Diss. ETH No. 16601

Multi-Modal Non-Rigid Registration of Volumetric Medical Images

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of

Doctor of Sciences ETH Zurich

presented by

ADRIAN STEFAN ANDRONACHE

M.Sc. Electrical Engineering,

University Politehnica of Bucharest

born 6th May 1976

nationality: Romanian

accepted on the recommendation of

Prof. Dr. Gábor Székely, examiner

Prof. Dr. Franjo Pernuš, co-examiner

Dr. Philippe Cattin, co-examiner

2006

Abstract

Due to recent advances in image guided surgery, the availability of efficient algorithms for the fusion of pre- and intra-operative image data is of increasing importance. Because of the unavoidable deformations in the anatomy the classical rigid registration techniques can only be used for very limited special cases. Hierarchical subdivision techniques allow, however, to decompose non-rigid matching problems into rigidly registering numerous sub-images of decreasing size. While Mutual Information has proven to be a very robust and reliable similarity measure for intensity-based registration of multi-modal images, numerous problems have to be faced if it is applied to small-sized images, compromising its usefulness for such subdivision schemes.

Within this dissertation, we examine and explain the rather unexpected behavior of Mutual Information when applied to structureless image regions or regions covering only a limited number of image pixels. Methods to overcome these limitations are presented and their performance tested using the hierarchical subdivision registration scheme.

A new method inspired from point pattern analysis is introduced to identify problematic regions not containing enough structural information to reliably find the proper correspondence between the images using Mutual Information. The proposed method not only improves the accuracy and robustness of the registration, but can also be used as a very efficient stopping criterion for the further subdivision of nodes in the hierarchy improving the computational complexity by a factor of 5 for 2D images.

To overcome the statistical consistency problem when estimating Mutual Information for small image regions, an intensity mapping technique is presented. This mapping transforms the initial multi-modal images into a common pseudo-modality and therefore allows to switch the similarity measure to the more robust Cross-Correlation. The integration of

this mapping technique brought about an additional reduction of the computational complexity by a factor of about 2 for 3D images.

Zusammenfassung

Aufgrund der aktuellen Fortschritte in der Bildgeführten Chirurgie ist die Verfügbarkeit von effizienten Algorithmen für die Registrierung von prä- und intraoperativen Bilddaten von grosser Wichtigkeit. Die klassischen rigiden Registrierungstechniken sind aber aufgrund der unausweichlichen Deformationen im Gewebe nur beschränkt einsetzbar. Im Gegensatz dazu erlaubt die hierarchische Subdivisionstechnik eine Zerlegung eines nicht-rigiden Matching Problems in mehrere rigide Registrationen von kleineren Unterbildern. Während sich Mutual Information als ein robustes und verlässliches Ähnlichkeitsmass für die intensitätsbasierte Registrierung multimodaler Bilder erwiesen hat, ergeben sich verschiedene Probleme bei der Verarbeitung von kleinen Unterbildern, was die Anwendbarkeit dieser Methoden kompromittiert.

In dieser Dissertation untersuchen wir das zum Teil unerwartete Verhalten von Mutual Information wenn es auf Bilder mit wenig strukturellem Inhalt oder auf Regionen mit wenigen Bildpixeln angewendet wird. Im weiteren werden Methoden vorgestellt, welche diese limitierenden Faktoren umgehen und ihre Leistungsfähigkeit anhand einer Implementation in einem hierarchischen Subdivisionsregistrierungssystem geprüft.

Eine neue auf der Punktmusteranalyse basierende Methode wird eingeführt um problematische Regionen zu identifizieren, welche nicht genügend Struktur für eine Registrierung mit Mutual Information enthalten. Die Methode verbessert aber nicht nur die Genauigkeit der Registration, sondern bildet auch ein sehr effizientes Stoppkriterum um weitere hierarchische Unterteilungen zu verhindern. Die benötigte Rechenzeit sinkt durch deren Einsatz um einen Faktor 5 für 2D Bilder.

Im weiteren wird noch eine Intensitätstransformation für kleine Unterbilder vorgestellt. Diese Methode erlaubt es durch transformieren der beiden Bilder in eine Pseudomodalität auch für multi-modale Bilder die

robustere Crosscorrelation als Aehnlichkeitsmass zu verwenden. Die Integration dieser Transformationstechnik hat eine weitere Reduktion der Rechenzeit um einen Faktor 2 erbracht.

Contents

List of Figures

List of Tables

1

Introduction

1.1 Image registration

The discovery of the X-rays in 1895 by Wilhelm Conrad Röntgen marked the beginning of a new era in medicine, because for the first time it was possible to see and analyze organs and tissues inside the human body without a surgical operation. The X-ray images were opening the way to better understand the human body's anatomy and functionality. Soon, the X-ray images became a useful tool for medical diagnosis, and even now, after more than one century this film-based medical images are still broadly used in radiological applications.

In the computer era, the film-based X-ray images are replaced by digital images, providing the spatial information about both anatomical and functionality of the human body in high resolution. The last decades of immense scientific research and technological progress resulted in the development of numerous new and performant imaging techniques. These techniques can be differentiated by their underlying imaging principles, such as the physical properties of radiation field and of the sensors. Each imaging technique leads to a different *image modality*, which is characterized by its capability of picturing in a particular manner the details of the different tissues. Depending on the characteristics of the irradiating field and the sensitivity of the acquiring sensor, the same material may appear differently in images of different modalities. For example, X-ray computed tomography (CT) images are sensitive to tissue density and atomic composition according to their X-ray attenuation coefficient, while magnetic resonance (MR) images are responsive to proton density, relaxation times, flow, and other parameters. Ultrasound (US) images,

            (a)   (b)   (c)   (d)
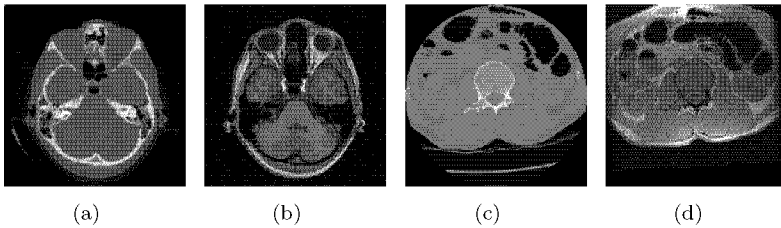
Figure 1.1: *Example of images acquired with different modalities. (a,b) CT and MR transversal scans of the head. (c,d) CT and MR transversal scans of the abdominal region*

detect subtle changes in acoustic impedance at tissue boundaries and discriminate the tissues by their different diffraction patterns. Also, endoscopy and surgical microscopy provides images of visible surfaces deep inside the body. Figure 1.1 shows how the same tissues are pictured in a different manner by the different acquiring image modality.

Even more, the two-dimensional (2D) X-ray images are now replaced with three-dimensional data (3D) or time series images which increase the possibility of spatially locate and analyze the human anatomy and physiology. These advances of the imaging technology and the ever improving quality of the generated images, has enormously increased the potential of medical imaging in the clinical practice.

Nowadays, medical imaging became a vital component of a large number of clinical applications. Developing extremely fast during the last decades, medical imaging is bringing new techniques to improve existing clinical procedures or to facilitate the establishment of completely new approaches. The most important component of this process is the quickly growing role of medical imaging in supporting therapy by delivering essential information for pre-operative planning and intra-operative navigation. All the related procedures are using a large number of digital images on a routine basis. Taken at different times or by different radiological modalities, these images are usually containing complementary information. Therefore, most of the clinical procedures are requiring a proper integration (fusion) of the information contained in the images of different modalities. The crucial step in this fusion process is the registration procedure which ensures that the images of interest are in

a sufficiently good spatial alignment by establishing the correspondence
of anatomical structures within the patient. The process of image fusion
helps the physician to monitor the evolution of diseases or the effect of
the different treatments applied by looking at the pathological changes
in the shape or internal structure of the organs. Prior knowledge can also
be integrated in the process of diagnosis by fusing the medical images
acquired with an anatomical atlas.

The process of accurate quantitative analysis of the structural and phys-
iological parameters from medical images is usually a difficult problem.
The complexity comes not only from the 3-D nature of the problem,
but also because of intrinsic limitations of the image acquisition pro-
cess itself. Often, the medical images present a poor contrast, a limited
resolution and artifacts generated by partial volume effects, noise, or in-
homogeneities. The ambiguity in the images results in intra- and inter-
observer variability in the measurements that may be of the same order
of magnitude as the signal to be quantified itself. Therefore, this may
deteriorate the reproducibility of the measurements and weaken the sig-
nificance of the clinical observations derived from the images. Improving
the reliability of image measurements requires robust and well-validated
strategies for automated, computer-based image analysis. However, the
robustness of these automated approaches often suffers from limiting as-
sumptions imposed on the data that are not always made explicit, from
reliance on object or image specific parameters, and from the use of
too simple models, such that a manual intervention is usually required.
Moreover, the validation and the experimental evaluation is often lacking
or insufficient to support reliable application on real-world problems.

In the same time, the clinical requirements for medical image analysis
have grown considerably, imposing new and more challenging demands
on the medical image processing algorithms involved, and consequently
on the accuracy of the image registration procedures. Often encountered
as image matching or alignment, image registration is the process of
aligning two or more images of the same scene from different viewpoints
and/or time such as the corresponding structural features are in the same
physical position. This term is also used for the alignment of images with
a computer model, or aligning image features with locations in physical
space. Basically, there is no restriction regarding the modality used to
acquire the images, and also regarding the time point of their acquisition.

Image registration is a crucial step in all image analysis tasks in which the final information is gained from the combination of various data sources like in image fusion, change detection and multichannel image restoration. Therefore, image registration has applicability in many fields, like remote sensing (multispectral classification, environmental monitoring, change detection, image mosaicing, weather forecasting, creating super-resolution images, integrating information into geographic information systems (GIS)), in cartography (map updating), computer vision (target localization or tracking, model based segmentation, automatic quality control) but this thesis only addresses the field of radiological imaging.

The main goal of the research presented in this thesis is the development of an automatic method for non-rigid registration of mono- and multi-modal medical images. While in the early phase of development the distance between image features has been dominantly used for characterizing the goodness of fit between the structures to be matched, the importance of purely intensity-based measures has grown very quickly during the past decade. This is due to the fact, that these approaches are not dependent on feature detection, which is not only preventing the influence of errors caused by a pre-processing step, but also making these methods more suitable for automatic processing. The introduction of Mutual Information (MI) as a similarity measure by Viola and Wells [Viola and W. M. Wells 1995] and Collignon et al. [Maes *et al.* 1996] has especially influenced the development of intensity-based image registration due to its inherent ability to deal with multi-modal matching problems. The decisive property in this respect is that MI does not depend on any linear relation between the different image intensities but it relies only on statistical properties of the two-dimensional histogram. The following years of intensive research were proving the advantages and the robustness of this approach, and MI started to be extensively used even in the field of mono-modal registration.

New robust and accurate non-rigid registrations algorithms are heavily required by the increasing importance of the medical imaging in supporting therapy, by delivering essential information for pre-operative planning and intra-operative navigation. Every medical application that necessitates an image registration procedure is imposing specific requirements concerning the accuracy, robustness, multi-modal capability and speed. In order to be as general as possible, we propose a new algorithm

that can be further used as a modular framework and that can be easily adapted to a wide area of applications. The new algorithm is developed such that it overcomes the limitations and problems noticed for many other registration algorithms already presented in the literature. Based on the idea presented in [Likar and Pernuš 2001], we developed an enhanced hierarchical image splitting strategy for non-rigid registration of mono- and multi-modal images. The hierarchical strategy covers a big variety of possible deformations by decomposing the non-rigid matching problem into rigidly registering numerous sub-images of decreasing size. The final continuous deformation field is calculated by regularization and interpolation of the local transformations. Problem specific constraints can be easily integrated into the algorithm. This thesis contributes to the development of more reliable automated non-rigid medical image registration tools for multi-modal images with broad applicability in clinical practice.

1.2 Classification of the registration methods

The big variety of applications with specific requirements and available technologies led to an impressive number of methods that have been proposed and developed during the past decades. Several surveys and even textbooks have been already published presenting a broad and complete overview of the techniques proposed in the literature [Brown 1992, van der Elsen *et al.* 1993, Hajnal *et al.* 2001, Hill *et al.* 2001, Zitova and Flusser 2003]. Therefore, in this section we present only a brief overview of the classification of the proposed methods and their underlying principles. The mathematics and the implementation details of those methods related to the subject of the present thesis are discussed in the next chapter.

The most complete set of classifications criteria was proposed by Maintz and Viergever in 1998. Following the terminology proposed in [Maintz and Viergever 1998], the registration procedures were classified according to the problem statement, the registration paradigm and the optimization procedure. The problem statement includes criteria such as the object or the region of interest, the modality and dimensionality of the

images involved, the source of the images (eg. the same or different patients) and also the nature of the transformation to be recovered. The registration paradigm involves the image features and the correspondence criterion used to compute the registration while the optimization procedure criterion regards the search strategy for determining the actual registration parameters and also the degree of the required user interaction.

According to the registration paradigm, the registration procedures can be classified as being *extrinsic* or *intrinsic*. The *extrinsic* registration procedures are based on foreign markers introduced into the image space. These methods rely on artificial objects (e.g. stereotactic frame, fiducial screw or skin markers, dental adapter, mold, *etc.*), usually attached to the patient and that are designed to be well visible and accurately detectable in all the imaging modalities. As such, the registration of the acquired images is fast and easily performed in an automatic manner. The main drawbacks of these methods are the prospective and the usually invasive character of marker placement. While, non-invasive markers can also be used, but they lead to less accurate results. Usually, the extrinsic methods are limited to rigid transformations (used in orthopedics and neurosurgery) and very few methods were proposed for non-rigid applications (only if a big number of markers can provide sufficient information about the local deformation).

The *intrinsic* registration methods are developed such that only the image information is used as it is acquired from the patient. In this case, the registration can be based on a limited set of identified salient points (*landmarks*), on the alignment of segmented binary structures (*segmentation based*), or directly onto measures computed from the image grey values (*voxel property based*). The landmarks are usually well identifiable points of the morphology of the visible anatomy, located automatically or interactively by the user. Similar to the extrinsic methods, the landmark based methods are also usually restricted to rigid or affine transformations, and the performance is heavily dependent on the accuracy of the landmarks extraction. Basically, a landmark identification procedure is similar to segmentation, only extracting simple points instead of lines, surfaces or volumes. Therefore, the accuracy of these methods is usually highly data and application dependent, and is limited by the segmentation error of the landmarks and the precision of their localization.

The voxel property based approaches represent a special class of the intrinsic methods because they operate directly on the image grey values without any prior data reduction by a segmentation procedure. There are two main approaches: on one hand to represent the information by a set of scalars and orientations (the *principal axes and moments based* methods), on the other hand to use the full image content. The latter approach is currently in the focus of research because it proves to be the most flexible and potentially the most accurate, as it uses all the information available in the image throughout the entire registration process. The main drawback of the full image content approach is the computational cost, but this is not necessarily an issue due to the steadily increase of available computational resources. Because this approach is also in the focus of the present thesis, more details about the voxel property based methods will be presented in the following chapter.

Another important criterion for classification of the registration procedures concerns the nature of the geometrical transformation to be recovered. The appropriate type of the transformation needed to correct the spatial misalignment is imposed by different factors, either related to the image content or to the scanning procedure. For example, the stiffness of the tissues to be registered or other physical properties of the anatomical structures contained in the images are imposing constraints to the choice for the transformation model. One can assume that a simple *rigid* body assumption can be appropriate for registering the skull in 3D scans of the head. This means that the appropriate spatial transformation is a combination of only three translations and three rotations. However, this rigid assumption holds only if the user can neglect the scanner miscalibration or geometric distortions in the images. More elaborate *affine* transformations can further compensate for image scaling and skew differences. *Projective* transformations are sometimes useful to register 2D X-rays to 3D scans.

As in most of the cases the anatomy shows unavoidable deformations between consecutive imaging processes or during the surgical procedure, *non-rigid* or *curved* transformations capable of dealing with more localized spatial changes are nowadays in the focus of research and development. The non-rigid transformations are also suitable to compensate for the anatomical differences existing either when the images are coming from different subjects or when an atlas registration is performed.

In contrast to the previous types of transformations that can be described only by a small set of parameters, the non-rigid transformations are usually described in terms of a vector displacement field (disparity) or as polynomial transformations. All these transformations can be applied *globally* to the entire image space, or *locally* to certain regions within the images. The local transformations are not very common because they may violate the continuity and the bijectiveness of the transformations. However, if more such local transformations are covering the entire image space, they can be integrated into a global curved transformation.

There are other important criteria of the classification of the registration methods that have to be mentioned. One of them concerns the imaging technology used to acquire the images. If the images to be registered result from the same or different scanning technology, the registration is called to be *mono-modal* or *multi-modal*, respectively. In preoperative procedures, the physician might be interested in a *modality to model* registration, such that he can compare the pathology of the patient with a standard computer model. There are different imaging devices available in the OR like endoscopes, echographs, microscopes or even video cameras which can provide images of the scene in real time. These real world images, displayed on a screen can be fused with a preoperative scan of an internal organ to provide guidance to the surgeon. These *patient to modality* registration methods are usually integrated into minimally invasive surgery procedures.

1.3   Organization of the thesis

In the following chapter we present the related work and the basic theory behind a rigid registration procedure, followed by an overview of the existing methods for non-rigid image registration. The chapter concludes by presenting the hierarchical image splitting strategy proposed by Likar and Pernuš which constitutes the starting point of our research.

Chapter 3 presents our investigation of the behavior of Mutual Information. It firstly describes the results of our research that revealed a disturbing peculiar behavior of MI. This drawback, together with the already known limitations can seriously perturb the accuracy of any hierarchical registration procedure that uses the matching capability of

MI. Consequently, this chapter presents several solutions we developed and proposed to overcome the limitations of MI.

After an overview of the existing methods for deformation field regularization, Chapter 4 describes our proposed method to approach this problem.

In Chapter 5 we present the entire enhanced hierarchical algorithm composed of the different methods and solutions presented in the previous chapters. In this chapter we provide details of the algorithm implementation and its modular architecture, and we discuss several possible extensions of the different modules.

Chapter 6 presents and analyzes the results of our non-rigid registration algorithm for mono- and multi-modal images.

The last chapter draws the conclusions and presents several directions for further research.

2

Related Work and Background Theory

The previous chapter introduced the general concepts and the classification of the proposed methods for image registration. This chapter presents the underlying theory and algorithms behind the main existing voxel-based methods for both rigid and non-rigid approaches. After establishing notation and terminology used in this thesis, this chapter continues by presenting the concepts behind a general 3D voxel-based rigid-body registration algorithm in detail. This algorithm is considered as an exemple, from which general rules and concepts can be extended for all the other registration applications, including 2D and 3D images, image referencing with physical space, and non-rigid intra- and inter-subject registration. The chapter will continue with an overview of these main approaches, especially for non-rigid registration, and conclude with a detailed description of the hierarchical image splitting strategy that represents the starting point of the research presented in this thesis.

2.1   Notation and terminology

A registration algorithm, as stated in the definition, has to find the proper spatial transformation that relates the information conveyed in an image to the another image or to the physical space. Therefore, a registration algorithm requires as input at least two images of the same scene, or an image and a computer model or the parameters of the physical space. The main goal of this work is to present a new algorithm

for multi-modal image registration. For simplicity reasons, the entire thesis considers only the case of two input medical images.

2.1.1 The discrete nature of the images

Medical images undergoing the registration process are in a digital format. They are formed by sampling and quantizing the real world scene in their field of view into a limited number of samples. For an image A, this field of view forms the image domain Ω_A and it can be expressed as a discretization of the finite continuous space of the imaged patient volume $\bar{\Omega}_A$ with an infinite sampling grid Γ_{ξ}^A. This sampling grid defines the spatial resolution of the medical image and it is characterized by the sample spacing $\vec{\xi}^A = \left(\xi_x^A, \xi_y^A, \xi_z^A\right)$ given by the properties of the scanning device. This sampling process is likely to be different for two different images and while the sampling is commonly uniform in a given direction, it may be anisotropic for the different directions within the image. In case of 2D images, its samples are called *pixels* (an abbreviation of "picture elements") and for 3D images the samples are called generally *voxels* (an abbreviation of "volume elements").

Within an image A, each voxel is associated with a spatial location $\vec{s}_A = (s_x^A, s_y^A, s_z^A)$ and an image intensity value $A(\vec{s}_A)$. The images are not only discretized in space, but also in intensity values (or levels). Very often in practice, the intensity values are quantized into n bit levels, (with n often 8, 16, 24 or 32). The entire domain of the possible intensity levels of an image is denoted with Ω_a:

$$\Omega_a = \{a = A\left(\vec{s}_A\right), \forall \vec{s}_A \in \Omega_A\}. \tag{2.1}$$

It is important to notice that a pair of images to be registered are derived from a real scene or object, i.e. the patient. These images have a limited field of view, that normally does not cover the entire patient. Furthermore, this field of view is likely to be different for the two images. As the images A and B represent the same object, taken with the same or a different modality, their image domains Ω_A and Ω_B might be different, but there is a relation between the spatial locations in A and

B. Sometimes it is useful to define the images within their field of view as a mapping of points in the patient to intensity values:

$$A : \vec{s}_A \in \Omega_A \longmapsto A(\vec{s}_A)$$
$$B : \vec{s}_B \in \Omega_B \longmapsto B(\vec{s}_B). \tag{2.2}$$

Therefore, within the common space covered by both image fields of view, any spatial location $\vec{s}$ of the imaged object will have a corresponding position $\vec{s}_A$ within the image A and $\vec{s}_B$ within the image B.

2.1.2 Registration transformation

The final goal of image registration is to recover the transformation that relates the spatial position of the features in one image with the position of the corresponding features in the other image. In geometrical terms, this transformation represents a mapping of every position $\vec{s}_B$ from Ω_B into the corresponding position $\vec{s}_A$ from Ω_A over the entire domain of interest. Denoted with T_{BA}, this spatial relation is called *the registration transformation*, and can be defined as:

$$T_{BA} : \vec{s}_B \in \Omega_B \longmapsto \vec{s}_A \in \Omega_A \iff T_{BA}(\vec{s}_B) = \vec{s}_A. \tag{2.3}$$

This registration transformation is valid only within the domain of interest represented by the overlapping part of the two image fields of view, denoted with Ω_{AB}^T:

$$\Omega_{AB}^T = \{\vec{s}_A \in \Omega_A | \vec{s}_A = T_{BA}(\vec{s}_B), \vec{s}_B \in \Omega_B\}. \tag{2.4}$$

In any registration procedure, one of the images is considered as the *reference image* (or *target image*) while the other image is considered as the *floating image* (or *source image*). The reference image is spatially fixed and its image domain represents the mapping target of the registration transformation. As such, the floating image is undergoing the registration transformation, from its own image domain into the reference image domain. During the entire thesis, image B will always be considered as the floating image that has to be registered with the reference image A. Therefore, the subscript from T_{BA} will also be omitted from the notation of the registration transformation ($T_{BA} = T$). However, it is sometimes useful to define the inverse transformation $T_{AB} = T^{-1}$ from the image

A to the image B, but care has to be taken because not all the geometrical transformations are invertible. The floating image B which is transformed into the reference image domain is denoted as $B^T = T(B)$ and mathematically it can be expressed as a mapping of all the points $\vec{s}_B$ from Ω_B into the overlapping domain Ω_{AB}^T:

$$B^T : T(\vec{s}_B) \in \Omega_{AB}^T. \tag{2.5}$$

Because Ω_{AB}^T is only a subset of the image domain Ω_A, the corresponding domain of intensity levels, denoted Ω_a^T, will also be a subset of Ω_a. Similarly, the transformed floating image B^T will have a corresponding domain of intensity levels Ω_b^T. However, because the transformation from one image domain into the other is likely to involve an intensity interpolation of the voxels that are falling in between the grid of the reference image domain, the transformed intensity domain Ω_b^T is not necessarily a subset of the initial intensities. The aspects involving intensity interpolation will be discussed later.

2.1.3   Image coordinate system

Due to the digital nature of the images, a two-dimensional (2D) image can be represented as a rectangular array whose elements are the image samples. Similarly, a 3D image is obtained by stacking together 2D images and can be represented as a 3D array. A spatial location $\vec{s}_A = (s_x^A, s_y^A, s_z^A)$ of a voxel within the image A is given by the natural indexing within the image array. This array defines a Cartesian coordinate system for the image, known as *the image coordinate system*. This coordinate system is associated to the image domain Ω_A, and conventionally is defined such that the x and y axes are chosen to be in the slice plane, and the z axis along the slice stack. The origin of the image coordinate system is considered as the first voxel of the image, the x and y coordinates denoting the column and row number, respectively, and the z coordinate denoting the slice number, assuming that the data is stored row by row and slice by slice.

2.1.4   World coordinate system

Due to the possible different orientation of the patient relative to the scanning device, a more general coordinate system needs to be defined with respect to the imaged scene. This is the so called *world coordinate system* conventionally defined such that the orientation of its axes is relative to the patient's body and its units are expressed in real-world units (millimeters). The world coordinate system is related to the image coordinate system by the origin of its axes that is defined in the center of the image volume. The convention is as follows: the x axis is running from the patient's right side to his left side (R→L), the y axis from the patient's anterior to his posterior (A→P), and the z axis form the patient's inferior towards his superior (I→S).

According to the properties of the image A, such as voxel size, orientation, and dimension, a mapping $\mathcal{T}_{iw}^A$ can be defined such that the entire image domain Ω_A can be transposed into the world coordinate system, and vice versa. This is similar to the process of mapping the digital image back into the continuous space of the originally scanned volume of the patient $\bar{\Omega}_A$. The sampling quantization makes the process of recovering the original volume practically impossible, so $\widetilde{\Omega}_A$ will denote the discrete image domain transformed into the world coordinate system:

$$\widetilde{\Omega}_A = \mathcal{T}_{iw}^A\left(\Omega_A\right) \tag{2.6}$$

where $\mathcal{T}_{iw}^A$ represents the transformation from the image coordinates to the world coordinates. Accordingly, by denoting the location of a point in image coordinates with $\vec{s}_A \in \Omega_A$, its corresponding position in world coordinates is denoted with $\vec{w}_A = (w_x^A, w_y^A, w_z^A) \in \widetilde{\Omega}_A$ and defined as:

$$\vec{w}_A = \mathcal{T}_{iw}^A\left(\vec{s}_A\right). \tag{2.7}$$

The transformation $\mathcal{T}_{iw}^A$ represents the mapping from $\mathbb{N}^3$ to $\mathbb{R}^3$ and the inverse transformation from world to image coordinates is $\mathcal{T}_{wi}^A$:

$$\mathcal{T}_{wi}^A = (\mathcal{T}_{iw}^A)^{-1}. \tag{2.8}$$

It is important to note, that this world to image coordinates transformation $\mathcal{T}_{wi}^A$ represents a mapping from $\mathbb{R}^3$ to $\mathbb{N}^3$, and it will take into consideration the discrete nature of the image domain. Therefore, this

transformation involves a voxel interpolation of those points falling in between the natural grid of the digital image, as it will be discussed later.

In order to differentiate between the registration transformation T that is mapping a point from the floating image domain into the reference image domain, in this world coordinate system we will denote the transformation with $\mathcal{T}$. The difference is the support of the two spatial mapping functions: $T : \mathbb{N}^3 \to \mathbb{N}^3$ and $\mathcal{T} : \mathbb{R}^3 \to \mathbb{R}^3$.

2.2   Rigid and affine registration algorithms

This section addresses both the rigid and the affine registration procedures with the generic name of *rigid registration*, because the only difference is the number of degrees of freedom allowed for the spatial transformation. Figure 2.1 delineates the four key elements of a standard voxel-based registration algorithm: the transformation model, the similarity metric, the voxel interpolation method and the optimization procedure. Firstly, an assumption has to be made on the appropriate transformation model that can reasonably compensate for the spatial misalignment between the different images of the scene of interest. At the same time, a suitable similarity metric has to be defined in order to numerically quantify the goodness of fit between the images, according to the scanning technology used to acquire the images to be registered. For the voxel-based registration algorithms the similarity metric is estimated directly from the image intensities and therefore its choice depends on the scanning modalities of the input images. The interpolation is required to estimate the intensity of the voxels which are falling in between the sampling grid points of the reference (target) image. During the entire registration procedure, the optimization algorithm is used to iteratively search for the correct spatial transformation, by making a series of successive trials and guesses in order to maximize the predefined similarity metric.

The rigid registration algorithm is one of the simplest registration procedures because it operates with the simple transformation model of a *rigid body*, which can only translate and rotate in 3D physical space. An *affine* registration procedure allows in addition scaling and skew, thus making
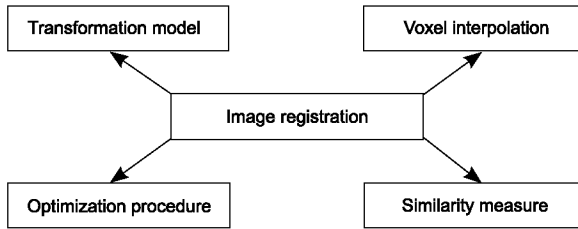
Figure 2.1: *The general scheme of an image registration algorithm*

a step towards the non-rigid models. More details about these transformation models will be given in the next subsection. A general flow chart of a rigid registration procedure is depicted in Fig. 2.2. It is tailored to follow the previous description of a rigid registration procedure.

To summarize, a rigid registration procedure starts with an initial estimate of the spatial transformation $\mathcal{T}_{\alpha}$, for which the corresponding similarity measure is calculated. The subscript α marks the dependency of the registration transformation on a set of parameters which are regularly expressed in the common world coordinates of the images. This α is a set of six parameters, as will be described later, representing the six degrees of freedom of a rigid body: three translation and three rotations. If there are no other means of initializing the set of transformation parameters α, then the best initial estimate is the simple alignment of the centers and the axes of the world coordinate systems associated to the images to be registered. In this case, the initial set of parameters α is set to zero and the corresponding spatial transformation is the identity matrix. Usually, this is a good initialization estimate, assuming that the images are focused on the objects to be registered. This alignment of the images in the world coordinate system is ensured by the individual associated image to world transformation matrixes $\mathcal{T}_{iw}^{A,B}$ that have to be calculated once at the beginning of the registration procedure. These image to world transformations compensate for different orientation, voxel anisotropy or differences in voxel size between images. Therefore, by combining all these transformations, the spatial transformation $T_{\alpha} = T_{BA}(\alpha)$ is estimated from the floating image domain Ω_B into the reference image domain Ω_A. Using this transformation matrix T_{α}, the floating image B is transposed into the reference image domain.
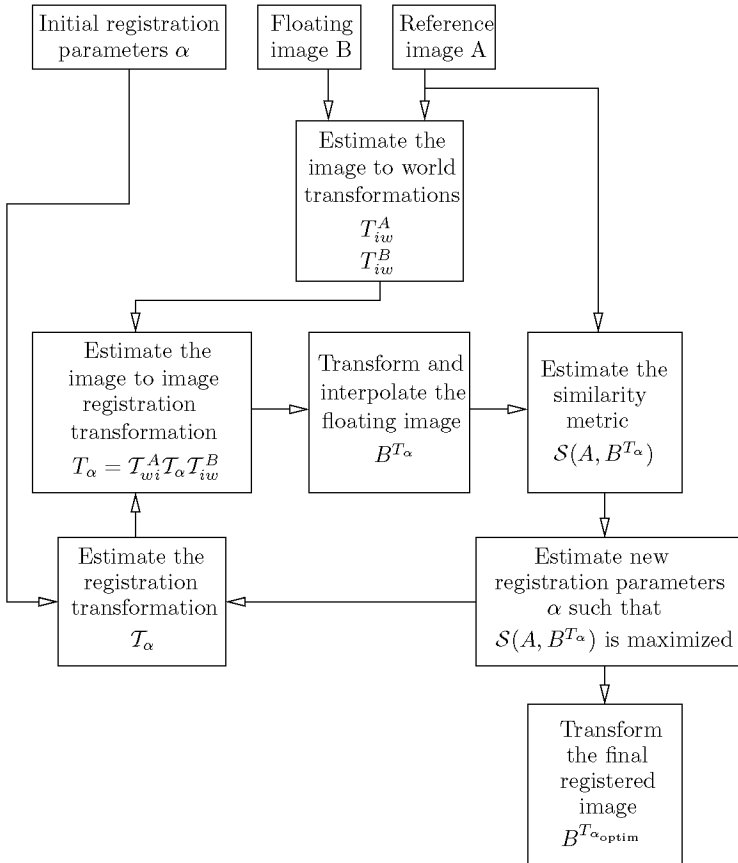
Figure 2.2: *The flow chart of a rigid registration algorithm*

This mapping includes the voxel interpolation method which ensures the correct reconstruction of the floating image on the new reference image sampling grid. Once both images, the reference image A and the transformed floating image B^{T_α}, are in the same image domain, then the similarity metric $\mathcal{S}(A, B^{T_\alpha})$ between the images can be estimated over the entire volume of overlap Ω_{AB}^T. Then, the optimization algorithm starts searching a new set of parameters $\alpha \rightarrow \alpha_{\mathrm{optim}}$ that would optimize the similarity metric. The algorithm repeats the sequences described above until the similarity measure reaches its optimum. Finally, according to this optimal set of parameters, the floating image B is transformed into the reference image domain. However, sometimes it is of interest to know only the corresponding transformation into the image coordinates $T_{\alpha_{\mathrm{optim}}}$ or the world coordinate frame $\mathcal{T}_{\alpha_{\mathrm{optim}}}$. These transformations can be directly used by subsequent image processing procedures, e.g. segmentation procedure, or image referencing with the physical space.

The following subsections present in detail the four main elements of the rigid registration algorithm, (1) the descrition of the spatial transformations, (2) the different choices of the most common similarity metrics, (3) a summary of three fast voxel interpolation methods, and (4) an overview of the optimization procedures.

2.2.1   Rigid and affine spatial transformations

In many practical applications, local non-rigid tissue deformations are negligible or irrelevant, so the geometrical relationship between the images to be registered can be modeled by a rigid or an affine linear transformation. The spatial transformation can be defined either between the image domains or directly in the world coordinate frame. The latter transformation, denoted with $\mathcal{T}$ is compensating for the spatial misalignment induced by the change of shape or pose of the imaged object in the real physical space. Apart from the real world transformation $\mathcal{T}$, the image to image transformation T includes the image to world individual transformations $\mathcal{T}_{iw}^{A,B}$ that are compensating for the possible different image characteristics induced by the scanning device. Therefore, the image to world coordinates mapping will be characterized first, followed by the description of the rigid and affine transformations in the world coordinate system.

Image to world transformation

Given an image A, the transformation $\mathcal{T}_{iw}^A$ that maps any location $\vec{s}_A$ of the image domain Ω_A into the corresponding location $\vec{w}_A$ in the world domain $\widetilde{\Omega}_A$ is *the image to world coordinate transformation*. This transformation $\mathcal{T}_{iw}^A$ can be expressed as a 4×4 matrix using homogeneous coordinates that is a composition between the image orientation $\mathcal{O}^A$ (or the relative position of the patient to the scanning device), the image voxel size $\vec{\xi}^A$ and the positioning of the origin of the world coordinate system (in the center of the image volume) expressed in image coordinates $\vec{c}_A = (c_x^A, c_y^A, c_z^A)$:

$$\mathcal{T}_{iw}^A = \mathcal{O}^A \cdot \begin{bmatrix} \xi_x^A & 0 & 0 & -c_x^A \\ 0 & \xi_y^A & 0 & -c_y^A \\ 0 & 0 & \xi_z^A & -c_z^A \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.9}$$

Some of the most common image orientation matrixes $\mathcal{O}^A$ are given in Table 2.1.

| | Image orientation | | |
	transversal	sagittal	coronal
$\mathcal{O}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Table 2.1: *The standard orientation matrix $\mathcal{O}$ relating the image coordinates to the world coordinates for different image orientations*

In order to reduce the effects of radiation, some CT scanners have a non-zero gantry tilt angle, leading to the non-orthogonality of the z axis to the slice plane. This γ^A angle has also to be taken into account when calculating the image to world coordinates transformation:

$$\mathcal{T}_{iw}^A = \mathcal{O}^A \cdot \begin{bmatrix} \xi_x^A & 0 & 0 & -\xi_x^A c_x^A \\ 0 & \xi_y^A \cos(\gamma^A) & 0 & -\xi_y^A c_y^A \cos(\gamma)^A \\ 0 & \xi_y^A \sin(\gamma^A) & \xi_z^A & -\xi_z^A c_z^A - \xi_y^A c_y^A \sin(\gamma^A) \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.10}$$

It is very useful to define the inverse transformation, from world to image coordinates $\mathcal{T}_{wi}^A$. This transformation is given by the simple inversion of the $\mathcal{T}_{iw}^A$ matrix:

$$\mathcal{T}_{wi}^A = (\mathcal{T}_{iw}^A)^{-1}. \tag{2.11}$$

Having defined both image to world transformations, the mapping of an image point $\vec{s}_A = (s_x^A, s_y^A, s_z^A) \in \Omega_A$ into the world coordinate $\vec{w}_A = (w_x^A, w_y^A, w_z^A) \in \widetilde{\Omega}_A$ can be expressed as:

$$(\vec{w}_A, 1)^T = \mathcal{T}_{iw}^A \cdot (\vec{s}_A, 1)^T, \tag{2.12}$$

and vice versa, from the world to image coordinates:

$$(\vec{s}_A, 1)^T = \mathcal{T}_{wi}^A \cdot (\vec{w}_A, 1)^T \tag{2.13}$$

where we used the following notations:

$$(\vec{w}_A, 1)^T = \begin{bmatrix} w_x^A \\ w_y^A \\ w_z^A \\ 1 \end{bmatrix}, \text{ and } (\vec{s}_A, 1)^T = \begin{bmatrix} s_x^A \\ s_y^A \\ s_z^A \\ 1 \end{bmatrix}. \tag{2.14}$$

Rigid transformation

The simplest registration transformation $\mathcal{T}$ is described in terms of a *rigid body* transformation which allows only translations and rotations. This transformation can be completely described in 3D space by six parameters (also called as degrees of freedom): three translations $\vec{t} = (t_x, t_y, t_z)^T$ along the x, y and z directions, and three Euler rotations $\vec{r} = (r_x, r_y, r_z)$ around the x, y and z axes respectively.

The rigid transformation of a point at the location $\vec{w} = (w_x, w_y, w_z)^T$ can be represented by a rotation matrix $\mathcal{R}$ and a translation vector $\vec{t} = (t_x, t_y, t_z)^T$:

$$\mathcal{T}_{\text{rigid}}(\vec{w}) = \mathcal{R} \cdot \vec{w} + \vec{t} \tag{2.15}$$

where the 3×3 rotation matrix $\mathcal{R}$ is constructed from the Euler rotation angles as follows:

$$\mathcal{R} = \begin{bmatrix} \cos_x \cos_z & \cos_x \sin_z + \sin_x \sin_y \cos_z & \sin_x \sin_z - \cos_x \sin_y \cos_z \\ -\cos_y \sin_z & \cos_x \cos_z + \sin_x \sin_y \sin_z & \sin_x \cos_z - \cos_x \sin_y \sin_z \\ \sin_y & -\sin_x \cos_y & \cos_x \cos_y \end{bmatrix} \tag{2.16}$$

with the notation $\sin_i = \sin(r_i), \forall i \in \{x, y, z\}$.

This rigid transformation is usually applied to the continuous common world coordinate system and can be expressed as a 4×4 matrix using homogeneous coordinates:

$$
\mathcal{T}_{\text{rigid}}(\vec{w}) = \begin{pmatrix} & & & t_x \\ & \mathcal{R} & & t_y \\ & & & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} w_x \\ w_y \\ w_z \\ 1 \end{pmatrix}. \tag{2.17}
$$

With this representation, a rigid transformation of a location $\vec{w}_B = (w_x, w_y, w_z)^T$ from $\widetilde{\Omega}_B$ into the corresponding location $\mathcal{T}_{\text{rigid}}(\vec{w}_B) = \vec{w}_A$ of the continuous domain $\widetilde{\Omega}_A$ is:

$$
(\vec{w}_A, 1)^T = \mathcal{T}_{\text{rigid}} \cdot (\vec{w}_B, 1)^T. \tag{2.18}
$$

By putting together all the translational and rotational parameters we can form a six element vector α which is called the registration parameters set $\alpha = (t_x, t_y, t_z, r_x, r_y, r_z)$.

Affine transformation

The rigid transformation can only be used in a limited number of cases, so an extension of the number of degrees of freedom for the spatial mapping is required. The affine transformation is able to compensate not only for spatial displacements between images, but also for differences in their voxel size or scale. The affine transformation consists of three-dimensional translations $\vec{t} = (t_x, t_y, t_z)$, rotations $\vec{r} = (r_x, r_y, r_z)$, scalings $\vec{s} = (s_x, s_y, s_z)$, and skews $\vec{g} = (g_x, g_y, g_z)$, preserving the parallelism of the lines but breaking their orthogonality. Denoting the skew matrix with $\mathcal{G}$ and the scale matrix with $\mathcal{S}$, the expression of an affine transformation $\mathcal{T}_{\text{affine}}$ is:

$$
\mathcal{T}_{\text{affine}} = \mathcal{T}_{\text{rigid}} \cdot \mathcal{G} \cdot \mathcal{S} = \mathcal{T}_{\text{rigid}} \cdot \begin{bmatrix} 1 & g_x \cdot g_z & g_x & 0 \\ g_y & 1 & 0 & 0 \\ 0 & g_z & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
$$
$$
\tag{2.19}
$$

where $\mathcal{T}_{\text{rigid}}$ represents the previously discussed rigid transformation containing the translations and the rotations parameters. A set of 12 parameters $\alpha = (\vec{t}, \vec{r}, \vec{g}, \vec{s})$ is used in this thesis todescribe the affine transformation. These are the 6 parameters specific to the rigid transformation, the 3 parameters of the skew and 3 of the scaling factors. Note that the expression of the affine transformation can be reduced to the trivial rigid body transformation by setting $s_x = s_y = s_z = 1$ and $g_x = g_y = g_z = 0$.

Image to image transformation

Given two images, the reference image A and the floating image B, which are related by the registration transformation $\mathcal{T}_{BA}$, and having defined the transformations between the image and the world coordinates, then the entire registration transformation T_{BA} from the floating image coordinates to the reference image coordinates can be expressed as a simple matrix multiplication:

$$T_{BA} = \mathcal{T}_{wi}^{A} \cdot \mathcal{T}_{BA} \cdot \mathcal{T}_{iw}^{B}. \tag{2.20}$$

Accordingly, the corresponding location $\vec{s}_A \in \Omega_A$ of a transformed location $\vec{s}_B \in \Omega_B$ is:

$$(\vec{s}_A, 1)^T = T_{BA} \cdot (\vec{s}_B, 1)^T = \mathcal{T}_{wi}^{A} \cdot \mathcal{T}_{BA} \cdot \mathcal{T}_{iw}^{B} \cdot (\vec{s}_B, 1)^T. \tag{2.21}$$

2.2.2 Similarity measures

In order to be able to register two images, a measure has to be defined to numerically quantify the goodness of fit between the images, namely *the similarity measure*. For all the voxel-based registration algorithms, this similarity measure is extracted directly from the voxel properties (the image intensities) either from the entire image volumes or from a region within. This is why these methods are referred to as *voxel-based* registrations. As already presented in the description of the rigid registration algorithm, the similarity measure is used as the objective function when searching for the optimal registration transformation. Usually, the voxel-based registration procedures ignore the geometrical features of the anatomical structures contained in the images. However, as will be presented later in Sec. 2.3.2, there are several approaches for voxel-based

non-rigid registration procedures which combine the similarity measure defined by the image intensities with a penalty term used to regularize the deformation field (e.g. in [Maintz *et al.* 1998, Rueckert *et al.* 1999, Guimond *et al.* 2001, Hellier *et al.* 2001], etc.).

The choice of the appropriate similarity measure is crucial for a successful image registration procedure, so the decisive criterion is the type of images to be registered. Therefore, depending on the type of the modalities used to acquire the images, the user can choose between several similarity measures. For example, in the case of a mono-modal application, the images are likely to render the corresponding structures with very similar intensities. Therefore, the sum of the absolute difference values or the sum of the squared intensity differences can be used for comparison. More elaborate similarity measures have been proposed because, beside the image noise, some additional variation of the intensities may appear between images or even within an image itself. These similarity measures are derived from the Cross-Correlation coefficient, and are intended to be insensitive to intensity changes generated by different physical factors during the scanning procedure.

In a multi-modal registration no direct relationship between the image intensities can be assumed in general. Therefore, a completely different approach is required in order to quantify the similarity between the registered images. While some attempts have been made to preprocess an image so that the images appear to be from the same modality, e.g. [van den Elsen *et al.* 1994], more recent approaches have been inspired by information theory ([Hajnal *et al.* 2001]). These similarity measures are relying on the probabilistic relations and the distributions of the intensities in the images.

In this section will be presented only a brief overview of the major similarity measures proposed in literature. For more details, the reader should refer to [Hajnal *et al.* 2001] and the references therein.

Similarity measures derived from intensity differences

One of the simplest voxel similarity measure is the sum of squared intensity difference (SSD) between the images to be registered. During the registration process the proper transformation $\mathcal{T}$ is estimated by iteratively minimizing the SSD calculated from the overlapping domain Ω_{AB}^{T}

of the reference image A and the transformed image B^T. If N is the total number of voxels from Ω_{AB}^T, then SSD can be expressed as:

$$SSD = \frac{1}{N} \sum_{\vec{s}_A \in \Omega_{AB}^T} \left(A(\vec{s}_A) - B^T(\vec{s}_A)\right)^2 . \qquad (2.22)$$

The normalizing factor $\frac{1}{N}$ makes the similarity measure invariant with the number of voxels N from the overlapping domain that may change during the registration.

It has been shown in [Viola 1995] that SSD is the best similarity measure for a registration procedure if the images involved differ only by Gaussian noise. However, this is not always the case and in practice the images may contain an intensity bias even if they come from the same scanning device. In this case, the SSD will fail to compare the images properly and therefore cannot lead to a correct registration. In such cases the sum of absolute differences (SAD):

$$SAD = \frac{1}{N} \sum_{\vec{s}_A \in \Omega_{AB}^T} \left| A(\vec{s}_A) - B^T(\vec{s}_A) \right| \qquad (2.23)$$

could be used as an alternative similarity measure. However, the problem connected to the intensity bias cannot be completely solved and therefore more robust similarity measures have been derived.

Similarity measures based on Cross-Correlation

A very robust similarity measure for mono-modal images is the Cross-Correlation coefficient (CC):

$$CC = \frac{\sum_{\vec{s}_A \in \Omega_{AB}^T} \left(A(\vec{s}_A) - \overline{A}\right) \cdot \left(B^T(\vec{s}_A) - \overline{B}\right)}{\left[\sum_{\vec{s}_A \in \Omega_{AB}^T} \left(A(\vec{s}_A) - \overline{A}\right)^2 \cdot \sum_{\vec{s}_A \in \Omega_{AB}^T} \left(B^T(\vec{s}_A) - \overline{B}\right)^2\right]^{1/2}} \qquad (2.24)$$

where $\overline{A}$ and $\overline{B}$ are the mean values of the voxel intensities within the overlapping domain Ω_{AB}^T of the images A and B^T. CC is a very good similarity measure for all the registration applications in which there is a linear relationship between the image intensities, especially because it overcomes the aforementioned intensity bias problem.

Variance of intensity ratios

[Woods *et al.* 1992] proposed a similarity measure based on the ratio image derived from the images to be registered. Initially proposed for registering PET and later on for MR images, the method consists of iteratively optimizing the transformation $\mathcal{T}$ which maximizes the uniformity of the ratio image. This uniformity is quantified as a normalized standard deviation of the voxels in the ratio image, thus the name of the measure is *variance of intensity ratios* (VIR). VIR is mathematically expressed, for the N voxels within the overlapping volume Ω_{AB}^T, as:

$$VIR = \frac{1}{\overline{R}} \sqrt{\frac{1}{N} \sum_{\vec{s}_A \in \Omega_{AB}^T} \left(R(\vec{s}_A) - \overline{R} \right)^2} \qquad (2.25)$$

where $\overline{R}$ is the mean value of the intensity ratios $R(\vec{s}_A)$ over the entire volume of overlap. An intensity ratio at the spatial position $\vec{s}_A \in \Omega_{AB}^T$ is defined as:

$$R(\vec{s}_A) = \frac{A(\vec{s}_A)}{B^T(\vec{s}_A)} \qquad \forall \vec{s}_A \in \Omega_{AB}^T. \qquad (2.26)$$

Later on, [Woods *et al.* 1993] used this measure after minor modifications for the first multi-modality registration of MR/PET image pairs. The multi-modality registration capability comes from the presumption that all the voxels which have a particular intensity value in the MR image represent the same tissue type and therefore the values of the corresponding PET voxels should be similar to each other. The algorithm partitions the MR image into 256 separate intensity levels and finds the spatially corresponding intensity values in the PET image. Therefore, a set of 256 bins (pairs of one MR intensity and the corresponding set of PET intensity levels) is constructed and the algorithm tries to maximize the uniformity of the PET voxel values within each bin by locally minimizing the normalized standard deviation. If A is the MR and B the PET image, then the algorithm minimizes the sum of the normalized standard deviations for all those intensity values $b \in \Omega_b^T$ that are transformed over each of the intensities $a \in \Omega_a^T$:

$$VIR_B = \sum_a \frac{n_a}{N} \frac{\sigma_B(a)}{\mu_B(a)} \qquad (2.27)$$

where:

$$n_a = \sum_{\Omega_a^T} 1$$
$$\mu_B = \frac{1}{n_a} \sum_{\Omega_a^T} B^T(\vec{s}_A) \qquad (2.28)$$
$$\sigma_B(a) = \frac{1}{n_a} \sum_{\Omega_a^T} \left(B^T(\vec{s}_A) - \mu_B(a) \right)^2 .$$

The algorithm was used for registering images of the head. Its performance heavily depends on the binning procedure and it might fail if this intensity binning is not uni-modal. To avoid this situation, the scalp firstly had to be removed from both the PET and MR image before the registration procedure. The method was not very successful with other multi-modal registration cases, but it considerably inspired the research towards the development of other multi-modal similarity measures.

Joint histogram and joint probability distribution

The VIR method proposed by Woods opened the way towards analyzing the behavior of the joint intensity histogram during registration. By partitioning the image intensities into a predefined number of levels, one could build the A and B^T *joint histogram*. The joint histogram is built in a deterministic manner by counting over the entire volume of overlap Ω_{AB}^T the number of the intensity pairs of the voxels falling in the same spatial position in both images. However, [Viola and W. M. Wells 1995] proposed an alternative probabilistic approach to estimate the joint histogram. Denoting the number of intensity levels within Ω_a^T and Ω_b^T with L_a and L_b, the joint intensity histogram $\mathcal{H}_{AB}^T$ is a $L_a \times L_b$ matrix whose elements can be expressed as:

$$\mathcal{H}_{AB}^T(a,b) = \sum_{\vec{s}_A \in \Omega_{AB}^T} \delta_{ab}^{AB^T}(\vec{s}_A), \forall a \in \Omega_a^T \text{ and } \forall b \in \Omega_b^T \qquad (2.29)$$

where $\delta_{ab}^{AB^T}$ represents the sampling function defined as:

$$\delta_{ab}^{AB^T}(\vec{s}_A) = \left\{ \begin{array}{l} 1, \text{ if } A(\vec{s}_A) = a \text{ and } B^T(\vec{s}_A) = b \\ 0, \text{ otherwise.} \end{array} \right. \qquad (2.30)$$

As for all other measures, the transformed image B^T usually has to be resampled according to the discrete reference image domain Ω_A, therefore an interpolation of the voxel intensities is implicitly involved.

By normalizing the $\mathcal{H}_{AB}^T$ to the number of voxels N within the volume of overlap Ω_{AB}^T, the joint histogram approximates the joint probability distribution function (PDF) of the image intensities $p_{AB}^T(a, b)$:

$$p_{AB}^T(a, b) \approx \frac{1}{N}\mathcal{H}_{AB}^T(a, b), \forall a \in \Omega_a^T \text{ and } \forall b \in \Omega_b^T. \qquad (2.31)$$

It is obvious that $p_{AB}^T(a, b)$ depends on the transformation $\mathcal{T}$. This property inspired several research groups to investigate similarity measures derived from the joint PDF of the images. By visualizing the 2D joint PDF for different transformations, one can notice a clustering of the distribution as the images become more aligned. In the special case of mono-modal registration, the joint PDF will be spread along the main diagonal of the $N_a \times N_b$ matrix. In the ideal case of registered images the joint histogram will be reduced to a single line, with some small deviations coming from the imaging noise. The first approaches derived from the joint histogram were attempting to minimize the variance of the intensity clusters, but the most successful similarity measures calculated from the joint PDF are rooted in the information theory.

Joint entropy

Introduced by [Shannon 1948], the concept of *entropy* as a measure of information revolutionized the theory of communication and data transmission. Described as the average information supplied by a set of symbols $\{s\}$ whose probabilities are given by $\{p(s)\}$, the Shannon entropy is defined as:

$$H = -\sum p(s) \log p(s). \qquad (2.32)$$

The entropy can also be considered as a measure of randomness of a signal because it reaches its maximum if all symbols have equal probabilities, and its minimum when all symbols have probability 0 except one. This property of the entropy, together with the properties of the joint intensity PDF inspired the use of entropy as a similarity measure for image registration. For example, [Studholme *et al.* 1995, Collignon *et al.* 1995] proposed to minimize the joint entropy calculated from histogram. This would lead to a minimization of the dispersion in the histogram while approaching the optimal registration position.

For two images A and B, the joint entropy $H(A, B^T)$ is given by:

$$H(A, B^T) = - \sum_{a \in \Omega_a^T} \sum_{b \in \Omega_b^T} p_{AB}^T(a, b) \log p_{AB}^T(a, b) \qquad (2.33)$$

where a and b represent the discrete intensity values of the two images, and therefore the bins of the joint PDF. It is also possible to reduce the number of bins by merging neighboring intensity values, leading to a reduction of the sparsity of the histogram. For example, a very common practice for MR and CT images is to use between 32 and 256 bins to calculate the joint histogram.

The method looks very similar to VIR, but it proved to be more stable and robust as it does not rely on the uni-modal nature of the distribution as the VIR algorithm. Nevertheless, strong limitations of this approach have also been noticed, especially in connection with the sensitivity of the joint entropy to changes in the volume of overlap.

Mutual Information

In 1995 two research groups were working independently on another similarity measure inspired by information theory, namely the *Mutual Information* (MI). In [Viola and W. M. Wells 1995, Maes *et al.* 1996], both groups presented basically the same strategy of *maximizing the MI*, but with different ideas of implementing the rigid registration procedures. Maes proposed to calculate the registration transformation by maximizing the MI using the Powell's direction-set method. In addition, the MI is calculated from the deterministic joint histogram of the images. On the other hand, Viola used a gradient search to maximize the MI. Their solution presents a stochastic Parzen windowing method to estimate the joint histogram and therefore they could express mathematically the MI gradient which is directly used by the optimization procedure.

The MI approach represents an extension of the joint entropy minimization algorithm, overcoming the limitations caused by the dependency on the actual image overlap. The idea was to extend the joint entropy with the individual information carried by the overlap Ω_{AB}^T of the two images A and B. Expressing this information in terms of the marginal entropies $H(A)$ and $H(B^T)$, the MI becomes:

$$I(A, B^T) = H(A) + H(B^T) - H(A, B^T). \qquad (2.34)$$

The marginal image entropies can be calculated as:

$$H(A) = -\sum_{a \in \Omega_a^T} p_A^T(a) \log p_A^T(a)$$
$$H(B^T) = -\sum_{b \in \Omega_b^T} p_B^T(b) \log p_B^T(b) \tag{2.35}$$

with p_A^T and p_B^T representing the marginal intensity distributions of A and B, within Ω_{AB}^T. They can either be calculated directly from the images by normalizing their individual histograms or from the two dimensional joint PDF, calculated as described previously.

It is important to remember that the marginal entropies and the joint entropy of the images are not constant during the registration process. Even though the information comprised in both images remains the same, their volume of overlap Ω_{AB}^T changes with the estimated transformation T. In addition, the voxel interpolation also affects the joint and the marginal PDFs of the images and therefore the histograms, too. Other limitations, related to the statistical consistency of the estimated PDFs are also to be considered, and will be discussed in the next chapter.

By considering the properties of the logarithm function, the definition of MI from Eq. 2.34 can be reformulated in terms of marginal and joint probability distributions as:

$$I(A, B^T) = \sum_{a \in \Omega_a^T} \sum_{b \in \Omega_b^T} p_{AB}^T(a, b) \log \frac{p_{AB}^T(a, b)}{p_A^T(a) p_B^T(b)}. \tag{2.36}$$

MI between two images can also be considered as the amount of information that one image contains about the other, and reaches its maximum when the images are aligned. This can be seen easier if we consider the conditional probabilities $p(b|a)$. The conditional entropy is:

$$H(B^T|A) = \sum_{a \in \Omega_a^T} \sum_{b \in \Omega_b^T} p_{AB}^T(a, b) \log p^T(b|a) \tag{2.37}$$

and using the conditional entropies, MI becomes:

$$I(A, B^T) = H(A) - H(B^T|A) = H(B^T) - H(A|B^T). \tag{2.38}$$

If the conditional entropy in Eq. 2.38 is zero, it means that knowing the intensities $A(\vec{s}_A)$ enables the perfect prediction of the corresponding

intensity value in B^T. Therefore, the registration by maximization of MI involves finding the transformation that makes the image A the best possible predictor of the image B^T within their region of overlap.

The MI was widely studied and a series of properties were proved in the literature [Vajda 1989]. Some of the most important ones are listed in Table 2.2.

$$
\begin{aligned}
\text{Non-negativity:} \quad & I(A, B) \geq 0 \\
\text{Independence:} \quad & I(A, B) = 0 \Leftrightarrow p_{AB}(a, b) = p_A(a) p_B(b) \\
\text{Symmetry:} \quad & I(A, B) = I(B, A) \\
\text{Self information:} \quad & I(A, A) = H(A) \\
\text{Data processing:} \quad & I(A, B) \leq I(A, B^T) \\
\text{Boundedness:} \quad & I(A, B) \leq \min(H(A), H(B)) \\
& I(A, B) \leq (H(A) + H(B))/2 \\
& I(A, B) \leq \max(H(A), H(B)) \\
& I(A, B) \leq H(A, B) \\
& I(A, B) \leq H(A) + H(B)
\end{aligned}
$$

Table 2.2: *Properties of Mutual Information*

MI is more robust than the joint entropy regarding the overlap problem, but it is not completely immune. The volume of overlap of the two images influences MI in two ways. Firstly, a decrease in overlap reduces the number of samples, which decreases the statistical power of the probability density estimation. Secondly, in [Studholme *et al.* 1997, Studholme *et al.* 1999] it is shown that with the increase of misregistration (usually coinciding with a decrease of the overlap volume) the MI measure may actually increase. This can occur when the relative areas of object and background even out and the sum of the marginal entropies increases faster than the joint entropy. Therefore, they proposed a normalized version of the MI which is less sensitive to the volume of overlap:

$$
NMI(A, B^T) = \frac{H(A) + H(B^T)}{H(A, B^T)}. \tag{2.39}
$$

Maes have suggested in [Maes *et al.* 1997] another form of normalization of MI, the *entropy correlation coefficient* (ECC) that is related to the NMI as:

$$ECC(A, B^T) = 2 - \frac{2}{NI(A, B^T)}. \tag{2.40}$$

The introduction of MI as a similarity measure has revolutionized image registration and has especially influenced the development of the intensity-based image registration techniques. The following years of research have proven the advantages and the robustness of this approach, and MI started to be extensively used, even in the field of mono-modal registration. MI assumes that there is a functional relationship between the intensities at the same location in different images. The decisive property of MI is that it does not depend on any linear relation between the image intensities, but it relies only on statistical properties of the two-dimensional histogram. An excellent overview of the methods developed using the MI concept is presented by Maes et al. in [Maes *et al.* 2003].

2.2.3   Voxel interpolation procedures

During the registration procedure, the floating image is transformed from its initial image domain Ω_B into the reference domain Ω_A, more precisely, into the overlapping volume $\Omega_{AB}^T \subseteq \Omega_A$:

$$B : \vec{s}_B \in \Omega_B \longmapsto B^T : T(\vec{s}_B) \in \Omega_{AB}^T \tag{2.41}$$

This image transformation is always required for optimizing the similarity metric between the images while searching for the best registration transformation. For a given transformation $\mathcal{T}$, the intersection of the discrete domains Ω_A and Ω_B is likely to be an empty set because no sample point will exactly overlap. Therefore it is necessary to make an estimation of the image intensities in between the sample positions, also compensating for differences in image samplings $\vec{\xi}^A$ and $\vec{\xi}^B$. This intensity estimation is called *voxel interpolation*. In contrast to the final reconstruction of the registered floating image, which requires highly accurate voxel interpolation, the estimation of the similarity measure during the optimization procedure needs a fast method in order to reduce the computational costs. Such computationally efficient algorithms

may, however, introduce blurring, ringing or other aliasing artifacts. Another important difference between these two cases is the target domain of the interpolation. In the case of the final transformation the interpolation is performed onto the intensity domain Ω_b while for estimating the similarity measure, the voxel interpolation is performed onto Ω_a.

In order to assess and to further use the result of the registration procedure, a final transformation of the registered floating image is usually required. This final transformation of the floating image into the reference image domain (also called floating image reconstruction) necessitates an accurate voxel interpolation method to calculate the intensities of the spatially transformed floating image. As the anatomical scans are not band limited, a perfect interpolation is not possible. The interpolation function proposed in [Hajnal *et al.* 1995] implies a *sinc* function truncated with a suitable Hamming window. Care has to be taken to ensure that the sum of the weights of the truncated kernel is 1, otherwise artifacts may result. Various modifications of the *sinc* interpolation have been proposed, either regarding the Hamming windowing function, or by approximating the windowed *sinc* function with B-spline interpolators. For more details about accurate voxel interpolation methods please refer to [Hajnal *et al.* 2001].

For fast voxel interpolation methods computationally less expensive methods are needed. Even if lacking of accuracy, they can be sufficient to calculate the similarity measure between the images undergoing the registration. While numerous interpolation methods have been proposed in the literature [Hajnal *et al.* 2001], three methods are dominantly used. Schematically depicted in Fig. 2.3 for 2D space, these three methods are: the *nearest neighbor interpolation* (NNI), the *trilinear interpolation* (TRI) and the *partial volume interpolation* (PVI). While the first two methods can be used to approximate the transformed voxel intensities needed to estimate any of the aforementioned similarity measures, the PVI was especially designed in [Collignon *et al.* 1995] to create the joint intensity histogram.

For a transformed position $\vec{s}_A = T(\vec{s}_B)$, the simple NNI technique (Fig. 2.3(a)) considers the intensity of the closest (nearest) voxel:

$$A(\vec{s}_A) \approx A(\vec{s}_A^{(x)}), \text{ where } \vec{s}_A^{(x)} = \min_x(|\vec{s}_A^{(x)} - \vec{s}_A|). \tag{2.42}$$
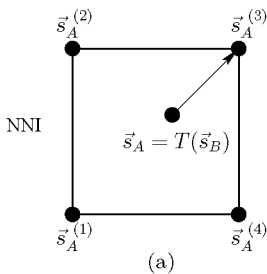
$$\vec{s}_A = T(\vec{s}_B)$$
$$a = A(\vec{s}_A) \approx A(\vec{s}_A^{(x)})$$
$$b = B(\vec{s}_B)$$
$$\mathcal{H}_{AB}^{T}(a,b)+ = 1$$
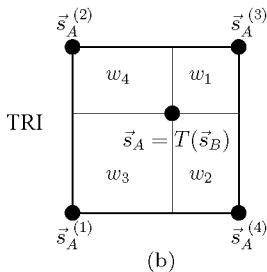
$$\vec{s}_A = T(\vec{s}_B) \sum_x w_x = 1$$
$$a = A(\vec{s}_A) \approx \sum_x A(\vec{s}_A^{(x)})$$
$$b = B(\vec{s}_B)$$
$$\mathcal{H}_{AB}^{T}(a,b)+ = 1$$
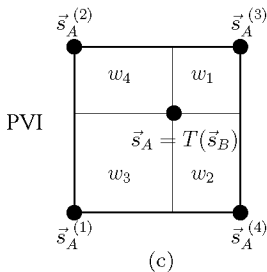
$$\vec{s}_A = T(\vec{s}_B) \sum_x w_x = 1$$
$$a_x = A(\vec{s}_A^{(x)}), \forall x \in \{1,2,3,4\}$$
$$b = B(\vec{s}_B)$$
$$\mathcal{H}_{AB}^{T}(a_x,b)+ = w_x, \forall x \in \{1,2,3,4\}$$

Figure 2.3: *Fast voxel interpolation techniques for 2D and the corresponding expressions used to update the joint image histogram; (a) nearest neighbor interpolation, (b) trilinear interpolation and (c) partial volume interpolation*

The NNI usually leads to significant errors, and therefore TRI is used more often, defining the intensity in a point $\vec{s}_A = T(\vec{s}_B)$ as a weighted combination of the intensities of its neighbors:

$$A(\vec{s}_A) \approx \sum_x w_x \cdot A(\vec{s}_A^{'(x)}), \text{ with } \sum_x w_i = 1. \qquad (2.43)$$

Depicted in Fig. 2.3(b), the TRI considers the weights as linearly dependent from the distance between the current point $\vec{s}_A = T(\vec{s}_B)$ and its neighbors $\vec{s}_A^{'(x)}$. In order to calculate the MI between two images, a more elaborate technique has been proposed, namely the PVI. Using weights which are similarly defined in the case of TRI, the PVI simultaneously updates all the histogram entries corresponding to the transformed point and each of its neighbors. Depicted in Fig. 2.3(c), the PVI technique creates smoother changes of the joint histogram for varying transformations and hence the goal function of the optimization then becomes smoother. Further, PVI does not artificially create new entries in the histogram which is very likely in the case of TRI and that would lead to unpredictable changes in the marginal distributions.

A serious problem related to all fast voxel interpolation methods is that they lead to artifacts in the similarity function. These interpolation induced artifacts may affect the accuracy and even the success of the registration procedure. These aspects will be described in more detail in the next chapter.

2.2.4   Optimization procedures

All voxel-based registration procedures involve the use of a similarity function defined on image intensities. Because no analytical solution exists, a numerical optimization has to be used in an iterative fashion to determine the optimal registration transformation.

As seen in Sec. 2.2.1, the spatial transformation is usually characterized by a certain number of degrees of freedom which correspond to a set of parameters α. This set fully describes the transformation and will be called the *parameter space*. This space is explored by the optimization algorithm such that the similarity measure as a function of α is optimized. Each point corresponds to a different estimate of the spatial

transformation and implicitly an individual estimate of the image similarity measure. The goal of the optimization algorithm is therefore to find the optimum location within this parameter space, given an initial estimate.

All the optimization procedures are based on an iterative approach in which the transformation is gradually refined by trial and error. For each iteration, the current estimate of the transformation is used to calculate the similarity measure. Then, the optimization algorithm searches a better estimate which increases the similarity measure. The stopping criterion of the optimization algorithm is given by predefining a minimum tolerance value for the changes of the similarity measure between consecutive iterations. Optimum search algorithms can be classified according to their dependency on the availability of derivative information. The derivatives of the similarity measure should be easily and efficiently computable in order to limit the computations lower needed by optimization. A detailed description of diverse optimization algorithms can be found in [Press *et al.* 1988].

The most important limitation of all the optimization procedures is the danger to converge to a *local optimum* of the similarity function. These local optima usually mask the global optimum of the function, and can seriously compromise the accuracy of the entire registration procedure. They are usually produced either by interpolation artifacts or by a good local match between the image intensities. Different techniques have been proposed to increase the robustness of the optimization against local optima. For example, in a hierarchical multi-resolution approach, the images are registered with a progressively increasing resolution while converging to the correct estimate of the transformation. A detailed overview of the proposed methods can be found in [Maintz and Viergever 1998, Hill *et al.* 2001, Hajnal *et al.* 2001]. In order to avoid problems with local optima, we also propose a method which will be described in detail in Chap. 4

2.3   Non-rigid registration algorithms

The development of the non-rigid registration algorithms was imposed by the need to compensate for the inherent deformation of the human

anatomy due to different physiological or physical factors. Even though many have tried to develop fixation devices to immobilize the human body in a certain position for successive scanning procedures, the non-rigid nature of the soft tissues cannot be completely overcame. The non-rigid registration procedures can compensate for the local deformations by finding the appropriate deformation field that properly registers the images.

While the rigid registration procedures consist of finding a spatial transformation that remains unchanged over the entire floating image volume, the non-rigid registration procedures are designed to find a dense field of deformation vectors covering the entire image domain, but which are locally adapted to compensate for both spatial displacement and shape differences. Basically, this field of deformation vectors maps each point from the floating image into the corresponding position in the reference image such that all objects in the warped image ideally will coincide precisely with the corresponding objects in the target image. Care has to be given to the pathologically induced changes that can produce a lack of correspondence in the images, such as a tumor growth, or cuttings of a surgical procedure. These changes are an important issue in a non-rigid registration as they can be handled in an inappropriate fashion resulting in big errors during the registration.

A classical voxel-based non-rigid registration procedure follows closely the algorithmic flow presented in Fig. 2.2 for a rigid registration procedure. The differences in implementation are mainly related to the strategy used to build the deformation field. The remaining of this section provides a brief overview of the most important methods proposed, emphasizing hierarchical approaches which are in the focus of this thesis.

2.3.1   Model-based approaches to non-rigid registration

The main difference between the methods proposed for non-rigid registration comes from the representation of the deformation field. The earliest approaches described very accurately the deformation field as a physical phenomena. The partial differential equations (PDE) describing the corresponding processes have been solved by numerical integration

techniques like the finite element method (FEM). Unfortunately, these procedures are computationally very expensive leading to the exploration of less accurate but computationally more efficient alternatives by simplifying the physical phenomena. Like many other problems, the non-rigid registration is often a tradeoff between the accuracy of approximating real behavior of the tissues and computational complexity.

Elastic registration techniques

The first elastic registration technique was proposed by Bajcsy et al. for matching a brain atlas with a CT image of a human patient. Presented in [Bajcsy and Kovacic 1989], the idea is to model the deformation of the source image into the target image as a physical process which resembles the stretching of an elastic material such as rubber. This process is governed by two forces, one *internal* and one *external*. The internal force is generated by the deformation of the elastic material and opposes the external force that deforms the elastic body from its equilibrium shape. As a consequence, the deformation of the elastic body stops if both forces are in equilibrium. The behavior of the body is described as a linear elastic deformation. The external forces applied are usually extracted from the gradient of the similarity measure that can be a local correlation coefficient based on intensities, intensity differences, or structural features such as edges or curvatures. An alternative choice for the similarity measure is the distance between the curves or surfaces of the corresponding anatomical structures. Many extensions of this elastic registration framework have been proposed, such as in [Davatzikos 1996] allowing for spatially varying elasticity parameters, enabling certain anatomical structures to deform more freely than others.

Fluid registration

The internal forces of an elastic transformation limit the registration by excluding large deformations or discontinuities (e.g. produced by a cut). In fluid registration techniques based on Navier-Stokes equations these constraints are relaxed over time. This property makes them especially attractive for inter-subject and atlas registration tasks which have to accommodate large deformations and large shape variability. However,

the probability of misregistrations increases because of the weaker regularization effects. The first algorithm was proposed by [Christensen *et al.* 1995] and later on was computationally improved by [Bro-Nielsen and Gramkow 1996].

FEM based approaches

The finite element method (FEM) is a classical engineering analysis technique for the solutions of PDEs. Basically, the FEM subdivides the domain of interest into an interconnected set of subregions or elements and approximates the solution by a locally continuous function. In the context of image registration, the FEM method was applied in [Bro-Nielsen 1998] and in a simplified version in [Edwards *et al.* 1998] to calculate tissue deformations for image-guided surgery. Later, they proposed a three component model to simulate the properties of rigid, elastic, and fluid structures. For this purpose, the floating image has to be divided into a triangular mesh with n connected nodes ϕ_i. Each node is labeled according to its physical properties of the underlying anatomical structures. While the nodes labeled as rigid are kept fixed, the nodes labeled as elastic or fluid are deformed by minimizing an energy function, which is composed of a number of different energy terms that constrain the deformation.

Optical flow methods

Another very well known registration technique is the so called optical flow method. Introduced in [Horn and Schunck 1981], the concept of optical flow was originally used in computer vision to recover the relative motion of an object and the viewer between two successive frames of a temporal image sequence. Its fundamental assumption is that the brightness of a particular point is constant in time, therefore the motion field can be extracted from the temporal image difference. Smoothness constraints are imposed on the extracted motion field in order to enforce a unique solution to the problem.

2.3.2 Hierarchical non-rigid registration approaches

As previously presented in Sec. 2.3.1, several methods have been pro-
posed to accurately solve the registration problem. These approaches
are estimating the underlying mapping based on a physical analogy like
deformations or flow. In the case that this physical phenomenon is the
underlying cause of the difference between the images and the necessary
parameters are available, these methods represent the ideal approach to
the non-rigid registration problem. However, in many practical cases,
one or both of these conditions are not fulfilled. Therefore, alternative
methods have been proposed for the description of a dense non-rigid
mapping between images. After introducing the basic concepts of piece-
wise image registration techniques, this section presents the most impor-
tant hierarchical strategies which decompose the non-rigid registration
into numerous local matching problems.

Piecewise image registration

The simplest approach proposed for non-rigid registration of the images
is to decompose it into numerous rigid matchings problems which are fi-
nally interpolated to preserve the continuity of the image content. [Little
et al. 1997] uses this strategy for the registration of the spine which is a
flexible structure to a certain extent. By decomposing the spine into indi-
vidual vertebrae which can be rigidly matched, a good approximation of
the deformation of the spine and its surrounding tissue can be obtained
by smoothly interpolating the spatial displacements between the rigid
bony structures. There are also other similar situations, where rigid ar-
ticulated structures are connected together into a deformable structure.
The piecewise approach seems to be the most appropriate and the most
simple method for such cases, when interested only in the vicinity of the
rigid structures.

Based on a similar approach, [Maintz *et al.* 1998] presented a technique
which can achieve a higher degree of deformation by progressively par-
titioning the initial volumes into smaller pieces that are locally rigidly
registered. The deformation can be estimated by interpolating these
local matchings into a dense deformation field. This basic approach
is extended and used in both multi-grid and hierarchical strategies for

non-rigid registrations. These methods are described in the following subsections.

Registration using free-form deformations

The classical voxel-based method using free-form deformations (FFD), presented in [Rueckert *et al.* 1999], was developed for the non-rigid registration of contrast-enhanced breast MR images. Apart from possible pathological factors (as tumor growth or surgery), the deformation of the breast can be induced either by possible changes in patient positioning or by physiological factors, such as respiration or heartbeat. Similar to [Lee *et al.* 1997], Rueckert *et al.* proposed a multi-resolution approach of coarse to fine refinement of the non-rigid motion $\mathcal{T}_{\mathrm{non-rigid}}$ of the breast. Furthermore, the global non-rigid deformation is modeled by an overall affine component combined with local non-rigid transformations, which are described in terms of a FFD based on B-splines [Lee *et al.* 1996]:

$$\mathcal{T}_{\mathrm{non-rigid}} = \mathcal{T}_{\mathrm{affine}} + \mathcal{T}_{\mathrm{B-splines}}. \tag{2.44}$$

The global affine transformation $\mathcal{T}_{\mathrm{affine}}$ compensates for the overall motion of the subject and follows the classical model with 12 degrees of freedom. The local transformations $\mathcal{T}_{\mathrm{B-splines}}$ model the local changes of the breast. The FFD proved to be a powerful tool for modeling 3D deformable objects and they have been previously also applied to tracking and motion analysis in cardiac images [Bardinet *et al.* 1996]. The basic idea of a FFD is to deform an object by manipulating an underlying mesh of control points. The resulting deformation changes the shape of the 3D object and produces a smooth and continuous transformation. Rueckert et al. proposed the multi-resolution approach to achieve maximal flexibility with reasonable computational cost. At every level l of the hierarchy, the resolution of the control mesh is increased, along with the image resolution, in a coarse to fine fashion. Consequently, along the hierarchy an entire series of local transformations are estimated which will finally define the non-rigid component of the deformation:

$$\mathcal{T}_{\mathrm{B-splines}} = \sum_l \mathcal{T}_{\mathrm{B-splines}}^{(l)}. \tag{2.45}$$

The cost function underlying the entire registration process is a combination of a classical voxel similarity measure with a smoothness term

resulting from the transformation model. The image similarity is given
by the classical normalized MI (Eq. 2.39), because it is more robust to
changes in the volume of overlap as shown in [Studholme *et al.* 1999].
Unfortunately, an empirical smoothness term is used having an increas-
ing importance with the increasing resolution of the control point mesh.

The adaptive bases algorithm

In order to reduce the computational complexity of the existing methods
for non-rigid registration based on maximization of MI over a regular grid
of splines, [Rohde *et al.* 2003] proposed a method that spatially adapts
the transformation. This spatial adaptation allows the reduction of the
number of degrees of freedom in the overall transformation, thus speeding
up the process and improving its convergence properties. Their method
is also relying on a hierarchical strategy in which the global nonrigid
registration problem is partitioned into several smaller ones. In order to
ensure the physical consistency of the deformation field, they developed
several constraining schemes derived from the Jacobian matrix. Finally,
the deformation field is expressed as a combination of symmetric basis
functions. Beside this deformation model, they introduced another key
element which is a method to identify regions that are poorly registered,
and whose transformation needs to be improved. This way they avoid
useless computation for those regions that are already correctly regis-
tered. The hierarchical approach they propose is basically similar to
the previous FFD method, except for the function used to express the
deformation. They also used a coarse to fine strategy to refine the defor-
mation field, which is again a sum of the intermediate transformations
obtained at each level of the hierarchy.

2.4   The hierarchical image subdivision strategy

Extending the approach of [Maintz *et al.* 1998], Likar and Pernuš pre-
sented a fully functional algorithm for registration of microscope images
of muscle fibres. Their approach named *the hierarchical image subdivi-
sion strategy* consists of decomposing the non-rigid registration problem
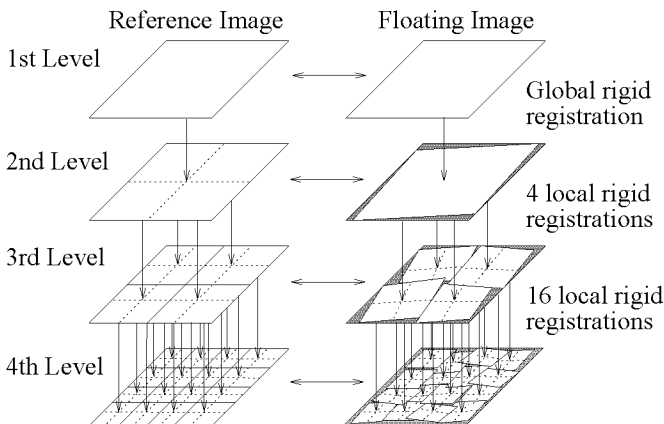
Figure 2.4: *The hierarchical subdivision scheme*

into numerous local rigid registrations of small sub-images. Figure 2.4 depicts the hierarchical strategy as presented in [Likar and Pernuš 2001].

Likar and Pernuš' idea was to reduce the complexity of the existing non-rigid registration methods that use complicated functions to model the spatial transformation. In all the aforementioned approaches (see Sec. 2.3), the deformation field is described in terms of a high number of parameters, allowing numerous degrees of freedom for the spatial transformation. Furthermore, the optimization procedure may get easily trapped into local optima of the similarity function. To overcome these problems, Likar and Pernuš proposed a local approach in which the images to be registered are progressively subdivided into smaller sub-images which are locally registered, by maximizing MI. While the local registrations are following the affine transformation model, the global continuous and smooth deformation field is finally estimated using thin plate splines interpolation (TPS). The hierarchical image subdivision they proposed creates successive resolution levels. Each level l consists of the image partitioning phase, the local registration of all sub-images, and a TPS interpolation which maps the floating image into the reference image domain. The partitioning phase is simultaneously done on both the reference and the floating images, resulting in $2^{2(l-1)}$ sub-image pairs. In order to avoid the problems of MI related to changes in the volume of

overlap, Likar and Pernuš proposed the normalized MI as a similarity measure, as it has been proven to have more robust behavior according to [Studholme *et al.* 1999].

The basic idea is simple and elegant, but additional constraints and control mechanisms had to be integrated at each hierarchical stage to ensure the physical consistency and the accuracy of the final deformation field. The performance of the entire procedure heavily depends on accuracy of the local registration of the sub-images as a local misregistration of a sub-image pair at a certain level can propagate further to successive levels due to the TPS interpolation. Moreover, the probability of local misregistrations is increasing along the hierarchy, as the matching capability of the NMI is decreasing with the size of the sub-image pairs ([Pluim *et al.* 2000]). The matching power of NMI relies on the statistical consistency of the local joint intensity histogram which, calculated in the deterministic manner presented in [Maes *et al.* 1996], will contain less and less image samples with the decreasing size of a sub-image pair. Consequently, the estimate of NMI for small sub-image pairs will deteriorate, due to both the imperfection of the voxel interpolation and the reduced number of samples used to estimate the joint histogram. The solution Likar and Pernuš proposed to overcome this problem was to combine the locally estimated distribution of the joint intensity probabilities with prior information. This increases the statistical consistency of the NMI, and therefore enables the correct registration of finer details. However, the use of prior information does not entirely ensure the accuracy of the local registration, and spurious local misregistrations are still possible. Therefore, to increase the overall registration accuracy, they propose a threefold local registration consistency test and a correction of the pixel intensity at every level of the hierarchy before performing the TPS image reconstruction.

2.4.1   Algorithmic implementation

The algorithm designed by Likar and Pernuš' was implemented and tested on 2D images. The local registration of all sub-images is done using the affine transformation maximizing NMI. The first level of the hierarchical strategy consists of a global registration of the floating image $B = B^{(0)}$ with the reference image A. After finding the optimal global

affine registration transformation T, the floating image is warped into the reference image coordinates. At the second level of the hierarchy, the reference image A and the floating image $T^{(1)}(B^{(1)})$ are divided into four equal sub-image pairs $\{A_i^{(2)}, T^{(1)}(B_i^{(2)})\}$, $i = 1..4$ that are individually registered. The individual affine transformations $T_i^{(2)}$ obtained after registering each sub-image pair is then used to warp the centers of the floating sub-images $T_i^{(2)}(B_i^{(2)})$, and to construct a set of control points for the following TPS interpolation which is used to build a dense deformation field $T^{(2)}$. This deformation field is finally applied on the floating image to reconstruct the second level registered image $T^{(2)}(B^{(2)})$. The hierarchical algorithm continues similarly for the subsequent levels, each stage starting with the image partitioning followed by local registrations and finalizing with the image reconstruction. The image partitioning of an individual level l of the hierarchy results in $2^{2(l-1)}$ sub-images. The entire hierarchical splitting is finished as reported in [Likar and Pernuš 2001] at the fourth level, where for input images of size 760×512 pixels are divided into 64 sub-images of 95×64 pixels.

The local joint and marginal entropies of every sub-image pair are calculated based on the normalized local joint histogram, which is calculated using 64 bins and the partial volume voxel interpolation within the overlapping region of the sub-image pair $\{A_i^{(l)}, T_i^{(l)}(B_i^{(l)})\}$, $i \in \{1..2^{2(l-1)}\}$.

2.4.2   Incorporating prior information

In order to overcome the loss of statistical consistency and the increasing artifacts during the hierarchical splitting, Likar and Pernuš proposed several adjustments of the local registration algorithm. The voxel interpolation induced artifacts can be minimized by using random resampling of one of the images in order to decorrelate the grid of the two images, which are the same because of the image reconstruction phase at each hierarchical stage. They also proposed to increase the statistical consistency of the MI by incorporating the prior information $p^{\star}(A, B^{\star})$ into the local intensity probability distribution $p_0(A_i^{(l)}, B_i^{(l)})$:

$$p_i^{(l)}(A_i^{(l)}, B_i^{(l)}) = \lambda \cdot p_0(A_i^{(l)}, B_i^{(l)}) + (1 - \lambda) \cdot p^{\star}(A, B^{\star}) \qquad (2.46)$$

where $\lambda \in [0, 1]$ is a weighting parameter defining the tradeoff between the floating and the prior probability. For $\lambda = 1$ the joint probability is

defined in the classical manner with no prior information. For $\lambda = 0$ the joint probability is not anymore a function of the transformation $T_{l,i}$, and the optimization of the MI is no longer possible. Therefore, a trade-off is proposed, including an adaptation of λ according to the size of the sub-image. The prior information $p^\star(A, B^\star)$ can be obtained once from a pre-registered training set of images if the intensity distributions of the training set is similar to the images undergoing the actual registration. However, this is not always the case, so they also proposed an alternative solution that estimates the prior information from the coarsely pre-registered images at the previous level.

2.4.3   Consistency of the local registrations

If no constraints are imposed on the local registrations, the consistency between nearby displacements could be enlarged. As the accuracy of the registration at each level of the hierarchy is crucial for the success of the local registrations at the subsequent level and consequently for the precision of the entire registration procedure, local misregistrations should be detected and eliminated prior to the elastic interpolation. Likar and Pernuš proposed a triple consistency test for the local registrations. In addition to the classical geometrical consistency test that can detect big spatial mismatches, they use a similarity and an optimum distinctiveness test. These tests detect sub-images whose similarity function is small and inconsistent, and analyze the graph of the similarity measure around the optimal position found.

2.4.4   Elastic interpolation

At every level of the hierarchical algorithm, TPS are used to define the overall transformation using the centers of those sub-images passing all the three consistency tests as control points.

Thin-plate splines represent a special family of splines based on radial basis functions [Bookstein 1989]. They approximate the physical bending

of a thin metal plate, given a set of control points. The spline is a linear combination of n radial basis functions $\theta(s)$:

$$t(\vec{w}) = \rho_1 + \rho_2 w_x + \rho_3 w_y + \rho_4 w_z + \sum_{i=1}^{n} \epsilon_i \theta(|\vec{\phi_i} - \vec{w}|) \qquad (2.47)$$

where $\vec{\phi_i}$ denotes the location of the n control points in the floating image. The most common selection of the basis function $\theta(s)$ is:

$$\theta(s) = \left\{ \begin{array}{ll} |s|^2 \log(|s|) & \text{in 2D} \\ |s| & \text{in 3D.} \end{array} \right. \qquad (2.48)$$

The Eq. 2.47 defines a mapping of a 3D point $\vec{w}$ into a 1D coordinate $t(\vec{w})$. By introducing three separate thin plate spline functions $\mathcal{T} = (t_x, t_y, t_z)^T$, a mapping between 3D images can be defined in which the coefficients ρ_i characterize the affine part of the transformation. These coefficients can be calculated by using the set of $3n$ linear equations given by the initial interpolation conditions, which impose that the control points $\vec{\phi_i}$ in the source image are mapped onto their corresponding counterpart $\vec{\phi_i'} = \mathcal{T}(\vec{\phi_i})$ in the target image. To determine the $3(n+4)$ coefficients uniquely, 12 additional equations are required. These 12 equations guarantee that the sum of the non-affine coefficients ϵ_i and their cross-product with $\vec{\phi_i}$ are zero. In matrix form this can be expresses as:

$$\left(\begin{array}{cc} \boldsymbol{\Theta} & \boldsymbol{\Phi} \\ \boldsymbol{\Phi}^T & \mathbf{0} \end{array} \right) \left(\begin{array}{c} \vec{\epsilon} \\ \vec{\rho} \end{array} \right) = \left(\begin{array}{c} \boldsymbol{\Phi}' \\ 0 \end{array} \right) \qquad (2.49)$$

where $\vec{\rho}$ is a 4×3 vector of the affine coefficients ρ_i, and $\vec{\epsilon}$ is a $n \times 3$ vector of the non-affine coefficients ϵ_i, $\boldsymbol{\Theta}$ is the kernel matrix comprising the basis functions with $\Phi_{ij} = \theta(|\phi_i - \phi_j|)$. Solving the system for $\vec{\rho}$ and $\vec{\epsilon}$ results in a thin-plate spline transformation allowing to interpolate the displacements of the control points.

Modeling the deformation by thin-plate splines has several advantages. For example, they can be used to incorporate additional constraints into the transformation model such as rigid bodies [Little *et al.* 1997]. It is also possible to use an approximation solution, where the degree of approximation at the landmark position depends on the confidence of the landmark localization [Rohr *et al.* 1996].

2.5   Discussions

The algorithm introduced by Likar and Pernuš offered a starting point for our development which builds on a few key elements of their method. Firstly, the hierarchical image subdivision strategy represents a good idea for decomposing the curved transformation, but certain disadvantages still need to be addressed. The statistical consistency of the joint histogram diminishes along the hierarchical splitting because of the reduction of the number of samples available in the sub-images. This problem, simultaneous with the increasing effects of the interpolation artifacts is only partially solved by including the prior information and by random resampling one of the images. This thesis will investigate possible methods for further improvement of the algorithms performance. Moreover, our investigation of the MI as a similarity measure has revealed a peculiar behavior that can seriously deteriorate the accuracy of any MI based registration procedure.

3

Mutual Information and the Hierarchical Image Splitting

This chapter presents an analysis of the difficulties with MI which emerge when using the hierarchical image subdivision scheme for non-rigid registration of multi-modal images. Several limitations have already been recognized in the literature in connection to either interpolation artifacts or the statistical consistency of MI. Our investigation of the MI behavior revealed additional problems when using it as a similarity measure for registering noisy patches showing no clear image structure. The study of the MI properties was done using a framework similar to the hierarchical image subdivision scheme proposed by Likar and Pernuš which was presented in the previous chapter. In this context, the drawbacks of the MI behavior become increasingly serious during the image subdivision process due to the decreasing number of samples used to compute the two-dimensional joint intensity histogram. We also propose solutions to overcome these limitations, solutions which lead to a considerable increase in the robustness and accuracy of the entire non-rigid registration procedure. Firstly, we propose a new method to detect noisy patches or regions of homogeneous intensity within an image. Based on this method, we developed a reliable stopping criterion for the entire hierarchical subdivision procedure. Secondly, we propose to switch from MI to the CC as a similarity measure at higher levels of hierarchy, when the MI becomes unstable and statistically unreliable. This change of the similarity measure is made possible by a new method for mapping the intensities of the initial multi-modal images into a common pseudo-modality that the CC can successfully register. Moreover, incorporating

these solutions in the hierarchical strategy for image registration can drastically reduce and simplify its computational complexity.

3.1   Limitations of Mutual Information

Even though many studies have proven the robustness and the advantages of the maximization of the MI criterion, problems have been identified in connection with either interpolation artifacts or inherent limitations of the MI's statistical consistency. After presenting an overview of those limitations that are already reported in the literature, this section describes the results of our study of the MI's properties in connection to the calculation of parameters for the discrete intensity probability distribution estimated by histograms.

3.1.1   Interpolation artifacts

A major accuracy limitation of an image registration procedure stems from the fact that images are not exact replicas of the real world, but discretely sampled versions. In the context of a voxel based registration procedure, the series of successive spatial transformations applied on the floating image to calculate the similarity measure requires an estimation of the reference image intensities in the continuous space. For this, a voxel interpolation method is used to calculate intensities of the reference image between the sampling grid. This leads to inaccuracies when estimating MI called *interpolation artifacts* and they appear as a pattern of successive local maxima and minima in the registration function, as can be seen in Fig. 3.1. As already broadly recognized and described in the literature by many authors (eg. in [Maes *et al.* 1997, Pluim *et al.* 1999, Pluim *et al.* 2003, Tsao 2003, Ji *et al.* 2003]), the interpolation artifacts prohibit sub-voxel accuracy of the registration procedure. Many researchers have tried to overcome this problem which affects the classical MI as in its normalized version in a similar fashion. Accordingly, in this chapter we will discuss only the case of the classical definition (Eq. 2.34) of MI.

In order to visualize the effects of the interpolation artifacts on MI, we consider the following simple registration scenario of a pair 2D transver-

sal slices from CT scans of the leg imaged in a perfectly identical position. Figure 3.1 depicts the test images, acquired before and after injecting contrast agent in order to better visualize pathologies of the blood vessels. The two CT images are of 512×512 pixels with 256 grey value levels. Figure 3.1(a) depicts the reference CT image A, and Fig. 3.1(b) depicts the floating image B. To visualize the interpolation artifacts, the experiment consists of calculating the MI from the entire overlapping region of the floating image B that is translated along the x and y direction over the reference image A. Figure 3.1(c) and 3.1(d) clearly show the interpolation artifacts of MI when using the trilinear and partial volume interpolation, respectively. These artifacts appear between the sampling grid positions as a local extrema (maxima or minima) of the similarity measure and they may disturb the optimization procedure compromising the entire registration accuracy.

These artifacts are due to the computation of the joint and marginal intensity probability distributions that are estimated from the joint intensity histogram of the images. Initially, when the images are aligned, the grid points of A and B also coincide. Therefore, no interpolation is needed when estimating the joint intensity histogram. At the same time, the dispersion of the histogram is minimal when the images are registered and therefore the joint entropy is minimal. By translating the floating image B with an integer number of the voxel dimension, the grid points of the two images will be again aligned avoiding the need for interpolation, but the dispersion of the joint histogram is increasing due to misregistration, reducing the MI accordingly. For all other translations, corresponding to some fraction of voxel dimension, the grid points of the images do not coincide anymore and therefore interpolation is required to estimate intensity values between grid positions of the reference image. For example, when PVI is used to create the joint histogram, several entries have to be updated each time a pair of grid points is not aligned. Consequently, the joint histogram is not only dispersed because of the image content and a possible misregistration, but it also contains an additional dispersion induced by the interpolation method. More dispersion implies a higher joint entropy value, which in turn decreases the MI. This is illustrated in Figure 3.1(d), where the floating image was translated up to 3 voxels in the image plane from the registered position. By repeating the same experiment using the TRI, an opposite behavior of the interpolation artifacts can be noticed on the

(a)

(b)

(c)

(d)

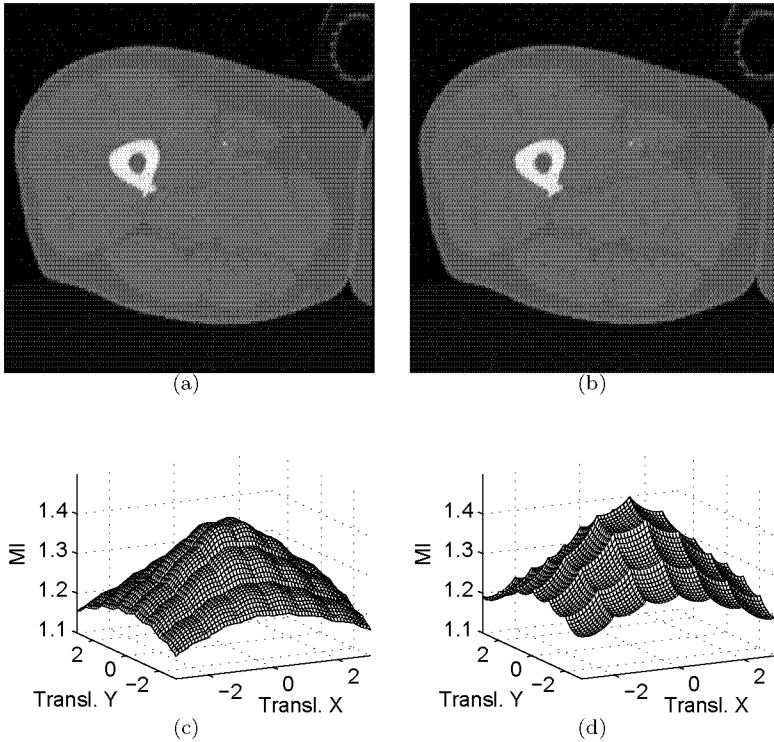**Figure 3.1**: (a) Reference CT image of the leg. (b) The floating perfectly aligned angiographic image. Interpolation artifacts in the MI estimated from the volume of image overlap when translating the floating image along the x and y axis over the reference image, (c) using trilinear interpolation and (d) using partial volume interpolation. The x and y displacements are given in voxel dimensions

response of MI. It means that when the grid points of the two images coincide, the joint entropy is minimal, while for all other translations it increases. This effect is induced by the trilinear nature of the interpolation process which leads to spurious intensity values, or may produce an artificial blurring of the reference image which tends to equalize the joint probability distribution and therefore increases the joint entropy.

The cause of these phenomena lies in the definition of the MI in Eq.2.34. This mathematical expression is based on the marginal and joint image entropies. The Shannon entropy is a nonlinear function that measures the disorder of a signal. Figure 3.2 shows the plot of a simple term in Eq. 2.32 as a function $f(p) = -p \cdot \log(p)$ together with the identity function $f_i(p) = p$. Due to the sharp fall of $f(p)$ near to $p = 0$, the entropy term $f(p)$ becomes smaller than the sum of the entropy terms obtained by dividing p into a multitude of probabilities $p = \sum p_i$. Accordingly, the entropy of a probability distribution will increase as the dispersion of the distribution increases. Having this in mind, and considering that in case of computing the Mutual Information as $MI(A, B) = H(A) + H(B) - H(A, B)$, the source of the artifacts can be easily explained. The entropy of the floating image $H(B)$ varies only because of changes in the volume of image overlap, no voxel interpolation is needed. Meanwhile, the entropy of the reference image $H(A)$ changes due to both interpolation effects and modifications in the volume of overlap. At the same time, these effects have an even greater influence on the joint entropy $H(AB)$, because it is calculated from a much more fragmented probability distribution then its marginal distributions. Therefore, the behavior of MI is dominated by the behavior of the joint entropy.

In a similar fashion, it has been shown in [Pluim *et al.* 2003, Tsao 2003, Ji *et al.* 2003] that the interpolation artifacts are also present in more complex voxel interpolation functions. The effect of the number of bins used for calculating the joint histogram is of big importance, as a reduction of the number of intensity levels used by the joint histogram may reduce these interpolation artifacts. However, the accuracy of the registration is also proportionally decreasing and therefore a compromise is needed depending on the nature of the images to be registered.

A very common solution proposed to minimize the interpolation artifacts is the random resampling of one of the images. In this way, the
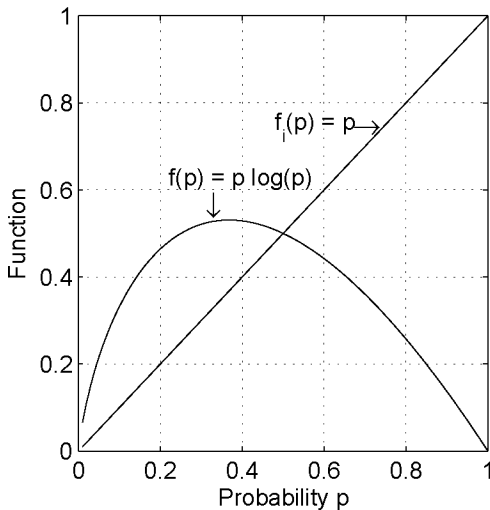
Figure 3.2: *The Shannon entropy term $f(p) = -p \cdot \log(p)$ as function of p and the identity function $f_i(p) = p$*

grid points of the images are not correlated anymore and according to the aforementioned observations the joint intensity histogram is always calculated by using voxel interpolation. Likar and Pernuš also adopted this solution for their hierarchical subdivision scheme. Further, because the size of the sub-images to be registered become smaller with each level of the hierarchical splitting, these interpolation artifacts are increasingly disturbing. Therefore, beside the random resampling strategy, they artificially increase the number of samples by incorporating prior information. However, our experiments showed that this prior information leads only to a scaling of both, the MI value and the interpolation artifacts. In addition, the use of the prior information is likely to affect even more seriously the accuracy of the local registrations because it may introduce false maxima in the MI function.

3.1.2   Mutual Information of small sub-images

As a consequence of the successive image splitting, patches of low structural content may appear. These structureless patches often lead to morphologically inconsistent local registrations due to a low MI response. Likar and Pernuš suggested to identify such patches by applying a threshold on the MI value and to exclude them from the local adjustment process. However, two problems are likely to arise when using this approach. Firstly, in this way structured sub-images with low MI values, that nevertheless have a clear optimum and can be registered in a morphologically consistent way will be prevented from becoming properly adjusted. Secondly, we have observed that MI significantly increases when structureless patches start to overlap a structure in the reference image. Therefore, these patches may not be eliminated by the simple threshold criterion and will deteriorate the registration performance.

Another consequence of the hierarchical image splitting is that the interpolation artifacts present in the MI function are also increasing. Figure 3.3 illustrates this effect on a simple CT/MR registration experiment. Figure 3.3(a) shows the initial reference CT and Fig. 3.3(b) the floating MR image. By translating the floating image horizontally over the reference image, the response of the MI is calculated and plotted in Fig. 3.3(c). The remainder of Fig. 3.3 depicts the same experiment by successively splitting the floating MR image (the central column) and calculating the response of the MI (the right column) as a function of horizontal translation over the reference CT region (the left column).

3.1.3   The effect of noise on Mutual Information

It is well known from information theory [Cover and Thomas 1991], that if two signals are statistically independent then their MI is reaching its minimum possible value, namely zero. Therefore, one would expect that by shifting a structureless sub-image around its initial position, the similarity measure has a small response. Surprisingly, experiments clearly demonstrated that even though MI is small, it starts to increase as soon as the structureless sub-image overlaps a region of higher structural content. This is a serious problem as these structureless sub-images are very
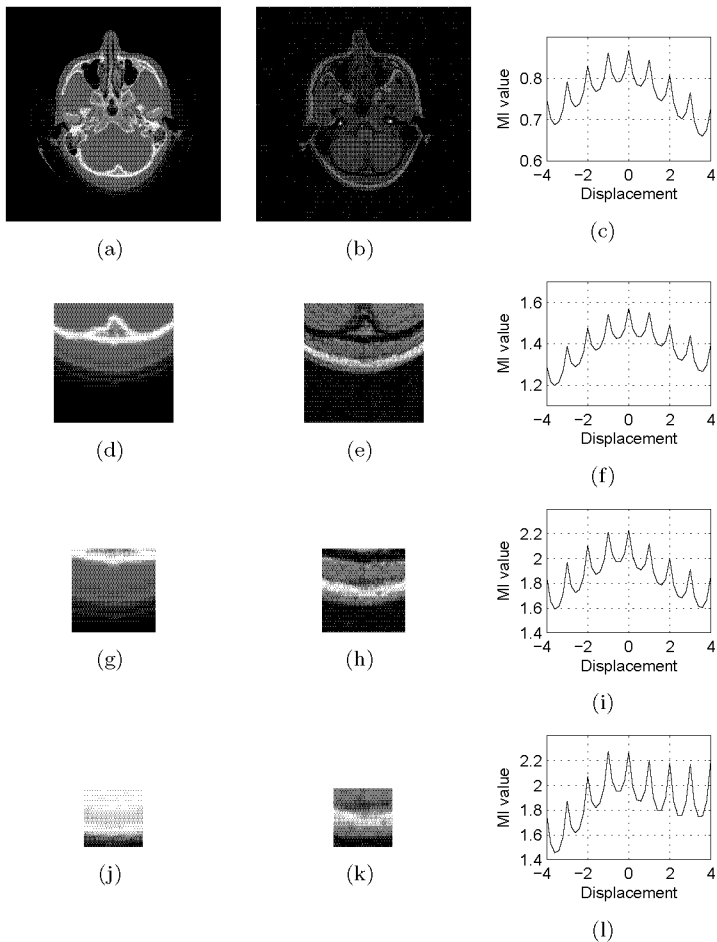
Figure 3.3: *Experiment showing the increasing PVI artifacts along the hierarchical image splitting. (a) The reference CT. (b) The floating MR image. (c) The response of MI when the floating MR image is horizontally translated over the reference CT image. The remainder of the figure shows the same experiment by successively splitting the floating MR image (the central column) and calculating the response of the MI (the right column) as a function of horizontal translation over the reference CT region (the left column)*

likely to appear at higher levels of the hierarchy when either the background or large tissue regions are partitioned into homogeneous patches. Therefore, wrong local registrations may be introduced that would compromise the accuracy of the non-rigid registration result. These local misregistrations are likely not to be detected and corrected by a deformation field regularization procedure based on simple tests as it is proposed in [Likar and Pernuš 2001]. This problem is even more pronounced in the context of multi-modal image registration when not all tissue details can be seen in all modalities. For example, in the case of CT/MR registration this problem is quite likely to appear, because the MR image shows structural details in soft tissue regions, while the corresponding area in the CT acquisition can be nearly or completely homogeneous due to minimal differences in X-ray absorption. The example shown in Fig. 3.4 illustrates how such structureless sub-images can induce important local misregistrations at the 6th level of the hierarchy. The experiment was performed on one selected region of interest of the CT/MR matching example shown in Fig. 3.3(a,b). Figures 3.4(a,b) show the region of interest around the sphenoid sinus in the left temporal bone in the reference CT and floating MR images, respectively. Figure 3.4(c) shows the result when all sub-images are undergoing the local registration. Figure 3.4(d) shows the result when only those sub-images having a clear structure are locally registered, while the structureless sub-images remain in their initial position.

In order to understand and explain this behavior of MI for structureless image patches, further experiments have been performed for 1D signals, as the observation on them can be easily extended for 2D or 3D images. The following experiment shows the behavior of MI in the presence of noise. Let us consider two signals A and B as depicted in Fig. 3.5. We generated the reference signal A by adding white noise to a step function. The floating signal B consists of white noise, and is statistically independent of A. Using Equation 2.34, we can calculate the MI between the two signals as a function of the displacement when the floating signal is translated along the reference signal.

The none-zero baseline of the MI, clearly identifiable in Fig. 3.5(b), can be explained by a combination of two different effects. One is rooted in the difficulty to achieve full independence between signals represented by a finite number of discrete samples. In the ideal case of statistically inde-
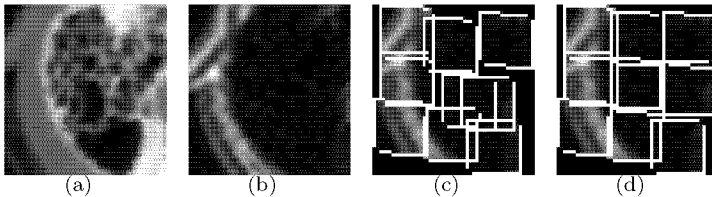
Figure 3.4: *Registration details of the sphenoid sinus in the left temporal bone at the 6$^{\text{th}}$ level of the hierarchy where the original floating image is divided in 32 × 32 sub-images of 16 × 16 pixels. (a) The examined region on the reference CT. (b) The examined region consisting of 3 × 3 sub-images of the floating MR image. (c) The final position of each sub-image after the local rigid registration. (d) The result after applying the local rigid registration only to those MR sub-images having a clear structure, keeping the position of the structureless sub-images unchanged. The structure consistency check clearly prevented the two middle patches from being pulled towards structures in the reference CT*
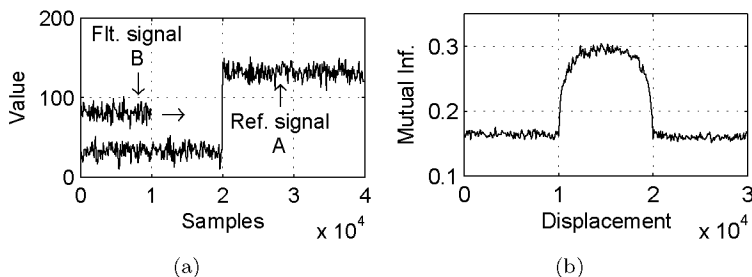
Figure 3.5: *One-dimensional experiment showing the behavior of MI in the presence of noise (a) Original test signals: the reference signal A and the floating signal B. (b) The response of MI when the floating signal B is shifted over the reference signal A*

pendent signals A and B the entropies would cancel out resulting in zero MI. This is, however, very unlikely for a real case scenario. On the other hand it is well known in information theory [Cover and Thomas 1991] that at the transition from a continuous differential entropy to a discrete entropy there is systematic bias by an error term $log_2(\Delta)$ depending on the size of the quantization bins used for histogram generation:

$$\lim_{\Delta \to 0} (H(x^\Delta) + log_2(\Delta)) = h(x) \tag{3.1}$$

where $H(x^\Delta)$ is the discrete entropy, $h(x)$ the differential entropy, and Δ the size of the quantization bins used for histogram generation. This theorem only applies to the marginal entropies $H(A)$ and $H(B)$. We are not aware of any results on deriving a similar relation for the joint entropy $H(A, B)$. Clearly for strictly independent signals A and B the quantization error of the discrete entropy would cancel out. This is, however, not the case if the independency condition is perturbed.

To further investigate the problem, numerical experiments have been performed showing the dependency of the entropies on the sample size. The results are shown in Fig. 3.6. The Cross-Correlation (CC) graph clearly shows, as expected, that the statistical independence between the two finite, discrete random signals improves with their sample length. On the other hand, the graph on the left shows a very interesting property of the entropies. While the sum of the marginal entropies $H(A) + H(B)$ is only slightly influenced by the sample length and very quickly reaches the theoretically predicted value for discrete entropy, the joint entropy $H(A, B)$ requires substantially more samples to show a similarly stable behavior. In other words, the mutual entropy is much more sensitive to deviations from independency between the signals to be matched.

Relating this observation to the test signals from Fig. 3.5 it is obvious, that once the floating signal B starts to overlap the step in A, we get a bi-modal distribution for both $H(A)$ and $H(A, B)$, while $H(B)$ remains constant. The number of available samples then needs to be distributed among two separate intensity clusters for the marginal entropy $H(A)$ and the joint entropy $H(A, B)$. As can be clearly seen from the above graph, the joint entropy $H(A, B)$ decreases much faster than the marginal entropy $H(A)$, thus leading to the observed strong increase of the MI.
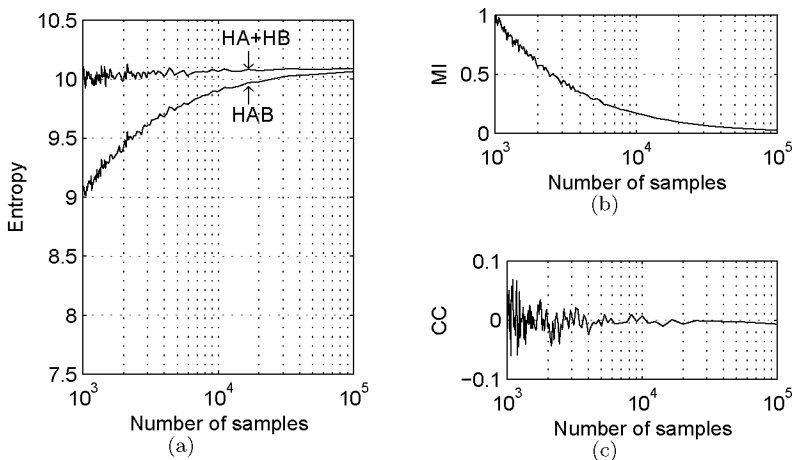
Figure 3.6: (a) The dependency of the entropies $H(A) + H(B)$ and $H(A, B)$ on the number of samples in the signal. The corresponding dependency of (b) MI and (c) CC on the number of samples in the signal

3.1.4   Including the prior global information

The previously described negative effects on MI depend on both the number of image samples and the image content and can be considerably reduced by including prior information in the actual joint probability distribution, as proposed by Likar and Pernuš. This method artificially increases the number of samples used to estimate the joint intensity histogram and consequently increases the statistical power of MI. However, care has to be taken when choosing the weighting parameter, because fake maxima can be created in the response of MI by suppressing the local statistics. Accordingly, they proposed an adaptive scheme for the weighting parameter, such that it equals the ratio between the current sub-image area and the entire image surface. Figure 3.7 shows in detail two representative cases for the local registration of CT/MR images of the head. Marked with a square on both the reference CT and the floating MR image, the selected sub-images are of 32 × 32 pixels and they correspond to the 5th hierarchical level. The first experiment consists of

registering sub-images containing clear anatomical structures, as can be seen in Fig. 3.7(c,d). The second experiment, depicted in Fig. 3.7(e,f), represents a difficult registration case where the sub-images have homogeneous intensities. While the first example should be an easy case of registration using MI because of the well defined structures present in the sub-images, the second example is one of the aforementioned difficult cases that revealed the peculiar behavior of MI. Both experiments consist of calculating the response of MI to horizontal translations of the selected floating sub-images over the reference image. The MI is estimated for both cases with and without including the prior information extracted from the images.

A comparison between Fig. 3.7(g) and Fig. 3.7(h) shows two advantages of including the prior information in the estimation of the joint intensity distribution. The first obvious advantage is that the interpolation artifacts are drastically reduced. At the same time, the MI statistical consistency has considerably increased. This second observation is less obvious in the two responses of the MI but is clearly visible in the colored images that show the final local registrations results overlayed. Depicted in Fig. 3.7(i,j), the colored images show the reference image in red, overlayed with the floating sub-image in green, translated according to the optimal MI value estimated without and with prior information, respectively.

The second experiment is similar to the previous one and the results are depicted in Fig. 3.7(k-n). Even though the previous experiment showed increased statistical consistency of MI when it includes the prior information, the approach is not completely solving the related problems. With or without prior information, the responses of MI depicted in Fig. 3.7(k,l) show statistical inconsistency when is estimated for signals with homogeneous intensity values. In the context of the hierarchical image splitting strategy, this disturbing inconsistency of the MI is likely to introduce local misregistrations which can seriously affect the performance of the final result. Therefore, we continued our investigation to find alternative solutions to overcome these drawbacks of MI.
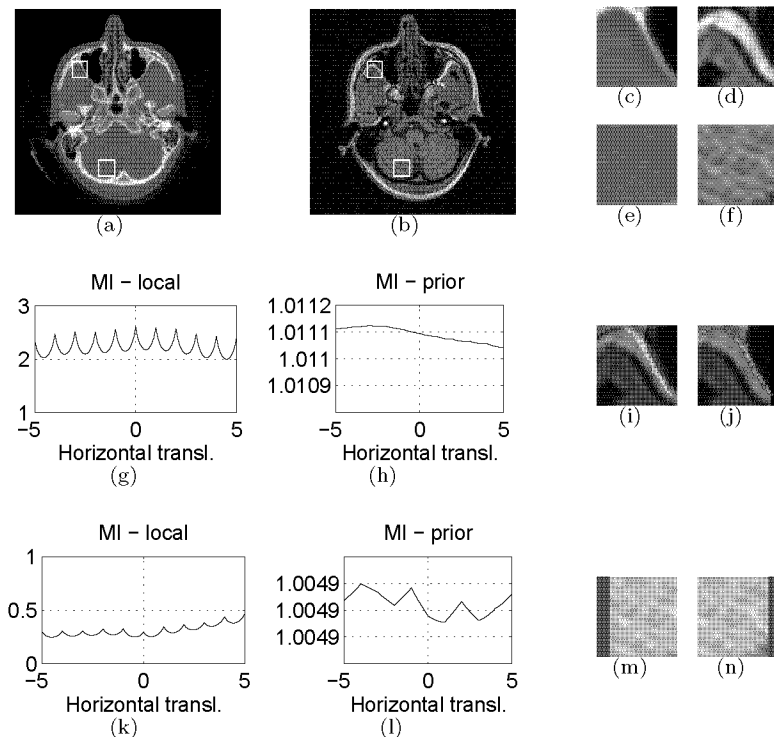
Figure 3.7: *Experiment showing the advantages of including the prior information when estimating the MI. (a,b) The reference CT and the floating MR image, respectively. (c,d) The regions of interest having clear structures and (e,f) with homogeneous intensities, cropped from the CT and MR images, respectively. (g) The response of MI estimated only from the local information and (h) including the prior information to horizontal translations up to ±5 pixels of the floating sub-image over the reference CT. (i,j) The colored overlayed images of the reference image (red) with the floating sub-image (green) transformed according to the optimal value of the MI depicted in (g,h), respectively. (k-n) Similar to (g,j), the experiment is performed with the sub-images from (e,f)*

3.2 Spatial autocorrelation coefficient

As already presented in the previous section, the interpolation artifacts and the statistical inconsistency of the MI can be reduced to a certain extent by using one of the various solutions proposed, such as random resampling, or including prior information. However, these methods will not overcome the peculiar behavior of MI for image patches containing no clear anatomical structure. In the context of the hierarchical strategy, this problem of MI is increasingly disturbing during subdivision leading often to regions of homogeneous intensities, showing mostly imaging noise. Therefore, we focused our research towards alternative solutions that can completely eliminate such problematic image regions. Inspired by point pattern analysis, the solution we propose is based on a new method to detect noisy patches or regions of homogeneous intensity within an image. We adapted and integrated this structure detection method into the hierarchical image splitting strategy, considerably increasing the registration robustness. Moreover, this method can also be used as a reliable stopping criterion for the entire hierarchical subdivision procedure.

3.2.1 Introduction in Point Pattern Analysis

In the field of geography and ecology one of the basic goals is to gain information about the spatial variation pattern in their data at different geographical locations within a region of interest. This research area is called point pattern analysis (PPA) and is focusing on detecting specific events and on answering questions about the distribution of their locations. Of particular interest is if these events are randomly or regularly distributed or clustered in the analyzed region. Most of the statistical tests describing such spatial characteristics are relying on *spatial autocorrelation*. A detailed overview of these methods can be found in [Cliff and Ord 1973]. Two spatial autocorrelation coefficients are especially intensively used due to their robustness ([Lee and Wong 2001]) *Moran's I* and *Geary's C* coefficient. They share common characteristics, but their statistical properties are different. However, both coefficients are based

on the comparison of the neighboring units[1] which will lead to a strong positive spatial autocorrelation if they have similar values. Otherwise, if the neighboring units are very dissimilar, then the spatial autocorrelation will become strongly negative. The analysis of the statistical properties of the two aforementioned spatial autocorrelation coefficients concluded that the Moran's I coefficient is favored by most of the analysts because it has highly desirable statistical properties ([Lee and Wong 2001]). More specifically, the variance of the I coefficient is less affected by the distribution of the sample data than that of Geary's C coefficient.

3.2.2   Moran's I coefficient of spatial autocorrelation

According to [Cliff and Ord 1973], for a data set $X = \{x_i\}$ of mean value $E(X) = \overline{x}$, the Moran's I coefficient is defined as:

$$I = \frac{N}{\sum_{i,j=1}^{N} w_{ij}} \cdot \frac{\sum_{i,j=1}^{N} w_{ij} \cdot (x_i - \overline{x}) \cdot (x_j - \overline{x})}{\sum_{i=1}^{N} (x_i - \overline{x})^2} \tag{3.2}$$

where $W = \{w_{ij}\}$ is called the contiguity matrix, representing the connectivity weights, or the amount of interaction between the locations i and j, and N stands for the number of observations in the analyzed data. Note, that I is similar to the classical form of an autocorrelation coefficient: the numerator term is a measure of covariance among the $\{x_i\}$ and the denominator term describes the signal variance. It is also obvious that Moran's I is based upon the cross-products of deviations of the x_i from the mean value $\overline{x}$. It varies in the interval $[-1, 1]$, the values close to the extremity of this interval indicating the presence of structure in the spatial distribution while random patterns are characterized by values close to zero.

As already mentioned, the contiguity matrix $W = \{w_{ij}\}$ represents the amount of interaction between the data elements and has a big impact in determining the desired spatial autocorrelation coefficient. The W allows the investigator to choose a set of weights which he finds most appropriate based on prior considerations. Further more, if different models are proposed regarding the degree of conectivity between neighboring

---

[1]The neighboring units are called *counties* in point pattern analysis. For our application they will represent the image pixels.

elements, alternative sets of weights might be used to investigate these hypotheses. It is important to stress that the weights must be chosen carefully in order to avoid spurious correlations. The factors which are most important will depend upon the study at hand. For more details regarding the choice of the connectivity weights please refer to [Cliff and Ord 1973] and [Lee and Wong 2001]. For image processing application, the most usual practice is to build the contiguity matrix with coefficients that are inversely proportional to the Euclidean distance between the centers of the corresponding image pixels.

In order to test the significance of Moran's I, Cliff and Ord proposed to evaluate its first two statistical moments under either of the following two assumptions:

(n) - *normality* (n) - the $\{x_i\}$ are the result of N independent drawings from a normal distribution;

(r) - *randomization* (r) - independently of the underlying distribution of the population, the variation of I over the set of all possible permutations of the elements of $\{x_i\}$ (there are $N!$ such combinations).

Using the subscripts n and r to distinguish between the normalization and randomization assumption respectively, the first order moment of I (its mean value) is:

$$E_n(I) = E_r(I) = \frac{-1}{(N-1)} \tag{3.3}$$

while the second order moment (its variance) is:

$$E_n(I^2) = \frac{N^2 S_1 - N S_2 + 3 S_0^2}{(N^2 - 1) S_0^2} \tag{3.4}$$

$$E_r(I^2) = \frac{N[(N^2 - 3N + 3)S_1 - N S_2 + 3 S_0^2]}{(N-1)(N-2)(N-3)S_0^2} - \frac{k[(N^2 - 2)S_1 - 2N S_2 + 6 S_0^2]}{(N-1)(N-2)(N-3)S_0^2} \tag{3.5}$$

where:

$$
\begin{aligned}
S_0 &= \sum_{i,j=1}^{N} w_{ij} \\
S_1 &= \frac{1}{2} \sum_{i,j=1}^{N} (w_{ij} + w_{ji})^2 \\
S_2 &= \sum_{i=1}^{N} \left(\sum_{j=1}^{N} w_{ij} + \sum_{j=1}^{N} w_{ji} \right)^2 \\
k &= \frac{\sum_{i=1}^{N} (x_i - \overline{x})^4}{\left(\sum_{i=1}^{N} (x_i - \overline{x})^2 \right)^2}.
\end{aligned}
\tag{3.6}
$$

It has been shown in [Cliff and Ord 1973] that the standard $Z_{n,r}$ value associated to the Moran's I coefficient:

$$
Z_{n,r} = \frac{I - E_{n,r}(I)}{E_{n,r}(I^2)}.
\tag{3.7}
$$

is asymptotically normally distributed as the number of data elements N increases and therefore its significance can be tested against the Gaussian distribution.

3.2.3   Image information consistency test

In order to use Moran's I as an indicator of the structural content for 2D sub-images, it has to be adapted regarding to the choice of the contiguity matrix and to the test for significance of the I value. It is logical to build the contiguity matrix using a weighting scheme inversely proportional to the Euclidean distance $d(\cdot, \cdot)$ between the currently inspected pixels and its neighbors. Furthermore, a maximum interaction distance $\vec{\mathcal{D}} = (\mathcal{D}_x, \mathcal{D}_y)$ has to be selected according to the minimum size of the structures to be detected in the image.

Denoting the spatial location of a pixel within a 2D image (i.e. image patch) $A^\star$ of size $N_x \times N_y$ with $\vec{s} = (s_x, s_y) \in \Omega_{A^\star}$, we can define the vicinity of size $\vec{\mathcal{D}}$ around $\vec{s}$ as:

$$
\mathcal{V}_{\vec{s}}^{\vec{\mathcal{D}}} = \{ \vec{p} = (p_x, p_y) \in \Omega_{A^\star}, \forall \mid s_x - p_x \mid \leq \mathcal{D}_x \text{ and } \mid s_y - p_y \mid \leq \mathcal{D}_y \}.
\tag{3.8}
$$

By changing the linearized index notation from Eq. 3.2 to image coordinates, then the contiguity matrix W can be expressed as:

$$
W = \begin{cases} w_{\vec{s}\vec{p}} = \frac{1}{d(\vec{s}, \vec{p})}, & \forall \vec{p} \in \mathcal{V}_{\vec{s}}^{\vec{\mathcal{D}}} \setminus \{\vec{s}\} \\ 0, & \text{otherwise.} \end{cases}
\tag{3.9}
$$

Denoting the intensity value of the image pixel located at the spatial position $\vec{s} \in \Omega_{A^\star}$ with $a_{\vec{s}} = A^\star(\vec{s})$ and the mean value of the image intensities with $\overline{a}$, the Moran I becomes:

$$I = \frac{1}{\sum_{\vec{p} \in \mathcal{V}_{\mathcal{D}}^{\mathcal{D}} w_{\mathcal{D}\vec{p}}} \cdot \frac{\sum_{\vec{s} \in \Omega_{A^\star}} \sum_{\vec{p} \in \mathcal{V}_{\vec{s}}^{\mathcal{D}}} w_{\vec{s}\vec{p}} \cdot (a_{\vec{s}} - \overline{a}) \cdot (a_{\vec{p}} - \overline{a})}{\sum_{\vec{s} \in \Omega_{A^\star}} (a_{\vec{s}} - \overline{a})^2}. \qquad (3.10)$$

Similarly with the original formulation, Moran's I applied to images varies in the interval $[-1, 1]$, where the random patterns are characterized by values close to zero. Before testing the significance of I, its first and second order moments can be estimated under the randomization assumption, using Eq. 3.3 and Eq. 3.5. Note, that for an image patch $A^\star$ of size $N_x \times N_y$, the number of data elements is the number of image pixels $N = N_x \cdot N_y$. The associated k and $S_i, i = 1..3$ factors are also computed similarly to Eq. 3.6:

$$
\begin{aligned}
S_0 &= \sum_{\vec{p} \in \mathcal{V}_{\mathcal{D}}^{\mathcal{D}}} w_{\mathcal{D}\vec{p}} \\
S_1 &= 2 \sum_{\vec{p} \in \mathcal{V}_{\mathcal{D}}^{\mathcal{D}}} w_{\mathcal{D}\vec{p}}^2 \\
S_2 &\sum_{\vec{p} \in \mathcal{V}_{\mathcal{D}}^{\mathcal{D}}} \left(\sum_{\vec{s} \in \mathcal{V}_{\mathcal{D}}^{\mathcal{D}}} w_{\vec{s}\vec{p}} + \sum_{\vec{s} \in \mathcal{V}_{\mathcal{D}}^{\mathcal{D}}} w_{\vec{p}\vec{s}} \right)^2 \\
k &= \frac{\sum_{\vec{s} \in \Omega_{A^\star}} (a_{\vec{s}} - \overline{a})^2}{\left(\sum_{\vec{s} \in \Omega_{A^\star}} (a_{\vec{s}} - \overline{a})^2 \right)^2}.
\end{aligned}
\qquad (3.11)
$$

The test for the significance of the Moran's I applied for images is performed by Student's t-test on the significance level of 95% using the randomization assumption.

Figure 3.8 shows the classification result of a 2D neuroradiological MR slice into regions having consistent structures and regions without structures when using the aforementioned test for significance of Moran's I coefficient. The initial MR image depicted in Fig. 3.8(a) is partitioned in 32×32 sub-images of 16×16 pixels. Figure 3.8(d) shows the spatial autocorrelation coefficient I evaluated for all the partitioned sub-images. By using the threshold of 1.96 on the corresponding standard Z values depicted in Fig. 3.8(e), the classification is shown in Fig. 3.8(b,c).

(a)   (b)   (c)

Moran I   Z value

(d)   (e)
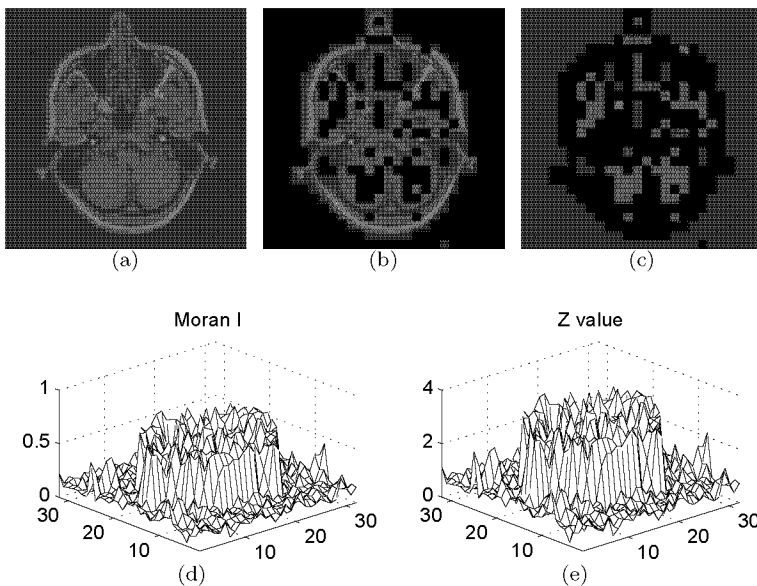
Figure 3.8: *Classification of image regions into consistent and inconsis-
tent structures when using the Moran's I consistency test. (a) The orig-
inal MR slice, splitted into 32×32 sub-images of 16×16 pixels. (b) Sub-
images classified as showing consistent anatomical structures. (c) Sub-
images classified as having inconsistent structures (noisy or structure-
less). (d) The Moran's I coefficient evaluated for all sub-images. (e) The
standard Z values associated to Moran's I values*

3.3   Further analysis of the hierarchical image splitting strategy

This section discusses the disadvantages of the binary splitting scheme of
the images, as originally proposed by Likar and Pernuš. Two alternative
image partitioning schemes are presented to increase the robustness of
the final deformation field resulted from the hierarchical strategy.

3.3.1 Binary splitting

The binary image splitting scheme proposed by Likar and Pernuš, divides the images to be registered into sub-image pairs with borders at fixed positions. The positioning of the borders is determined exclusively by the image geometry and completely ignores the image content. In addition to the problems raised by the loss of statistical consistency of the MI, this partitioning scheme may lead to additional problems like distributing single structures between neighboring sub-images. If in the remainder of one of these sub-images there is another well defined structure, then this may dominate the entire local rigid registration. An example can be seen in Fig. 3.4, where the temporal bone around the sphenoid sinus is partitioned into numerous sub-images. A possibly undetectable local misregistration of any of the sub-images may induce deformations to the nearby structure, which is not in agreement with the underlying physical phenomena. Therefore, several schemes to overcome the aforementioned problem of the binary splitting have been explored such as an *overlapping windows strategy* or *adaptive splitting*. We propose to use the later method, as to our experience in combination with the Moran consistency test, this will lead to an increased robustness of the entire hierarchical procedure for non-rigid registration.

3.3.2 Overlapping sub-images

An alternative solution to the binary image subdivision is the overlapping sub-image splitting scheme proposed by [Xu and Dony 2004]. By defining additional splitting borders through the centers of the sub-images resulted from the binary splitting, this scheme creates overlapping sub-images. They claim that their method helps to overcome the problem of the edge splitting, and that the registration results are improved to some limited extent leading to much better visual result. Of course, the computational complexity increases due to both the registration of the new sub-images introduced and the computation of TPS using an increased number of control points. An example of overlapping sub-images is sketched in Fig. 3.9. An overlapping image division at the lth level of the hierarchy generates $(2 \cdot 2^{l-1} - 1)^2$ sub-images, instead of $2^{2(l-1)}$ sub-images resulting from a binary splitting.
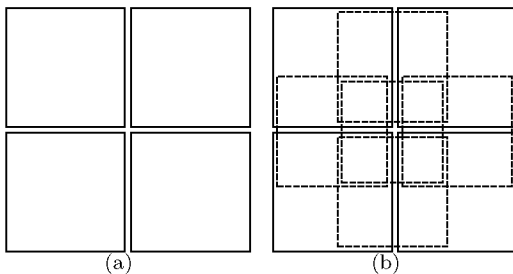
Figure 3.9: *Schematic examples of image partitioning schemes: (a) the binary splitting and (b) the overlapping sub-images scheme*

Basically, the overlapping windows method is close to the binary splitting, being susceptible to similar limitations, by the fixed position of the sub-images within the original image. However, possible distortions are reduced by the proximity of the TPS control points.

3.3.3 Adaptive image splitting strategy

Following any of the aforementioned image splitting schemes, one can reach the limit case of sub-images containing one single pixel. For this extreme case obviously no image similarity measure can give a consistent response. The usual procedure that many have adopted is to limit the possible size of the sub-images to a minimum predefined value. This leads of course to limitations on the accuracy and performance of the non-rigid image matching. These limitations concern mainly the registration of those structured details that are smaller than the minimum predefined sub-image size. In addition, the original hierarchical strategy partitions both images after a TPS reconstruction of the floating image. This interpolation is likely to induce significant changes on both the structures and the intensities of the floating image. Furthermore, this image reconstruction may lead to a possible decrease of contrast and details within the floating image, and therefore the accuracy of the registration may decrease. Even if visual inspection of the registered image may be satisfactory, this may only be caused by the lack of image details due to the repetitive TPS interpolations.

In order to overcome the aforementioned limitations, we propose an adaptive scheme for the hierarchical image splitting, which is fully automatic and permits the registration of fine details present in both images. Using the image information consistency test presented in Sec. 3.2, the adaptive subdivision scheme allows the registration of sub-images with the size of the contiguity matrix defined to compute the Moran's I coefficient. Moreover, in contrast to the original hierarchical image subdivision scheme, we propose to partition only the floating image without any reconstruction between consecutive hierarchical levels. As such, the newly partitioned sub-images will inherit the registration parameters from their parents registered at the previous hierarchical level. This parameter inheritance is done using a method that will be presented in the next chapter. In this way, the details of the floating image are kept according to their original resolution and consequently, the accuracy of the local registration is considerably increased.

The underlying principle of our image subdivision scheme is to locally decide whether a new partitioned sub-image contains significant structures and therefore if it is possible to perform a consistent local registration. This decision regarding the sub-image structural content is done using the Moran's I coefficient by testing both the current floating sub-image and its corresponding area on the reference image. This double test done on both images ensures that the local registration is done between significant structures simultaneously present in both images. As such, at every level of the hierarchical image splitting, the Moran I can test for the presence of structures in all the newly formed sub-images. All those sub-images passing the Moran test, can be further registered and subdivided at the next level of the hierarchy. Those sub-images failing the Moran test are omitted from both the local registration procedure and also from further subdivisions. In this way, the entire subdivision scheme is locally adaptive and increases the robustness of the registration procedure by not only eliminating possible inconsistent local matches of structureless image patches, but also by registering only those sub-images which have consistent structured information. The same principle makes the Moran's I an important part of the final stopping criterion of the hierarchical image splitting, which is discussed later in Chapter 5.

In order to visualize the evolution of the adaptive image subdivision along the hierarchical non-rigid registration, we performed the non-rigid

registration experiment between the CT images of the leg presented in Fig. 3.1. Figure 3.10 shows the evolution of the floating image splitting grid projected undistorted onto the reference CT image without contrast agent. Figure 3.10(a) depicts the splitting grid corresponding to the 2nd hierarchical level and the subsequent images show the following levels up to the 7th (Fig. 3.10(f)) level. At this final stage, the size of the smallest sub-image is 8×8 pixels and all those sub-images with a bigger size have previously failed the Moran's consistency test. Note, that in the original hierarchical registration procedure as proposed by Likar and Pernuš, the maximum reported hierarchical level is 4, having sub-images of 64×64 pixels.

3.4   Intensity mapping

The hierarchical image splitting strategy for non-rigidly registering medical images can be applied successfully not only for multi-modal images but also for mono-modal cases. As already presented in the first subsection of this chapter, the statistical consistency of MI is decreasing along the hierarchy, weakening the performance of the entire registration and limiting the maximum level of the hierarchical subdivision. However, for mono-modal image registration there are alternative similarity measures and CC is favored by most researchers as it was proven to be robust and insensitive to possible changes of illumination or image shadings. Cross-correlation based similarity measures showed an increased consistency and robustness also compared to MI, especially in those situations where small sub-images were to be registered. We therefore investigated the possibility to extend the usability of CC as a similarity measure to multi-modal cases. Such a method could enable to replace MI with a more robust CC for higher levels in the hierarchy.

In the past few years, several researchers have investigated the usability of CC in multi-modal registration cases. However, CC assumes a linear relationship between the image intensities. The approach commonly proposed in the literature tries to find a functional relationship between the intensities of the different modalities such that one modality is transformed into the contrast space of the second image. As such, several methods have been proposed either for estimating a functional relation-
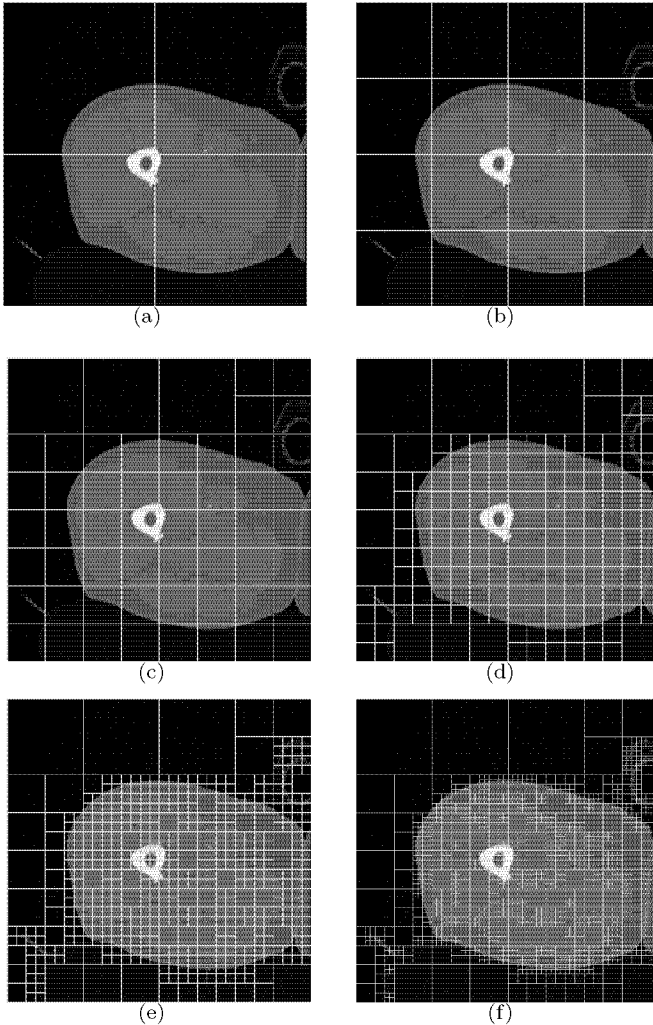
Figure 3.10: *The evolution of the adaptive floating image splitting grid along the hierarchical non-rigid registration procedure. The grid is projected undistorted on the reference image, from the 2nd (a) up to the 7th (f) hierarchical level*

ship between the intensities of images from different modalities or for
the direct estimation of similarity measures which integrate this func-
tionality in their definition. For example, the VIR criterion presented
by Woods in [Woods *et al.* 1993] proved to be efficient for matching
PET with MR images. In [Nikou *et al.* 1998] an extension was pre-
sented that removed the need for manual segmentation and extended
the method's applicability to other modality combinations. Another
extension of Woods' VIR criterion called *correlation ratio* is described
in [Roche *et al.* 1998]. Later on, in [Guimond *et al.* 2001], an adaptive
intensity correction was proposed that combines the correlation ratio
with the demons algorithm [Thirion 1996]. A completely different ap-
proach for CT-MR cross-registration described in [van den Elsen *et al.*
1994] bases on a simple intensity mapping of the original CT image such
that bone and air have identical appearance as in an MR image. All the
proposed methods, however, lead to the appearance of fake structures
within the mapped image, which strongly limits their usability. These
ghost features caused by imaging details which are not visible in both
modalities can lead to ambiguities that result in misregistrations.

This section presents a local intensity mapping that allows to switch
from MI to the more robust CC at higher hierarchical levels. In con-
trast to the already existing approaches that estimate the functional
relationship from one image modality to the other, we propose to build
a common intermediate *pseudo-modality*. The intensities in both images
are mapped simultaneously onto a common contrast space, which is not
necessarily one of the two source intensities, but rather a combination
of them. Although the transformed images may locally resemble one
of the modalities, on an overall scale this is not true. The proposed
mapping algorithm is demonstrated on CT/MR image registration but
is generally applicable for any combination of modalities.

3.4.1   Cross-Correlation versus Mutual Information in mono-modal registration

Figure 3.11 depicts two relevant examples demonstrating the advantages
of using the CC instead of MI as a similarity measure in a realistic mono-
modal registration scenario. The results of these experiments are in
complete accordance with with the previous observations from Sec. 3.1.3,

where Fig. 3.6 shows that the CC gives a more consistent response than MI when it is estimated for signals having a small number of samples.

The two examples are taken from a registration experiment between a contrast enhanced and a native CT image of the leg, both images of size 512×512 pixels with 256 intensity levels. Figure 3.11(a) depicts the reference image without contrast agent, and Fig. 3.11(b) shows the floating CT image with contrast agent. For the experiments, two representative image patches were manually selected from the floating image and horizontally translated from their initial position over the reference image. The selected image patches are of size 32×32 pixels, equivalent to sub-images of the 5th level of the hierarchy. In the first experiment, the floating image patch shows a clear structure due to muscle boundaries. By using Eq. 2.36 and Eq. 2.24, the responses of MI (Fig. 3.11(d)) and CC (Fig. 3.11(f)) are calculated for a horizontal displacement of up to 10 pixels. The second experiment is done for an image patch with almost uniform intensities, containing no relevant structure. The corresponding responses of MI and CC for a horizontal shift for up to 10 pixels are depicted in Fig. 3.11(c) and Fig. 3.11(e) respectively.

These experiments demonstrates the superior performance of CC over MI. Firstly, the interpolation artifacts are drastically reduced. Secondly, for the image patch having almost homogeneous intensity values with no visible structure, the CC shows more robust response than MI. The CC decreases when translating the structureless image patch from its initial position because on both sides of the reference image there are more pronounced structures uncorrelated with the floating image patch. Meanwhile, in complete accordance with the discussion presented in Sec. 3.1.3, MI increases as the structureless sub-image starts to overlap the bone, pretending a better fit between the two image regions when the structure represented by the bone-muscle interface lies somewhere in the noisy patch. Note, that MI shows significant responses in both experiments and a simple threshold based approach of the MI responses, as proposed in [Likar and Pernuš 2001], might fail in detecting this obvious outlier.

3.4.2   Local intensity mapping

The proposed mapping strategy bases on the observation that the performance of a registration algorithm will not increase if one of the images

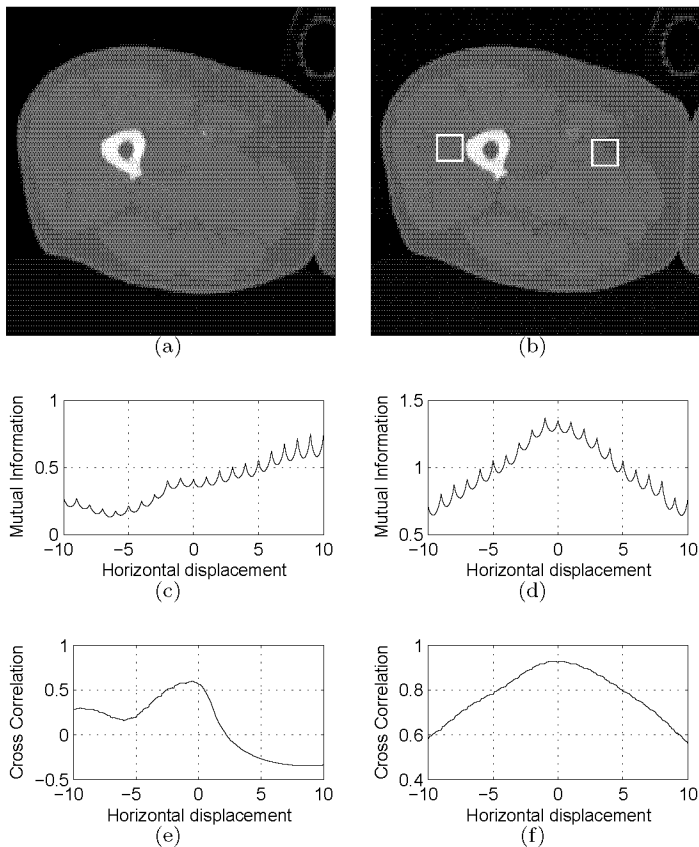**Figure 3.11**: *(a) The reference CT. (b) The floating CT of the leg. The right mask covers an image patch having clear structures, the left one is over a structureless image patch. The MI response when the structureless image patch (c) and the structured region (d) are horizontally translated up to 10 pixels from their initial position. (e,f) The CC response for the same horizontal translations of the aforementioned image patches*
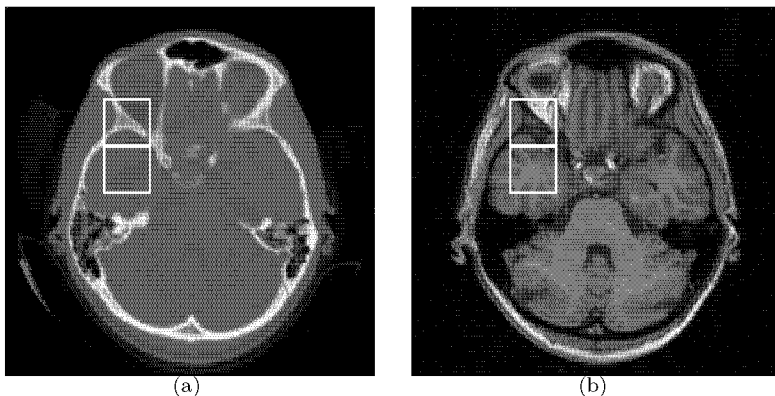
Figure 3.12: *Transversal slices of rigidly registered (a) CT and (b) MR acquisitions of the head*

contains more structural details than the other. On the contrary, details visible in only one of the images can lead to ambiguities by inducing misleading optima in the similarity measure. The performance of the registration procedure thus only depends on those image features which exist simultaneously in both modalities. The proposed mapping procedure builds an intermediate pseudo-modality of the images that will show only the common image features and drop additional details prominent in only one of the modalities.

Figure 3.12 depicts two corresponding transversal slices from rigidly registered 3D volumes of $512 \times 512 \times 50$ voxels of size $0.39 \times 0.39 \times 0.6$ mm^3. Obvious differences can be noticed not only in the intensities of most of the structures but also in the visibility of details of the tissues. These differences make that the mapping between the intensities is neither linear, nor invertible.

We propose to estimate the functional relationship between the intensities of the different modalities by using the information contained in the joint histogram of the coarsely registered images. In a first step, the mean values and the variances of all MR image (B) voxels are calculated, which correspond to the normalized intensity values $a \in \Omega_a \subseteq 0..255$ in the CT image (A) using the joint histogram $\mathcal{H}_{AB}$:

$$\mu_A(a) \quad = \quad \frac{\sum_{b \in B} \mathcal{H}_{AB}(a,b) \cdot b}{\sum_{b \in B} \mathcal{H}_{AB}(a,b)} \tag{3.12}$$

$$\sigma_A^2(a) \quad = \quad \frac{\sum_{b \in B} \mathcal{H}_{AB}(a,b) \cdot (b - \mu_A(a))^2}{\sum_{b \in B} \mathcal{H}_{AB}(a,b)} \tag{3.13}$$

and likewise for the MR image voxels $b \in \Omega_b \subseteq 0..255$:

$$\mu_B(b) \quad = \quad \frac{\sum_{a \in A} \mathcal{H}_{AB}(a,b) \cdot a}{\sum_{a \in A} \mathcal{H}_{AB}(a,b)} \tag{3.14}$$

$$\sigma_B^2(b) \quad = \quad \frac{\sum_{a \in A} \mathcal{H}_{AB}(a,b) \cdot (a - \mu_B(b))^2}{\sum_{a \in A} \mathcal{H}_{AB}(a,b)}. \tag{3.15}$$

For each of the histogram bins of the CT and MR, a flag f_a, f_b and a counter c_a, c_b is defined. The flag encodes whether an intensity value should be mapped by the corresponding μ function or should be kept unchanged:

$$f_a, f_b = \begin{cases} 1 & \text{map to the other modality} \\ 0 & \text{undefined in the joint histogram} \\ -1 & \text{keep the value unchanged.} \end{cases} \tag{3.16}$$

The counters are increased by one, whenever an intensity from one modality is mapped onto its bin. As it is our aim to suppress structures in a patch which are not visible in the other modality, the flags select intensities according to the image with smaller variance. The counters are also updated according to this decision:

$$\forall a, \begin{cases} \text{if } \sigma_A(a) < \sigma_B(\mu_A(a)) \to f_a = 1, \text{ inc. counter } c_{\mu_A(a)} = c_{\mu_A(a)} + 1 \\ \text{if } \sigma_A(a) > \sigma_B(\mu_A(a)) \to f_a = -1, \text{ inc. counter } c_a = c_a + 1 \end{cases} \tag{3.17}$$

and likewise for the MR image:

$$\forall b, \begin{cases} \text{if } \sigma_B(b) < \sigma_A(\mu_B(b)) \to f_b = 1, \text{ inc. counter } c_{\mu_B(b)} = c_{\mu_B(b)} + 1 \\ \text{if } \sigma_B(b) > \sigma_A(\mu_B(b)) \to f_b = -1, \text{ inc. counter } c_b = c_b + 1. \end{cases} \tag{3.18}$$

As can be seen in the schematic joint histogram in Fig. 3.13(a+b) three different regions can be distinguished according their variance. For the

regions where $f_a > f_b$ and $f_a < f_b$ the mapping direction is unambiguous and indicated with an arrow. In Fig. 3.13(b) the same regions are labeled according to the flag notation. If $\sigma_A \approx \sigma_B$ no clear decision can be made. For the intensities in this ambiguous region (see Fig. 3.13(c)) it is very likely that the value b_l will be mapped to a_i and a_i which in turn is associated with b_k. Accordingly, the mapping function $\nu_A(a)$ for A can be written as:

$$\forall a = 0..255, \ b = \nu_A(a) = \begin{cases} \text{if } f_a > f_{\mu_A(a)} & \rightarrow & \mu_A(a) \\ \text{if } f_a < f_{\mu_A(a)} & \rightarrow & a \\ \text{if } f_a = f_{\mu_A(a)} & \rightarrow & \text{ambiguity.} \end{cases} \quad (3.19)$$
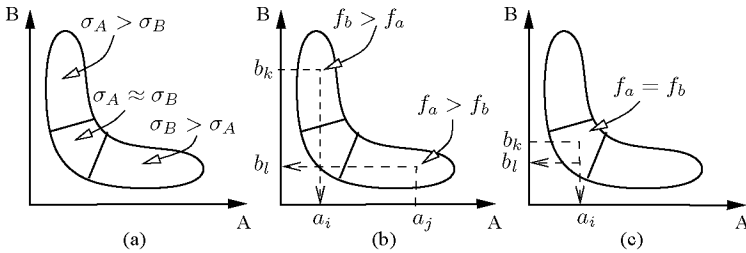
Figure 3.13: *Schematic joint histogram with (a) three regions defined by their variances, (b) the three regions labeled with the flag notation, and (c) illustration of ambiguities*

Two different ambiguous cases can be distinguished: (1) $f_a = f_{\mu_A(a)} = 1$, i.e. the intensities in both modalities are to be changed and (2) $f_a = f_{\mu_A(a)} = -1$ i.e. both of them should be kept. The mapping counters c_a, $c_{\mu_A(a)}$ are used to resolve such situations:

$$\forall f_a = f_{\mu_A(a)}, \ b = \nu_A(a) = \begin{cases} \text{if } c_a > c_{\mu_A(a)} \rightarrow a \\ \text{if } c_a < c_{\mu_A(a)} \rightarrow \mu_A(a). \end{cases} \quad (3.20)$$

Corresponding formulas for Eq. 3.19/3.20 are used for $\nu_B(b)$. As such, the mapping function $\nu_B(b)$ for B can be written as:

$$\forall b = 0..255, \ a = \nu_B(b) = \begin{cases} \text{if } f_b > f_{\mu_B(b)} & \rightarrow & \mu_B(b) \\ \text{if } f_b < f_{\mu_B(b)} & \rightarrow & b \\ \text{if } f_b = f_{\mu_B(b)} & \rightarrow & \text{ambiguity} \end{cases} \qquad (3.21)$$

and the ambiguities are solved similarly by using the counters c_b, $c_{\mu_B(b)}$:

$$\forall f_b = f_{\mu_B(b)}, \ a = \nu_B(b) = \begin{cases} \text{if } c_b > c_{\mu_B(b)} \rightarrow b \\ \text{if } c_b < c_{\mu_B(b)} \rightarrow \mu_B(b). \end{cases} \qquad (3.22)$$

These ambiguities have to be resolved iteratively until no further changes in the resulting mappings are detected. Figure 3.14 and 3.15 show examples of CT/MR mappings using the proposed method.

3.4.3   Preliminary results and conclusions

Two representative examples were chosen to demonstrate the advantages of integrating this intensity mapping procedure into the hierarchical registration, such that CC can be used as the similarity measure instead of MI after a certain level of the hierarchy has been reached. Two regions of interest have been selected for illustration, marked with white squares on Fig. 3.12. All patches are of $64 \times 64 \times 17$ voxels, equivalent to the 4$^{\text{th}}$ level of the subdivision.

The first experiment was performed with an image pair (upper white squares in Fig. 3.12) containing rich structural details. Figure 3.14 shows the original patches, their intensity mapped versions and the behavior of MI (on the original) and CC (on the intensity mapped images) for horizontal displacements up to ± 10 pixels. It can be seen that for regions having sufficient structural information, both similarity measures are sufficiently stable for finding the correct registration position.

A different region of interest (lower white squares in Fig. 3.12) has been used for the same experiment. While the corresponding CT patch is almost uniform, the MR image shows significant contrast within the brain tissue covered. This is a classical case in which MI generally fails to find the correct registration position, see [Andronache *et al.* 2005]. Figure 3.15 shows the original and intensity mapped patches together
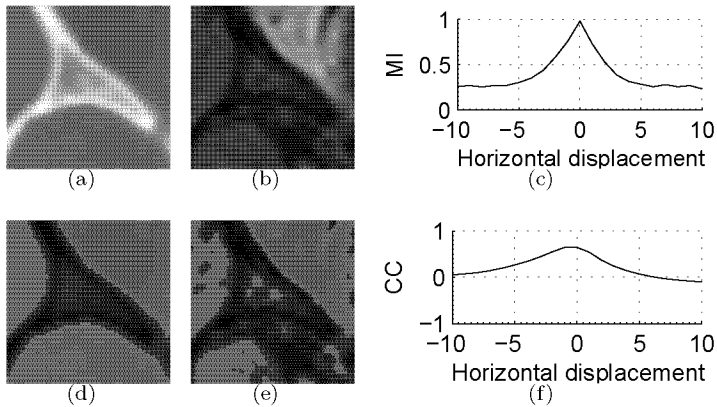
Figure 3.14: *(a,b) Initial patches showing rich structural details and (d,e) their intensity mapped versions. (c) The response of MI on the original and (f) CC on the intensity mapped images to horizontal translations*
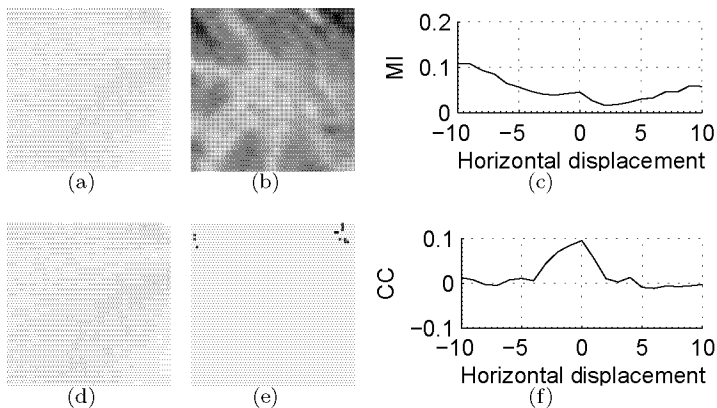
Figure 3.15: *(a,b) Patches with major differences of tissue contrast in CT and MR. (d,e) Intensity mapped versions of the images. (c) The response of MI on the original and (f) CC on the intensity mapped images to horizontal translations*

with the comparison between the MI and the CC responses to horizontal translations. While CC remains robust for this region, too, MI shows highly unreliable behavior.

As has been previously discussed, MI shows unsatisfactory behavior for the matching of structureless or small image patches due to the lack of statistical consistency caused by the small number of available image samples. CC proved to be more robust, but it can not be directly used for cases, where the intensity relation between the modalities is non-linear. The mapping strategy presented in this section enables the combination of both similarity measures for multi-modal registration procedures relying on a hierarchical subdivision strategy. At the first levels of the hierarchy, where the partitions are still relatively large, MI can be used to coarsely register the corresponding patches. After this stage, the images can be transformed to a pseudo-modality using the presented mapping technique and the similarity measure can be switched to the more robust CC.

With the proposed hybrid approach that uses MI for the first levels and CC for the last few levels, two important properties of these similarity measures can be seamlessly combined in a unique manner, namely the multi-modal capabilities of MI with the robustness of CC without increasing the computational complexity of the underlying algorithm.

3.5   3D registration

Many clinical procedures today involve the usage of 3D images. The original hierarchical non-rigid registration, developed for the registration of 2D microscopic images will therefore have to be generalized to 3D.

It is, however, very straightforward to extend the hierarchical image splitting strategy from 2D images to 3D volumes. Fundamentally, the principle is the same in 3D as in 2D, but there are some minor modifications and adaptations to be made. These modifications are imposed by the additional dimensionality and also by the usual anisotropy of the voxels. While similarity measures such as MI benefit from the increase in structural information and in the number of image samples, the higher computational complexity due to the increase of the number of degrees of freedom will have to be handled.

3.5.1   Mutual Information behavior for 3D data

In Sec. 3.1, a series of limitations of MI as a similarity measure were presented, especially for sub-images of small size. Although the discussions were based on 2D images, most of the problems can also be observed in 3D volumes. These problems are strongly related to either the limited number of available image samples or the lack of structural context of the sub-images. Due to the inherent increase of the available information due to the 3D nature of the images, MI has higher reliability at the corresponding hierarchical levels. In this case, the hierarchical image splitting will lead to sub-volumes and the chance of local misregistrations is smaller as the number of image samples increases. However, the same negative tendencies in the behavior of MI can be noticed, like increasing interpolation artifacts or diminishing statistical consistency when dealing with structureless sub-images.

3.5.2   Extending Moran's consistency test to 3D

In order to avoid the spurious local mis-registrations due to a lack of structural content within the partitioned sub-images, the same spatial autocorrelation coefficient can be used to determine the nature of the partitioned image regions. A few adaptations have to be made to the 2D version of the consistency test as presented in Sec. 3.2. These concern mainly the definition of the contiguity matrix and the computation of the standard Z value.

According to the definition, the contiguity matrix contains information regarding the amount of interaction between the neighboring image voxels. Following the same concept as in 2D, the W matrix will be determined according to a maximum distance interaction $\vec{\mathcal{D}} = (\mathcal{D}_x, \mathcal{D}_y, \mathcal{D}_z)$ between two image voxels, equivalent to the minimum size of the image structure to be detected. For 3D images, the contiguity matrix will also be selected inversely proportional to the Euclidean distance between the inspected voxels. The anisotropy of the voxel dimension has to be taken into consideration when calculating the voxel distance.

For a 3D image patch $A^\star$, a spatial location of a voxel is denoted with $\vec{s} = (s_x, s_y, s_z)$ within the image domain $\Omega_{A^\star}$. The vicinity of size $\vec{\mathcal{D}}$ around $\vec{s}$ is therefore defined accordingly as:

$$\mathcal{V}_{\vec{s}}^{\vec{\mathcal{D}}} = \{\vec{p} = (p_x, p_y, p_z) \in \Omega_{A^\star}, \forall \mid s_i - p_i \mid \leq \mathcal{D}_i \text{ with } i \in \{x, y, z\}\}. \tag{3.23}$$

Having expressed the vicinity of a voxel, the contiguity matrix can be calculated using Eq. 3.9 and the Moran I coefficient is calculated using Eq. 3.10. Note, that now in Eq. 3.9 and Eq. 3.10 3D image coordinates have to be used. Once the autocorrelation coefficient is calculated, the estimation of the first two moments $E(I)$ and $E(I^2)$ can be done using the equations 3.3 - 3.5. Once more, the structural content of the image is establish by testing the significance of the Moran's I coefficient by the Student's t-test.

3.5.3   Image splitting strategy in 3D

Another modification imposed by the 3D nature of the image concerns the partitioning procedure. If in 2D cases, the images are very likely to present an unitary or close to 1 aspect ratio on the x and y axes, for 3D images the situation is different in most of the cases as usually the slice distance is significantly higher than the pixel size. This results in an image volume with strongly imbalanced aspect ratio along the 3D directions.

For 3D images we use a splitting strategy which does not necessary enforce a simultaneous partitioning in the three directions of the image volume. Instead, we try to achieve cube shaped sub-images in the world coordinate system. Sub-images partitioned along one direction are divided in the corresponding dimension in two equal parts. The difference lies in the adaptive selection of the directions on which the splitting will be applied. This decision is based on a simple test of the aspect ratio of the edge lengths of the sub-volume. To achieve the cube shaped sub-volumes, the division is performed only in those directions that have a ratio ≥ 1.5 compared to the smaller ones. If all dimensions are comparable (all possible ratio combinations are < 1.5), the splitting is done simultaneously on all directions. In this way, the image splitting strategy may result in a variable number of partitioned sub-volumes at each level

of the hierarchy. Note, that no sub-voxel partitioning is applied during the subdivision process.

This adaptation of the splitting strategy for 3D images is applicable to image volumes consisting of a small number of slices. In this case, the initial volume is divided into sub-volumes composed of single slices having a decreasingly imbalanced aspect ratio along the 3D directions after a few hierarchical levels.

4

Deformation Field Regularization

All the non-rigid image registration algorithms determine a dense deformation field which is supposed to be sufficiently flexible to correctly match the images everywhere in their image domain. As already presented in Sec. 2.3, the research is currently focused on approaches which usually derive the deformation field from a series of local transformations that are found according to different strategies. These local transformations are often independently found and constraints must be imposed to ensure their continuity over the whole warped domain. As these constraints are very likely to be insufficient to ensure physically realistic deformations, an additional regularization procedure is typically applied. At the same time, the deformations can eventually be extrapolated to those image regions that are without salient image features. The regularization procedure usually enforces the deformation field to be a homeomorphism (continuous, one-to-one, and topology preserving). However, this assumption may be violated in the case of pathologically induced changes that can produce a lack of dense correspondence in the images (e.g. tumor growth) or discontinuities (like surgical procedures).

This chapter gives a brief overview of the state-of-the-art for deformation field regularization, followed by the description of a deformation field regularization adapted to our proposed adaptive hierarchical strategy.

4.1   An overview of the existing methods

Various regularization methods have been proposed in the literature
(see [Hajnal *et al.* 2001, Maes *et al.* 2003, Stefanescu *et al.* 2004]
and the references therein), staring with the representation of the defor-
mation field as a weighted sum of smooth basis functions, or free-form
approaches incorporating diverse smoothness constraints, to parameter-
ized deformation models physical-based inspired or constructed using
statistical analysis.

In the context of the hierarchical image splitting strategy, the local rigid
registrations of the partitioned sub-images are optimized completely in-
dependently of the neighboring sub-images, without enforcing any con-
straints. In its original version ([Likar and Pernuš 2001]), every level of
the hierarchy uses TPS to reconstruct the complete floating image. In
order to avoid the propagation of possible local misregistrations at sub-
sequent levels, they proposed a threefold local registration consistency
test and a correction of the pixel intensity at every level of the hierar-
chy before performing the TPS image reconstruction. In a first stage, a
geometrical test is used to detect and correct possible large mismatches.
Additionally, to detect and eliminate mismatches with only small geo-
metrical displacement, they used two additional tests on the similarity
measure. Firstly, they detect the sub-images whose similarity function is
small and inconsistent with the other sub-images. Secondly, they detect
possible errors of the optimization due to local maxima by analyzing the
graph of the similarity measure around the identified optima.

4.2   The proposed deformation field regu-
larization

Even though our proposed algorithm uses a hierarchical strategy similar
to [Likar and Pernuš 2001], important differences stem from our adap-
tive image splitting scheme. In contrast to their method in which at
each level both the reference and the TPS reconstructed floating image
are partitioned simultaneously, our splitting strategy is applied only on
the floating image, more precisely on the sub-images registered at the

previous level. The biggest advantage of this approach is due to the combination of the Moran's information consistency test with a scheme for inheriting registration parameters to the newly partitioned sub-images. This combination eliminates most of the problems in the original hierarchical strategy and also allows simpler heuristics to regularize the deformation field. Even though Moran's test removes all possible local misregistrations produced by structureless sub-images, some spurious local mismatches may still appear due to additional factors. For example, the optimization procedure might get trapped in a local maximum of the similarity function, caused by either voxel interpolation artifacts or a locally good match induced by repetitive patterns present in the image scene. In both cases the similarity measure presents multiple maxima that are locally consistent but globally incorrect, leading to erroneous overlap of neighboring sub-images.

In order to achieve robust and accurate registration, the accuracy of the individual local matches has to be ensured. For this, we propose a couple of tests to regularize the deformation field. Firstly, we *test the similarity function for the presence of local maxima* induced by interpolation artifacts. Secondly, we perform an *outlier detection* to remove possible mismatches induced by similar structures in the neighborhood.

4.2.1   Test for local maxima the similarity function

As already presented in Subsect. 3.1.1, MI is highly sensitive to voxel interpolation artifacts. The results of various experiments reported by numerous researchers (e.g. [Pluim *et al.* 2003, Tsao 2003, Ji *et al.* 2003, Hajnal *et al.* 2001]) in connection with rigid transformations showed that the sensitivity of MI is larger for image translations than for rotations. Therefore, we propose to sample the MI response on a regular grid of size N_{test} around the currently found optimal position. If a higher maximum is found, the registration parameters are changed according to this position and the optimization procedure is restarted. We selected the MI sampling grid of size N_{test} equal to the size of the contiguity matrix defined in the Moran test.

4.2.2   Outlier detection

It is possible that sub-images have rich structural information in the form of a repetitive pattern. This can induce a series of local maxima in the the similarity function, which locally seem to be consistent but may lead to globally inconsistent matches. To such cases we use a two stage outlier detection. Firstly, the test detects sub-images that moved too far away from their initial inherited positions, followed by checking the position of every sub-image relative to its neighbors. All those sub-images failing in any of these two stages are marked as *outliers* and their associated registration parameters are corrected as will be described in the following section.

The elimination of the large geometrical mismatches is performed similar manner as by [Likar and Pernuš 2001]. The translation d_i of each sub-image $B_i^{(l)}$ resulting from the local registration is compared with a predefined maximum allowed distance R_1. Denoting the center of the sub-image in world coordinates with $\phi_i^{(l)}$, then the distance d_i between the spatial positions before and after the local registration at the lth hierarchical level is defined by the Euclidean distance as:

$$d_i = \|\mathcal{T}_i(\phi_i^{(l)}) - \widetilde{\mathcal{T}}_i(\phi_i^{(l)})\| \tag{4.1}$$

where $\mathcal{T}_i$ is the transformation corresponding to the initially inherited set of parameters $\alpha_i^{(l)}$ and $\widetilde{\mathcal{T}}_i$ corresponds to the currently optimized parameter set $\widetilde{\alpha}_i^{(l)}$. In contrast to the original method proposed by Likar and Pernuš, we allow more flexibility for the local displacement of a sub-image, by setting this maximum distance R_1 to be up to 50% of the currently tested sub-image diagonal. All the sub-images with a distance $d_i > R_1$ are declared *outliers* and their associated registration parameters are corrected using the interpolation method described in the next section.

Up to this stage all sub-images were handled independently of their neighboring sub-images. Therefore, the second stage iteratively checks the consistency of the position of every sub-image $B_i^{(l)}$ relative to its direct neighbors $B_k^{(l)}$. More precisely, this test verifies that the center of the currently tested sub-image lies within a region of confidence defined by all its adjacent sub-images. This region of confidence represents an ellipsoid with main axes (D_2) equal to 50% of the size of

the currently tested sub-image. The center of this ellipsoid is the centroid of the set of masses inversely proportional with the area/volume of the adjacent sub-images and placed at their corresponding centers. A sub-image whose center lies outside the region of confidence is labeled as *temporal outlier*. This second stage of the geometrical test is performed iteratively because undetected outliers might be included in the evaluation of the region of confidence. At every iteration, all the previously marked temporal outliers are eliminated from the calculation of the region of confidence, but tested again for their geometrical consistency. The procedure iterates until the list of temporal outliers remains unchanged. Once no new temporal outliers are being added or eliminated from the list, all the sub-images in the list are marked as outliers. Note, that the sub-images located at the border of the floating image lack certain neighboring sub-images. To virtually create these missing neighbors, the existing neighbors are mirrored relative to the horizontal and vertical axes across the center of the bordering sub-image.

Fig. 4.1 depicts a 2D schematic example of the two stage geometric test. As the center of the sub-image marked with the square with thick continuous borders (C_1') is outside the circle of confidence with radius R_1 around its initial position C_1, the sub-image is marked as an outlier. Consequently, its position is corrected resulting in the thick dotted square, with the center C_1''. Note, that C_1'' lies right to its initial position C_1, due to the influence of the adjacent sub-image marked with the thin dotted square which was taken into account while correcting the registration parameters. In this schematic example no more outliers will be detected by the second stage of the geometrical test, as the sub-image marked with the thin dotted square has its center C_2 inside the region of confidence. This region of confidence with diagonal D_2 and center C_m is derived from all adjacent sub-images, including the previously detected and corrected outlier.

4.2.3   Parameter inheritance and outlier correction

All the outliers detected by the aforementioned tests require a new set of interpolated registration parameters. This correction of the registration parameters is done using the inverse distance weighted interpolation method presented in [Shepard 1968]. This method estimates an interpo-
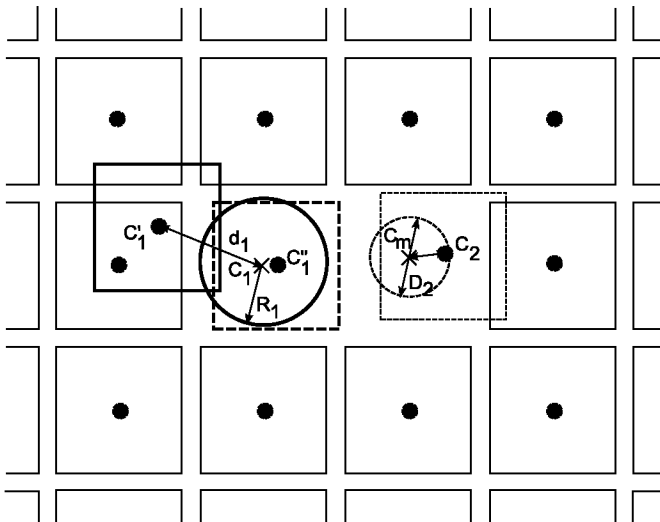
Figure 4.1: *A schematic example showing the performance of the two stage geometrical test. The sub-image marked with the square with thick continuous borders is detected as outlier by the first stage of the geometrical test and consequently its position is corrected in the square with thick dotted borders. The second stage of the geometrical test finds no more outliers (see text for more details)*

lated rigid transformation at a desired spatial position, given the set of rigid transformations at a series of control points.

Adapted to the hierarchical notation, the new set of parameters $\alpha_i^{(l)}$ of a detected outlier $B_i^{(l)}$ can be extracted from the newly estimated transformation matrix $\widetilde{\mathcal{T}}_i$, interpolated at the spatial position $\vec{s}_B = \phi_i^{(l)}$ which represents the center of the outlier. Assuming that the outlier has n_0 adjacent sub-images that are not marked as outliers with centers $\phi_k^{(l)}$ and associated transformations $\mathcal{T}_k$, then:

$$\widetilde{\mathcal{T}}_i(\vec{s}) = \sum_{k}^{n_0} w_k(\vec{s}) \cdot \mathcal{T}_k. \tag{4.2}$$

The interpolation weights w_k are defined as:

$$w_k(\vec{s}) = \frac{q_k(\vec{s})}{\sum_j^{n_0} q_j(\vec{s})} \tag{4.3}$$

where:

$$q_k(\vec{s}) = \frac{1}{\|\phi_k^{(l)} - \vec{s}\|^2} \tag{4.4}$$

such that:

$$\sum_k^{n_0} w_k(\vec{s}) = 1. \tag{4.5}$$

The same interpolation strategy is used for calculating the inherited registration parameters of the partitioned sub-images at the transition from one level to another of the hierarchical algorithm. At any level $l \geq 2$ the set of control points is formed by the parents (at the level $l - 1$) of all the neighboring sub-images.

4.2.4   The final dense deformation field

Similar to Likar and Pernuš, we use TPS to estimate the final dense deformation field. Having defined an automatic stopping criterion for the subdivision procedure as described in the next chapter, TPS interpolates the centers of all the sub-images at the last hierarchical level. Following the description presented in Sec. 2.4.4, we estimate the $3(n^{(l_{\text{final}})} + 4)$ TPS interpolation parameters ρ_i and ϵ_i using the non-linear radial basis function given in Eq. 2.48 and solving the system of Eq. 2.49. The necessary set of control points are the centers ϕ_i of all the $n^{(l_{\text{final}})} \leq 2^{2(l_{\text{final}}-1)}$ sub-images of the final hierarchical level, after performing the regularization of the deformation field. Hence, using Eq. 2.47, the dense smooth deformation field can be estimated at any point $\vec{s}_B$ of the floating image.

5

The Enhanced Hierarchical Multi-modal Non-rigid Registration Algorithm

This chapter presents the entire enhanced hierarchical algorithm we developed for non-rigid registration of multi-modal images. By integrating the previously discussed methods in a single framework, we overcome several limitations of the original version of Likar and Pernuš. In addition, we extended the algorithm to handle volumetric images. This chapter presents the algorithmic flow in detail and continues with the description of an alternative approach changing the similarity measure from MI to CC at higher levels of the hierarchy.

5.1 Description of the registration algorithm

The hierarchical algorithm follows a coarse-to-fine gradual approximation of the non-rigid deformation field to compensate spatial misalignment between two mono- or multi-modal medical images. By progressively subdividing the floating image (see Fig. 2.4), an image pyramid is constructed, comprising at every level of an exponentially growing number of sub-images whose centers and their related transformation parameters constitute the control points of the deformation field. The initial registration parameters of every sub-image are inherited from the previous hierarchical level. The new partitioned sub-images are then

checked by Moran's information consistency test and only sufficiently structured image patches are passed to the registration stage. For all the sub-images that fail this consistency test, the hierarchical splitting is stopped. The maximization of MI is then used to find the best individual registration parameters that locally match every sub-image to the reference image. The optimization procedure uses the classical Powell multi-dimensional direction set method and Brent's one-dimensional optimization algorithm described in [Press *et al.* 1988]. To further increase the robustness of MI, a prior distribution is combined with the local joint intensity histogram, using a weighting parameter equal to the ratio of the areas of the actual sub-image and the remaining of the floating image. Possible outliers due to local maxima of the goal function are detected by performing an exhaustive search for the global MI maximum in a small vicinity around the estimated optimal position. If this check confirms that the optimization converged to a local maxima, the registration procedure is reinitialized and restarted. The following regularization of the deformation field is ensuring the geometrical consistency between neighboring sub-images. Therefore, spatial constraints are imposed on the centroids of the sub-images by defining a region of confidence according to their neighbors. The registration parameters of the identified outliers are then corrected by interpolation. The stopping criterion checks whether any sub-images are left that can be further subdivided according to Moran's test. Otherwise, the dense deformation field of the final result is calculated using TPS, densely interpolating the local transformations of the sub-images over the whole image domain.

During the detailed description of the individual stages, we will use the following notation. The current level of the hierarchy is denoted with $l \geq 1$, and the initial input images are d dimensional, with $d = \{2, 3\}$. The reference image is denoted with A and the floating image with B. The adaptive splitting strategy partitions only the floating image in $n^{(l)}$ sub-images $B_i^{(l)}$. Every sub-image has an associated set of rigid transformation parameters $\alpha_i^{(l)} = (\vec{t}, \vec{r})$ and an information consistency flag $F_i^{(l)}$, specifying whether a sub-image passed ($F_i^{(l)} = 1$) or failed ($F_i^{(l)} = 0$) the Moran test.

Following the flow-chart depicted in Fig. 5.1, our hierarchical algorithm starts with a global rigid registration of the input images. As such,

the global misalignment between the reference and the floating image is compensated, providing a good initialization for the subsequent levels.

All the succeeding $l \geq 2$ hierarchical levels are iteratively following the same procedure which is composed of six major stages S1 - S6. Once the stopping criterion (S6) is fulfilled, the hierarchical iterations are stopped and the algorithm finalizes with the estimation of the dense deformation field (S7) and the reconstruction of the registered floating image (S8).

S1: **The image partitioning** (Sec. 3.3.3) is performed at the current hierarchical level l on all the sub-images $B_k^{(l-1)}$ of level $l-1$ that successfully passed Moran's consistency test. Each of these $B_k^{(l-1)}$ are partitioned into 2, 4 or 8 equally sized new sub-images $B_i^{(l)}$, such that the aspect ratios along the x, y and z directions are as close to 1 as possible (Sec. 3.5.3). Apart from the newly partitioned sub-images $B_i^{(l)}$, the current level l directly inherits all the structureless sub-images $(F_i^{(l-1)} = 0)$ from the previous level, i.e. the union of all the sub-images at the current level l reconstructs the entire floating image.

S2: **The inheritance of the registration parameters** follows the strategy for inverse distance weighted interpolation, described in Sec. 4.2.3. These inherited registration parameters $\alpha_i^{(l)}$ are estimated in the current sub-image center $\vec{\phi}_i^{(l)}$ by interpolating the set of parameters $\{\alpha_k^{(l-1)}\}$ placed in the centers $\vec{\phi}_k^{(l-1)}$ of the parents $B_k^{(l-1)}$ of all the neighboring sub-images.

S3: **Moran's information consistency test** (Sec. 3.2.3) is performed on all newly partitioned sub-images $B_i^{(l)}$ defining their associated consistency flags $F_i^{(l)}$. It checks the information content of both the floating sub-image $B_i^{(l)}$ and its corresponding region in the reference image $A_i^\star$. If any of these fails the information consistency test, then the floating sub-image $B_i^{(l)}$ is marked as *structureless* by setting its associated information consistency flag $F_i^{(l)} = 0$. Once detected, a *structureless* sub-image is automatically excluded from all subsequent registrations and subdivisions.

S4: **The local rigid registration** follows the procedure described in Sec. 2.2 and is performed on all newly partitioned sub-images with
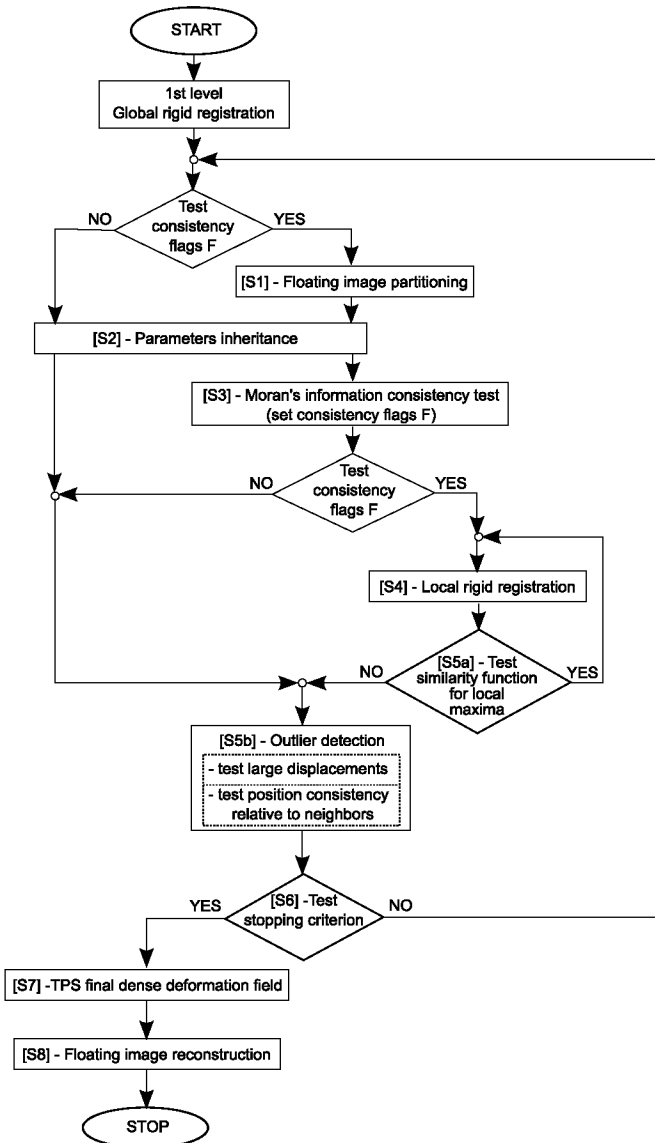
Figure 5.1: *The flow-chart of the enhanced hierarchical non-rigid registration algorithm (see text for more details)*

$F_i^{(l)} = 1$. Starting with the inherited set of transformation parameters $\alpha_i^{(l)}$, the Powell's optimization algorithm estimates a new set $\widetilde{\alpha}_i^{(l)}$ which maximizes MI estimated from the local joint intensity histogram incorporating prior information. The local histogram is built in a deterministic manner from the overlapping region of $\mathcal{T}_i(B_i^{(l)})$ with the reference image A. The prior information is estimated from all the sub-images except the current one, using their initial registration parameters as estimated in step S2. For combining local and prior information, we use a weighting parameter λ equal to the ratio between the number of image samples contained in the current sub-image $B_i^{(l)}$ and the remainder of the floating image.

S5: **The deformation field regularization** (Sec. 4.2) is composed of two consecutive tests:

   (a) **The test for local maxima in the similarity function** (Sec. 4.2.1) decides that during the previous local rigid registration stage S4, the optimization algorithm was not trapped in a local maximum of the similarity function, induced by interpolation artifacts. This test is performed on all sub-images that were undergoing the local rigid registration phase. For each of these sub-images, the response of the similarity function is analyzed on a regular grid centered around the optimized position. If another local maximum higher than the original one is found, the registration parameters are changed accordingly and used for re-optimizing the registration in step S4.

   (b) **Outlier detection** (Sec. 4.2.2) is a two stage geometrical test which ensures the consistency of the deformation field at the current hierarchical level l. In the first stage, each sub-image is tested whether its center has spatially displaced by more than half of its diagonal compared to the previous local registration. If necessary, the set of registration parameters of a detected outlier is corrected using the method described in Sec. 4.2.3. The second stage of the outlier detection iteratively checks the consistency of the spatial position of each sub-image relative to its neighbors. The center of the cur-

rently inspected sub-image is tested whether it lies within a region of confidence, represented by an ellipsoid whose diagonals are equal to 50% of the size of the currently inspected sub-image and which is centered in the centroid defined by its neighboring sub-images. The method mentioned above is used to correct the registration parameters of a detected outlier.

S6: **The stopping criterion** decides whether the hierarchical subdivision can be finished. This decision is taken at each level by testing the consistency flags of all the sub-images. If any of them has a consistency flag $F_i(l) = 1$, a new hierarchical level with all its phases is performed. Otherwise, the entire registration algorithm continues with stage S7 and calculates the final dense deformation field. This decision will implicitly enforce a minimum possible size of the partitioned sub-images corresponding to the size of the contiguity matrix used in the Moran test.

S7: **The final dense deformation field** is estimated (Sec. 4.2.4) using the TPS technique for spatial interpolation of the centers of all the sub-images at the last hierarchical level. The corresponding transformations between the floating and the reference image are defined by the last optimized registration parameters.

S8: **The reconstruction of the final registered image** is performed by warping the floating image into the domain of the reference image by using the previously estimated deformation field.

5.2 Registration algorithm using both Mutual Information and Cross-Correlation

In Sec. 3.4 we presented an alternative solution to overcome the statistical instability of MI for small sub-images. We proposed an intensity mapping strategy which enables the combination of both MI and CC as similarity measures for achieving the local registration. The basic idea is to use MI at the first levels of the hierarchy, where the partitioned sub-images are relatively large and MI is statistically consistent. As the

images are coarsely registered, the proposed intensity mapping can be used to robustly transform the multi-modal registration into a pseudo-mono-modal registration scenario, and thus allowing to use CC to perform the local matchings for subsequent hierarchical levels, increasing the robustness of the registration algorithm.

The implementation of this strategy requires minor changes in the previously described algorithmic flow. Basically, the changes affect only stage S4. Currently we enable the switch from MI to CC at a predefined hierarchical level l_{switch}. From this level on, the intensity mapping procedure is applied for each sub-image passing the Moran test. Once the pseudo-modality images are estimated, the local registration becomes a standard rigid registration procedure which uses CC as the similarity function. All the other stages of the hierarchical algorithm remain unchanged with the exception that the deformation field regularization procedure S5(a) will test CC for the presence of local maxima.

6

Results and Validation

In this chapter we investigate the performance of our proposed enhanced hierarchical non-rigid registration algorithm (shortly called ENRA) which was developed and implemented in C++ according to the description given in Chapter 5. The implementation was done within a general framework capable to handle with 2D and 3D images from any input modality. Also, the implementation allows to choose the similarity function used to perform the local rigid registration of the partitioned sub-images. This chapter presents first the results of 2D registration experiments with of mono- and multi-modal images and continues with results of a 3D data. If nothing else specified, the experiments are performed using MI as the similarity function. The chapter concludes with a discussion regarding the validation of the previously presented results.

6.1 Experiments in 2D

This section presents 2D registration experiments on pairs of mono-modal and multi-modal images. All images used were transversal slices extracted from a volumetric data set. In order to ensure that the two selected slices are in correspondence, the initial volumetric data were rigidly registered before extracting the 2D images. The algorithm used to pre-register the volumetric data was a classical implementation of a rigid registration procedure as described in Sec 2.2.

For all the experiments, the ENRA was using the following general setup:

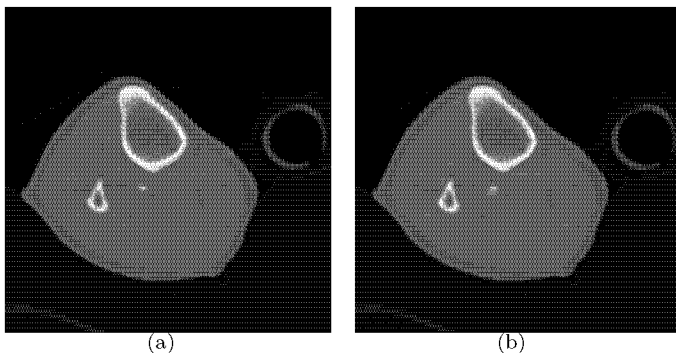(a) The MI was estimated from the joint intensity histogram which was generated using 256 bins;

(a)   (b)

Figure 6.1: *2D CT slices through a leg used for the mono-modal experiment. (a) The reference native data. (b) The floating contrast enhanced image*

(b) The weighting factor λ for the incorporation of the prior into the MI estimation was selected as the ratio between the area of the current sub-image undergoing the local registration and the remainder of the area of the floating image;

(c) The Moran test was performed with a contiguity matrix with maximum interaction distance $D = (3, 3)$ pixels;

(d) The threshold for the magnitude of the standard Z value of Moran's I coefficient used in the information consistency test was set to 1.96.

The first experiment was performed with a pair of corresponding transversal slices extracted from volumetric CT scans of a leg which were acquired for an angiographic study. Figure 6.1(a,b) depicts the initial 2D images undergoing this registration experiment. The reference image is a native CT which has 512×512 pixels of size 0.3144×0.3144 mm^2. The floating image is a contrast enhanced CT and has the same dimensions as the reference image.

Even if the selected data are mono-modal, the presence of the contrast agent in the floating image locally induces a multi-modal aspect in this registration experiment. Therefore, we compared the performance of
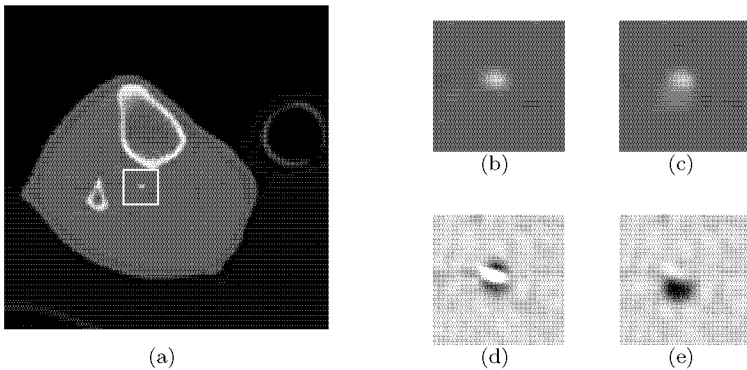
Figure 6.2: *Details of two non-rigid registration results of a pair of transversal CT slices of a leg when using CC and MI as similarity functions. (a) The region of interest on the reference CT. The corresponding sub-image in the reference native CT (b) and the floating contrast enhanced CT (c). Difference image after registration with ENRA when using CC (d) and MI (e)*

MI and CC for local rigid registrations. Globally, the two registration results were very similar, except a few image details that were locally misregistered when using CC. These image details are the blood vessels, such as the one marked with the white square on the floating image in Fig. 6.2(a). Figure 6.2(b,c) show this region as it appears in both the reference and the floating images. The contrast agent changed the appearance of the blood vessels significantly between the two acquisitions and therefore CC fails to correctly register those sub-images containing such a non-linear intensity change. This problem only appears at higher levels of the hierarchy, when the blood vessels cover a significant surface of a sub-image. In the same time, MI can find the correct registration position, due to robust statistical features in the presence of structures. Fig. 6.2(d,e) show in detail the final registration results when using CC and MI, respectively.

The previous results were illustrating that MI can perform better than CC in case of complex intensity relationship between the images to be registered. Therefore, we further describe the performance of the ENRA when using MI as the similarity measure to perform local rigid regis-
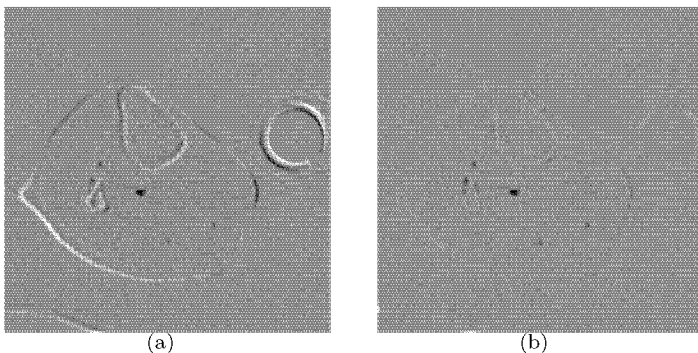
(a)   (b)

Figure 6.3: *Difference images after rigid (a) and non-rigid (b) registration of the 2D mono-modal images. The remaining dark dots in the difference image are caused by the presence of the contrast agent in the blood vessels of the floating image*

trations. Figure 6.3(a) shows the difference image between the initial CTs (Fig. 6.1), as they were extracted from the rigidly pre-registered volumes. In this initial difference image, large misalignments due to different contraction status of the muscles are clearly visible. However, these deformations were successfully compensated by the ENRA as can be seen in the final difference image depicted in Fig. 6.3(b).

In this experiment the entire non-rigid registration procedure has been stopped after the 7th hierarchical level, where the floating image was partitioned into sub-images of at least 8×8 pixels. Figure 6.4 shows the evolution of ENRA by interpolating the individual sparse deformation grid after each level of the hierarchy and then reconstructing the floating CT. The difference images are calculated after each of the 2nd up to the last (7th) hierarchical level.

In order to better visualize the evolution of the adaptive hierarchical splitting strategy using Moran's information consistency test, Fig. 6.5 depicts the partitioning grid of the floating image schematically projected onto the aforementioned set of difference images. Figure 6.5(a) depicts the 2nd hierarchical level, consisting of four partitioned sub-images. The following images show the evolution of the partitioning grid, up to the 7th hierarchical level, which concluded the entire procedure. This figure
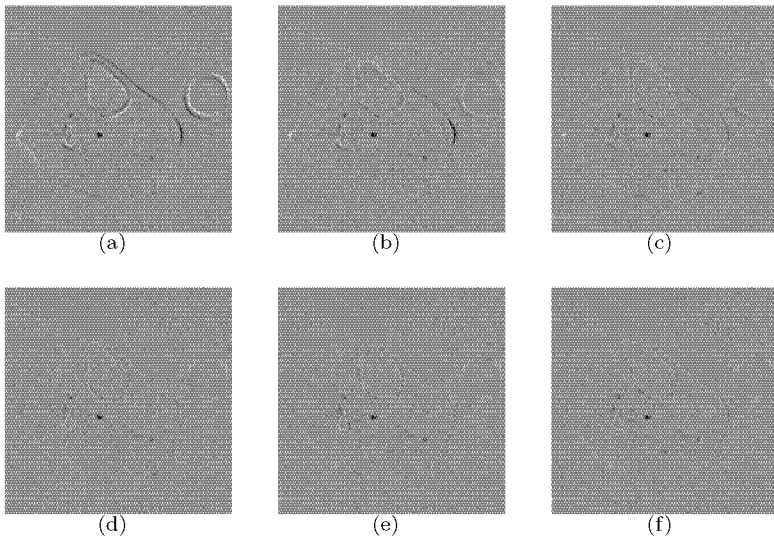
Figure 6.4: *Difference images of the intermediate results obtained after the 2nd (a) up to the 7th (f) hierarchical level*

clearly shows how the image splitting omits those sub-images failing Moran's test, and therefore each level of the hierarchy contains sub-images of variable size.

For this mono-modal experiment, we compared the performance of ENRA to the original method proposed in [Likar and Pernuš 2001]. The original algorithm was stopped after the 5th hierarchical level, because spurious local misregistrations were appearing, without being detected by their proposed triple consistency test as described in Sec. 4.1. Their algorithm not only performed less accurately, but was also 5 times slower as our method. On a Sun Blade 2000 machine, our algorithm needed 10 minutes to complete all the 7 hierarchical levels compared to 52 minutes for only 5 hierarchial levels used by the original method. This big difference is only partially explained by the obviously increased number of sub-images to register at each level of the hierarchy. Another reason lies in the similarity consistency test which enforces the re-optimization of an
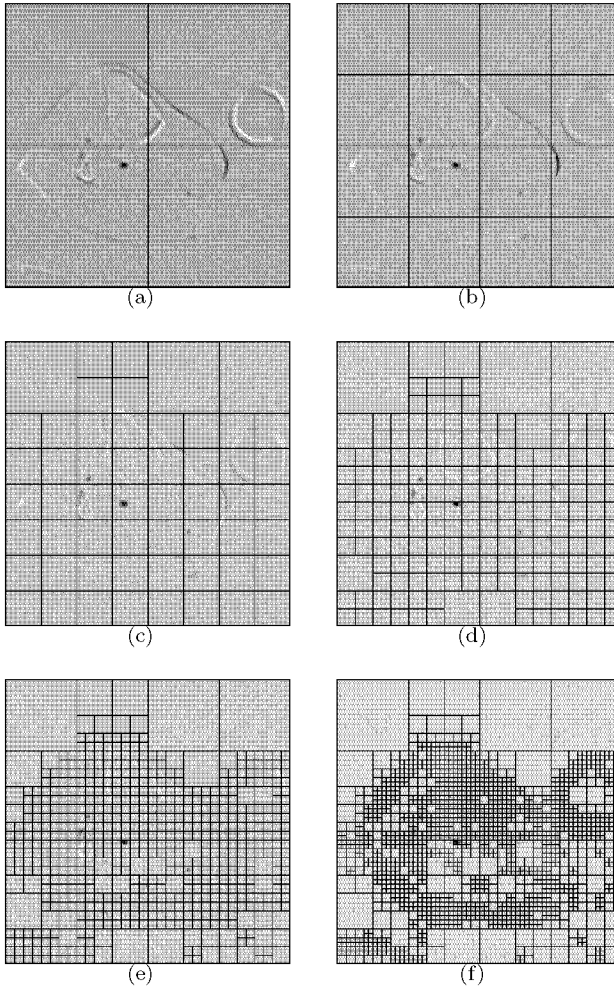
Figure 6.5: *Evolution of the adaptive hierarchical splitting along the non-rigid registration procedure of 2D CT transversal slices of the leg. The partitioning grid was schematically projected onto the difference images between the reference and the partially registered floating CT image after each of the hierarchical levels, from the 2nd (a) up to the 7th (f)*

increased number of structureless sub-images, whose similarity measures are inconsistent and show spurious maxima.

Further experiments have been performed for non-rigidly register multi-modal images. Figure 6.6 shows the performance of ENRA in comparison to the original method for a 2D experiment of a CT/MR cross registration of the head. Similar to the previously presented experiment, the 2D sample images are corresponding slices extracted from the rigidly pre-registered 3D initial volumes. Figure 6.6(a) depicts the reference CT and Fig. 6.6(b) the floating MR slice. Both images have 512×512 pixels of the size 0.5×0.5 mm^2. A comparison between Fig. 6.6(e) and (f) clearly shows the favorable effect of using Moran's test when the local registration is dealing with structureless sub-images. In addition, the Moran's consistency test incorporated into the enhanced algorithm allows us to go further with the hierarchical subdivision down to sub-images of 16×16 pixels, while Likar and Pernuš reported a minimum sub-image size of 64×64 pixels. Further, depending on the number of detected structureless sub-images, our proposed algorithm performed five times faster than the original method (9 minutes for ENRA and 50 minutes for the original algorithm, on a Sun Blade 2000 machine).

The reason for choosing the CT as reference and the MR as floating image lies in the fact that geometric distortions are minimal for a CT scan while the MR will show susceptibility artifacts in the vicinity of air filled cavities of the skull. Consequently the spatial accuracy of the MR scan is expected to be worse. This limited accuracy is an additional important reason that, beside the possible tissue deformations, a non-rigid registration procedure is required for a CT/MR registration.

Figure 6.7 shows the result of the previous experiment as a colored overlay image. The red channel corresponds to the CT and the green to the MR image. The non-rigid registration has been stopped at the 6$^{\text{th}}$ hierarchical level, where the initial floating image was partitioned into sub-images of at least 16×16 pixels. Figure 6.7(a) shows the overlay after the rigid registration and Fig. 6.7(b) shows the final result of the ENRA. A simple visual inspection shows the ability of our method to compensate for large elastic deformations. For example, the algorithm was successfully recovering the strong deformations of the left ear produced by the fixation device used during the CT acquisition. However, the method worked less satisfactory for the right ear which still shows a
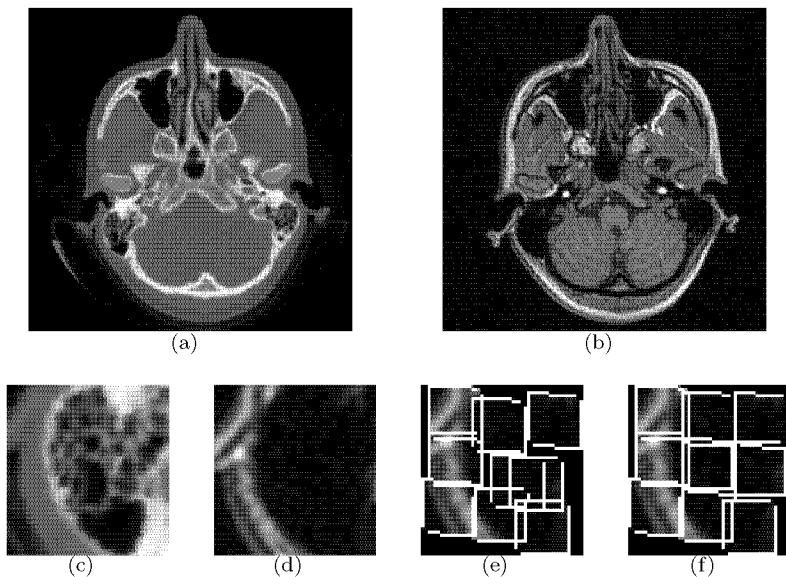
Figure 6.6: Registration details of the sphenoid sinus in the left temporal bone at the 6th level of the hierarchy, where the floating image is divided into 32×32 sub-images of 16×16 pixels. (a) The reference CT. (b) The floating MR image. The examined region, ((c) on the CT and (d) on the MR) consists of 3×3 sub-images. (e) Depicts the final position of each image patch after the local rigid registration, while (f) shows the result after applying the local rigid registration only to those MR patches which passed the Moran test. The consistency check clearly prevented the two middle patches from being pulled towards structures in the reference CT

spatial misalignment. The investigation of this remaining misalignment revealed that the problem was caused by an unfortunate combination of two factors. On one hand, the fix positioning of the partitioned sub-images was not compatible with the image structure in this region. On the other hand, the regularization constraints were too restrictive for reconstructing the actual deformation. Therefore, beside more relaxed geometrical constraints for the deformation field, further improvement allowing adaptive splitting conforming to the structural content of the images will be necessary to improve the performance of the method.

6.2 Experiments in 3D

This section presents the performance of ENRA for registering mono- and multi-modal 3D images. MI was used as the similarity measure for all the experiments presented in this section, even though some of the registration examples are mono-modal.

The parameter setting was basically the same as for the 2D experiments, except for the definition of the contiguity matrix used by Moran's test. In order to use a contiguity matrix with balanced aspect ratio between the three directions, the anisotropy of the acquisitions has to be corrected. For images with voxel dimension $\vec{\xi} = (\xi_x, \xi_y, \xi_z)$, the maximum interaction distance becomes therefore $D = (3, 3, d_z)$ voxels, where $d_z = \max(1, \lfloor 3\xi_x/\xi_z \rfloor, \lfloor 3\xi_y/\xi_z \rfloor)$, with $\lfloor \cdot \rfloor$ being the usual rounding operator.

One of the 3D experiments was performed on the previously mentioned pair of angiographic CT scans of a leg. Figure 6.8(a,b) shows orthogonal sections through the initial volumes. Both data set have $512 \times 512 \times 100$ voxels of the size of $0.3144 \times 0.3144 \times 1$ mm^3. For this registration experiment, the native CT was considered as the reference and the contrast enhanced one was the floating image. In order to accurately detect the calcified plaques on the walls of the blood vessels, a precise registration between these two volumes was required, such that the vessels from the two data sets correspond. Even though the patient's legs were immobilized on the scanner table, the global rigid registration of the two CT volumes shows significant misalignments that only a non-rigid registration can compensate. Figure 6.8(c) illustrates the remaining misalignments

(a)
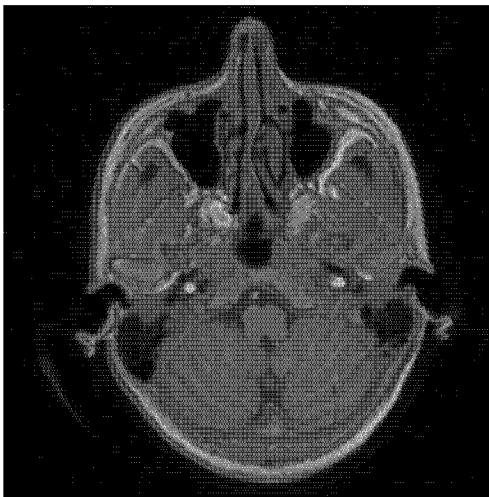
(b)

Figure 6.7: *Rigid (a) and elastic (b) registration results of a neuroradiological CT and MR slices. The reference CT is shown in red while the floating MR image is shown in green*

on orthogonal slices through the difference image after the global rigid registration. These originate from deformations due to differences in leg muscles stresses between the two acquisitions. By applying ENRA, the corresponding elastic deformations have been successfully compensated. Figure 6.8(d) shows the difference image between the native CT scan and the final non-rigidly registered contrast enhanced data. The only visible structures are dark regions, caused by the contrast agent which changed the intensities of the blood vessels between the two volumes. As discussed in the previous section, this mono-modal registration experiment was performed using MI as the similarity measure in order to correctly deal with the local intensity changes induced by the contrast agent.

Additional registration experiments were performed with multi-modal images using the neuroradiological CT/MR image pair mentioned before. A global registration was performed to compensate for the initial big misalignment between the two scans. Figure 6.9 depicts orthogonal sections through the globally rigidly registered images. The initial CT scan is of $512 \times 512 \times 133$ voxels of size $0.309 \times 0.309 \times 0.6 \, \text{mm}^3$ and the MR scan is of $512 \times 512 \times 111$ voxels of size of $0.5 \times 0.5 \times 1 \, \text{mm}^3$. After the rigid registration, in order to use the data set for further experiments, the floating MR image was reconstructed on both the floating and the reference image domains. Therefore, in the last case, the registered MR image became of the same dimension as the reference CT.

The first multi-modal experiment shows the capability of ENRA to correctly perform the non-rigid registration in difficult situations were both small image distortions and big elastic deformations are present in close proximity within an image. The test CT and MR images are cropped from the aforementioned rigidly pre-registered data set. The reference CT has $512 \times 512 \times 46$ voxels of $0.39 \times 0.39 \times 0.6 \, \text{mm}^3$ dimension. The floating MR image is of $512 \times 512 \times 28$ voxels of the size $0.5 \times 0.5 \times 1.0 \, \text{mm}^3$. Figure 6.10 visualizes details of our non-rigid registration method in a region where elastic deformation is needed to correct for both MR susceptibility artifacts (i.e. within the left sphenoid sinus) and tissue deformation (i.e. the left ear). In order to better compare the results between the global rigid registration and after applying the enhanced hierarchical process, Fig. 6.10(b,c) show the outline of the head and of the left sphenoid sinus extracted from the MR volume overlaid on the CT. The

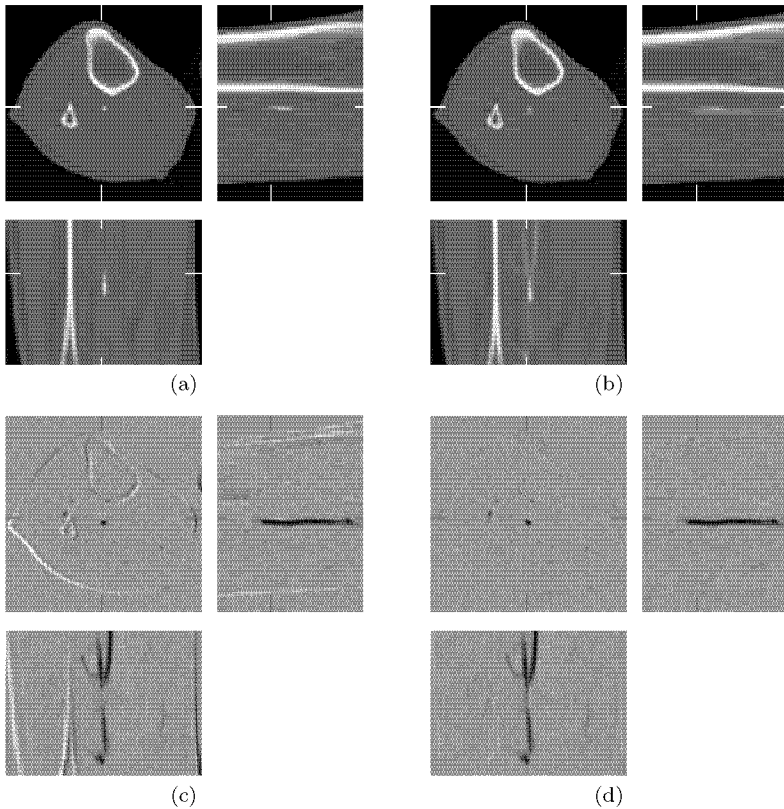**Figure 6.8**: *Rigid and non-rigid registration results for 3D CT scans of a leg. (a) Orthogonal sections through the reference native CT and (b) through the floating contrast enhanced CT. Difference image after their global rigid (c) and hierarchical non-rigid (d) registration. The white marks placed on the borders of the images denote the corresponding cutting planes through the 3D volumes*
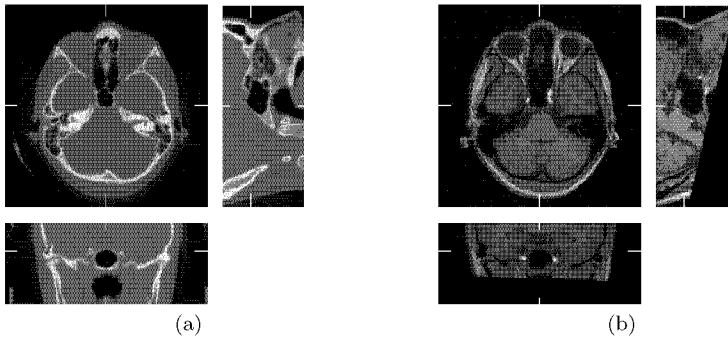
(a)   (b)

Figure 6.9: *Orthogonal sections through (a) the reference CT scan and (b) the floating MR scan after rigid registration. The white marks placed on the borders of the images denote the corresponding cutting planes through the 3D volumes*

remaining deviations of the two contours are caused by both the spatial constraints imposed by the regularization of the deformation field and the size of the smallest sub-image ($16 \times 16 \times 8$ voxels) which was still successfully passing the Moran test.

The following experiment represents the registration of another region of interest extracted from the aforementioned rigidly pre-registered images (Fig. 6.9). The cropping was performed to exclude the lower part of the head, which is not fully covered by the MR image. In this experiment, the reference CT, shown in Fig. 6.11(a), contains 50 slices of 512×512 pixels. The floating MR image is illustrated by Fig. 6.11(b). By using the same colore overlap scheme as described before, the performance of our non-rigid registration algorithm can be seen on Fig. 6.11. The initial overlay of the rigidly registered 3D images is shown on Fig. 6.11(c) while the overlay of the final non-rigidly registered images is demonstrated by Fig. 6.11(d).

6.3   Results validation

Before an image registration algorithm can be used in practice for a clinical procedure, a thorough evaluation has to be performed to assess
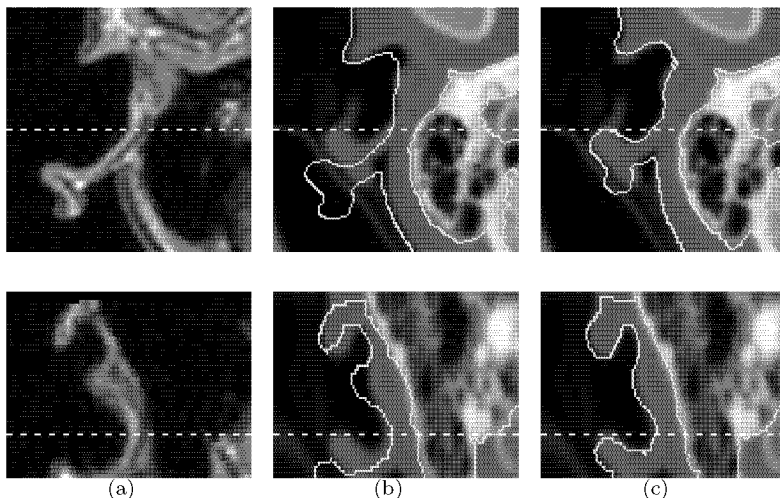
(a)   (b)   (c)

Figure 6.10: *Result of a CT/MR cross-registration. (a) Transversal and coronal sections through the region of interest in the initial floating MR volume. (b,c) Corresponding sections in the reference CT volume, overlaid with the contours of the head and of the sphenoid sinus after a global rigid (b) and after the full hierarchical (c) registration. The dashed lines mark the position of the cutting planes*

its performance. Firstly, a technical validation phase has to estimate the algorithm's robustness and the accuracy of its results. Secondly, the physicians have to clinically validate the new algorithm, by evaluating its usefulness and contribution to improving clinical diagnosis and planning procedures. However, the validation of a registration algorithm is very often a problematic issue, because the ground truth for the spatial correspondences is unknown.

The robustness of a registration algorithm can be characterized by its behavior when external factors affect the input images. The algorithm should handle artificial perturbations induced either in the form of a known amount of misregistration, noise or simulated artifacts in the images. In addition, a robust algorithm should be able to handle data acquired with a wide range of different imaging modalities of different subjects.

(a)   (b)

(c)   (d)
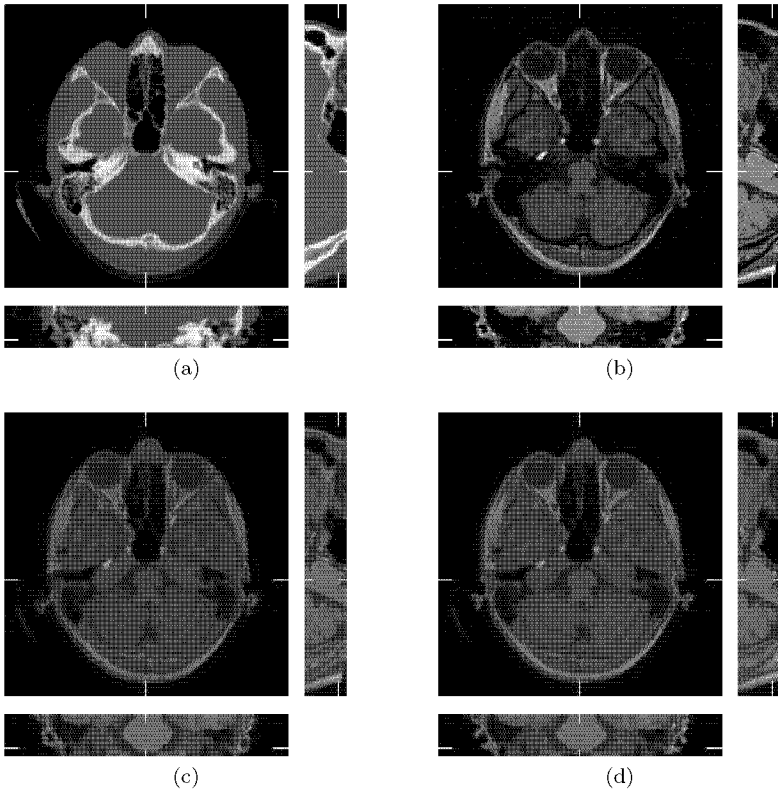
Figure 6.11: *Orthogonal sections through (a) the reference CT and (b) the floating MR image. Orthogonal sections through the colored overlayed images after (c) rigid and (d) non-rigid registration. The CT is colored in red and the MR image in green. Note: The white and yellow lines on the borders of the images mark the corresponding cutting planes through the initial 3D volumes*

The assessment of the accuracy of the registration results is a widely debated issue. Often, the lack of ground truth of the actual spatial misalignment between the images to be registered makes an accurate evaluation of the registration results impossible[1]. The most widely used method for validating the registration result is the visual inspection of the registered images. However, this method is subjective, strongly dependent on the perception and interpretation of the researcher or physician. Clearly, if the images look wrong then the result is most probably wrong. However, if the images look right, the result might still be wrong because very small and localized misregistrations are usually difficult to see. Therefore, to eliminate the subjectiveness of the visual inspection, a mathematical framework to numerically evaluate the correctness of the registration result is required. Many research groups have developed different validation methods to assess the registration accuracy. The most important approaches are based on phantom images, physical markers, segmentation and labeling of different regions or comparison with calculated tissue deformations provided e.g. by the finite element method.

In order to validate our registration algorithm we rely on phantom images. The validation setup consists of an artificial registration scenario to recover a predefined deformation field. To circumvent the problem of the missing ground truth, we chose a triplet of two CT and one MR dataset of a cadaver. The two CT were scanned successively using different radiation energies and we assume them to be perfectly aligned. The third dataset was acquired using a MR scanner and it was registered with the CT data using the hierarchical algorithm. The pre-registered CT/MR datasets have a size of $512 \times 512 \times 60$ voxels of dimension $0.47 \times 0.47 \times 1.25$ mm^3, see Fig. 6.13(a,b). The scheme of the entire validation setup is described in Fig. 6.12. A predefined deformation field

---

[1] J. Michael Fitzpatrick was initiating an international project which aims to evaluate the accuracy of non-invasive, retrospective, rigid, inter-modality image registration techniques for images of the human head, particularly with regard to CT/MR and PET/MR registration. Entitled "Evaluation of Retrospective Image Registration" [West *et al.* 1997], this project was accessible by research groups from all over the world and offered a database of test images for which the correct rigid registration parameters were known, and considered as gold standard. The researchers could download the test images and send back their registration parameters obtained with their own method. Then, a testing protocol for blind evaluation of the results' accuracy was performed by comparing to the gold standard registrations. Unfortunately, the project was closed before we could use its facilities.

$\mathcal{D}_0$ was applied on one of the two CT dataset. For this, the initial volume (CT_1) was split in 4 equally sized blocks of size $256 \times 256 \times 60$ voxels. Two of these blocks were rigidly rotated as shown in Fig. 6.13(c) and then the deformed volume was reconstructed using TPS and partial volume interpolation. The deformed volume $CT_1^{\mathcal{D}_0}$ was then considered as the reference image for the following validation experiments. The second CT volume (CT_2) and the MR image were then successively registered to $CT_1^{\mathcal{D}_0}$ using the following two approaches:

1. MI as the similarity function during the entire hierarchical registration procedure (Sec 5.1);

2. The intensity mapping algorithm described in Sec. 3.4 and switching MI to CC at the 4th hierarchical level (Sec 5.2).

The recovered deformation fields were then compared to the known artificial deformation field $\mathcal{D}_0$.

Table 6.1 summarizes the average $(mean)$, the standard deviation (std) and the maximum deviation (max) of the registration error between $CT_1^{\mathcal{D}_0}$ and CT_2. The registration error was calculated for the entire volume and for a region of interest marked with a white square in Fig. 6.13(a). The region of interest is of size $64 \times 64 \times 40$ voxels. As the gain of switching from MI to CC only applies to a small number of sub-images, the average registration error over the entire volume remains basically in the same range for the two strategies. However, for the selected region of interest, where MI generally tends to fail, the registration error improved considerably when switching from MI to CC. The algorithm using both MI and CC shows a more stable result than in the case of using only MI. However, it can be seen that the out of plane error is smaller in the case of using only MI. This can be explained by the influence of the interpolation artifacts, which in case of partial volume interpolation appear as local maxima in the similarity measure when the voxels perfectly overlap. However, the registration error is in both cases considerably smaller than a voxel dimension, justifying our assumption that the two initial CT dataset are aligned.

Similarly, Table 6.2 summarizes the multi-modal registration errors between the $CT_1^{\mathcal{D}_0}$ and the pre-registered MR image. The increased value
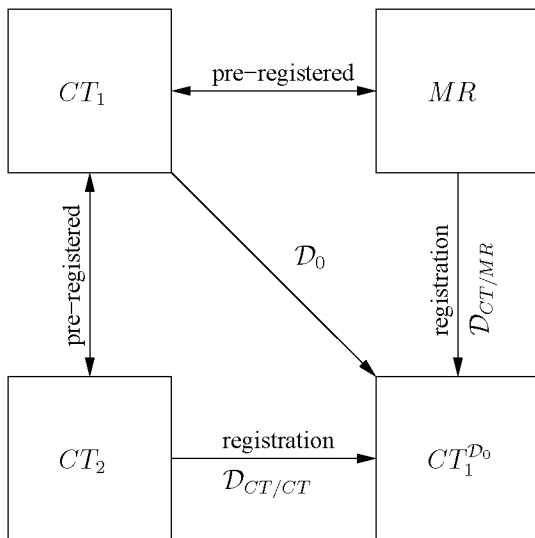
Figure 6.12: *The scheme of the validation setup. The initial CT_1, CT_2 and MR images are pre-registered. A known deformation field $\mathcal{D}_0$ is applied onto CT_1. The resulting $CT_1^{\mathcal{D}_0}$ is considered as reference image for the subsequent registration experiments: $CT_1^{\mathcal{D}_0} \leftrightarrow CT_2$ and $CT_1^{\mathcal{D}_0} \leftrightarrow MR$. The resulting deformation fields are then compared to the initial, known deformation field $\mathcal{D}_0$*
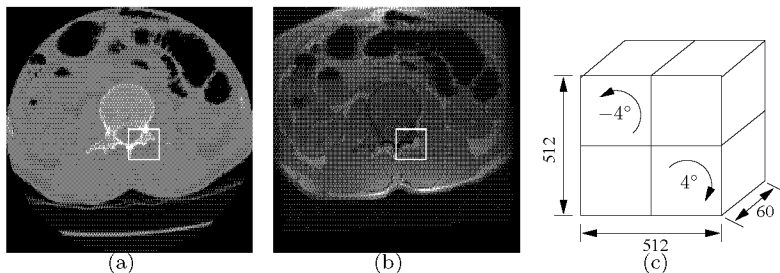

Figure 6.13: *(a) CT and (b) MR sample slice of the spine volume used for the validation tests (c) schematic visualization of the artificial deformation field $\mathcal{D}_0$*

$CT_1^{\mathcal{D}_0} \leftrightarrow CT_2$					
Global Results		MI [mm]		MI+CC [mm]	
Level	Direction	$mean \pm std$	max	$mean \pm std$	max
4	in plane (xy)	0.14 ± 0.12	1.36	0.18 ± 0.15	1.56
4	out of plane (z)	0.04 ± 0.11	1.56	0.16 ± 0.21	1.56
4	overall (xyz)	0.15 ± 0.15	1.65	0.27 ± 0.23	1.96
5	in plane (xy)	0.12 ± 0.10	1.22	0.13 ± 0.13	1.88
5	out of plane (z)	0.03 ± 0.10	1.56	0.11 ± 0.15	1.88
5	overall (xyz)	0.13 ± 0.13	1.64	0.18 ± 0.19	2.20
6	in plane (xy)	0.12 ± 0.10	1.22	0.12 ± 0.12	1.53
6	out of plane (z)	0.03 ± 0.10	1.56	0.09 ± 0.13	1.56
6	overall (xyz)	0.13 ± 0.14	1.65	0.17 ± 0.16	1.64
Local Results					
4	in plane (xy)	0.05 ± 0.04	0.38	0.10 ± 0.06	0.33
4	out of plane (z)	0.01 ± 0.01	0.14	0.05 ± 0.04	0.25
4	overall (xyz)	0.06 ± 0.04	0.41	0.11 ± 0.07	0.39
5	in plane (xy)	0.13 ± 0.08	0.39	0.06 ± 0.05	0.34
5	out of plane (z)	0.01 ± 0.01	0.10	0.03 ± 0.05	0.46
5	overall (xyz)	0.13 ± 0.08	0.41	0.08 ± 0.06	0.46
6	in plane (xy)	0.12 ± 0.06	0.35	0.06 ± 0.04	0.38
6	out of plane (z)	0.01 ± 0.01	0.08	0.03 ± 0.04	0.44
6	overall (xyz)	0.12 ± 0.06	0.35	0.08 ± 0.05	0.52

Table 6.1: *Evolution along the hierarchy of the CT/CT registration error calculated for the entire volume as well as for the local region marked with a white square in Fig. 6.13(a)*

of the registration error compared to the previous mono-modal CT/CT case can be expected. Firstly, one has to consider that these results include the registration errors of the pre-registration (CT_1/MR). Secondly, multi-modal registration cases are normally more difficult and less accurate than mono-modal registrations. This reduced accuracy is often induced by the lack of correspondence between the imaged structures. Additionally, the registration accuracy is reduced by interpolation artifacts while compensating for the differences in the spatial resolution of the images. However, the results look promising, as the registration error is within voxel dimension.

Another advantage of the hybrid algorithm that switches the similarity measure from MI to CC is a considerable increase of the computational efficiency. Using a 3GHz Intel Pentium PC for the previous mono-modal registration example, the algorithm execution time was 3h49' when using only MI and 3h03' when switching to CC. For the multi-modal registration example, the execution time was 7h22' in case of using MI and 4h30' when switching to CC. The difference in the execution time between the two versions of the algorithm comes from two factors: (1) the calculation of CC is faster than of MI and (2) more sub-images need to be reoptimized when MI is used as a similarity measure. At the same time, more reoptimizations were required for the multi-modal example compared to the mono-modal one, explaining the difference in execution time between the two registration experiments.

$CT_1^{\mathcal{D}_0} \leftrightarrow MR$					
Global Results		MI [mm]		MI+CC [mm]	
Level	Direction	$mean \pm std$	max	$mean \pm std$	max
4	in plane (xy)	0.43 ± 0.39	4.68	0.54 ± 0.50	6.93
	out of plane (z)	0.39 ± 0.58	6.08	0.35 ± 0.44	4.01
	overall (xyz)	0.63 ± 0.66	7.59	0.71 ± 0.60	6.93
5	in plane (xy)	0.48 ± 0.40	4.70	0.46 ± 0.47	5.16
	out of plane (z)	0.41 ± 0.55	7.31	0.38 ± 0.43	4.37
	overall (xyz)	0.70 ± 0.62	8.64	0.65 ± 0.58	5.23
6	in plane (xy)	0.51 ± 0.41	4.44	0.47 ± 0.46	4.95
	out of plane (z)	0.42 ± 0.56	7.18	0.39 ± 0.44	4.14
	overall (xyz)	0.73 ± 0.62	8.35	0.67 ± 0.58	4.98
Local Results					
4	in plane (xy)	0.18 ± 0.10	1.01	0.17 ± 0.09	0.95
	out of plane (z)	0.29 ± 0.27	1.86	0.13 ± 0.10	0.61
	overall (xyz)	0.36 ± 0.26	2.12	0.24 ± 0.10	0.96
5	in plane (xy)	0.23 ± 0.10	0.95	0.17 ± 0.09	0.94
	out of plane (z)	0.30 ± 0.25	1.65	0.29 ± 0.21	1.76
	overall (xyz)	0.41 ± 0.23	1.89	0.36 ± 0.19	1.76
6	in plane (xy)	0.36 ± 0.22	1.94	0.23 ± 0.17	2.00
	out of plane (z)	0.27 ± 0.27	1.45	0.23 ± 0.20	1.20
	overall (xyz)	0.49 ± 0.29	2.02	0.36 ± 0.23	2.06

Table 6.2: *Evolution along the hierarchy of the CT/MR registration error calculated for the entire volume as well as for the local region marked with a white square in Fig. 6.13(a,b)*

7

Conclusions and Outlook

Modern clinical care increasingly depends on efficient algorithms of high accuracy for the fusion of temporal and multi-modal image datasets. These fused images are indispensable for pre-operative planning and intra-operative navigation. In this dissertation we developed a new algorithm for automatic non-rigid registration of multi-modal images.

The introduction of MI as a measure of similarity revolutionized image registration, as it extended the applicability of intensity based methods to multi-modal registrations. The considerable research conducted during the last decade on the properties of MI revealed, however, several limitations in connection with interpolation artifacts and its statistical consistency. Our investigation disclosed two additional problems which have not been addressed in the literature before but nevertheless can seriously deteriorate registration accuracy. One the one hand, MI shows a peculiar behavior for image patches having no clear structural content, and on the other hand MI proved to be statistically inconsistent when is estimated for a limited number of image samples. As we relied on a previously proposed approach using hierarchical image splitting strategy to decompose the global non-rigid registration into numerous local rigid registrations of sub-images of decreasing size, these effects become increasingly serious along the hierarchical splitting. The image subdivision process inherently reduces the available number of image samples within the sub-images and at the same time increases the probability of structureless sub-images to appear. This thesis presents several new methods we developed to overcome these limitations. Integrated into a new, enhanced, non-rigid registration algorithm, the proposed solutions try to identify the problematic structureless regions and, on the other

hand, try to replace MI with a more robust similarity measure for small sub-images.

The first improvement we proposed was the adaptive subdivision scheme, where local decision is taken whether registration and further subdivision are needed. Inspired by point pattern analysis, we used Moran's I spatial autocorrelation coefficient to quantify the structural content within an image region. Based on this measure we developed a test for the structural consistency of an image patch. The Moran test eliminates all those image regions that lack structural content from any further local registrations and subdivisions as these are very likely to lead to incorrect results. This has not only led to a considerable increase of the reliability and accuracy of the entire registration, but also enabled to match smaller details by reaching higher levels of the hierarchical splitting. Moreover, depending on the structural content of the images, it considerably decreased the computational load of the registration compared to the original approach proposed by Likar and Pernuš. In addition, this strategy does not necessitate complex heuristics for the deformation field regularization procedure, but can rely on simple geometrical tests.

The second improvement we proposed was to switch to a more robust similarity measure when the statistical consistency of MI reaches its limit due to the restricted number of available image samples. As CC proved to be more robust for such cases, our research was directed towards an intensity mapping allowing to transfer multi-modal registration to a mono-modal problem, which can be handled by cross correlation. The newly proposed method uses the joint intensity histogram of the coarsely registered images to estimate an intensity mapping that transforms the images from different modalities into a common intermediate pseudo-modality. This method was used at later hierarchical levels of the proposed non-rigid registration algorithm. This strategy offers several distinct advantages. By using MI at the first levels of the hierarchy we retain the multi-modal capability of the registration algorithm. At the same time, by coarsely registering the images, we increase the reliability of the estimated intensity mapping functions. At the same time, the robustness of the registration is increased by avoiding the spurious local misregistrations induced by MI at later stages of the hierarchy.

There are several directions for further research to improve the robustness of the non-rigid registration algorithm for multi-modal images. One

possible extension addresses the adaptive image splitting strategy. Currently, the positioning of the splitting boundaries is purely driven by the image lattice, completely ignoring the image content. Adjustment of splitting boundaries according to the anatomical structures would, however, not only reduce the computational load but also improve image registration quality.

Other issues that should be further investigated are related to the estimation of the deformation field. Firstly, constraints could be imposed on the registration parameter correction by determining the reliability of the local registrations along the different directions using the gradient of the similarity measure. Secondly, prior knowledge about the physical properties of the tissue (e.g. stiffness) could be integrated into the matching process if available. This would require a preprocessing phase for the segmentation of the different tissues. As an example, bony structures could impose rigid transformation constraints to the deformation field.

Finally, the computational speed of the algorithm can further be improved by using a distributed computing strategy. Since at every level of the hierarchy there are multiple local registrations which are completely independent, the hierarchical algorithm can be easily parallelized. This would considerably decrease calculation time.

Acknowledgements

First of all I would like to thank my supervisor Prof. Dr. Gábor Székely for accepting me as a doctoral student and for suggesting this challenging topic of research in the field of computer-assisted medical imaging. I am grateful to him for his continuous support and guidance, which contributed enormous to the successful completion of my Ph.D. studies. Also I would like to thank Prof. Dr. Franjo Pernuš, my co-examiner, for his support.

Likewise, I wish to express my heartfelt gratitude to a special friend, Dr. Philippe Cattin, who closely supervised my research and patiently provided me inestimable support along the last three years. I wish to thank him for all the valuable ideas and discussions we had together, for the countless hours he spent correcting and commenting this manuscript and all the other published papers.

Further I would like to express special thanks to all my friends from all over the world, who gave me all the affective support one could dream off and help me go through exciting and invaluable experiences during all my sojourn in Zurich.

Also, I want to thank Sonia Paget for correcting my English. Further I would like to thank all my BIWI colleagues for their friendship, technical support and interesting discussions.

Last, but not least, I want to thank my parents and my sister who accompanied and supported me warmly through all these years I was so far away from them.

Bibliography

[Andronache *et al.* 2005] Adrian Andronache, Philippe Cattin, and Gábor Székely. Adaptive subdivision for hierarchical non-rigid registration of multi-modal images using mutual information. In James S. Duncan and Guido Gerig, editors, *MICCAI (2)*, volume 3750, pages 976–983. Springer-Verlag GmbH, Sept 2005.

[Bajcsy and Kovacic 1989] Ruzena Bajcsy and Stane Kovacic. Multiresolution elastic matching. *Comput. Vision Graph. Image Process.*, 46(1):1–21, 1989.

[Bardinet *et al.* 1996] E. Bardinet, L. D. Cohen, and N. Ayache. Tracking and motion analysis of the left ventricle with deformable superquadrics. *Med. Image Anal.*, 1:129–149, 1996.

[Bookstein 1989] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans Pattern Anal Mach Intell*, 11:567–585, June 1989.

[Bro-Nielsen and Gramkow 1996] M. Bro-Nielsen and C. Gramkow. Fast fluid registration of medical images. In R. Hohne, K.H.; Kikinis, editor, *Visualization in Biomedical Computing. 4th International Conference, VBC '96 Proceedings*, volume 1131, pages 267–276. Springer-Verlag, 1996.

[Bro-Nielsen 1998] M. Bro-Nielsen. Finite element modeling in medical vr. *IEEE*, 86:490–503, 1998.

[Brown 1992] Lisa Gottesfeld Brown. A survey of image registration techniques. *ACM Computing Surveys (CSUR)*, 24(4):325 – 376, December 1992.

[Christensen *et al.* 1995] G. E. Christensen, M. I. Miller, J. L. Marsh, and M. W. Vannier. Automatic analysis of medical images using a

deformable textbook. *Computer assisted radiology*, pages 146–151, 1995.

[Cliff and Ord 1973] A.D. Cliff and J.K. Ord. *Spatial Autocorrelation.* ISBN 0-85086-036-9. Pion Limited: R.J. Chorley and D.W. Harvey, 207 Brondesbury Park, London NW2 5JN, 1973.

[Collignon *et al.* 1995] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. *Information Processing in Medical Imaging*, pages 263–74, 1995.

[Cover and Thomas 1991] T.M. Cover and J.A. Thomas. *Elements of Information Theory.* John Wiley & Sons, 1991.

[Davatzikos 1996] C. Davatzikos. Nonlinear registration of brain images using deformable models. *Mathematical methods in biomedical image analysis*, pages 94–103, 1996.

[Edwards *et al.* 1998] P.J. Edwards, D.L.G. Hill, and D.J. Hawkes. A three-component deformation model for image-guided surgery. *Med. Imag. Anal.*, 2:355–367, 1998.

[Guimond *et al.* 2001] Alexandre Guimond, Alexis Roche, Nicholas Ayache, and Jean Meunier. Three-dimensional multimodal brain warping using the demons algorithm and adaptive intensity corrections. *IEEE Transactions on Medical Imaging*, 20(1):58–69, January 2001.

[Hajnal *et al.* 1995] J.B. Hajnal, N. Saeed, E.J. Soar, A. Oatridge, I.R. Young, and G.M. Bydder. A registration and interpolation procedure for subvoxel matching of serially acquired mr images. *Comput. Assist. Tomogr.*, 19:289–296, 1995.

[Hajnal *et al.* 2001] Joseph V. Hajnal, Derek L.G. Hill, and David J. Hawkes. *Medical Image Registration.* CRC Press, 2001.

[Hellier *et al.* 2001] P. Hellier, C. Barillot, E. Memin, and P. Perez. Hierarchical estimation of a dense deformation field for 3-d robust registration. *Medical Imaging, IEEE Transactions on*, 20(5):388–402, May 2001.

[Hill *et al.* 2001] Derek L G Hill, Philipp G Batchelor, Mark Holden, and David J Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46, 2001.

[Horn and Schunck 1981] B. Horn and B. Schunck. Determining optical flow. *Artif. Intell.*, 17:185–203, 1981.

[Ji *et al.* 2003] Jim Xiuquan Ji, Hao Pan, and Zhi-Pei Liang. Further analysis of interpolation effects in mutual information-based image registration. *IEEE Transactions on Medical Imaging*, 22(9):1131–1140, September 2003.

[Lee and Wong 2001] Jay Lee and David W.S. Wong. *Statistical Analysis with ArcView GIS*. ISBN 0-471-34874-0. John Wiley & Sons, Inc., 2001.

[Lee *et al.* 1996] S. Lee, G. Wolberg, K.-Y. Chwa, and S. Y. Shin. Image metamorphosis with scattered feature constraints. *IEEE Trans. Visualization Comput. Graph.*, 2:337–354, 1996.

[Lee *et al.* 1997] S. Lee, G. Wolberg, and S. Y. Shin. Scattered data interpolation with multilevel b-splines. *IEEE Trans. Visualization Comput. Graph.*, 3:228–244, 1997.

[Likar and Pernuš 2001] Bostjan Likar and Franjo Pernuš. A hierarchical approach to elastic registration based on mutual information. *Image and Vision Computing*, 19:33–44, January 2001.

[Little *et al.* 1997] J. A. Little, D. L. G. Hill, and D. J. Hawkes. Deformations incorporating rigid structures. *Computer Vision Image Understanding*, 66(2):223–232, May 1997.

[Maes *et al.* 1996] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multi-modality image registration by maximization of mutual information. *Proc IEEE Workshop Mathematical Methods in Biomedical Image Analysis*, pages 14–22, June 1996.

[Maes *et al.* 1997] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, April 1997.

[Maes *et al.* 2003] Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Medical image registration using mutual information. *Proc IEEE*, 91(10):1699–1722, October 2003.

[Maintz and Viergever 1998] J. Maintz and M. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

[Maintz *et al.* 1998] J. Maintz, E. Meijering, and M. Viergever. General multimodal elastic registration based on mutual information. In *Proc. SPIE Vol. 3338, p. 144-154, Medical Imaging 1998: Image Processing, Kenneth M. Hanson; Ed.*, pages 144–154, June 1998.

[Nikou *et al.* 1998] C. Nikou, F. Heitz, J.P. Armspach, , and I.J. Namer. Single and multimodal subvoxel registration of dissimilar medical images using robust similarity measures. *SPIE Conference on Medical Imaging*, 1998.

[Pluim *et al.* 1999] Josien P. W. Pluim, J.B. Antoine Maintz, and Max A. Viergever. Mutual information matching and interpolation artefacts. *Proc SPIE Medical Imaging*, 3661:56–65, 1999.

[Pluim *et al.* 2000] JPW Pluim, JB Antoine Maintz, and MA Viergever. Interpolation artefacts in mutual information-based image registration. *Computer Vision and Image Understanding*, 77(2):211–232, February 2000.

[Pluim *et al.* 2003] Josien P. W. Pluim, J.B. Antoine Maintz, and Max A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22:986–1004, 2003.

[Press *et al.* 1988] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes in C - The Art of Scientific Computing.* ISBN 0-521-35465-X. Cambridge University Press, 1988.

[Roche *et al.* 1998] Alexis Roche, Gregoire Malandain, Xavier Pennec, and Nicholas Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In W.M. Wells et al., editor, *MICCAI*, 1998.

[Rohde *et al.* 2003] G.K. Rohde, A. Aldroubi, and B.M. Dawant. The adaptive bases algorithm for intensity-based nonrigid image registration. *Medical Imaging, IEEE Transactions on*, 22(11):1470 – 1479, November 2003.

[Rohr *et al.* 1996] K. Rohr, H. S. Stiehl, R. Sprengel, W. Beil, T. M. Buzug, J. Weese, and M. H. Kuhn. Point-based elastic registration of medical image data using approximating thin-plate splines. In *Visualization in Biomedical Computing, Lecture Notes in Computer Science*, volume 1131, pages 297–306, 1996.

[Rueckert *et al.* 1999] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Transactions on Medical Imaging*, 18:712–721, 1999.

[Shannon 1948] C. E. Shannon. A mathematical theory of communication. Technical report, Bell Laboratories, 1948.

[Shepard 1968] Donald Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, pages 517–524. ACM Press, 1968.

[Stefanescu *et al.* 2004] R. Stefanescu, X. Pennec, and N. Ayache. Grid powered nonlinear image registration with locally adaptive regularization. *Medical Image Analysis*, 8(3):325–342, September 2004. MICCAI 2003 Special Issue.

[Studholme *et al.* 1995] C. Studholme, D.L.G. Hill, and D.J. Hawkes. Multiresolution voxel similarity measures for mr-pet registration. *Information Processing in Medical Imaging*, pages 287–98, 1995.

[Studholme *et al.* 1997] C. Studholme, D.L.G. Hill, and D.J. Hawkes. Automated 3d registration ofmrand pet brain images by multiresolution optimization of voxel similarity measures. *Med. Phys.*, 24:25–35, 1997.

[Studholme *et al.* 1999] C. Studholme, D.L.G. Hill, and D.J. Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 32:71–86, 1999.

[Thirion 1996] J.P. Thirion. Non-rigid matching using demons. In *Int. Conf. Computer Vision and Pattern Recognition*, 1996.

[Tsao 2003] Jeffrey Tsao. Interpolation artifacts in multimodality image registration based on maximization of mutual information. *IEEE Transactions on Medical Imaging*, 22(7):854–864, July 2003.

[Vajda 1989] Igor Vajda. *Theory of Statistical Inference and Information*. B. Kluwer Academic, 1989.

[van den Elsen *et al.* 1994] P. A. van den Elsen, E.-J. D. Pol, T. S. Sumanaweera, P. F. Hemler, S. Napel, and J. R. Adler. Grey value correlation techniques for automatic matching of CT and MR brain and spine images. *Visualization in Biomed. Comp.*, 2359:227–237, 1994.

[van der Elsen *et al.* 1993] P.A. van der Elsen, E.J.D. Pol, and M.A. Viergever. Medical image matching - a review with classification. *IEEE Eng. in Medicine and Biology*, 12(1):26–39, 1993.

[Viola and W. M. Wells 1995] P. Viola and III W. M. Wells. Alignment by maximization of mutual information. In *ICCV '95: Proceedings of the Fifth International Conference on Computer Vision*, ISBN 0-8186-7042-8, page 16. IEEE Computer Society, 1995.

[Viola 1995] Paul A. Viola. *Alignment by Maximization of Mutual Information*. PhD thesis, Massachusetts Institute of Technology, 1995.

[West *et al.* 1997] Jay West, J. Michael Fitzpatrick, Matthew Y. Wang, Benoit M. Dawant, Calvin R. Maurer, Robert M. Kessler, Robert J. Maciunas, and et al. Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21:554–566, 1997.

[Woods *et al.* 1992] R.P. Woods, S.R. Cherry, and J.C. Mazziotta. Rapid automated algorithm for aligning and reslicing pet images. *Comput. Assist. Tomogr.*, 16:620–633, 1992.

[Woods *et al.* 1993] R. P. Woods, J. C. Mazziotta, and S. R. Cherry. MRI-PET registration with automated algorithm. *Comput. Assist. Tomogr.*, 17(4):536–546, 1993.

[Xu and Dony 2004] Xiaoyan Xu and Robert D. Dony. Evaluation of hierarchical elastic medical image registration method. *Proceedings of the 2004 Canadian Conference on Electrical and Computer Engineering*, May 2004.

[Zitova and Flusser 2003] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.

Curriculum Vitae

Personal Data

Name	Adrian Stefan Andronache
Date of birth	06 May 1976
Place of birth	Targoviste, Romania
Citizenship	Romanian

Education and Degrees

2002 - 2006	Doctor of Science, Swiss Federal Institute of Technology (ETH) Zurich, Switzerland, Department of Information Technology and Electrical Engineering, Computer Vision Laboratory
2000 - 2001	Master of Science in Electrical Engineering, Politehnica University of Bucharest, Romania, Faculty of Electronics and Telecommunications, Section Wide-Band Residential Communications Systems
1995 - 2000	Bachelor of Science in Electrical Engineering, Politehnica University of Bucharest, Romania, Faculty of Electronics and Telecommunications, Section Numerical Communications, Multimedia and Hi-Fi Systems