jordan and R.A. Jacobs. Hierarchical Mixtures of Experts and the EM Algorithm. *Neural Computation*, 6:181–214, 1994.

[110] S. Joshi, M. Miller, and U. Grenander. On the Geometry and Shape of Brain Sub-Manifolds. *Pattern REcognition and Artificial Intelligence*, 11:1317–1343, 1997.

[111] S. Joshi, S. Pizer, P. Thomas Fletcher, A. Thall, and G. Tracton. Multi-scale 3-D Deformable Model Segmentation Based on Medial Description. In *Proceedings of IPMI (IPMI'01)*, pages 64–77, 2001.

[112] B. Kamgar-Parsi and B. Kamgar-Parsi. Rejection with Multiplayer Neural Networks: Automatic Generation of the Training Set. In *Proceedings of the World Congress on Neural Networks (WCNN'95)*, pages 174–177, 1995.

[113] Takashi Kanai, Hiromasa Suzuki, and Fumihiko Kimura. Metamorphosis of Arbitrary Triangular Meshes. *IEEE Computer Graphics and Applications*, 20(2):62–75, 2000.

[114] P.A. Kay, R.A. Robb, D.G. Bostwick, and J.J. Camp. Robust 3-D Reconstruction and Analysis of Microstructures from Serial Histologic Sections, with Emphasis on Microvessels in Prostate Cancer. In K. H. Höhne and R. Kikinis, editors, *Visualisation in Biomedical Computing*, volume 1131 of *Lecture Notes in Computer Science*, pages 129–134, Hamburg (Germany), 1996. Springer.

[115] A. Kelemen, G. Szekely, and G. Gerig. Three-Dimensional Model-based Segmentation of Brain MRI. *IEEE Trans. on Medical Imaging*, 18(10):838–849, 1999.

[116] D. Kendall. Shape Manifolds, Procrustean Metrics and Complex Projective Spaces. *Bulletin of the London Mathematical Society*, 16:81–121, 1984.

[117] B. Kim, K.A. Frey, S. Mukhopadhyay, B.D. Ross, and C.R. Meyer. Co-Registration of MRI and Autoradiography of Rat Brain in Three-Dimensions Following Automatic Reconstruction of 2D Data Set. In N. Ayache, editor, *Computer Vision, Virtual Reality and Robotics in Medicine*, volume 905 of *Lecture Notes in Computer Science*, pages 262–266, Nice (France), 1995. Springer.

[118] Ulrik Kjems, L. K. Hansen, and C. T. Chen. A Non-linear 3-D brain co-registration method. In P. C. Hansen, editor, *Proceedings of the Interdisciplinary Inversion Workshop 4*, Lyngby, Denmark, 1996. IMM, Technical University of Denmark.

[119] J. Koenderink and A. Van Doorn. Dynamic shape. *Biological Cybernetics*, 53:383–396, 1986.

[120] K. Koffka. *Principles of Gestalt psychology.* Harcourt Brace Jovanovic, 1935.

[121] V.A. Kovalev, F. Kruggel, H.-J. Gertz, and D. Yves von Cramon. Three-Dimensional Texture Analysis of MRI Brain Datasets. *IEEE transactions on Medical Imaging*, 20(5):424–433, 2001.

[122] R. K. Kwan, A.C. Evans, and G.B. Pike. An Extensible MRI Simulator for Post-Processing Evaluation. In *Proceedings of the International Conference in Visualization in Biomedical Computing (VBC'96)*, pages 135–140, 1996.

[123] K.I. Laws. *Textured Image Segmentation.* Ph.d. thesis, University of Southern California, 1980.

[124] M. Leventon, E. Grimson, and O. Faugeras. Statistical Shape Influence in Geodesic Active Contours. In *Proceedings of Computer Vision and Pattern Recognition (CVPR'00)*, pages 4–11, 2000.

[125] M. Leyton. Symmetry-curvature duality. *Computer Vision, Graphics, and Image Processing*, 38:327–341, 1987.

[126] M. Leyton. Inferring causal history from shape. *Cognitive Science*, 13:357–387, 1989.

[127] H. Li, R. Deklerck, B. De Cuyper, A. Hermanus, E. Nyssen, and J. Cornelis. Object Recognition in Brain CT-Scans: Knowledge-Based Fusion of Data from Multiple Feature Extractors. *IEEE transactions on Medical Imaging*, 14(2):212–229, 1995.

[128] J.M. Links, L.S. Beach, B. Subramana, M.A. Rubin, J.G. Hennessey, and A.L Reiss. Edge complexity and partial volume effects. *Journal of Computer Assisted Tomography*, 22(3):450–458, 1998.

[129] P. Lipson, A.L. Yuille, D. O'RKeefe, J. Cavanaugh, J. Taaffe, and D. Rosenthal. Deformable Templates for Feature Extraction from Medical Images. In *Proceedings of First European Conference on Computer Vision (ECCV'90)*, pages 477–484, 1990.

[130] J. A. Little, D. L. G. Hill, and D. J. Hawkes. Deformations incorporating rigid structures. *Computer Vision and Image Understanding*, 66(2):223–232, 1997.

[131] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001, 1998.

[132] D.G. Lowe. *Perceptual Organization and Visual Recognition.* Kluwer Academic, 1985.

[133] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, March 1998.

[134] G. Malandain and E. Bardinet. Fusion of autoradiographies with an MR volume using 2-D and 3-D linear transformations. In *Proceedings of Information Processing in Medical Imaging (IPMI'03)*, pages 487–498, 2003.

[135] J. Mao, K. Mohiuddin, and A.K. Jain. Minimal Network Design and Feature Selection through Node Pruning. In *Proceedings of International Conference on Pattern Recognition (ICPR'94)*, pages 622–624, 1994.

[136] D. Marr. A theory for cerebral neocortex. In *Proceedings of the Royal Society of London*, pages 161–234, 1970.

[137] D. Marr. Early processing of visual information. In *Proceedings of the Royal Society of London*, pages 483–519, 1976.

[138] D. Marr and E. Hildreth. Theory of edge detection. In *Proceedings of the Royal Society of London*, pages 187–217, 1980.

[139] D. Marr and H.K. Nishihara. Visual Information Processing: Artificial Intelligence and the Sensorium of Sight. *Technology Review*, 81(1):2–23, 1978.

[140] D. Marr and T. Poggio. A computational theory of human stereo vision. In *Proceedings of the Royal Society of London*, pages 301–328, 1979.

[141] A. Materka and M. Strzelecki. Texture Analysis Methods - A Review. *Technical University of Lodz, COST B11 report*, 1998.

[142] M. Matesin, S. Loncaric, and D. Petravic. A Rule-Based Approach to Stroke Lesion Analysis from CT Brain Images. In *Proc. of Second International Symposium on Image and Signal Processing and Analysis*, pages 219–223, 2001.

[143] C.R. Maurer, R. Qi, and V. Raghavan. A Linear Time Algorithm for Computing Exact Euclidean Distance Transforms of Binary Images in Arbitrary Dimensions. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 25(2):265–270, 2002.

[144] T. McInerney and D. Terzopoulos. Deformable Models in Medical Image Analysis: A Survey. *Medical Image Analysis*, 1(2):91–108, 1996.

[145] D. Metaxas and D. Terzopoulos. Shape and Nonrigid Motion Estimation through Physics-Based Synthesis. *IEEE transactions On Pattern Analysis and Machine Intelligence*, 15(6):580–591, 1993.

[146] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, 1999.

[147] J. Montagnat and H. Delingette. Globally constrained deformable models for 3D object reconstruction. *Signal Processing*, 71(2):173–186, 1998.

[148] J. Montagnat, H. Delingette, and N. Ayache. A review of deformable surfaces: topology, geometry and deformation. *Image and Vision Computing*, 19:1023–1040, 2001.

[149] H. Muehlenbein, M. Gorges-Schleuter, and O. Kraemer. Evolution algorithms in combinatorial optimization. *Parallel Computing*, 7:65–88, 1998.

[150] S.K. Nadkarni, D.R. Boughner, M. Drangova, and A. Fenster. Three-dimensional echocardiography: assessment of inter- and intra-operator variability and accuracy in the measurement of left ventricular cavity volume and myocardial mass. *Phys. Med. Biol.*, 45(5):1255–73, 2000.

[151] K.L. Narr, P.M. Thompson, T. Sharma, J. Moussai, A.F. Cannestra, and A.W. Toga. Mapping Morphology of the Corpus Callosum in Schizophrenia. *Cerebral Cortex*, 10(1):40–49, 2000.

[152] A. Obenaus, C.J. Yong-Hing, K.A. Tong, and G.E. Sarty. A Reliable Method for Measurement and Normalization of Pediatric Hippocampal Volumes. *Pediatric Research*, 50:124–132, 2001.

[153] P.P. Ohanian and R.C. Dubes. Performance evaluation for four classes of texture features. *Pattern Recognition*, 25(8):819–833, 1992.

[154] B. O'Sullivan and J. Shah. New TNM staging criteria for head and neck tumors. *Semin. Surf. Oncol.*, 21(1):30–42, 2003.

[155] S. Ourselin. *Recalage d'images médicales par appariement de régions - Application à la construction d'atlas histologiques 3D*. Ph.d. thesis, Université de Nice Sophia-Antipolis, 2002.

[156] S. Ourselin, A. Roche, G. Subsol, X. Pennec, and N. Ayache. Reconstructing a 3D Structure from Serial Histological Sections. *Image and Vision Computing*, 19(1-2):25–31, January 2001.

[157] W. Patola and B. Coulter. *MRI artifacts.* 1997. (http://www1.stpaulshosp.bc.ca/stpaulsstuff/MRartifacts.html).

[158] X. Pennec. Multiple registration and mean rigid shape - Application to the 3D case. In K.V. Mardia, C.A. Gill, and Dryden I.L., editors, *Image Fusion and Shape Variability Techniques (16th Leeds Annual Statistical Workshop)*, pages 178–185. University of Leeds, UK, July 1996.

[159] A. Pitiot, A.W. Toga, and P.M. Thompson. Adaptive Elastic Segmentation of Brain MRI via Shape-Model-Guided Evolutionary Programming. *IEEE TMI*, 21(8):910–923, 2002.

[160] S.M. Pizer, D.S. Fritsch, P. Yushkevich, V. Johnson, and E.L. Chaney. Segmentation, Registration, and Measurement of Shape Variation via Image Object Shape. *IEEE TMI*, 10(18):851–865, 1999.

[161] Plato. *Meno (380 BC) in W.Lamb (ed.), Plato in Twelve Volumes, vol. 3.* Harvard University Press, 1967.

[162] T. Poggio and V. Torra. Ill-posed Problems and Regularization Analysis in Early Vision. In *Proceedings of AARPA Image Understanding Workshop*, pages 257–263, 1984.

[163] M.I. Posner. *Foundations of Cognitive Science.* MIT Press, 1989.

[164] F. Poupon, J.-F. Mangin, V. Frouin, and Industrial Metrology. 3D Multi-object Deformable Templates Based on Moment Invariants. In M. Frydrych, J. Parkkinen, and A. Visa, editors, *10th Scandinavian Conference on Image Analysis*, pages 149–156. Pattern Recognition Society of Finland, June 1999.

[165] R.J. Prokop and A.P. Reeves. A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP: Graphics Models and Image Processing*, 54(5):483–460, 1992.

[166] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature-selection. *PRL*, 15(11):1119–1125, November 1994.

[167] Anand Rangarajan, Haili Chui, Eric Mjolsness, Suguna Pappu, Lila Davachi, Patricia Goldman-Rakic, and James Duncan. A robust point-matching algorithm for autoradiograph alignment. *Medical Image Analysis*, 1(4):379–398, 1997.

[168] R. A. Rauch and J. R. Jinkins. Variability of corpus callosal area measurements from midsagittal MR images: effect of subject placement within the scanner. *American Journal of Neuroradiology*, 17(1):27–28, 1996.

[169] N. Raynaud. *Approche Statistique pour la Segmentation d'Images Tridimensionelles du Foie.* Rapport de DEA, Ecole Normale Supérieure de Cachan, 2000.

[170] D. Rey, G. Subsol, H. Delingette, and N. Ayache. Automatic Detection and Segmentation of Evolving Processes in 3D Medical Images: Application to Multiple Sclerosis. *Medical Image Analysis*, 6(2):163–179, June 2002.

[171] W. Richards and A. Polit. Texture matching. *Kybernetic*, 16:155–162, 1974.

[172] A. Roche, G. Malandain, and N. Ayache. Unifying Maximum Likelihood Approaches in Medical Image Registration. *International Journal of Imaging Systems and Technology: Special Issue on 3D Imaging*, 11(1):71–80, 2000.

[173] F. Rosenblatt. The Perceptron: a Perceiving and Recognizing Automaton. *Report 85-460-1, Project PARA, Cornell Aeronautical Laboratory, New York*, 1957.

[174] M.A. Roula, A. Bouridane, and P. Miller. Multispectral Analysis of Chromatin Texture for Automatic Grading of Prostatic Neoplasia. In *Proceedings of Medical Image Understanding and Analysis (MIUA'02)*, 2002.

[175] P.J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 79:871–880, 1984.

[176] H.A. Rowley, S. Baluja, and T. Kanade. Neural Network-Based Face Detection. *IEEE transactions on Pattern Analysis and Maching Intelligence*, 20(1):23–38, 1998.

[177] S. Ruan, C. Jaggi, J. Xue, J. Fadili, and D. Bloyet. Brain Tissue Classification of Magnetic Resonance Images Using Partial Volume Modeling. *IEEE transactions on Medical Imaging*, 19(12):1179–1187, 2000.

[178] Y. Rubner, C. Tomasi, and L. Guibas. A metric for distributions with applications to image databases. In *Proc. of International Conference on Computer Vision (ICCV'98)*, Bombay, India, 1998. IEEE Computer Society, Narosa.

[179] M. Rydmark, T. Jansson, C.H. Berthold, and T. Gustavsson. Computer assisted realignment of light micrograph images from consecutive sections series of cat cerebral cortex. *Journal of Microscopy*, 165:29–47, 1992.

[180] N. Sarkar and B.B. Chaudhuri. An efficient approach to estimate fractal dimension of textural images. *Pattern Recognition*, 25(9):1035–1041, 1992.

[181] B. Scholkopf, C. Burges, and A. Smola. *Advances in Kernel Methods: Support Vector Learning*. MIT Press, 1999.

[182] T.B. Sebastian, J.J. Crisco, P.N. Klein, and B.B. Kimia. Constructing 2D Curve Atlases. In *Proc. of CVPR*, pages 70–77, 2000.

[183] M.V. Senat, J.P. Bernard, M. Boulvain, and Y. Ville. Intra- and interoperator variability in fetal nasal bone assessment at 11-14 weeks of gestation. *Ultrasound in Obstetric and Gynecology*, 22(2):138–41, 2002.

[184] A.C. Shortall, E. Harrington, H.B. Patel, and P.J. Lumley. A pilot investigation of operator variability during intra-oral light curing. *Br. Dent. J.*, 193(5):276–280, 2002.

[185] W. Siedlecki and J. Sklansky. On automatic feature selection. *International Journal of Pattern Recognition and Artificial Intelligence*, 2(2):197–200, 1988.

[186] A. Singh and P. Allen. Image-flow computation: an estimation-theoretic framework and a unified perspective. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 56(2):152–177, 1992.

[187] J. Sklansky. Image Segmentation and Feature Extraction. *IEEE transactions on Systems, Man, and Cybernetics*, 8:237–247, 1978.

[188] J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on Medical Imaging*, 17(1):87–97, 1998.

[189] L.H. Staib and J.S. Duncan. Boundary Finding with Parametrically Deformable Models. *IEEE Trans. on PAMI*, 14(11):1061–1075, 1992.

[190] A. Stocker, O. Sipila, A. Visa, O. Salonen, and T. Katila. Stability Study of Some Neural Networks Applied to Tissue Characterization of Brain Magnetic Resonance Images. In *Proceedings of the 13th Int. Conference on Pattern Recognition*, pages 472–477, 1996.

[191] C. Studholme, D. L. G. Hill, and D. J. Hawkes. Incorporating connected region labelling into automated image registration using mutual information. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA'96)*, pages 23–31. IEEE Computer Society Press, 1996.

[192] M. Styner and G. Gerig. Medial Models Incorporating Object Variability for 3-D Shape Analysis. In *Proceedings of IPMI (IPMI'01)*, pages 502–516, 2001.

[193] G. Subsol. A Scheme for Automatically Building 3D Morphometric Anatomical Atlases Based on Feature Lines: A List of References. *DIKU International Summer School'98 Shape Variation*, 1998.

[194] G. Székely, A. Kelemen, C. Brechbühler, and G. Gerig. Segmentation of 2-D and 3-D Objects from MRI Volum Data using Constrained Elastic Deformations of Flexible Fourier Surface Models. *Medical Image Analysis*, 1(1):–, 1996.

[195] G. Taubin. An improved algorithm for algebraic curve and surface fitting. In *Proc. of International Conference on Computer Vision (ICCV'93)*, pages 658–665, Berlin, Germany, May 1993. IEEE Computer Society Press.

[196] P. M. Thompson, R. P. Woods, M. S. Mega, and A. W. Toga. Mathematical/computational challenges in creating deformable and probabilistic atlases of the human brain. *Human Brain Mapping*, 9(2):81–92, 2000.

[197] P.M. Thompson, J. Moussai, S. Zohoori, A. Goldkorn, A.A Khan, M.S. Mega, G.W. Small, J.L. Cummings, and A.W. Toga. Cortical Variability and Asymmetry in Normal Aging and Alzheimer's Disease. *Cerebral Cortex*, 8(6):492–509, 1998.

[198] P.M. Thompson, K.L. Narr, R.E. Blanton, and A.W. Toga. Mapping Structural Alterations of the Corpus Callosum during Brain Development and Degeneration. In *Proceedings of the NATO ASI on the Corpus Callosum*. Kluwer Academic Press, 1997.

[199] P.M. Thompson and A.W. Toga. Detection, Visualisation and Animation of Abnormal Anatomic Structure with a Deformable Probabilistic Brain Atlas Based on Random Vector Field Transformations. *Medical Image Analysis*, 1(4):271–294, 1997.

[200] F. Tomita and S. Tsuji. *Computer Analysis of Visual Textures*. Kluwer Academic Publishers, 1990.

[201] T.Randen and J.H. Husoy. Filtering for Texture Classification: a Comparative Study. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 21(4):291–310, 1999.

[202] Alain Trouvé and Laurent Younes. Diffeomorphic Matching Problems in One Dimension: Designing and Minimizing Matching Functionals. In *Proc. of ECCV*, pages 573–587, 2000.

[203] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[204] F. Ulupinar and R. Nevatia. Inferring shape from contour for curved surfaces. In *Proceedings of the International Conference on Pattern Recognition (ICPR'90)*, pages 147–154, 1990.

[205] P.A. van den Elsen, E.J.D. Pol, and M.A. Viergever. Medical image matching - a review with classification. *IEEE transactions on Engineering in Medicine and Biology*, 12(4):26–39, 1993.

[206] M.M. van Hulle and T. Tollenaere. A modular artificial neural network for texture processing. *Neural Networks*, 6(1):7–32, 1993.

[207] P.J. van Otterloo. *A Contour-Oriented Approach to Shape Analysis*. Prentice Hall, 1992.

[208] J. Lévy Véhe. Introduction to the Multifractal Analysis of Images. In Y. Fisher, editor, *Fractal Image Encoding and Analysis*. 1996.

[209] R. C. Veltkamp. Shape matching: similarity measures and algorithms. In *Proceedings of the International Conference on Shape Modeling and Applications*, pages 188–199, 1991.

[210] R.C. Veltkamp and M. Hagedoorn. State of the Art in Shape Matching. *Technical Report UU-CS-1999-27*, 1999.

[211] B.C. Vemuri, A. Radisavljevic, and C. Leonard. Multiresolution 3-D Stochastic Hybrid Shape Models for Image Segmentation. In *Proceedings of Information Processing in Medical Imaging (IPMI'93)*, pages 62–76, 1993.

[212] H. Voorhees and T. Poggio. Detecting textons and texture boundaries in natural images. In *Proceedings of the International Conference on Computer Vision (ICCV '87)*, pages 250–258, 1987.

[213] P. Walter, I. Elsen, H. Muller, and K.-F. Kraiss. 3D Object Recognition with a Specialized Mixtures of Experts Architecture. In *Proceedings of the International Join Conference on Neural Networks (IJCNN'99)*, pages 3563–3568, 1999.

[214] Y. Wang, T. Adali, and S.Y. Kung. Quantification and segmentation of brain tissues from MR images - a probabilistic neural network approach. *IEEE transactions on Image Processing*, 7(8):1165–1181, 1998.

[215] Y. Wang, B.S. Peterson, and L.H. Staib. Shape-Based 3D Surface Correspondence using Geodesics and Local Geometry. In *Proc. of CVPR*, pages 644–651, 2000.

[216] J. Weber and J. Malik. Rigid Body Segmentation and Shape Description from Dense Optical Flow Under Weak Perspective. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 19(2):139–143, Feb 1997.

[217] W. M. Wells, P. Viola, H. Atsumi, and S. Nakajima. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.

[218] H.H. Wen, W.C. Lin, and C.T. Chen. Knowledge-Based Medical Image Registration. In *Proceedings of the IEEE Engineering in Medicine and Biology Society*, pages 1200–1201, 1996.

[219] P.J. Werbos. *The roots of backpropagation: from ordered derivatives to neural networks and political forecasting.* Wiley, 1994.

[220] B. Weyn, G. Van de Wouwer, M. Koprowski, A. Van Daele, K. Dhaene, P. Scheunders, W. Jacob, and E. Van Marck. Value of morphometry, texture analysis, densitometry, and histometry in the differential diagnosis and prognosis of malignant mesothelioma. *Journal of pathology*, 189:581–589, 1999.

[221] R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, and J. C. Mazziotta. Automated image registration: I. general methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, 22(1):141–154, 1998.

[222] A.J. Worth and D.N. Kennedy. Segmentation of magnetic resonance brain images using analogue constraint satisfaction neural networks. *Image Vision Computing*, 12(6):345–354, 1994.

[223] L. Xu, A. Krzyzak, and C.Y. Suen. Associative Switch for Combining Multiple Classifiers. *Journal of Artificial Neural Networks*, 1:77–100, 1994.

[224] P.P.C. Yip and Y.H. Pao. Combinational optimization with use of guided evolutionary simulated annealing. *IEEE transactions on Neural Networks*, 6:290–295, 1995.

[225] A.L Yuille, P.W. Hallinan, and D.S. Cohen. Feature Extraction from Faces using Deformable Templates. *International Journal of Computer Vision*, 8:99–111, 1992.

[226] W. Zhao, T.Y. Young, and M.D. Ginsberg. Registration and Three-Dimensional Reconstruction of Autoradiographic Images by the Disparity Analysis Method. *IEEE transactions on Medical Imaging*, 12(4):782–791, 1993.

[227] D. Ziou and S. Tabbone. Edge Detection Techniques - An Overview. *International Journal of Pattern Recognition and Image Analysis*, 8(4):537–559, 1998.

[228] R. Zoroofi, Y. Sato, S. Tamura, H. Naito, and L. Tang. An improved method for MRI artifact correction due to translation motion in the imaging plane. *IEEE Trans. on Medical Imaging*, 14(3):471–479, 1995.

[229] K.H. Zou, W.M. Wells III, M. Kaus, R. Kikinis, F.A. Jolesz, and S.K. Warfield. Statistical Validation of Automated Probabilistic Segmentation against Composite Latent Expert Ground Truth in MR Imaging of Brain Tumors. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI'02)*, pages 315–322, 2002.

[230] S.W. Zucker. Toward a model of texture. *Computer Graphics and Image Processing*, 5:190–202, 1976.

[231] S.W. Zucker and K. Kant. Multiple-level Representations for Texture Discrimination. In *Proceedings of the IEEE Conference on Pattern Recognition and Image Processing*, pages 609–614, 1981.

[232] L. Zusne. *Visual Perception of Form.* Academic Press, 1970.

Appendix A

Shape in Plato

Below is an excerpt from the dialogue between Socrates and Meno in Plato's Meno [161], in which Socrates attempts the explain the concept of shape to a particularly incredulous listener.

Socrates. Well, I will try and explain to you what figure is. What do you say to this answer? - *Figure is the only thing which always follows colour*. Will you be satisfied with it, as I am sure that I should be, if you would let me have a similar definition of virtue?

Meno. But, Socrates, it is such a simple answer.

Soc. Why simple?

Men. Because, according to you, figure is that which always follows colour.

(Soc. Granted.)

Men. But if a person were to say that he does not know what colour is, any more than what figure is-what sort of answer would you have given him?

Soc. I should have told him the truth.[...]You will acknowledge, will you not, that there is such a thing as an end, or termination, or extremity?-all which words use in the same sense, although I am aware that Prodicus might draw distinctions about them: but still you, I am sure, would speak of a thing as ended or terminated-that is all which I am saying-not anything very difficult.

Men. Yes, I should; and I believe that I understand your meaning.

Soc. And you would speak of a surface and also of a solid, as for example in geometry.

Men. Yes.

Soc.: Well then, you are now in a condition to understand my definition of figure. I define *figure to be that in which the solid ends; or, more concisely, the limit of solid.*

Appendix B

Delineation protocols

We report here, *verbatim*, the various delineations protocols associated to the selected target structures, which serve as delineation guidelines to the human operators.

B.1 Ventricles

- Starting at the most anterior slice where the lateral ventricles are visible, the gray-CSF and white-CSF tissue boundaries of the lateral ventricles are followed with a mouse-driven cursor. The signal intensity of T1-weighted scans may be enhanced, although it should be consistent between subjects and preferably set to a standard. The lateral ventricles are divided into superior, posterior and inferior horns and should be contoured on coronal brain slices while referencing orthogonal planes to clarify neuroanatomical boundaries if necessary. CSF/tissue interfaces are digitized separately and include the superior; lateral, and medial CSF/tissue interfaces for the superior horn. Inferior and posterior horn CSF/tissue surfaces include only lateral and medial surfaces.

- The direction for tracing the lateral ventricle boundaries must be consistent. That is, each segment should start at the same neuroanatomical point for each brain scan. For the superior horn, the superior segment should start medially in each hemisphere and traced laterally. The superior horn medial segment should start at the same point as the superior segment and be traced ventrally on coronal sections. The lateral segment of the superior horn should be traced beginning at the termination point of the medial segment. The medial and lateral segments of the posterior horn should continue in the directions of the medial and lateral segments of the superior horn. For the inferior horn, superior and inferior segments are traced from medially to laterally.

- The choroid plexus is excluded in lateral ventricle traces even if there is no visible separation between choroid plexus and the floor of the ventricle.

B.2 Corpus Callosum

The corpora callosa are traced in the midsagittal section of each MRI. These are first brought into register by a 10 point piecewise affine registration and verified by confirming the presence of the falx cerebri, septum pellucidum, and the vertical orientation of the interhemispheric fissure in all three planes. The corpus callosum is traced in each magnified (4×) brain volume by following white matter tissue boundaries with a mouse-driven cursor. The used segmentation software allowed voxel locations to be recorded at a quarter of a voxel spatial resolution in all three planes.

In the midsagittal plane, the presence of the falx cerebri and septum pellucidum may blur tissue boundaries of the corpus callosum. Therefore, to obtain the most accurate possible measure of midsagittal callosal area, the three most medial brain slices were included (1 mm slice thickness). In addition, three lateral slices on each side were also included.

B.3 Hippocampus and amigdala

The hippocampus is traced in the oblique coronal viewing plane. Images are first reoriented to the long axis of the hippocampus by selecting the most anterior and posterior limits of the hippocampus and reorienting the images such that the anterior and posterior limits of the hippocampus are in parallel along the long axis. This has been previously reported to promote the clear and easy identification of hippocampal anatomy.

Once images are reformatted, the hippocampus is subdivided and traced as anterior and posterior segments in successive oblique coronal slices proceeding in a posterior to anterior manner.

- The posterior hippocampus is first traced on the slice in which the crus of the fornix can be delineated. As the fornix is the major efferent pathway of the hippocampus this has proved to be a reliable landmark that is based on hippocampal anatomy rather than structures unrelated to the hippocampus.

- Following the identification of the crus of the fornix, the hippocampus is traced using the alveus as the superior boundary and the white matter of the parahippocampal gyrus as the inferior boundary. The inferior temporal horn of the lateral ventricle is used as the lateral boundary and the ambient cistern the medial boundary.

- The hippocampus tracings include the head of the hippocampus and the subiculum. The shape of the hippocampus in posterior sections is one of a ball of gray matter that comprises the head and a straight segment of gray matter that is the subiculum. As one moves in an anterior direction, the hippocampus begins to fold over on to itself and is the level at which you see the beginning of the hippocampal sulcus.

- The last section in which to trace the posterior hippocampus is at the level that the crus cerebri is still connected to the pons.

- The next slice is the first slice to start tracing the anterior hippocampus. The anterior hippocampus is traced using the same boundaries as the posterior hippocampus. The difference is that the amygdala, the other mesial temporal lobe anatomic structure, generally is observable at this level and is situated superior or above the hippocampus.

- In beginning tracings of the anterior hippocampus the inferior temporal horn of the lateral ventricle clearly separates the hippocampus from the amygdala. However, as you move more anteriorly, the inferior temporal horn moves from a superior position to a more lateral position.
- The alveus is then used to separate hippocampus from amygdala. In some cases the amygdala and hippocampus cannot be separated from each other using the temporal horn of the lateral ventricle and in these circumstances the alveus is also used as the boundary. If the alveus cannot be delineated then a straight line is traced from the middle of the inferior temporal horn of the ventricle to the ambient cistern.
- In the most anterior sections, the anterior hippocampus is termed the pes hippocampi and generally does not have the folded shape that is characteristic in earlier slices. The hippocampus also becomes incrementally smaller in size and tends to move medially along with the temporal horn of the lateral ventricle.
- The last section in which to trace the anterior hippocampus is the slice that the inferior temporal horn has moved completely from a lateral position to a medial one and has also moved completely beneath or under the amygdala.

B.4 Caudate nucleus

The caudate nucleus is traced in successive 1 mm slices in the coronal plane. The caudate is subdivided into three defined subregions: the body, head, and caudate/putamen complex. Anatomical anterior and posterior boundaries for each sub-region of the caudate are traced as follows:

Body:

- anterior - slice at which one can differentiate the foramen of monroe
- posterior- slice where the junction between the left and right thalamus is first visible

Head:

- anterior- slice at which one can last distinguish the internal capsule
- posterior- slice anterior to the one in which the foramen of monroe can be distinguished

Caudate/Putamen Complex:

- anterior - slice where the caudate/putamen complex is no longer visible
- posterior- the slice anterior to the one in which you can last distinguish the internal capsule

Résumé

Nous proposons avec cette thèse un système de segmentation et de recalage automatique pour les images cérébrales. Nous avons mis l'accent sur la conception d'une méthodologie qui s'appuie au maximum sur l'expertise médicale a priori.

Nous appréhendons le problème de la recherche des contours de structures cibles sous l'angle du processus d'appariement d'un groupe de patrons déformables aux frontières des structures. Ces patrons évoluent en parallèle, sous la supervision d'un ensemble de règles dérivées à la fois de l'analyse de la dynamique des patrons et de l'expertise médicale. Ils sont soumis à une variété de contraintes, conçues à partir d'informations a priori sur la texture et la forme des structures.

L'information texturale est extraite par un classificateur non-linéaire qui prend la forme d'un réseau de neurones à 2 étages. Cette architecture hybride, liée à une phase d'apprentissage dynamique, permet de produire de meilleures cartes de classification et donc de meilleures contraintes.

Une approche originale, par apprentissage, du problème de l'appariement dense d'objets n-D permet l'introduction de connaissances a priori dans l'élaboration des modèles de forme des structures cibles. Nous présentons également un nouveau descripteur de forme, le descripteur "observed transport", dont la robustesse au bruit et le pouvoir de discrémination accru en font un bon candidat pour notre stratégie d'appariement.

Enfin, un modèle plus fidèle des transformations induites par les processus histologiques, l'approche affine par morceau, permet la conception d'un algorithme de recalage biomédical mieux adapté à la reconstruction de volumes histologiques 3-D.

Mots clés: *segmentation, recalage, information explicite, affine par morceau, descripteur transport observé*

Abstract

We propose in this thesis an automated segmentation and registration system for medical images. We have focused on devising a methodology where maximum use is made of the available a priori medical expertise.

We approach the issue of boundary finding as a process of fitting a group of deformable templates to the contours of the target structures. Those templates evolve in parallel under the supervision of a series of rules derived from the analysis of both the template's dynamics and medical experience. We subject the templates to a variety of constraints also designed based on prior medical knowledge about the textural and shape properties of the target structures.

Textural information is extracted by a non-linear classifier in the form of a 2-stage neural network whose hybrid architecture and dynamic learning phase help produce a better classification map.

An original learning approach to the dense matching of n-D objects enables the introduction of prior knowledge in the computation of shape models for the target structures. We also introduce a novel shape descriptor, the observed transport descriptor, whose noise robustness and enhanced discriminating power make it a good candidate for our matching strategy.

Finally, a more faithful model of the transformations induced by histological processes, the piecewise affine approach, allows the design of a biomedical registration algorithm better suited to the reconstruction of 3-D histological volumes, a first step towards hybrid combined MRI/histology atlases.

Keywords: *segmentation, registration, explicit information information, piecewise affine, observed transport descriptor*