
Robust Algorithms for Registration of 3D Images of Human Brain

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Jiří Čížek

aus České Budějovice,
Tschechische Republik

Köln 2004

Berichterstatter:

Prof. Dr. Rainer Schrader
Prof. Dr. med. Karl Herholz

Tag der mündlichen Prüfung: 31.1.2005

Abstract

This thesis is concerned with the process of automatically aligning 3D medical images of human brain. It concentrates on rigid-body matching of Positron Emission Tomography images (PET) and Magnetic Resonance images (MR) within one patient and on non-linear matching of PET images of different patients.

In recent years, mutual information has proved to be an excellent criterion for automatic registration of intra-individual images from different modalities. We propose and evaluate a method that combines a multi-resolution optimization of mutual information with an efficient segmentation of background voxels and a modified principal axes algorithm. We show that an acceleration factor of 6-7 can be achieved without loss of accuracy and that the method significantly reduces the rate of unsuccessful registrations. Emphasis was also laid on creation of an automatic registration system that could be used routinely in clinical environment.

Non-linear registration tries to reduce the inter-individual variability of shape and structure between two brain images by deforming one image so that homologous regions in both images get aligned. It is an important step of many procedures in medical image processing and analysis. We present a novel algorithm for an automatic non-linear registration of PET images based on hierarchical volume subdivisions and local affine optimizations. It produces a C^2 -continuous deformation function and guarantees that the deformation is one-to-one. Performance of the algorithm was evaluated on more than 600 clinical PET images.

Kurzzusammenfassung

Diese Arbeit beschäftigt sich mit der automatischen Korregistrierung dreidimensionaler Datensätze des menschlichen Gehirns, wie sie bei der Verwendung moderner bildgebender Verfahren gewonnen werden. Sie konzentriert sich auf das Ausrichten von Datensätzen aufeinander, die mittels Positronenemissionstomographie (PET) und Magnetresonanztomographie (MR) des gleichen Probanden erzeugt wurden sowie auf die nichtlineare Verformung von PET-Bildern, die von unterschiedlichen Probanden stammen.

In den letzten Jahren hat sich das “Mutual Information”-Konzept als ein exzellentes Kriterium für die automatisierte Korregistrierung medizinischer Bilddatensätze unterschiedlicher Modalitäten erwiesen. In dieser Arbeit wird eine Methode entwickelt, die eine Optimierung von “Mutual Information” in verschiedenen granularen Auflösungen mit einer effizienten Segmentierung des Hintergrundes und einer Modifikation des “Principal Axes”-Algorithmus verknüpft. Wir zeigen, daß eine 6-7fache Beschleunigung des Registrierungsprozesses ohne Verlust an Präzision möglich ist und dabei noch die Rate fehlerhafter Korregistrierungen deutlich gesenkt wird.

Zweck der nichtlinearen Korregistrierung ist es, die interindividuelle Variabilität zu reduzieren, die die Form und Struktur unterschiedlicher menschlicher Gehirne aufweist. Dabei wird ein Bild räumlich deformiert und an das andere Bild angepasst. Dies ist Ausgangspunkt vieler Prozeduren der medizinischen Bildverarbeitung und Bildanalyse. Wir stellen einen neuen Algorithmus für die automatische nichtlineare Korregistrierung von PET Bildern vor, der auf einer hierarchischen Teilung des Bildvolumens und lokalen affinen Optimierungen basiert. Der Algorithmus erzeugt eine C^2 -kontinuierliche Deformationsfunktion und garantiert, daß die Deformation eineindeutig ist. Die Leistung des Algorithmus wurde auf einer Anzahl von über 600 klinischen PET Bildern validiert.

Contents

1	Introduction	1
1.1	Review of medical image registration	3
1.1.1	Landmark based methods	4
1.1.2	Voxel based methods	4
1.2	Overview of the text	5
2	Structure and transformations of 3D medical images	7
2.1	Medical images	7
2.1.1	Positron emission tomography	7
2.1.2	Magnetic resonance imaging	9
2.2	Voxel and world coordinate systems	10
2.3	Image transformations	11
2.3.1	Affine transformation	12
2.3.2	Rigid-body transformation	14
2.3.3	Conversion between coordinate systems	14
2.3.4	Resampling to different image dimensions	15
2.3.5	Transformation of a 3D image	15
2.3.6	Interpolation	19
2.3.7	Decomposition of transformation matrix	20
3	Similarity measures	23
3.1	Introduction	23
3.2	Image overlap	25
3.3	Sum of squared differences	25
3.4	Correlation coefficient	26
3.5	Mutual information	27
3.5.1	Image histogram	28
3.5.2	Joint image histogram	30
3.5.3	Shannon entropy	31
3.5.4	Joint entropy	31
3.5.5	Mutual information	32
3.5.6	Normalized mutual information	32
3.6	Interpolation schemes for similarity measures	33
3.7	Complexity of similarity evaluation	33
4	Rigid-body registration	35
4.1	Introduction	35
4.2	Optimization	35
4.2.1	A note about computational efficiency	36
4.3	Multi-scale approach	37

4.3.1	Smoothing	37
4.3.2	Scale schemas	38
4.4	Summary of the algorithm	39
4.5	Segmentation of background	40
4.5.1	Thresholding	41
4.5.2	Iterative optimal threshold selection	42
4.5.3	Morphological cleaning	45
4.5.4	Performance of the morphological segmentation	47
4.5.5	Similarity measures with background masking	47
4.6	Global misalignment estimation	48
4.6.1	Moments and the principal axes transformation	50
4.6.2	Image registration using principal axes	52
4.6.3	Corrections for PET and MR images	54
4.7	How many bins to use for computation of mutual information?	55
4.7.1	Statistical background	55
4.7.2	Histogram size and mutual information	58
4.8	Experimental setup	59
4.8.1	Simulated PET images	60
4.8.2	Clinical PET and MR images	61
4.9	Test of the influence of histogram size	62
4.9.1	Results and discussion	63
4.10	Evaluation of masking and global displacement estimation	70
4.10.1	Results and Discussion	70
5	Non-linear registration	75
5.1	Introduction	75
5.2	Rationale for a new approach	76
5.3	Hierarchical Free-Form Block Matching	78
5.3.1	Similarity measure	78
5.3.2	Affine part	79
5.3.3	Non-linear part	79
5.3.4	Free-form Deformation	80
5.3.5	Lattices of control points	81
5.3.6	B-spline approximation	82
5.3.7	One-to-one mapping	83
5.3.8	Hierarchical subdivisions	85
5.3.9	Linear regularization	88
5.3.10	Preprocessing	90
5.3.11	Summary of the algorithm	90
5.3.12	Complexity of the algorithm	92
5.3.13	Spatial normalization	92
5.4	Evaluation	94

5.4.1	Binary geometric images	94
5.4.2	Normalization of FDG-PET and FMZ-PET images . . .	96
5.4.3	Normalization of high-resolution HRRT PET images .	101
5.4.4	Evaluation on clinical images	102
5.5	Spatial normalization of images with lesions	106
5.5.1	Creation of lesion mask	107
5.5.2	Masking within the hierarchical framework	108
5.6	Discussion	110
6	Implementation	113
6.1	Graphical user interface	113
6.2	Multi-threading	114
7	Conclusions	117
7.1	Contributions	117
7.1.1	Rigid-body registration	117
7.1.2	Non-linear registration	119
7.1.3	User interface	120
7.2	Future directions	120
	References	121

Introduction

Medical imaging is a rapidly developing field of medical diagnostics. It allows physicians to see into human body without the necessity to surgically open it and to study internal structures and function of human body in a non-invasive way. The principle of medical imaging is a measurement of certain physical properties of a real object (patient's body) in order to delineate functional or anatomical characteristics of the object.

Various physical properties of the human body are exploited, e.g.:

- Measurement of radioisotope density in tissue as a result of radio-tracer uptake reveals *functional information*. The radioisotope density is measured in a positron emission tomograph (PET) or in a single-photon emission tomograph (SPET).
- Measuring of proton spin density and relaxation times in a magnetic resonance scanner (MR) provides a precise anatomical view of *soft tissue*.
- Measuring of X-ray attenuation using a computed tomograph (CT) detects anatomy of *rigid tissue* (e.g. bones).

Modern scanners are able to deliver three-dimensional images reflecting distribution of the measured property. Such images allow a detailed inspection of structural and functional relationships in three dimensions and they initiated a rapid development of many new applications in neurology, neurosurgery, radiotherapy, cardiology and other areas. The applications now occur throughout the whole clinical track of events: diagnostics, therapy planning, treatment (e.g. surgical operations, radiotherapeutical treatment) and therapy follow-up.

Multi-modal registration

Each modality delineates and emphasizes different property of the measured object. A large information gain can be achieved by combining complementary information about the same object acquired by different modalities.

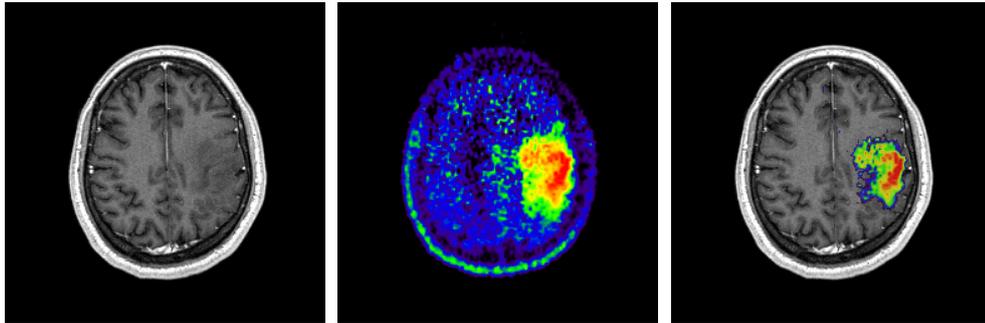


Figure 1.1: MR and PET image of a patient with a tumour. MR image (left) shows rather homogenous changes of the tissue whereas the high uptake of ^{11}C -methionin in the PET image (middle) indicates presence of a tumour that can be precisely localized and analyzed by combining the information in both images (right) (Kracht et al., 2004).

For example, MR images provide a high-resolution information about the anatomy of different types of soft tissue and PET images provide an information about specific functions of different structures but mostly with poor anatomical detail and lower spatial resolution. Figure 1.1 shows a single slice of a PET brain image revealing a spot with an abnormally high level of glucose metabolism indicating presence of a tumour. Combining the PET and the MR image allows a precise anatomical localization of the tumour site.

In practice, the process of acquiring multiple images of different properties cannot be done simultaneously and using the same imaging system. Each image modality is acquired in a different scanner, very often in an interval of several days and it is mostly impossible to assure that patients are always scanned in the same position and using the same settings. One of the important tasks in medical imaging is the alignment of different image modalities of a single patient so that they can be directly compared. This task is often referred to as a single-subject *multi-modal registration*. The images are taken from the same person which means that rigid parts of body, e.g. head, can be brought into an alignment just using rigid-body transformations, i.e. rotations and translations.

Inter-subject registration

It is often desirable to compare images from a number of individuals acquired using the same modality. For example, a comparison of a PET brain image of one patient with a group of normal PET images allows to detect abnormal changes in function that would not be otherwise detectable in the image alone. It is also useful for studying various generic properties of human body over a number of individuals or for creation of atlases and template images.

There are usually considerable differences in the size and shape of brains of different individuals and that is why affine or non-linear deformations must be used to bring images into alignment, a task usually termed the inter-subject *non-linear registration* or *spatial normalization*. Non-linear registration in three-dimensions is very difficult to perform manually and reliable automated procedures are thus crucial.

In this thesis we present two contributions to the development of faster and more reliable automated registration procedures: a fast algorithm for alignment of multi-modal images and a robust algorithm for inter-subject matching with one-to-one deformations. Both algorithms have a large applicability in clinical practice. Before we proceed with this main topic we shortly review the many different approaches to image registration.

1.1 Review of medical image registration

The problem of aligning 3D images is nearly as old as the 3D imaging techniques themselves. A comprehensive survey of registration methods was presented in [Maintz and Viergever \(1998\)](#). According to this survey, each registration procedure can be decomposed into three major parts:

- *Problem statement* determines the type of images involved (multi-modal or just a single modality), the type of required alignment (intra-subject, inter-subject or atlas registration) and the nature of transformation to be recovered (rigid-body, affine, non-linear, etc.).
- *Registration paradigm* involves the image features and the correspondence criterion exploited to compute the registration.
- *Optimization procedure* determines the search strategy for the selected criterion and the degree of user interaction required.

Registration methods can be divided according to the registration paradigm into extrinsic or intrinsic.

Extrinsic methods make use of external markers fixed onto the scanned person specifically with the intention to facilitate the subsequent registration. For example, in stereotactic neurosurgery a frame with a set of markers of known geometry is rigidly attached to the patient's skull and the markers are filled with a contrast medium that is clearly visible and distinguishable in the resulting image. Registration of images with such markers is fast, usually accurate and can be easily automated without the need for sophisticated registration algorithms. The necessity of mounting of markers is, however, very uncomfortable for patients and also puts a considerable burden on the clinical crew. Moreover, external markers are by their nature mostly restricted to

rigid-body transformations and cannot be used for matching of images from different patients where non-linear deformations are necessary. For these and other reasons, most images are acquired without the use of external markers and intrinsic methods must be employed in order to align them.

Intrinsic methods make use solely of the image contents. They can be landmark based or voxel based.

1.1.1 Landmark based methods

Landmark based methods rely on identification and extraction of homologous objects in both images that are subsequently brought into alignment. These can be points, lines, surfaces, curvatures etc. The set of objects is usually sparse which allows a relatively fast optimization procedures but the issue of a reliable landmark identification poses large problems.

Landmarks can be identified manually by an experienced physician, automatically by segmentation procedure or in a semi-automatic way. Several years ago, manual point based methods were frequently used for retrospective intra-subject registration ([Pietrzyk et al., 1994](#)) but they are labor-intensive, time-consuming and their accuracy depends on physician's experience.

Landmarks have often been used for non-linear registration of intra-subject images based on deformable models. In these methods an extracted structure from one image, mostly surface or curve, is elastically deformed to fit a corresponding object in the second image (e.g. [Johnson and Christensen, 2002](#); [Bookstein, 1997](#); [Thompson and Toga, 1996](#); [Collins et al., 1994](#)). The registration accuracy is limited to the accuracy of the segmentation step. These methods are usually automated except for the segmentation step which is mostly performed semi-automatically. An extensive survey on these methods was presented in [McInerney and Terzopoulos \(1996\)](#).

Extraction of landmarks in a (semi-)automatic way is a highly data and application dependent task that requires good quality images with sufficiently distinct pattern (e.g. magnetic resonance images). Automatic identification of landmarks in functional images like PET is not reliable. Landmark-based methods are therefore not suitable for the purpose of this thesis.

1.1.2 Voxel based methods

Voxel based methods use the full image content and are among the most promising methods of current research. They are theoretically the most flexible methods since they use all available information throughout the registration process. Their enormous expansion in the last years has been made possible by the increasing performance of modern computers. Voxel based methods can be applied in almost any medical application and using various

types of transformation.

The main principle of this class of methods is an iterative optimization of some measure of similarity between the registered images. One of the images is static and a suitable transformation is sought for the other image that aligns it with the static image. For each transformation, the similarity measure gives a number computed from corresponding voxels in both images that evaluates the quality of alignment. An optimization algorithm is applied to find a transformation that maximizes the similarity.

If the registered images come from the same modality, the matching criterion can simply compare values of corresponding voxel in both images. The situation becomes complicated when the input images come from different modalities. Such images look very different and a precise relationship between the modalities is unknown. Mutual information has proved to be an excellent criterion for registration of intra-individual images from different modalities (Maes et al., 1997; Pluim et al., 2001a; Studholme et al., 1997; Thurfjell et al., 2000; West et al., 1997). A multi-resolution approach was proposed by several researchers for acceleration of image registration (Maes et al., 1999; Pluim et al., 2001a; Thurfjell et al., 2000). They showed that a suitable multi-resolution schema is able to improve registration speed by a factor of more than 2 without loss of accuracy.

Methods for affine and non-linear registration of inter-individual images deal almost exclusively with images from one modality, mostly MR images. A frequent approach is to specify a parametric transformation model and a similarity measure and to perform an iterative optimization of parameters of the model in order to maximize similarity. Various transformation models were proposed, for example low-order polynomials (Woods et al., 1998) or cosine basis functions (Ashburner and Friston, 1999). A common problem in these methods is ensuring of a one-to-one mapping between images. Linear regularization is usually incorporated to improve robustness and topological consistency (Ashburner, 2000) but it does not prevent production of a non-injective mapping.

Interested reader may consult Pluim et al. (2003) and Maintz and Viergever (1998) for a more detailed information and a comprehensive survey of literature about each particular registration type.

1.2 Overview of the text

The text is organized as follows. Chapter 2 sets the scene by introducing the basic concepts of tomographical imaging and handling of 3D medical images. Section 2.3 describes the used coordinate system and introduces matrix representation of affine transformations and conversion between coordinate systems of two images with anisotropic voxel sizes. The chapter is concluded

with a discussion of efficient ways of transforming 3D images.

Chapter 3 is concerned with voxel similarity measures. Sum of squared differences and correlation coefficient are introduced as convenient measures for estimation of alignment of images from the same modality. In the last years, mutual information proved its superiority among measures estimating alignment of images from different modalities. The concept of mutual information and joint histograms is described in detail in this chapter.

The next two chapters present the main contributions of this thesis. Chapter 4 deals with registration of images from a single patient where the matching transformation can be sufficiently described using rigid-body transformation. This includes both single-modality registration (e.g. motion correction of a sequence of image frames of one patient) and multi-modality registration (e.g. PET-MR alignment). A method is described that brings images into alignment by optimizing a suitable similarity measure as a function of transformation parameters. The question of a suitable histogram size for computation of mutual information is discussed in detail and an empirical solution is suggested. An efficient segmentation method is proposed for speeding-up of the registration process. The occurrence of failed registrations is reduced by including an automated detection of a suitable initial transformation estimate based on a modification of the principal axes algorithm. The developed method is evaluated on a number of simulated and real images and the results are discussed at the end of the chapter.

Chapter 5 is about non-linear registration of images from different patients. A novel algorithm is proposed which enables matching of images of differing shapes. The method is designed mainly for matching of low-resolution PET images of human brain. Performance of the method is evaluated on matching of generated geometrical objects, on two control groups of normal images and on a large database of diseased brain images. In addition, a novel approach is presented that enables reliable matching of images with large lesions.

Implementation details are given in chapter 6. A graphical user interface developed for an easy utilization in clinical applications is described.

Chapters 4 and 5 include their own discussion sections. The chapter 7 summarizes the work and discusses possible future directions.

Structure and transformations of 3D medical images

In this chapter we introduce the fundamentals of two types of medical image acquisition that are considered in the subsequent chapters: images from positron emission tomographs (PET) and from magnetic resonance scanners (MR). The aim is to provide a general background information necessary to understand various issues that are inherent in medical applications with PET and MR.

We will then describe the structure of 3D medical images, introduce two convenient coordinate systems and the matrix representation of affine transformations and discuss transformations of anisotropic 3D volumes in detail.

2.1 Medical images

2.1.1 Positron emission tomography

Positron emission tomography is a non invasive, diagnostic imaging technique for measuring the metabolic activity of cells in the human body. PET is a unique technique in that it produces images of function of the body. Traditional diagnostic techniques, such as CT scans or MRI, produce images of anatomy or structure. The premise with these techniques is that disease induces a change in structure or anatomy that can be seen in the images. However, biochemical processes are also altered with disease and may occur even before there is a change in gross anatomy. PET is an imaging technique that is used to detect and visualize some of these functional processes and their changes. Even in diseases such as Alzheimer's disease or Parkinson's disease, where there is no gross structural abnormality, PET is able to reveal a functional alteration. PET is often used in combination with MR images, allowing a precise anatomical localization.

The principle of the positron emission tomography is following. A very small amount of radionuclide-labelled compound (called radiopharmaceutical or radiotracer) is introduced into a patient usually by intravenous injection. The radionuclide decays in the patient's body and some products of the decay

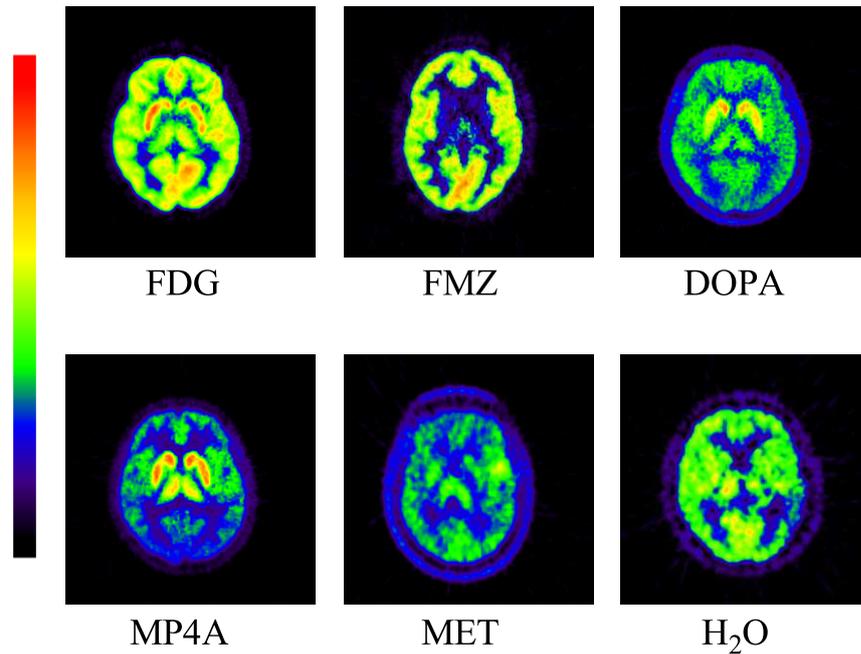


Figure 2.1: Single slices of PET images acquired using different tracers.

are detected in a scanner, allowing measurement of concentration of the tracer in tissue. In PET, radiopharmaceuticals are labelled with radionuclides that emit a positron which, after travelling a short distance (1-5 mm), encounters an electron from the surrounding environment. The two particles interact and annihilate which results in the emission of two gamma rays of 511 keV each in opposite directions. These two photons then travel through the tissue and air and are eventually detected by a field of detectors located in rings around the body. If the photons reach the detector rings they are detected by opposite detectors and approximately at the same time. Each such detected coincidence means that somewhere on the line connecting the two detectors an annihilation is likely to have occurred. The coincidences are recorded and afterwards reconstructed into a 3D image. Each image voxel carries information about the approximate number of emissions that occurred at that position which is closely related to the concentration of the injected tracer at that location.

The radiotracers used in PET are chosen to enable the study of particular metabolic and physiological processes in the living body. There is a number of tracers that can be used, each revealing a different type of functional (metabolic) information. For example, glucose labelled with ^{18}F accumulates in tumours, so that its high concentration indicates a possible presence of a tumour. Tracers commonly used in PET imaging include 2- ^{18}F -fluoro-

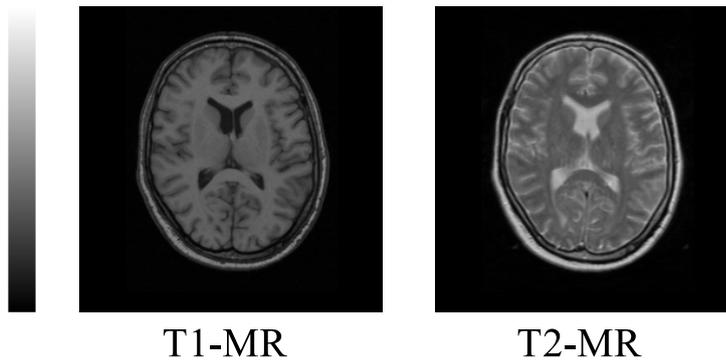


Figure 2.2: Two common types of MR images: T1-MR and T2-MR.

2-deoxy-D-glucose (^{18}F -FDG), ^{15}O -water, ^{11}C -methionin (^{11}C -MET), ^{11}C -flumazenil (^{11}C -FLUMA), 6- ^{18}F -fluoro-L-dopa (^{18}F -DOPA), N- ^{11}C -methyl-4-piperidyl acetate (^{11}C -MP4A). Figure 2.1 shows a slice of PET images of each type. For more details about the tracers, their applications and the PET acquisition process see for example [Herholz et al. \(2004\)](#); [Toga and Mazziotta \(1996\)](#); [Holmes \(1995\)](#); [Wienhard et al. \(1994\)](#).

2.1.2 Magnetic resonance imaging

MRI produces a map of hydrogen distribution in the body. Hydrogen is the simplest element known, the most abundant in biological tissue, and one that can be magnetized. This property is used in MRI scanners that consist to a large part of a very strong superconducting magnet that is able to align spins of hydrogen atoms in the patient's body.

Once hydrogen atoms have been aligned in the magnet, pulses of very specific radio wave frequencies are used to bring them out of alignment. The hydrogen atoms alternately absorb and emit radio wave energy, vibrating back and forth between their resting (magnetized) state and their agitated (radio pulse) state.

The MRI equipment records the duration, strength, and source location of the signals emitted by the atoms as they relax and translates the data into an image. In some cases, chemical agents such as gadolinium can be injected to improve the contrast between healthy and diseased tissue. Acquisition with different parameters (e.g. at different relaxation times) results in distinct types of MR images, the most common being the T1-MR and T2-MR, presented in Figure 2.2. For details about the MRI acquisition see [Toga and Mazziotta \(1996\)](#).

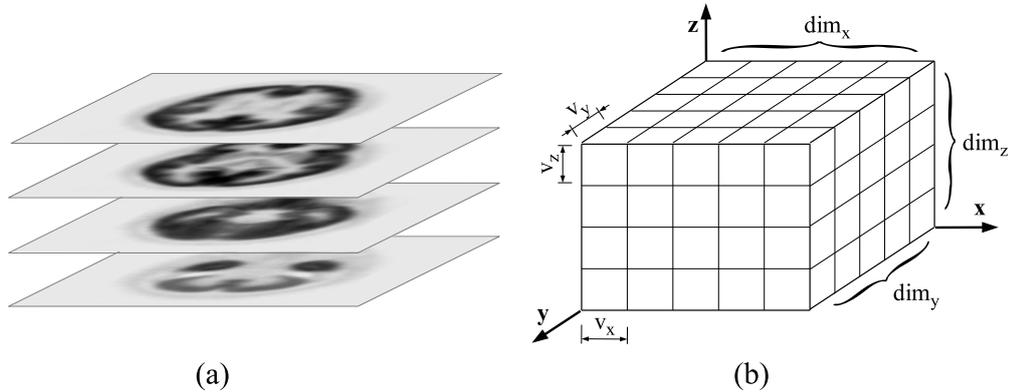


Figure 2.3: The output from a scanner is a set of 2D slices, each containing $dim_x \times dim_y$ pixels (a). For the purpose of image registration they are regarded as a 3D volume of $dim_x \times dim_y \times dim_z$ voxels with anisotropic sizes v_x, v_y, v_z mm (b).

2.2 Voxel and world coordinate systems

The output from a PET or MR scanner is a set of 2D slices¹ that together constitute a discrete image of measured tracer activity (Fig. 2.3a). For the purpose of methods presented in this thesis it is sufficient to regard the image data as a 3D volume of voxels with anisotropic sizes (Fig. 2.3b). The volume is thus characterized by the number of voxels in each dimension (dim_x, dim_y, dim_z) and the voxel size in millimeters (v_x, v_y, v_z).

Due to the anisotropic voxel sizes we need to work with two types of coordinate systems:

Voxel coordinate system 3D images are defined on a rectangular lattice and the voxel coordinate system provides a direct 3D representation of the image data stored in memory. Integer coordinates $(i, j, k) \in \mathbb{Z}^3$ correspond to the zero-based order of voxels in each axis and are related to the voxel centres. For example, the voxel with coordinates $(3, 7, 1)$ is the 4-th voxel in the x -axis, 8-th in the y -axis and 2-nd in the z -axis. The intensity value of a voxel is assigned to the corresponding integer coordinate. Non-integer coordinates do not have an intensity assigned to them and the intensity must be interpolated from the neighbouring integer coordinates. Interpolation is discussed in section 2.3.6.

The voxel coordinate system does not take voxel sizes into account. The

¹In the last years, PET images are often reconstructed in a fully 3D reconstruction process. However, the resolution in the “in-slice” dimensions is mostly better than in the remaining axis and we can still regard the image as a set of 2D slices.

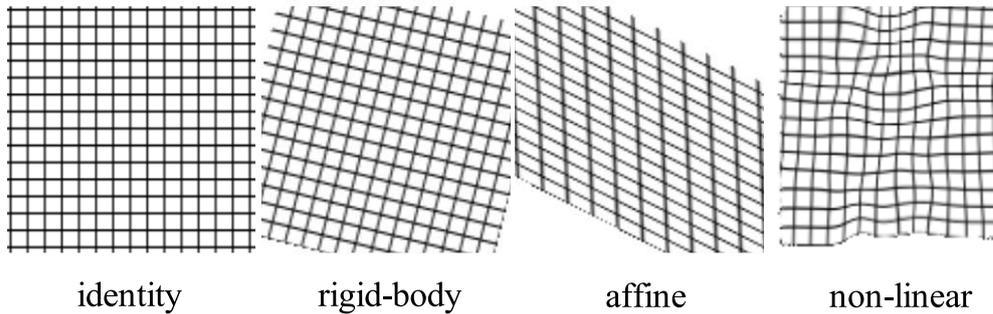


Figure 2.4: Geometric transformations of a regular rectangular grid.

legal range for a coordinate (i, j, k) is $i \in \langle 0, dim_x - 1 \rangle$, $j \in \langle 0, dim_y - 1 \rangle$, $k \in \langle 0, dim_z - 1 \rangle$. We will denote $\mathcal{S}(i, j, k)$ the intensity value of image \mathcal{S} at the coordinate (i, j, k) . When the 3D position of a voxel is not important, the notation $\mathcal{S}(i)$ will be used. $\mathcal{S}(i)$ is the intensity value of the i -th voxel (in an arbitrary fixed ordering), $i \in \langle 0, dim_x \cdot dim_y \cdot dim_z - 1 \rangle$.

World coordinate system: Coordinates of the world coordinate system $(x, y, z) \in \mathbb{R}^3$ are expressed in millimetres and take voxel sizes into account. The origin $(0, 0, 0)$ is in the exact centre of the image volume. The directions of axes are the same for both the voxel and the world coordinate system.

Every image has its own voxel coordinate system that depends on its dimensions and voxel size. The world coordinate system is common to all images and independent of dimensions and voxel sizes of individual images. Image transformations applied during a registration therefore take place in the world coordinate system.

In the whole text we will use the characters (i, j, k) for coordinates in the voxel coordinate system and the characters (x, y, z) for coordinates in the world coordinate system, unless otherwise stated.

2.3 Image transformations

The methods presented in this thesis work with three types of transformations: rigid-body, affine and non-linear (see Figure 2.4). Affine transformations and their subset, rigid-body transformations, are the topic of this section. Non-linear transformations are described in chapter 5 as a part of the non-linear registration algorithm.

2.3.1 Affine transformation

Affine transformations constitute one of the corner-stones of the registration methods presented in this thesis, imposing mathematical constraints on the types of geometric distortions that can be applied during the process of registration. They represent a subset of the linear transformations and they project, by definition, parallel lines onto parallel lines. An overview of affine transformations with respect to medical applications and with practical examples can be found in [Woods \(2000\)](#).

Affine transformations in a three dimensional space are defined by a set of 12 independent parameters and can be conveniently expressed using matrix notation by a 4×4 matrix M :

$$\mathbf{M} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The elegance and usefulness of the matrix formulation is that a sequence of transformations (i.e. transformation matrices) can be combined together via matrix multiplication to form a single matrix. Points in the three dimensional space must be expressed in homogenous coordinates in order to transform them using this matrix notation. Homogenous coordinates are a common technique that enables us to represent a full affine transformation of a point using matrix notation². For the purpose of this thesis it is sufficient to state that a point $\mathbf{p} = (x, y, z)$ in 3D Cartesian coordinates corresponds to a point $\mathbf{p}_h = (x, y, z, 1)$ in 3D homogenous coordinates and that a point $\mathbf{p}_h = (x, y, z, w)$, $w > 0$, can be transformed back to the common 3D coordinates as $\mathbf{p} = (x/w, y/w, z/w)$. For more information about homogenous coordinates see [Woods \(2000\)](#).

In the rest of this section we implicitly assume the use of homogenous coordinates in places where a matrix transformation of a *point* is concerned. Transformation of vectors, e.g. coordinate axes, is performed using just the upper left 3×3 matrix without involving homogenous coordinates.

Most of the parameters of the transformation matrix \mathbf{M} do not have a direct obvious meaning and therefore various other parametrizations are often used. A popular parametrization expresses the matrix \mathbf{M} as a sequence of the primitive operations translation, rotation, scaling and shear. In this thesis we will use a slightly different parametrization in order to facilitate reverse matrix decomposition (section 2.3.7). In this parametrization the transformation matrix is composed as:

²Translation can not be represented without the help of homogenous coordinates. Rotation, scaling and shear can be fully represented just by common 3D coordinates.

$$\mathbf{M} = \mathbf{T} \cdot \mathbf{R} \cdot \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{U}^T$$

\mathbf{T} represents translation by t_x , t_y , t_z millimetres in the x , y and z axis:

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

\mathbf{R} is a rotation matrix and can be further decomposed as:

$$\mathbf{R} = \mathbf{R}_x \cdot \mathbf{R}_y \cdot \mathbf{R}_z$$

The elementary matrices \mathbf{R}_x , \mathbf{R}_y , \mathbf{R}_z represent rotation by r_x , r_y and r_z radians around the x , y and z axis:

$$\mathbf{R}_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos r_x & \sin r_x & 0 \\ 0 & -\sin r_x & \cos r_x & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{R}_y = \begin{pmatrix} \cos r_y & 0 & \sin r_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin r_y & 0 & \cos r_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{R}_z = \begin{pmatrix} \cos r_z & \sin r_z & 0 & 0 \\ -\sin r_z & \cos r_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

\mathbf{U} is a rotation matrix as well, described by parameters u_x , u_y , u_z . It represents so called “stretch rotation” (Shoemake and Duff, 1992).

The matrix \mathbf{S} stands for scaling with factors s_x , s_y , s_z :

$$\mathbf{S} = \begin{pmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

A negative scaling factor represents mirroring across the respective axis. Mirroring is not of interest here and we will therefore limit the scaling parameters to positive values.

For each of the described elementary transformation there is a simple expression for inverse transformation parameters and for composition of an

inverse transformation matrix (see, e.g., Woods (2000)). The inverse \mathbf{M}^{-1} of the whole affine transformation matrix \mathbf{M} is computed using the elementary inverse transformation matrices as:

$$\mathbf{M}^{-1} = (\mathbf{U}^T)^{-1} \cdot \mathbf{S}^{-1} \cdot \mathbf{U}^{-1} \cdot \mathbf{R}^{-1} \cdot \mathbf{T}^{-1}$$

A point $\mathbf{p} = (x, y, z, 1)^T$ in a 3D coordinate space is transformed to a new point $\mathbf{p}' = (x', y', z', 1)^T$ as:

$$\mathbf{p}' = \mathbf{M} \cdot \mathbf{p}$$

Instead of working with transformation matrices, the registration methods described in the following chapters work internally with vectors of 12 independent transformation parameters $\boldsymbol{\alpha}$ where

$$\boldsymbol{\alpha} = (r_x, r_y, r_z, t_x, t_y, t_z, s_x, s_y, s_z, u_x, u_y, u_z)$$

This has several good reasons. In contrast to the coefficients of transformation matrices, these parameters have an obvious meaning and are suitable for interaction with users (e.g. displaying of registration progress, entering of initial transformation estimates, etc.). Tuning of optimization methods that work on vectors of parameters is easier. Another reason is that the composition of a transformation matrix from a vector of transformation parameters is simple and *unique* whereas decomposition of a matrix into independent parameters is ambiguous and requires utilization of numerical methods. The transformation matrix $\mathbf{M}_{\boldsymbol{\alpha}}$ corresponding to $\boldsymbol{\alpha}$ is composed and used solely during the transformation of image.

Note that a transformation matrix uniquely describes a certain transformation whereas a vector of transformation parameters must be accompanied by an information about the way of composing elementary matrices from parameters and about the *order* of these elementary transformations.

2.3.2 Rigid-body transformation

Rigid-body transformation is a special case of affine transformations that consists of rotation and translation only. It does not change Euclidean distances between any two coordinates, i.e. only position and orientation of objects can change whereas shape and size remain unchanged.

2.3.3 Conversion between coordinate systems

Voxel coordinates of an image are transformed to the world coordinates by first shifting the centre of the image to its origin and then scaling by the voxel size.

Let \mathbf{C} be the image centre expressed in the voxel coordinates, i.e.

$$\mathbf{C} = (c_x, c_y, c_z) = \left(\frac{\dim_x - 1}{2}, \frac{\dim_y - 1}{2}, \frac{\dim_z - 1}{2} \right)$$

The shift to the origin is accomplished by the centering matrix \mathbf{T}_c , the scaling by voxel size is provided by the matrix \mathbf{S}_w :

$$\mathbf{T}_c = \begin{pmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 1 & -c_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{S}_w = \begin{pmatrix} v_x & 0 & 0 & 0 \\ 0 & v_y & 0 & 0 \\ 0 & 0 & v_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The voxel-to-world and the world-to-voxel transformation matrices are then given by:

$$\mathbf{M}_{v \rightarrow w} = \mathbf{S}_w \mathbf{T}_c \quad \mathbf{M}_{w \rightarrow v} = \mathbf{T}_c^{-1} \mathbf{S}_w^{-1}$$

Note that transforming a voxel $\mathbf{p} = (i, j, k) \in \mathbb{Z}^3$ from the voxel coordinate space to a point $\mathbf{p}' = (x, y, z) \in \mathbb{R}^3$ in the world coordinate space assigns the precise intensity value of \mathbf{p} to \mathbf{p}' . A reverse transformation from the world coordinates in \mathbb{R}^3 to the voxel coordinates in \mathbb{Z}^3 can not get along without some kind of interpolation of intensity values in the image grid.

2.3.4 Resampling to different image dimensions

It is often necessary to resample an image to other dimensions, different voxel sizes or to express a coordinate in one image using coordinates of another image. A matrix that provides a conversion between the voxel coordinates v_1, v_2 of two images can be expressed using the matrices defined in the previous text:

$$\mathbf{M}_{v_1 \rightarrow v_2} = \mathbf{M}_{w \rightarrow v_2}^{-1} \cdot \mathbf{M}_{v_1 \rightarrow w}$$

2.3.5 Transformation of a 3D image

Consider a transformation of a 3D image \mathcal{S} (source) using an affine transformation matrix \mathbf{M} and subsequent resampling into a 3D image \mathcal{T} (target) with a different dimension and voxel size. This task is accomplished by a common backward mapping in two steps:

1. **Geometrical transformation:** The target volume \mathcal{T} is passed voxel by voxel, each voxel is converted to the world coordinate system, transformed using the inverse matrix \mathbf{M}^{-1} and then converted to the coordinate system of \mathcal{S} . The whole transformation can be combined into a single matrix $\mathbf{M}_{\mathcal{S},\mathcal{T}}$:

$$\mathbf{M}_{\mathcal{S},\mathcal{T}} = \mathbf{M}_{w \rightarrow v_{\mathcal{S}}} \cdot \mathbf{M}^{-1} \cdot \mathbf{M}_{v_{\mathcal{T}} \rightarrow w}$$

2. **Interpolation:** Intensity at the resulting location is interpolated from neighbouring voxels and assigned back to the original voxel in \mathcal{T} . If the resulting location falls outside of the volume of \mathcal{S} than we assign it some predefined value, e.g. zero or NaN.

Suppose that *interpolate* is an interpolation function that takes an image and a point in its voxel coordinate system as arguments and returns the interpolated intensity value at this point. Algorithm 1 outlines the principle of the image transformation.

Algorithm 1: Transformation of a 3D Image

Input: image \mathcal{S} , transformation matrix \mathbf{M}

Output: image \mathcal{T}

create space for the output volume \mathcal{T}

$\mathbf{M}_{\mathcal{S},\mathcal{T}} \leftarrow \mathbf{M}_{w \rightarrow v_{\mathcal{S}}} \cdot \mathbf{M}^{-1} \cdot \mathbf{M}_{v_{\mathcal{T}} \rightarrow w}$

foreach voxel $\mathbf{p} \in \mathcal{T}$ **do**

1 $\mathbf{p}' \leftarrow \mathbf{M}_{\mathcal{S},\mathcal{T}} \cdot \mathbf{p}$

2 $\mathcal{T}(\mathbf{p}) \leftarrow \text{interpolate}(\mathcal{S}, \mathbf{p}')$

end

This algorithm is not very efficient, the inefficient part being the first line of the *for* loop which forces matrix multiplication and requires 12 multiplications and 9 additions for each voxel, i.e. a total of $\dim_x \cdot \dim_y \cdot \dim_z \cdot (12 \text{ mul} + 9 \text{ add})$. We can achieve a better performance if we just transform the origin \mathbf{p}_0 of the voxel coordinate system of \mathcal{T} into a point \mathbf{p}'_0 in the coordinate system of \mathcal{S} and also the unit vectors $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z$ of the main axes of \mathcal{T} into vectors $\mathbf{e}'_x, \mathbf{e}'_y, \mathbf{e}'_z$. Transformed position $\mathbf{p}' \in \mathcal{S}$ of every point $\mathbf{p} = (x, y, z) \in \mathcal{T}$ can then be expressed as a linear combination of the transformed origin and main axes (see Figure 2.5):

$$\mathbf{p}' = \mathbf{p}'_0 + x \cdot \mathbf{e}'_x + y \cdot \mathbf{e}'_y + z \cdot \mathbf{e}'_z$$

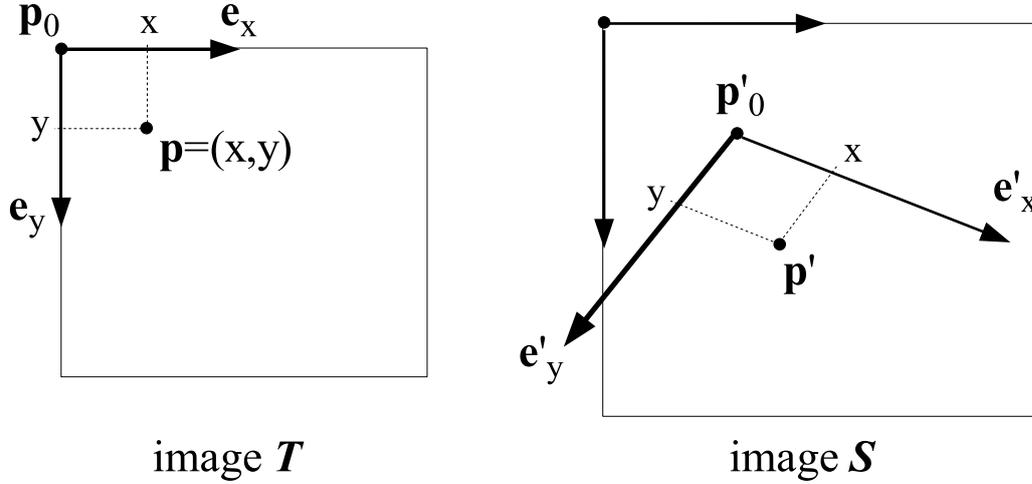


Figure 2.5: A 2D illustration of a transformation of a point $\mathbf{p} \in \mathcal{T}$ into a point $\mathbf{p}' \in \mathcal{S}$ as a linear combination of the transformed origin \mathbf{p}'_0 and main axes $\mathbf{e}'_x, \mathbf{e}'_y$.

If we transform the origin and vectors of main axes once at the beginning, the transformation part of the *for* loop then makes do just with additions. This approach corresponds to scanline algorithms often used in computer graphics (Wolberg, 1990) and is outlined in the Algorithm 2.

The initial transformation of the origin and main axes requires 39 multiplications and 27 additions. The *for* loops make do just with addition, the number of which is:

$$3 \cdot \dim_x \cdot (\dim_y \cdot (\dim_z + 1) + 1) \text{ add} = 3 \cdot (\dim_x \cdot \dim_y \cdot \dim_z + \dim_x \cdot \dim_y + \dim_x) \text{ add}$$

It is apparent from this expression that we can further reduce the number of operations by choosing the smallest image dimension for the outmost *for* loop and the largest dimensions for the innermost loop.

Table 2.1 compares the number of operations and computational time for two common image dimensions using both described algorithms. The number of operations in Algorithm 1 is approx. 7 times higher than that of Algorithm 2. Since multiplication usually takes slightly more time than addition, this factor is in reality larger (approx. 11), as indicated by the measured computational times³.

³The computational times were measured on a PC with an AMD Athlon 1 GHz processor and represent the mean value of a sequence of 100 successive transformations.

Algorithm 2: Scanline Transformation of a 3D Image

Input: image \mathcal{S} , transformation matrix M

Output: image \mathcal{T}

create space for the output volume \mathcal{T} ;

$M_{T,S} \leftarrow M_{w \rightarrow v_S} \cdot M^{-1} \cdot M_{v_T \rightarrow w};$

//transform origin

$\mathbf{p}'_0 \leftarrow M_{T,S} \cdot \mathbf{p}_0;$

//transform main axes

$\mathbf{e}'_x \leftarrow M_{T,S} \cdot \mathbf{e}_x;$

$\mathbf{e}'_y \leftarrow M_{T,S} \cdot \mathbf{e}_y;$

$\mathbf{e}'_z \leftarrow M_{T,S} \cdot \mathbf{e}_z;$

$\mathbf{p}' \leftarrow \mathbf{p}'_0;$

for $x \leftarrow 0$ **to** $dim_x - 1$ **do**

$\mathbf{p}'' \leftarrow \mathbf{p}';$

for $y \leftarrow 0$ **to** $dim_y - 1$ **do**

$\mathbf{p}''' \leftarrow \mathbf{p}'';$

for $z \leftarrow 0$ **to** $dim_z - 1$ **do**

$\mathcal{T}(x, y, z) \leftarrow interpolate(\mathcal{S}, \mathbf{p}''');$

$\mathbf{p}''' \leftarrow \mathbf{p}''' + \mathbf{e}'_z;$

$\mathbf{p}'' \leftarrow \mathbf{p}'' + \mathbf{e}'_y;$

$\mathbf{p}' \leftarrow \mathbf{p}' + \mathbf{e}'_x;$

	(d_x, d_y, d_z)	(128, 128, 47)		(256, 256, 200)	
	No. of operations	No. op.	Time	No. op.	Time
Algo 1	$d_x d_y d_z (12 \text{ mul} + 9 \text{ add})$	$16.2 \cdot 10^6$	0.81 s	$27.5 \cdot 10^7$	13.85 s
Algo 2	$39 \text{ mul} + 27 \text{ add} +$ $3(d_x d_y d_z + d_x d_y + d_x) \text{ add}$	$2.3 \cdot 10^6$	0.07 s	$3.9 \cdot 10^7$	1.12 s
$\frac{\text{Algo 1}}{\text{Algo 2}}$		6.95	11.57	6.97	12.37

Table 2.1: The number of multiplications (*mul*), additions (*add*) and the computational time needed for transformation of a 3D image with target dimensions (d_x, d_y, d_z) .

An important note to the transformation algorithms

A necessary presumption in using the scanline transformation algorithm is that the transformation can be fully described as a linear combination of a transformed origin and main axes of the coordinate system of the source image. The algorithm is thus limited to the class of linear transformations (affine+perspective). Non-linear transformations must be in general computed using Algorithm 1 where the first line of the *for* loop can be replaced by any kind of transformation. This fact was one of the important impulses in development of the non-linear registration algorithm HBM described in chapter 5. In a registration process, image volumes must be transformed many times (> 100) during a search for an optimum match. In order to allow a fast non-linear matching, the HBM algorithm was designed to use piecewise affine transformations with the fast scanline algorithm to approximate the non-linear nature of matching at the *optimization* stage and to use non-linear transformations just for formation of intermediate or resulting images.

2.3.6 Interpolation

Interpolation is a reverse (not inverse) process to acquisition of image data in a scanner. A tomographic scanner records a three-dimensional information about a *continuous* object (e.g. intensity of a radio-tracer in a human brain) by sampling it in *discrete* intervals which results in a 3D discrete grid of intensity samples. For further processing and analysis of the data it is usually necessary to know the intensity in between the discrete samples. This is done by interpolation from neighbouring samples.

A comprehensive overview of interpolation types and related issues can be found in [Thevenaz et al. \(2000\)](#); [Eddy and Young \(2000\)](#). A thorough evaluation and discussion of interpolation issues related to medical imaging

was published in [Pluim et al. \(2000\)](#). The computational demand raises with increasing quality of interpolation and the choice of the most suitable interpolation type depends on the specific application.

Nearest neighbour interpolation

Nearest-neighbour interpolation is the simplest type of all. An interpolated point (x, y, z) is assigned the value of the closest neighbouring voxel. The main benefit of this kind of interpolation is its simplicity that enables a very efficient implementation. It preserves the original voxel intensities. The price to pay is a considerable degradation of the resulting image (“blocky” appearance).

Tri-linear interpolation

A very popular approach in medical imaging is the tri-linear interpolation. The value of an interpolated point (x, y, z) is computed as the average value of its eight neighbouring voxels, weighted by their distance to the interpolated point. It is only slightly slower than nearest-neighbour, it produces a continuous interpolation and its implementation is simple. On the other hand, tri-linear interpolation results in a strong attenuation of high frequencies in the image. It is nevertheless a good compromise between quality and speed for many applications and we use it in all applications in this thesis.

Higher-order interpolation

A smoother interpolation can be achieved with higher-order polynomials, e.g. cubic B-splines, or with the sinc function. Computation of these interpolation functions involves much more neighbouring voxels and is significantly slower than the tri-linear interpolation. The result is smooth (at least the first partial derivatives are continuous) and high frequencies are not that much attenuated.

2.3.7 Decomposition of transformation matrix

As mentioned in section [2.3.1](#), composition of a transformation matrix from a vector of transformation parameters is simple and straightforward and it is of advantage to work with parameter vectors wherever possible. In some cases the reverse process might be required: decomposition of a transformation matrix into a parameter vector. It is needed, for example, for interaction with other software packages that pass a transformation matrix or use a different order of elementary transformations.

Every affine transformation of an object can be described by a matrix

with 12 variable coefficients m_{11} - m_{34} (section 2.3.1). The question is, how to decompose the matrix into a parameter vector α . Note that no decomposition procedure is unique because different parameter vectors may describe the same resulting transformation. We have chosen the polar matrix decomposition method of Shoemake and Duff (Shoemake and Duff, 1992; Heckbert, 1994, chap. 3). Polar decomposition factors a transformation matrix \mathbf{M} into primitive components \mathbf{T} , \mathbf{R} , \mathbf{U} and \mathbf{S} , defined in the section 2.3.1.

Similarity measures

3.1 Introduction

Consider two images, a target image \mathcal{R} and a source image \mathcal{S} , each containing an object of interest (e.g. brain images). Let t_{α} denote a transformation corresponding to some transformation parameters α . The aim of a registration is to transform the image \mathcal{S} so that the objects within the two images are aligned. In other words, we want to find transformation parameters α that match the object in \mathcal{R} to the object in $t_{\alpha}(\mathcal{S})$. This process is illustrated on an example of registration of a target MR image and a source PET images in Figure 3.1.

One possible approach is to maximize some measure of similarity between the images. We can formulate this mathematically as an optimization of some cost function $f(\mathcal{R}, t_{\alpha}(\mathcal{S}))$ of the two images. The cost function, or *similarity measure*, evaluates the quality of alignment of the object in image \mathcal{R} and the object in the transformed image $t_{\alpha}(\mathcal{S})$. The goal is to find a transformation α^* for which the similarity measure is maximal, i.e.

$$\alpha^* = \arg \max_{\alpha \in \mathbf{A}} f(\mathcal{R}, t_{\alpha}(\mathcal{S}))$$

where \mathbf{A} is the space of possible transformation parameters.

The crucial questions are:

1. How do we determine that two images are similar or well aligned?
2. How do we find an optimum transformation?

In this chapter we will deal with the first question. Finding a suitable transformation is the topic of the next chapters.

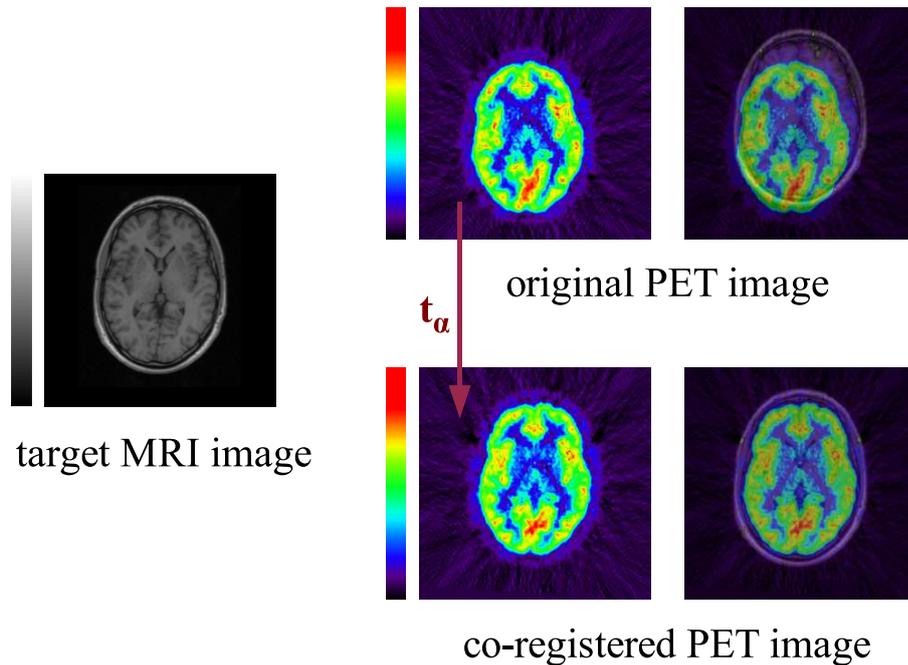


Figure 3.1: Schema of the registration process.

A similarity measure f should have the following properties:

- It should have a global maximum for the transformation that best aligns the images.
- The function should have as few local maxima as possible within a sufficiently large vicinity of the global maximum to avoid capture in a local maximum (i.e. a large capture range).

The right choice of a plausible function depends on the modalities of the registered images. There are, however, several general characteristics that most of them have in common. In contrast to the landmark-based methods reviewed in section 1.1, these voxel-based similarity measures do not rely on extracting specific features from images but work on the full image contents. More specifically, on a set of pairs of corresponding voxels $(\mathbf{v}, t_{\alpha}(\mathbf{v}))$ where $\mathbf{v} \in \mathcal{R}$ and $t_{\alpha}(\mathbf{v}) \in \mathcal{S}$.

When the images come from the same modality (e.g. both are PET images), they have a similar appearance, pattern and mostly a similar range of intensity values. Similarity of such images can be measured by a direct comparison of intensity values in corresponding voxels. Two of the most commonly used measures for this purpose are the *Sum of Squared Differences* and the *Correlation Coefficient*.

Similarity of images of different modalities (e.g. PET and MR) is in general difficult to assess since the precise relationships between the values and patterns are unknown. More sophisticated methods must be used to estimate the similarity. One of the most successful measures is the *Mutual Information*.

3.2 Image overlap

Before getting into a detailed discussion of similarity measures, attention must be paid to the question of spatial correspondence and image overlap. The target image \mathcal{R} and the source image \mathcal{S} generally do not have the same dimensions and voxel sizes and might encompass areas of different size in the world coordinates. Moreover, the overlap of the images varies according to the transformation t_α .

The similarity measures presented in the following sections work on pairs of corresponding voxels $(\mathbf{v}, t_\alpha(\mathbf{v}))$ where $\mathbf{v} \in \mathcal{R}$ and $t_\alpha(\mathbf{v}) \in \mathcal{S}$. Only voxels in the overlap of \mathcal{R} and \mathcal{S} are taken into account. To simplify definitions of similarity measures, we will define a sub-image \mathcal{R}^o as the part of the image \mathcal{R} that overlaps with the transformed image \mathcal{S} .

$$\mathcal{R}^o = \{\mathbf{v} | \mathbf{v} \in \mathcal{R}, t_\alpha(\mathbf{v}) \in \mathcal{S}\} \quad (3.1)$$

Note that voxels $\mathbf{v} \in \mathcal{R}$ are discrete samples in the voxel coordinates of \mathcal{R} and values $\mathcal{R}(\mathbf{v})$ are precise whereas voxels $\mathbf{v}' = t_\alpha(\mathbf{v})$ and values $\mathcal{S}(\mathbf{v}')$ must be interpolated in the image grid of \mathcal{S} . An abbreviation $\mathcal{S}_\alpha(\mathbf{v})$ will be used in place of $\mathcal{S}(t_\alpha(\mathbf{v}))$ in the following sections.

3.3 Sum of squared differences

Consider now two images \mathcal{R}, \mathcal{S} of the same modality. If they originate from the same patient and were acquired using the same scanner with the same settings, the vast majority of voxels in the aligned images differs only by noise. A very simple measure of similarity of two corresponding voxels \mathbf{v} and $t_\alpha(\mathbf{v})$ is the absolute difference of their values $|\mathcal{R}(\mathbf{v}) - \mathcal{S}_\alpha(\mathbf{v})|$. According to [Hill and Hawkes \(2000\)](#), the *sum of squared differences* (SSD) should theoretically be the optimum measure in the case where the two images differ only by Gaussian noise. The sum of squared differences between the overlapping parts of images is computed as:

$$SSD(\mathcal{R}, \mathcal{S}_\alpha) = \frac{1}{N} \sum_{\mathbf{v} \in \mathcal{R}^o} (\mathcal{R}(\mathbf{v}) - \mathcal{S}_\alpha(\mathbf{v}))^2 \quad (3.2)$$

where $N = |\mathcal{R}^o|$ is the number of voxels in the area of overlap in \mathcal{R} .

SSD of identical images is zero and SSD of non-identical or misaligned images is always greater than zero. The intrinsic assumption for using this measure is that the voxel values in both images are calibrated to the same intensity range. In such case the SSD function will have a global minimum at alignment. Note that it is important to take average in the equation 3.2, i.e. to divide by N . Otherwise the measure would depend on the number of voxels in the overlap and prefer solutions that decrease the area of overlap. Consequently, the global minimum of SSD would not necessarily be at alignment.

In practice, even the fact that two images originate from the same scanner does not guarantee that their intensity ranges are in concordance because the calibrations are often different. In order to accommodate to differing calibrations, we can add a global scaling parameter q into the expression of SSD which adjusts the intensity range of image \mathcal{S} to that of \mathcal{R} :

$$SSD(\mathcal{R}, \mathcal{S}_\alpha) = \frac{1}{N} \sum_{\mathbf{v} \in \mathcal{R}^o} (\mathcal{R}(\mathbf{v}) - q \cdot \mathcal{S}_\alpha(\mathbf{v}))^2 \quad (3.3)$$

The parameter q can be either fixed and estimated prior to registration or it can play a role of an additional optimized parameter. Choosing a fixed parameter is problematic since a bad estimate may spoil the whole registration. Adding of the calibration parameter to the set of optimized transformation parameters is a more flexible solution. However, an additional optimization parameter naturally increases computational demand of optimization. Moreover, another transformation dimension is added along which the optimization may fail. In the next section we will describe a more flexible measure that implicitly accounts for a linear relationship between image values.

3.4 Correlation coefficient

Consider again two images \mathcal{R} and \mathcal{S} originating from the same modality and suppose that their intensity ranges differ due to an unequal calibration. A simple measure of similarity that accounts for this linear relationship is the *cross correlation*. The measure is expressed as the sum of the product of all corresponding voxel pairs in the image overlap:

$$C(\mathcal{R}, \mathcal{S}_\alpha) = \sum_{\mathbf{v} \in \mathcal{R}^o} \mathcal{R}(\mathbf{v}) \cdot \mathcal{S}_\alpha(\mathbf{v}) \quad (3.4)$$

Cross correlation is frequently used in signal processing. It is also quite often applied in medical imaging although the justification of its applicability is limited and its usage problematic. One assumption for a successful application of this measure is that the number of voxels used for computation

does not change. In image registration this assumption is violated since the area of image overlap changes for different transformations of the image \mathcal{S} . As a result, the value of cross correlation generally tends to be higher for transformations that produce a larger overlap. This problem can be easily avoided by dividing cross correlation by the number of voxels in the overlap N , as in the computation of the *SSD* measure in equation 3.2. Still, there remains another unwanted property that the cross correlation bears: the dependency on the overall intensity of voxels in the overlap. The measure tends to prefer transformation that produce overlaps with higher intensity values even though other transformations might give a better alignment.

A measure that overcomes both problems is the *correlation coefficient*:

$$CC(\mathcal{R}, \mathcal{S}_\alpha) = \frac{\sum_{\mathbf{v} \in \mathcal{R}^o} (\mathcal{R}(\mathbf{v}) - \bar{R})(\mathcal{S}_\alpha(\mathbf{v}) - \bar{S})}{\sqrt{\sum_{\mathbf{v} \in \mathcal{R}^o} (\mathcal{R}(\mathbf{v}) - \bar{R})^2} \cdot \sqrt{\sum_{\mathbf{v} \in \mathcal{R}^o} (\mathcal{S}_\alpha(\mathbf{v}) - \bar{S})^2}} \quad (3.5)$$

where

$$\bar{R} = \frac{1}{N} \sum_{\mathbf{v} \in \mathcal{R}^o} \mathcal{R}(\mathbf{v}) \quad \bar{S} = \frac{1}{N} \sum_{\mathbf{v} \in \mathcal{R}^o} \mathcal{S}_\alpha(\mathbf{v})$$

are the mean intensity values of images \mathcal{R} and \mathcal{S}_α in their overlap.

In practice, we use a squared version of the correlation coefficient CC^2 in order to avoid the necessity to compute square roots. The range of values of CC^2 is $CC^2 \in \langle 0, 1 \rangle$ where $CC^2 = 0$ means no correlation and $CC^2 = 1$ means a perfect correlation.

The correlation coefficient is suitable for images with a wide range of intensity values like PET images. It is however not applicable for images with one or a few intensity values that repeat throughout the image.

3.5 Mutual information

In the previous sections we have presented two measures suitable for images of the same modality. In the following text we describe a measure applicable to images from different modalities: *mutual information*.

The roots of mutual information go back to year 1948 when Claude Shannon defined basic measures of information (e.g. the *entropy*) together with fundamental laws of data compression and transmission (Shannon, 1948). On this basis a whole new field - the theory of information - has evolved which plays an important role in many fields of computer science, physics, statistics, linguistics, biology, etc. In the works of Viola and Wells (Viola and Wells, 1995; Viola, 1995; Wells et al., 1996; Viola and Wells, 1997), Collignon et al. (Collignon et al., 1995a; Maes et al., 1997) and Studholme et al.

(Studholme et al., 1995, 1999) the joint entropy and the mutual information entered the field of image registration.

Entropy-based measures require estimation of the underlying intensity distribution in the images. There are basically two ways how to estimate the unknown intensity distribution: *image histograms* and *kernel density estimation*. The latter is a computationally demanding approach that is rarely used in medical image registration and was not pursued in this work.

In the next sections we will describe the concept of the marginal and joint image histograms and review the most basic definitions of the theory of information. On this basis we will then define the mutual information measure.

3.5.1 Image histogram

Histogram $H_{\mathcal{R}}$ of an image \mathcal{R} can be regarded as a way to estimate the unknown probability density of the intensity distribution in the image \mathcal{R} . It is obtained by dividing the intensity range of the image into equally spaced discrete bins.

Let r_{min} , resp. r_{max} be the minimum and maximum value in the image \mathcal{R} and $B_{\mathcal{R}}$ the number of bins in the histogram. A voxel \mathbf{v} with intensity $\mathcal{R}(\mathbf{v})$ falls into a bin k given as

$$k = \left\lfloor \frac{\mathcal{R}(\mathbf{v}) - r_{min}}{r_{max} - r_{min}} \cdot B_{\mathcal{R}} \right\rfloor \quad (3.6)$$

This mapping will be further marked $\mathcal{R}(\mathbf{v}) \mapsto k$. Note that this mapping is surjective and therefore not invertible. Each intensity maps to a certain bin but a certain bin corresponds to a range of intensities.

The bin width w is given as

$$w = \frac{r_{max} - r_{min}}{B_{\mathcal{R}}} \quad (3.7)$$

and a single histogram bin k corresponds to all intensities in the range

$$\left\langle k \cdot w + r_{min}, (k + 1) \cdot w + r_{min} \right\rangle \quad (3.8)$$

The value of the histogram $H(k)$ in a bin k is the number of voxels whose intensity falls into this bin:

$$H_{\mathcal{R}}(k) = \left| \{ \mathbf{v} | \mathcal{R}(\mathbf{v}) \mapsto k \} \right| \quad (3.9)$$

Figure 3.2 shows an example of a PET image and its histogram created using 128 bins.

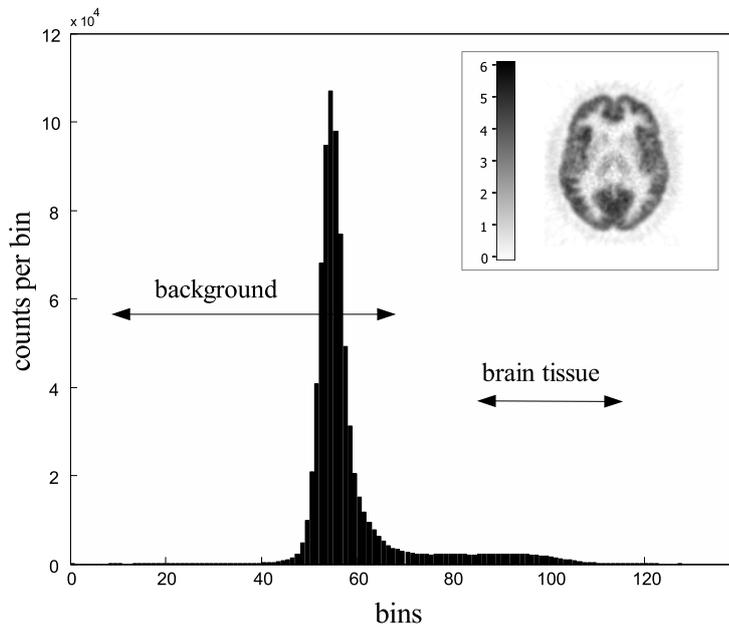


Figure 3.2: PET image and its image histogram (128 bins). The high peak in the lower bins corresponds to the background. Small bars in the higher bins represent brain tissue.

A histogram can be used to estimate the probability density function of the image. The probability of an intensity falling into bin k can be approximated as

$$P_{\mathcal{R}}(k) = \frac{H_{\mathcal{R}}(k)}{\sum_{i=0}^{B_{\mathcal{R}}-1} H_{\mathcal{R}}(i)} \quad (3.10)$$

where the denominator corresponds to the number of voxels in the image \mathcal{R} .

Removal of outliers

A characteristic property of PET images is that their intensity distribution contains a small number of outliers that originate from the reconstruction process where a 3D volume is computed from a set of 2D projections. These outliers are voxels with very high or very low intensity in comparison to the rest of the image. In the Figure 3.2 they correspond approximately to bins 0-40 and 120-128. They together represent about 1-3% of the whole volume but have considerable influence on the histogram in that they occupy many bins and “squeeze” the actually important data into a smaller number of bins. This can be easily avoided by setting the value of the lowest 1% of

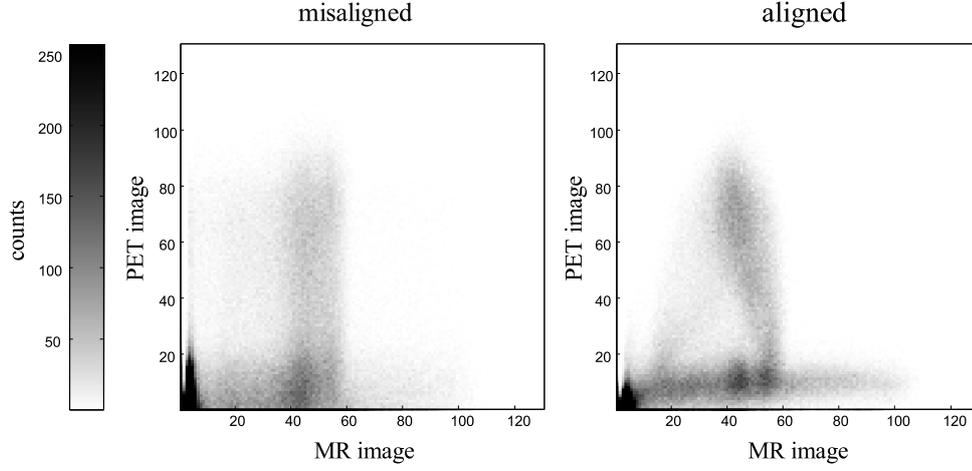


Figure 3.3: Joint histogram (128 x 128 bins) of a misaligned and aligned pair of PET and MR images.

all voxels to the minimum value of the remaining voxels and analogously by setting the value of the highest 1% of all voxels to the maximum of the rest. In the remaining sections and chapters we will implicitly assume that PET images have undergone this procedure. MR images usually do not have this problem.

3.5.2 Joint image histogram

Joint histogram $H_{\mathcal{R},\mathcal{S}}$ of images \mathcal{R} , \mathcal{S} is an estimate of the joint intensity distribution of the two images. $H_{\mathcal{R},\mathcal{S}}(p, q)$ is the number of voxels \mathbf{v} with intensity $\mathcal{R}(\mathbf{v})$ falling into bin p and intensity $\mathcal{S}(\mathbf{v})$ falling into bin q :

$$H_{\mathcal{R},\mathcal{S}}(p, q) = \left| \{ \mathbf{v} | \mathcal{R}(\mathbf{v}) \mapsto p \wedge \mathcal{S}(\mathbf{v}) \mapsto q \} \right| \quad (3.11)$$

The joint probability distribution $P_{\mathcal{R},\mathcal{S}}$ of intensity in images \mathcal{R} , \mathcal{S} can be assessed with the help of the joint histogram $H_{\mathcal{R},\mathcal{S}}$ as a probability of intensities in \mathcal{R} , resp. \mathcal{S} falling into bin k , resp. l :

$$P_{\mathcal{R},\mathcal{S}}(k, l) = \frac{H_{\mathcal{R},\mathcal{S}}(k, l)}{\sum_{i=0}^{B_{\mathcal{R}}-1} \sum_{j=0}^{B_{\mathcal{S}}-1} H_{\mathcal{R},\mathcal{S}}(i, j)} \quad (3.12)$$

where the denominator corresponds to the number of voxels in the overlap of \mathcal{R} and \mathcal{S} . If the number of bins used for the joint histogram is equal to the number of bins in the respective marginal histograms then the latter can be easily computed from the joint histogram by a summation.

An example of a joint histogram of a PET and MR images is shown in Figure 3.3. The histogram is more dispersed when the images are misaligned

than in case of an alignment.

3.5.3 Shannon entropy

Shannon Entropy of a set of N symbols $X = \{x_1, x_2, \dots, x_N\}$ is defined as:

$$E(X) = - \sum_{i=1}^N P_X(i) \log P_X(i) \quad (3.13)$$

where $P_X(i)$ is the probability of appearance of symbol x_i in the set. The Shannon entropy expresses the average information supplied by the set of symbols.

Entropy of an image \mathcal{R} can be estimated from its probability distribution $P_{\mathcal{R}}$:

$$E(\mathcal{R}) = - \sum_{k=0}^{B_{\mathcal{R}}-1} P_{\mathcal{R}}(k) \log P_{\mathcal{R}}(k) \quad (3.14)$$

Note that although the Shannon entropy is by far the most commonly used measure of information there are many other possible definitions. For example, generalizations of the Shannon entropy like the Rényi entropy or f -information metrics have been used by several authors for computation of generalized mutual information (Capek et al., 2001; Plum et al., 2001a; He et al., 2003; Bardera et al., 2004). Some of these measures seem to improve registration precision in special applications. Nevertheless, the generalized measures are more difficult to optimize and so far none of the tested variations persuasively outperformed the mutual information measure based on the Shannon entropy. For the purpose of our work, Shannon entropy seems to be the best choice.

3.5.4 Joint entropy

Joint entropy of two images $\mathcal{R}, \mathcal{S}_{\alpha}$ is given by:

$$E(\mathcal{R}, \mathcal{S}_{\alpha}) = - \sum_{k=0}^{B_{\mathcal{R}}-1} \sum_{l=0}^{B_{\mathcal{S}}-1} P_{\mathcal{R}, \mathcal{S}_{\alpha}}(k, l) \log P_{\mathcal{R}, \mathcal{S}_{\alpha}}(k, l) \quad (3.15)$$

The joint entropy measures the amount of information in the combined images. As noted in Studholme (1997), if two images with similar structures are misaligned, then in their combined image these structures will be duplicated: “For example, when a transaxial slice through the head is misaligned, there may be four eyes and four ears. As the images are brought

into alignment the duplication of features is reduced and the combined image is simplified.” Registration can thus be seen as a process of reducing the amount of information in the combined image which can be achieved by minimizing of joint entropy.

The performance of the joint entropy in image registration has been evaluated by several researchers (e.g. Collignon et al., 1995b; Studholme et al., 1995) and was found to be unreliable. The joint entropy has a problem with registration of images that only partially overlap: it tends to prefer larger overlaps even though at perfect registration the images might overlap only partially. This is due to the fact that the marginal entropies $H(\mathcal{R})$ and $H(\mathcal{S})$ are not constant during the registration process because of the changing area of overlap.

3.5.5 Mutual information

We can overcome these inconveniences by using the *mutual information* measure which takes the changing marginal entropies into account. It was proposed independently by Collignon et al. (Collignon et al., 1995a; Maes et al., 1997) and Viola and Wells (Viola and Wells, 1995; Viola, 1995). Mutual information is defined as:

$$MI(\mathcal{R}, \mathcal{S}_\alpha) = E(\mathcal{R}) + E(\mathcal{S}_\alpha) - E(\mathcal{R}, \mathcal{S}_\alpha) \quad (3.16)$$

By substituting from the equations 3.14, 3.15 and after some adjustments it can be written as

$$MI(\mathcal{R}, \mathcal{S}_\alpha) = \sum_{k=0}^{B_{\mathcal{R}}-1} \sum_{l=0}^{B_{\mathcal{S}}-1} P_{\mathcal{R}, \mathcal{S}_\alpha}(k, l) \log \frac{P_{\mathcal{R}, \mathcal{S}_\alpha}(k, l)}{P_{\mathcal{R}}(k)P_{\mathcal{S}_\alpha}(l)} \quad (3.17)$$

3.5.6 Normalized mutual information

Mutual information is less sensitive to overlap than the joint entropy but not completely immune. Studholme et al. showed that for larger misregistrations the mutual information measure may actually increase with increasing misregistration (Studholme, 1997). They proposed a normalized version of MI, the *normalized mutual information* (NMI), and found a distinct improvement in robustness with respect to varying overlap (Studholme, 1997; Studholme et al., 1999):

$$NMI(\mathcal{R}, \mathcal{S}_\alpha) = \frac{E(\mathcal{R}) + E(\mathcal{S}_\alpha)}{E(\mathcal{R}, \mathcal{S}_\alpha)}$$

Maes et al. proposed a similar measure, called the *entropy correlation coefficient* (Maes et al., 1997):

$$ECC(\mathcal{R}, \mathcal{S}_\alpha) = \frac{2 \cdot MI(\mathcal{R}, \mathcal{S}_\alpha)}{E(\mathcal{R}) + E(\mathcal{S}_\alpha)}$$

NMI and *ECC* are related in the following manner (Pluim et al., 2003):

$$ECC = 2 - \frac{2}{NMI}$$

3.6 Interpolation schemes for similarity measures

The SSD, resp. CC measure of two images \mathcal{R} and \mathcal{S}_α is evaluated by passing through every voxel \mathbf{v} of the target image \mathcal{R} , transforming it to a point $t_\alpha(\mathbf{v})$ in \mathcal{S} , interpolating its value from neighbouring voxels in \mathcal{S} and then adding the contribution of the pair $(\mathbf{v}, t_\alpha(\mathbf{v}))$ to the value of the respective measure. Trilinear interpolation (section 2.3.6) is used because it is continuous and very efficient.

The situation is different in computation of MI, resp. NMI. The trilinear interpolation could in principle be used exactly in the same way as with CC and SSD. In such case, however, intensity values are distributed into discrete histogram bins and a single voxel pair $(\mathbf{v}, t_\alpha(\mathbf{v}))$ contributes to exactly one bin. It implicates that even a small change of transformation parameters α may result in a discontinuous change in the histograms, regardless of the used interpolation type. There are many ways how to make the MI vary continuously with α . Some authors use Parzen windows for computation of the mutual information (Thevenaz and Unser, 2000; Viola and Wells, 1997). We use the efficient partial volume interpolation (PV) that was proposed in Collignon et al. (1995a) and Maes et al. (1997). For every pair $(\mathbf{v}, t_\alpha(\mathbf{v}))$, the eight nearest neighbours \mathbf{v}'_i in the other image are determined and the histogram entries of the eight pairs $(\mathbf{v}, \mathbf{v}'_i)$ are enlarged by a fraction inversely proportional to the distance between \mathbf{v} and \mathbf{v}'_i , as in trilinear interpolation.

3.7 Complexity of similarity evaluation

The most time consuming part of registration is the evaluation of similarity measure. The time spent on the optimization algorithm itself is negligible.

All three presented similarity measures are in principle evaluated in the same way, by passing through voxels of the target image volume, transforming each voxel \mathbf{v} onto a point \mathbf{v}' in the source image and adding contribution of the pair $(\mathbf{v}, \mathbf{v}')$ to the computed value of similarity measure. The transformation is carried out by the scan-line algorithm 2 (page 18). The time

required for evaluation of a similarity measure of two images is directly proportional to the number of samples taken from the target image.

Rigid-body registration

4.1 Introduction

The topic of this chapter is registration of images of the *same subject* acquired by the *same or different modalities* (intra- or inter-modality registration). Although the intra- and inter-modality cases introduce diverse challenges, the basic transformation and optimization approach remains same and they can be handled in the same framework. Apart from differences originating from acquisition by different modalities, images of the same subject usually differ only in position and orientation whereas the subject's shape remains uniform. This implies the use of rigid-body transformations (translation and rotation) in the matching procedure.

In chapter 3 we have defined several measures of similarity that evaluate quality of alignment of two images. Registration task consists in searching for a transformation α^* for which the selected similarity measure is maximal, i.e.

$$\alpha^* = \arg \max_{\alpha \in \mathbf{A}} f(\mathcal{R}, t_{\alpha}(\mathcal{S}))$$

where \mathbf{A} is the space of transformation parameters. In the following sections we will describe a convenient way of optimizing the similarity measures and propose a technique that reduces computational time and increases robustness of registration.

The chapter concentrates on a multi-modal registration of PET and MR images using mutual information as similarity measure. However, most of the text applies to a single-modal registration as well and in fact constitutes the basis for the non-linear registration method presented in the next chapter 5.

4.2 Optimization

The aim of an optimization is to find a set of transformation parameters α for which the similarity measure function $SM(\alpha)$ is maximized. In case of a 6-parameter rigid-body transformation there is a 6-dimensional space of possible parameters to seek through which clearly is not possible to manage

with an exhaustive search. A common approach is to make an initial estimate of transformation parameters (e.g. identity transformation or manually estimated parameters) and start an iterative search from there. At each iteration, the optimized function (similarity measure) is evaluated for the current parameter estimate and depending on its value a new estimate is constituted. The optimization stops on achieving some convergence criterion.

The optimization algorithm we have chosen here is the *downhill simplex minimization* by Nelder and Mead (Nelder and Mead, 1965). This method does not require computation of a gradient of similarity measure and is comparable in speed and accuracy to other commonly used algorithms like the Powell’s minimization or the conjugate gradient method, as demonstrated for registration of MR and CT images in Maes et al. (1999). Moreover, they show that the downhill simplex method is even slightly more efficient than the Powell’s method when a multi-scale approach (section 4.3) is used. The competitive performance of the simplex method and the relative simplicity of its implementation and tuning (no gradient estimation needed) makes it the method of choice here.

For an N -dimensional minimization, the method begins with $N + 1$ initial points defining a geometrical object called *simplex* in N -dimensional space of transformation parameters. The simplex is then iteratively transformed using a predefined sequence of transformations that aim at moving the vertices of the simplex towards the minimum. The implementation here is based on the code presented in Press et al. (1992). Offsets of the vertices of the initial simplex are +10 mm for translation and $+5^\circ$ for rotation. Convergence of optimization is declared when the fractional decrease of the optimized function is smaller than some threshold. The threshold is set to 10^{-5} (an empirically chosen value) in all presented tests unless otherwise stated.

Note that the simplex algorithm searches for minimum of a function. It means that the measures which require maximization, i.e. correlation coefficient and mutual information, must be used with negative sign.

4.2.1 A note about computational efficiency

The most time consuming part of registration is the evaluation of similarity measure. The time spent on the optimization algorithm itself is negligible. All three presented similarity measures are in principle evaluated in the same way: by passing through voxels of the target image volume, transforming each voxel \mathbf{v} onto some point \mathbf{v}' in the source image and adding the pair $(\mathbf{v}, \mathbf{v}')$, in a respective manner, to the computed value of the similarity measure. This implies that the time required for evaluation of similarity measure is directly proportional to the number of voxels in the target image. We therefore always use the image with smaller number of voxels as the target image.

In case of a PET-MR registration this is usually the PET image. Should a registration of a PET image to a target MR image be requested, then we perform a MR-to-PET registration first and then apply the inverse of the resulting transformation to the PET image.

4.3 Multi-scale approach

The computational demand can be significantly reduced without loss of accuracy with the help of a multi-scale coarse-to-fine optimization approach. First, optimization is performed at a low resolution using a small subset of image volume. After reaching convergence it proceeds at higher resolutions using the transformation computed at lower resolution as initial estimate.

We construct the subsampled image subsets as in [Maes et al. \(1999\)](#) by sampling the target image with integral factors $[f_x, f_y, f_z]$ along the x, y and z dimension using the nearest-neighbour interpolation. In this implementation, factor f_x means that every f_x -th voxel in the x-direction is used for computations (Fig. 4.1). No new intensities are introduced in the sampling process. Evaluation of a similarity measure with sub-sampling factors $[f_x, f_y, f_z]$ is theoretically $f_x \cdot f_y \cdot f_z$ times faster in comparison to the full resolution. If the optimization at lower resolutions converges close to the full-resolution optimum, it is likely that the number of iterations required to reach convergence at higher resolutions will be much smaller and the total computational time will be shortened.

4.3.1 Smoothing

Another option is to include smoothing prior to sampling of the images ([Pluim et al., 2001b](#)). In this way the subsampled subset should be less susceptible to noise in the original image. [Maes](#) speculated that smoothing may change the shape of iso-intensity objects on which the registration is based which could induce differences and inconsistencies in the optimal registration position at different resolution levels ([Maes, 1998](#)). On the other hand, one could argue that reducing the number of samples without smoothing might increase the influence of noise in the subsampled image and introduce new local minima into the similarity function.

In my experience, smoothing rather increases registration errors in mutual information-based multi-modal registrations whereas it reduces errors in inter-subject single-modal registrations. A possible explanation is that the process of computing mutual information already implicitly involves a certain kind of smoothing, namely the reduction of the intensity range into discrete bins. Nevertheless, smoothing may sometimes be of benefit in multi-modal registrations where images with poor signal are involved, e.g. H_2O -PET im-

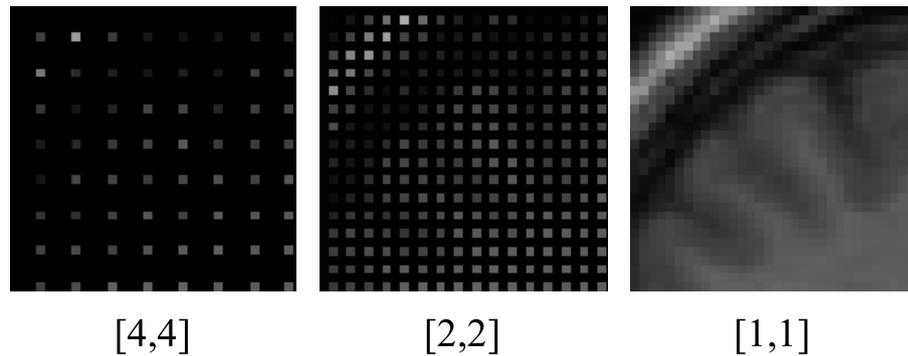


Figure 4.1: Illustration of the multi-scale approach on a magnified part of a MR image.

age frames or SPECT images, see [Thurfjell et al. \(2000\)](#). The single-modality similarity measures (SSD and CC) do not include any kind of smoothing and that is why an explicit reduction of noise is often helpful in these cases. Validation tests in this work use smoothing only when mutual information is not involved.

4.3.2 Scale schemas

The gain of the multi-scale approach evidently depends on the right choice of suitable number of levels and subsampling factors because subsampling may significantly influence the shape of the similarity function. Too large factors could introduce new local extremes and cause the optimization to end up too far from optimum. Such low-resolution optimization would then provide no good starting estimate for a higher-resolution optimization. In an extreme case this could make the multi-resolution approach even slower than the single full-resolution optimization due to additional optimizations at lower level that prove useless.

Several authors evaluated utilization of multi-scale optimization for mutual information based multi-modality registration ([Maes et al., 1999](#); [Thurfjell et al., 2000](#); [Pluim et al., 2001b](#)).

[Maes et al. \(1999\)](#) evaluated various parameters of a multi-scale optimization for registration of MR and CT images. The images used for evaluations had in-plane resolution of 256×256 voxels (approx. 1×1 mm) and 52-128 planes with 1.5-3mm separation. They used a fixed number of histogram bins at all levels (256 bins) and came to the conclusion that the best results can be achieved with subsampling factors up to 4, e.g. a three level scheme with factors $[4, 4, 1]$ in the first level followed by a level with factors $[2, 2, 1]$ and a full-resolution level $[1, 1, 1]$. Sampling factors larger than 8 did not bring any advantages according to their tests.

Thurfjell et al. (2000) evaluated multi-scale registration for simulated and clinical SPET (single-photon emission tomography) and MR images. They used different histogram sizes depending on the sub-sampling factors. They came to the conclusion that a SPET image (dimensions 64 x 64 x 62 voxels, 3 x 3 x 3.44 mm) can be matched to a corresponding MR image (dimensions 128 x 128 x 124 voxels, 1.5 x 1.5 x 1.75 mm) with sub-voxel accuracy using schemas with factors up to 8. Moreover, they showed that a similar precision can be acquired for SPET images even without the full-resolution level which brings an additional acceleration. Smoothing of images with gaussian filter of FWHM = 10 mm prior to registration increased robustness and accuracy in their tests.

Pluim et al. (2001b) inspected the effects of multi-scale approaches for matching of high-resolution MR and CT images and low-resolution PET, MR or CT images. They infer that multi-scale schemas are suitable only for high-resolution images whereas low-resolution images should be better co-registered at the full resolution only. They also did not find any advantage of using gaussian smoothing prior to registration. Their results clearly contrast with the results of Thurfjell et al..

The discrepancies in the published results can be partially ascribed to the differences in experimental setups but they very likely stem from the different choices of histogram sizes as well. We will see later in sections 4.7 and 4.9 that the number of histogram bins plays an important role. Unfortunately, only a few from the large number of articles dealing with mutual information-based registrations actually mention the number of bins that has been used which precludes comparisons and general inferences.

The aim of this work was not an evaluation of multi-scale approaches. For a PET-MR co-registration we use an empirically efficient three-level optimization schema with factors [4, 4, 1], [2, 2, 1] and (optionally) [1, 1, 1] which corresponds to the schema suggested in Thurfjell et al. (2000) and Maes et al. (1999). The z -axis is always sampled at the full resolution at all scale levels because the z -dimensions of PET images is usually much smaller than the x - and y -dimensions, the resolution in the z -axis is worse and the z -direction is empirically more difficult to align because it is less anatomically constrained.

4.4 Summary of the algorithm

The described registration algorithm largely reflects the most common framework of voxel-based algorithms for a rigid-body registration of medical images. We will further refer to this algorithm as the MMM algorithm (multi-modal matching). It is summarized in Algorithm 3 where $SM_{\mathcal{R},\mathcal{S}}(\alpha, l)$ is a similarity measure (i.e. NMI, SSD or CC) computed from images \mathcal{R} and \mathcal{S}_α at a scale level l .

Algorithm 3: Voxel-based rigid-body registration

Input: images \mathcal{R} , \mathcal{S} , initial transformation estimate α_I
Output: transformation parameters α_R that align \mathcal{R} and \mathcal{S}

```

// $\alpha_R$  are the transformation parameters to compute
 $\alpha_R \leftarrow \alpha_I$ 
//scale levels
 $L = \{[4, 4, 1], [2, 2, 1], [1, 1, 1]\}$ 

//loop over scale levels
forall  $l \in L$  do
    //iterative optimization
    while not converged do
         $sm \leftarrow$  value of  $SM_{\mathcal{R},\mathcal{S}}(\alpha_R)$ 
        create a new estimate of  $\alpha_R$  according to  $sm$ 
    end
end

```

4.5 Segmentation of background

In this section we will extend the registration framework by an efficient segmentation method that significantly improves computation time of registration. The basic idea is similar to the multi-scale approach described in the previous section, namely to increase the registration speed by using less voxels for computation. In contrast to the multi-scale approach where the acceleration is achieved by using voxels across the *whole volume* with a sparse sampling, the idea here is to limit the set of used voxels to the *area of interest*, i.e. the brain or head volume. In common PET and MR images of human brain this volume of interest occupies around 20-50% of the whole volume and the rest is background. It can be assumed that omitting background voxels from computation brings a significant acceleration.

In the following text we propose a method that segments a 3D volume into voxels of interest and background voxels. The aim was to develop a method that is fast, reliable and robust enough to be applied to various image modalities. Another important requirement was that the method is fully automatic, without the necessity of any tuning by user. For the purpose of registration the segmentation does not need to be very precise which allows the use of a simple and efficient technique. The method combines an intensity-based thresholding with an automatically detected threshold followed by a set of binary morphological operations. It is applicable to both

PET and MR images.

In ideal case the intensity in the background area would be zero, reflecting absence of the measured quantity, e.g. radioactivity from tracer molecules in PET images. In reality the background areas contain a substantial level of noise, in PET images combined with reconstruction artefacts. It can be assumed that omitting of background from computation would decrease the probability of a misregistration. This assumption has been validated in our previous article (Cizek et al., 2004) where we demonstrated that apart from a significant acceleration, a removal of background in noisy images can increase robustness of a registration. However, it was empirically observed that when the initial misalignment is very large then a removal of background may actually be a cause of a failed registration. This happens when the initial overlap of the two images is too small to allow a robust similarity estimation. We therefore suggest and evaluate a procedure for a coarse pre-alignment that increases robustness and further reduces registration time. It is an analytical (and thus very efficient) technique based on the principal axes transformation and it is described in section 4.6.

4.5.1 Thresholding

One of the easiest ways to segment an image into two parts, background and object, is to select a certain intensity value and mark all voxels with intensities lower than this value as background. With a well chosen threshold it is possible to correctly segment most (not all) background and object (i.e. head) voxels in PET and MR images. In our previous work (Cizek et al., 2004), the ratio q of the background (non-head) volume to the whole image volume was used for selection of the suitable intensity threshold. The intensity threshold corresponding to a certain ratio q was defined as the q -% quantile in the intensity histogram of the whole volume (i.e. the lowest intensity higher than $q\%$ of all image voxels). Thresholding resulted in a 3D binary mask where zero voxels corresponded to the voxels in the original image with an intensity lower than the intensity threshold.

An example of such a thresholded image is given in Figure 4.2. When a low threshold is used, many background voxels are not removed (Fig. 4.2 left). On the other hand, a higher threshold unavoidably removes some portion of brain voxels (Fig. 4.2 right). An implicit assumption of our previous approach was that the ratio q does not vary much for a given scanner and acquisition protocol. It was thus necessary to estimate the threshold manually for every scanner. To overcome this inconvenience, a simple automatic procedure has been developed that combines two common ways of threshold detection - an iterative algorithm that assumes no knowledge about the intensity distributions of background and object voxels, and an expression for

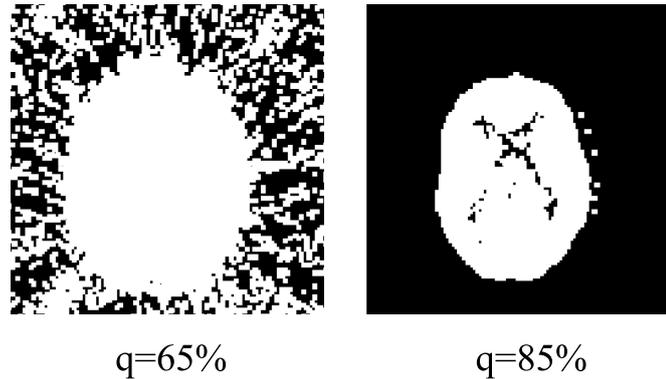


Figure 4.2: A binary mask of a PET image created by thresholding at 65%- and 85%-quantile.

a minimum-error threshold for distributions with known parameters.

4.5.2 Iterative optimal threshold selection

Let Ω denote the set of all voxels in the image. The goal is to separate the set Ω into two disjoint classes ω_b (background) and ω_o (object) such that $\Omega = \omega_b \cup \omega_o$. A simple iterative algorithm, summarized in Algorithm 4, can be used for an automatic threshold selection in an image with a bimodal histogram with unknown parameters (Sonka et al., 1994).

Algorithm 4: Iterative estimation of a threshold

- 1 Set the initial estimate of the threshold t^1 to the median value of all voxel intensities.
 - 2 Step k :
Assign all voxels with intensity lower than t^k to the background and the remainder to the object.
Compute mean intensities μ_b^k and μ_o^k of background and object voxels.
 - 3 $t_{k+1} \leftarrow \frac{\mu_b^k + \mu_o^k}{2}$
 t_{k+1} is a new threshold estimate.
 - 4 Stop if $|t^{k+1} - t^k| < \epsilon$. Otherwise continue with step 2.
-

It works well for images where the standard deviations of background and object voxel intensities are similar. In case of medical images this assumption is not valid. This is illustrated in Figure 4.3 showing a typical histogram of a PET image. The high narrow peak corresponds to the background and has a much smaller standard deviation than the bins of the object. As a

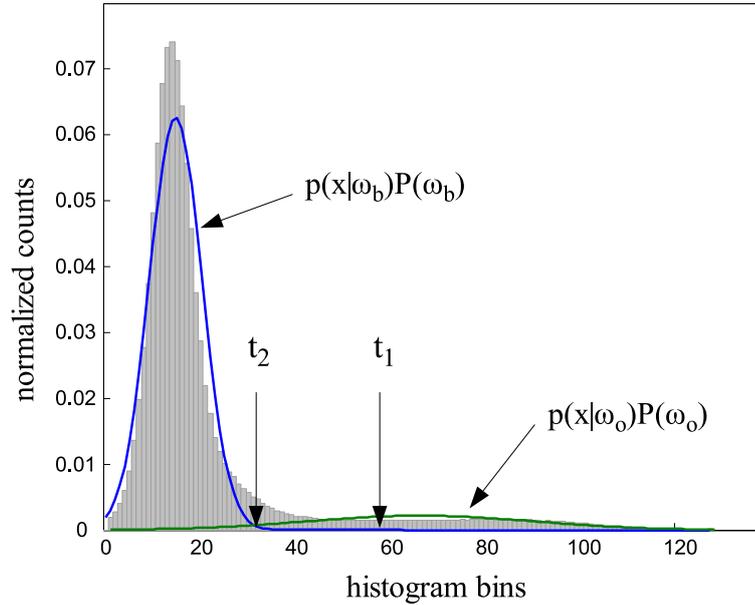


Figure 4.3: Histogram of a PET image. The thresholds t_1 and t_2 are results of the iterative algorithm 4 without and with the use of the Gaussian mixture density model.

result, the simple procedure would choose a wrong threshold (threshold t_1 in Figure 4.3) causing many object voxels to be marked as background which in turn would make subsequent morphological postprocessing unreliable.

We can avoid this problem if we replace the step 3 with a simple fitting of two probability density functions, one modelling the distribution of the background voxels and one for the object voxels, similar to the approach described in [Gonzales and Woods \(1993\)](#). Both classes of voxels (background ω_b , object ω_o) are thus characterized by class-specific probability density functions and by prior probabilities that together constitute a model of the whole image histogram, in general:

$P(\omega_i)$	prior probability for class ω_i ($\sum_i P(\omega_i) = 1$)
$p(x \omega_i)$	probability density function of voxel intensity x given class ω_i ($\int p(x \omega_i)dx = 1$)
$p(x)$	image histogram computed as a sum of conditional probability density functions weighted by prior probabilities: $p(x) = \sum_i P(\omega_i) \cdot p(x \omega_i)$

The prior probabilities represent ratios of the number of voxels in each class. If N_b , resp. N_o represent the number of voxels initially marked as

background, resp. object then $P(\omega_b) = N_b/N$ and $P(\omega_o) = N_o/N$ where $N = N_b + N_o$ is the total number of image voxels.

Normal (Gaussian) distribution was chosen to model the distribution of background and object intensity $p(x|\omega_b)$, $p(x|\omega_o)$. This approximation is roughly valid for the purpose of segmentation of PET and MR images, a precise fit is not required here. The model of the image histogram $p(x)$ is then:

$$p(x) = P_b \cdot \frac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{(x-\mu_b)^2}{2\sigma_b^2}} + P_o \cdot \frac{1}{\sqrt{2\pi}\sigma_o} e^{-\frac{(x-\mu_o)^2}{2\sigma_o^2}} \quad (4.1)$$

Using the histogram model $p(x)$ and the prior probabilities is possible to separate the image voxels x according to their intensity using the posterior probability of their belonging to a certain class. The *a posteriori* probability is given by the Bayes Rule:

$$P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)} \quad (4.2)$$

A voxel x will be marked as background if $P(\omega_b|x) > P(\omega_o|x)$, otherwise it will be marked as object. Consequently, the new threshold in the step 3 of the algorithm 4 is the intensity t for which $P(\omega_b|t) = P(\omega_o|t)$ (see threshold t_2 in Figure 4.3). This equation yields a minimum error estimate, as described in [Gonzales and Woods \(1993\)](#). Substituting both sides with the equation 4.2 yields an exponential equation:

$$p(x|\omega_b)P(\omega_b) = p(x|\omega_o)P(\omega_o) \quad (4.3)$$

which gives the quadratic equation

$$at^2 + bt + c = 0$$

where

$$\begin{aligned} a &= \sigma_b^2 - \sigma_o^2 \\ b &= 2(\mu_b\sigma_o^2 - \mu_o\sigma_b^2) \\ c &= \sigma_b^2\mu_o^2 - \sigma_o^2\mu_b^2 + 2\sigma_b^2\sigma_o^2 \ln\left(\frac{\sigma_o P_b}{\sigma_b P_o}\right) \end{aligned}$$

In case of two solutions, the new threshold is the value t for which $\mu_b < t < \mu_o$.

Empirically, about 10-20 iterations are sufficient for a good threshold estimate. The procedure is efficient, taking a fraction of a second for a common PET image (dimensions 128 x 128 x 47 voxels) and about 1-2s for

a common MR image (dimensions 256 x 256 x 150 voxels). The algorithm works well for PET modalities that exhibit separate histogram modes for background and object, e.g. *FMZ*-PET, *FDG*-PET, *MET*-PET, *MP4A*-PET. It cannot handle types of images where the background and object modes are nearly merged, e.g. *H₂O*-PET or *DOPA*-PET frames with a short acquisition time. These images can be thresholded with the quantile-based procedure described in [Cizek et al. \(2004\)](#).

4.5.3 Morphological cleaning

Intensity-based thresholding unavoidably marks some object voxels as background and vice versa (see Figure 4.4b). Binary 3D morphological operations can be applied to “clean” background areas and to fill “holes” in the object in the thresholded image. In general, morphological operations can simplify image data and eliminate irrelevant objects while preserving shape of large enough objects. The operators make use of *structuring elements* that determine the effect of the operator and its extent. A detailed description of the binary mathematical morphology and structuring elements can be found in [Seul et al. \(2001\)](#); [Parker \(1997\)](#); [Sonka et al. \(1994\)](#).

In our application, we first use an *erosion* operator with a spherical structuring element to remove all background voxels that were incorrectly marked as object voxels. The size of the structuring element should be large enough to assure removing of all background voxels. Radius of 3 voxels for PET images and 1 voxel for MR images is sufficient. Apart from cleaning of background, erosion also removes a certain part of the object area itself (Figure 4.4c). We recover the object by a sequence of *dilations*. Empirically, a sequence of two dilations with radius of 3 voxels for PET and 1 voxel for MR images joins up the object area (Figure 4.4d). In this way, background voxels can be efficiently removed while the head volume is preserved.

An advantage of this approach is that the automatic threshold selection does not need to be very precise. As demonstrated in Figure 4.5, the differences between binary masks created using thresholds at 65%, 75% and 85% quantile are quite significant. When erosion and dilation are used after thresholding, the resulting masks differ only slightly in the brain outline. This also justifies the simple threshold selection procedure we described in the previous section.

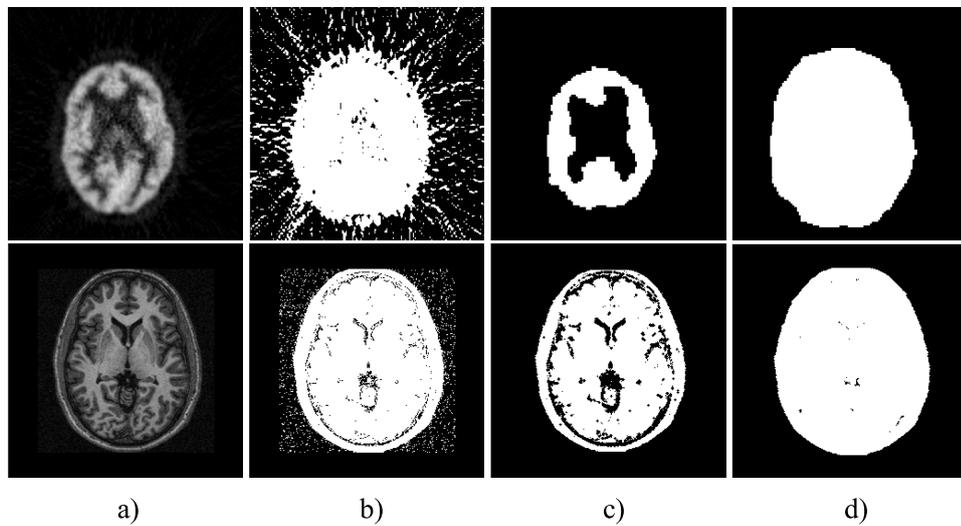


Figure 4.4: Computation of a mask for a PET (upper row) and MR (bottom row) image. Original image (a) is first thresholded into a binary image (b). Small spots in the background areas are removed by erosion (c). Holes and gulfs in the eroded image are then filled by dilation (d).

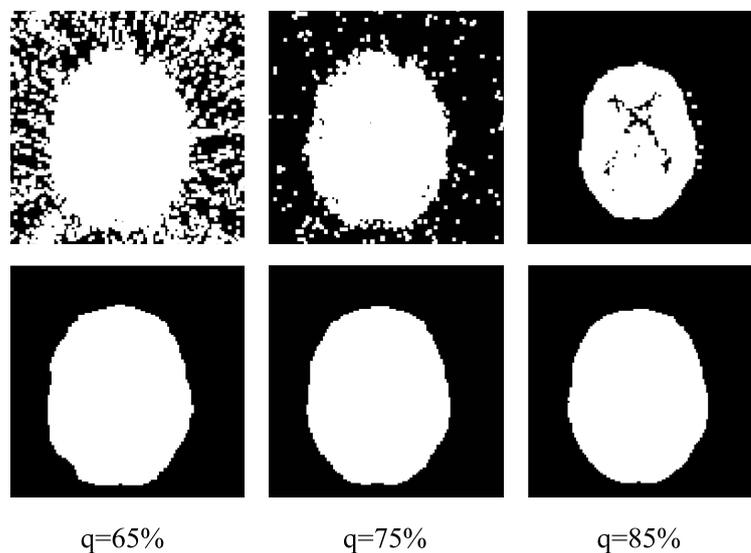


Figure 4.5: Binary masks created from a PET image using thresholding only (upper row) and thresholding followed by erosion and dilation (bottom row).

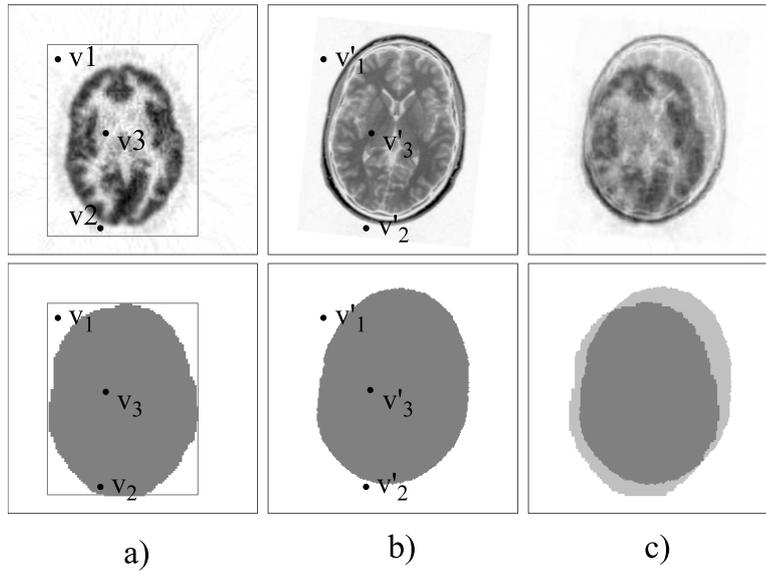


Figure 4.6: Reduction of image volumes in computation of mutual information using binary masks.

4.5.4 Performance of the morphological segmentation

The additional time needed to erode and dilate a 3D PET or MR image volume is about 2s for PET images and 3-10s for MR images¹ and is small in comparison with the gain in speed due to the masking, which we will demonstrate in section 4.10. The segmentation method has been tested on a set of 46 PET images acquired with various radiotracers and on a set of 30 T1-MR images. The results were checked visually, all segmentations gave sufficiently good results with respect to the intended application. In 2 PET images and 3 MR images there remained a few small separate clusters of voxels in the background area. Their influence on robustness or precision is certainly negligible compared to the large number of voxels in the object area. They only can potentially hamper computational speed by prohibiting a tight alignment of bounding box (see section 4.5.5 below).

4.5.5 Similarity measures with background masking

Recall that a similarity measure of two images \mathcal{R} , \mathcal{S} is evaluated by passing through voxels of the target image \mathcal{R} , transforming each voxel \mathbf{v} onto some point \mathbf{v}' in the source image \mathcal{S} and adding the pair $(\mathbf{v}, \mathbf{v}')$, in a respective manner, to the computed value of similarity measure.

¹All computational times mentioned in this thesis were measured on a PC with a 1.1 GHz AMD Athlon processor, unless otherwise stated.

Suppose now that binary masks were created for both images, dividing each image into background and object voxels. Evaluation of similarity measure runs in the following way:

1. The smallest rectangular box (bounding box) containing all object voxels is determined in the *target* volume (Figure 4.6 top left) using the target image mask (Figure 4.6 bottom left). Samples are taken only from this box. Note that the morphological “cleaning” of background in the segmentation procedure is an important step because it allows the bounding box to be tightly matched to the object.
2. The volume within the bounding box in the target image is sampled, background voxels are skipped (e.g. the voxel \mathbf{v}_1 in Figure 4.6). After this, all background voxels in the target image have been removed from the computation without the need for a transformation.
3. Sample voxels \mathbf{v} that pass the two preceding steps are transformed to corresponding samples \mathbf{v}' in the source volume. If \mathbf{v}' belongs to background in the source volume, the pair \mathbf{v}, \mathbf{v}' is removed from the computation (e.g. the voxel pair $\mathbf{v}_2, \mathbf{v}'_2$ in Figure 4.6).

Voxels that pass all these steps (e.g. the voxel pair $\mathbf{v}_3, \mathbf{v}'_3$ in Figure 4.6) are then used for computation of similarity measure. This approach largely reduces the number of necessary transformations - only voxels that pass steps 1 and 2 need to be transformed.

4.6 Global misalignment estimation

It is not possible for any optimization algorithm to ensure finding of a *global* optimum in a multi-dimensional search space of transformation parameters. The downhill simplex method used in this thesis as well as other commonly used methods (Powell’s direction set, gradient methods, etc.) are in their nature *local*. They start searching from a certain initial estimate and find a local optimum within their capture range. If there is no *a priori* knowledge about the mutual position of the registered images the initial estimate is usually the identity transformation. If the estimate is far from global optimum the registration is more likely to fail. It is therefore important to try and make a good initial transformation estimate that is close to the true global optimum.

One solution is to estimate the transformation manually and pass it on the registration algorithm. This option was implemented in the registration tool described in chapter 6 where users can make a fast manual pre-alignment using a convenient graphical interface. By experience, such pre-registration

nearly always helped in cases where the automatic registration procedure failed.

Another option is to use some kind of “global” stochastic optimization algorithm like simulated annealing (Capek et al., 2001), genetic algorithms, etc. These algorithms are, however, much more computationally demanding and existing results in the field of medical image registration have not persuasively demonstrated an increase of robustness that would justify the considerably high computational time.

Jenkinson et al. (2002) described an approach where transformation parameters, sampled in fixed intervals across the whole transformation space, were sequentially tested using a voxel similarity criterion. The potential problem of this approach is that the similarity criterion is used to evaluate alignment for very large misalignments. Even the normalized versions of mutual information cannot remove all local optima in the whole range of expected values. It means that there could appear a configuration of a total misregistration that will nevertheless give better value than a relatively small misregistration. It is therefore not possible to recover any extent of misregistration just using voxel-based similarity measures like NMI. I made this observation several times while experimenting with various global optimization approaches.

The approach proposed here attempts at making a close initial estimate without using the similarity criterion in the first place. The aim was to design an automatic procedure that would preferably make a global search over the whole space of possible transformation parameters but without spending too much time on it. The basic idea is to use the binary masks extracted in the process of background removal. The masks reflect the global shape of the registered images and can thus be used for making a rough initial alignment in a robust geometrical way. The proposed technique is a modification of *the principal axes transformation* which uses low-order moments to compute the registration transformation analytically.

Note that the aim is not recovering of *any* extent of rotation and translation. A reasonable assumption is that the head in both input images is roughly in the same global orientation (transversal, coronal or sagittal), e.g. increasing z -direction corresponds to the bottom-to-top direction in head in both images. The information about orientation of slices is usually stored with images and it is therefore easy to make the reorientation while loading the images. In other words, we do not expect that orientation of head in the input images differs by 90° or 180° in any axis but it is fairly possible that there will be a difference up to $30-40^\circ$ and the aim is to make the procedure more robust within this range of expected parameters.

In the following sections, we will first summarize the basic notions about the image moments and the principal axes transformation and then present

a modification suitable for registration of PET and MR images.

4.6.1 Moments and the principal axes transformation

The moment theory is a classical theory in physics. The utilization of various moments, such as the centre of mass or the moments of inertia, is a commonplace in many disciplines. Moments are frequently applied in digital image processing for description, classification and recognition of shapes (Seul et al., 2001; Parker, 1997; Sonka et al., 1994; Gonzales and Woods, 1993). They are useful features for determining location and orientation of an object in an image, and can be also applied to matching of image patterns. A possible application of moments for registration of CT, MR and PET images was presented in Alpert et al. (1990). In their work the image shapes had to be delineated manually which was a time consuming process.

Moments

For a 3D continuous function $f(x, y, z)$, the moment of order $(p + q + r)$ is defined as

$$m_{pqr} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q z^r f(x, y, z) dx dy dz \quad (4.4)$$

In digitized 3D images the function $f(x, y, z)$ corresponds to voxel intensities and the moments are evaluated using sums:

$$m_{pqr} = \sum_x \sum_y \sum_z x^p y^q z^r f(x, y, z) \quad (4.5)$$

These moments are dependent on scaling, translation, rotation and intensity transformation of the function $f(x, y, z)$. The zero-order moment m_{000} represents the total image power and the first-order moments determine the image centroid $(\bar{x}, \bar{y}, \bar{z})$:

$$(\bar{x}, \bar{y}, \bar{z}) = \left(\frac{m_{100}}{m_{000}}, \frac{m_{010}}{m_{000}}, \frac{m_{001}}{m_{000}} \right) \quad (4.6)$$

Image centroid determines the position of the image in space. Translation invariance can be achieved by *central moments* μ_{pqr} which are defined with the help of the image centroid as:

$$\mu_{pqr} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q (z - \bar{z})^r f(x, y, z) dx dy dz \quad (4.7)$$

and in digitized images as:

$$\mu_{pqr} = \sum_x \sum_y \sum_z (x - \bar{x})^p (y - \bar{y})^q (z - \bar{z})^r f(x, y, z) \quad (4.8)$$

Matrix of inertia and principal axes

Second-order central moments characterize the image orientation in space. They together constitute the symmetric *matrix of inertia* \mathbf{I} :

$$\mathbf{I} = \begin{pmatrix} \mu_{200} & \mu_{110} & \mu_{101} \\ \mu_{110} & \mu_{020} & \mu_{011} \\ \mu_{101} & \mu_{011} & \mu_{002} \end{pmatrix} \quad (4.9)$$

In an intuitive sense, the diagonal moments (*moments of inertia*) determine the rotation moment of the object around each coordinate axis and the off-diagonal moments (*products of inertia*) roughly correspond to torque forces that work on the object when it rotates around each coordinate axis.

The matrix of inertia is dependent on the object's orientation in the cartesian coordinate space. The *Principal axes theorem* states that for every rigid body it is possible to chose a cartesian system and an origin so that the matrix of inertia becomes diagonal. This can happen when one aligns the axes of the local reference frame in such a way that the mass of the object evenly distributes around the axes, thus, the off-diagonal product-of-inertia terms all vanish. Such coordinate axes are called the *principal axes* of the object. Their origin coincides with the object's centroid. For more information see, e.g., [Fowles and Cassiday \(1993\)](#).

For illustration, a principal axis of a rigid body corresponds to an axis where there are no torque and vibration forces on the body when it rotates around the axis. For simple geometrical shapes, the principal axes coincide with the axes of symmetry and, in general, form an orthogonal coordinate system with the origin in the centre of mass of the object.

The matrix of inertia \mathbf{I} can be expressed using a similarity transformation as:

$$\mathbf{I} = \mathbf{S} \mathbf{I}_a \mathbf{S}^T \quad (4.10)$$

where \mathbf{I}_a represents the (diagonal) matrix of inertia in the principal axes coordinate system and the matrix \mathbf{S} is a rotation matrix of eigencolumns, orthonormal vectors directed along the principal axes. Equation 4.10 can be interpreted as a rotation of the inertia matrix \mathbf{I}_a relative to the original image coordinates.

4.6.2 Image registration using principal axes

Suppose that images \mathcal{R}_1 , \mathcal{R}_2 represent the same object in different orientations and that the two objects have been shifted so that their centroids coincide. The similarity transformations for each image

$$\mathbf{I}_1 = \mathbf{S}_1 \mathbf{I}_a \mathbf{S}_1^T \quad \text{and} \quad \mathbf{I}_2 = \mathbf{S}_2 \mathbf{I}_a \mathbf{S}_2^T$$

imply that the inertia matrices \mathbf{I}_1 , \mathbf{I}_2 are related by

$$\mathbf{I}_2 = \mathbf{S}_2 \mathbf{S}_1^T \mathbf{I}_1 \mathbf{S}_1 \mathbf{S}_2^T. \quad (4.11)$$

where $\mathbf{S}_2 \mathbf{S}_1^T$ represents the rotation matrix that aligns the coordinate axes of the object in \mathcal{R}_1 with the coordinate axes of \mathcal{R}_2 . This equation is the basis of the registration based on principal axes.

If the centroids of the objects in \mathcal{R}_1 and \mathcal{R}_2 do not coincide then it is necessary to include the translation matrices \mathbf{C}_1 , \mathbf{C}_2 that shift the image centroids to the origin of the coordinate system. That is, they shift the image by $(-\bar{x}_1, -\bar{y}_1, -\bar{z}_1)$, resp. $(-\bar{x}_2, -\bar{y}_2, -\bar{z}_2)$ so that their centroids coincide when the rotation is applied. The resulting transformation matrix \mathbf{M} that registers the image \mathcal{R}_1 to \mathcal{R}_2 is then:

$$\mathbf{M} = \mathbf{C}_1 \mathbf{S}_2 \mathbf{S}_1^T \mathbf{I}_1 \mathbf{S}_1 \mathbf{S}_2^T \mathbf{C}_2^{-1}. \quad (4.12)$$

We use the Jacobi's method for computation of eigenvalues and eigenvectors of the symmetric matrix \mathbf{I} in equation 4.10. It is not the most efficient method but efficiency is not an issue for a 3×3 matrix and the Jacobi's method is fairly robust and simple to implement. The implementation used here is based on the code presented in [Press et al. \(1992\)](#). The moments are computed from voxels within the respective masks.

Direction of the principal axes

The eigenvectors of the diagonalized inertia matrix define the direction of the principal axes but not their orientation. It is thus necessary to check the correct orientation. Since we do not attempt recovering of rotational differences larger than approx. 40° we can check the orientation of corresponding pairs of axes from both images by checking the sign of the scalar product of the vectors.

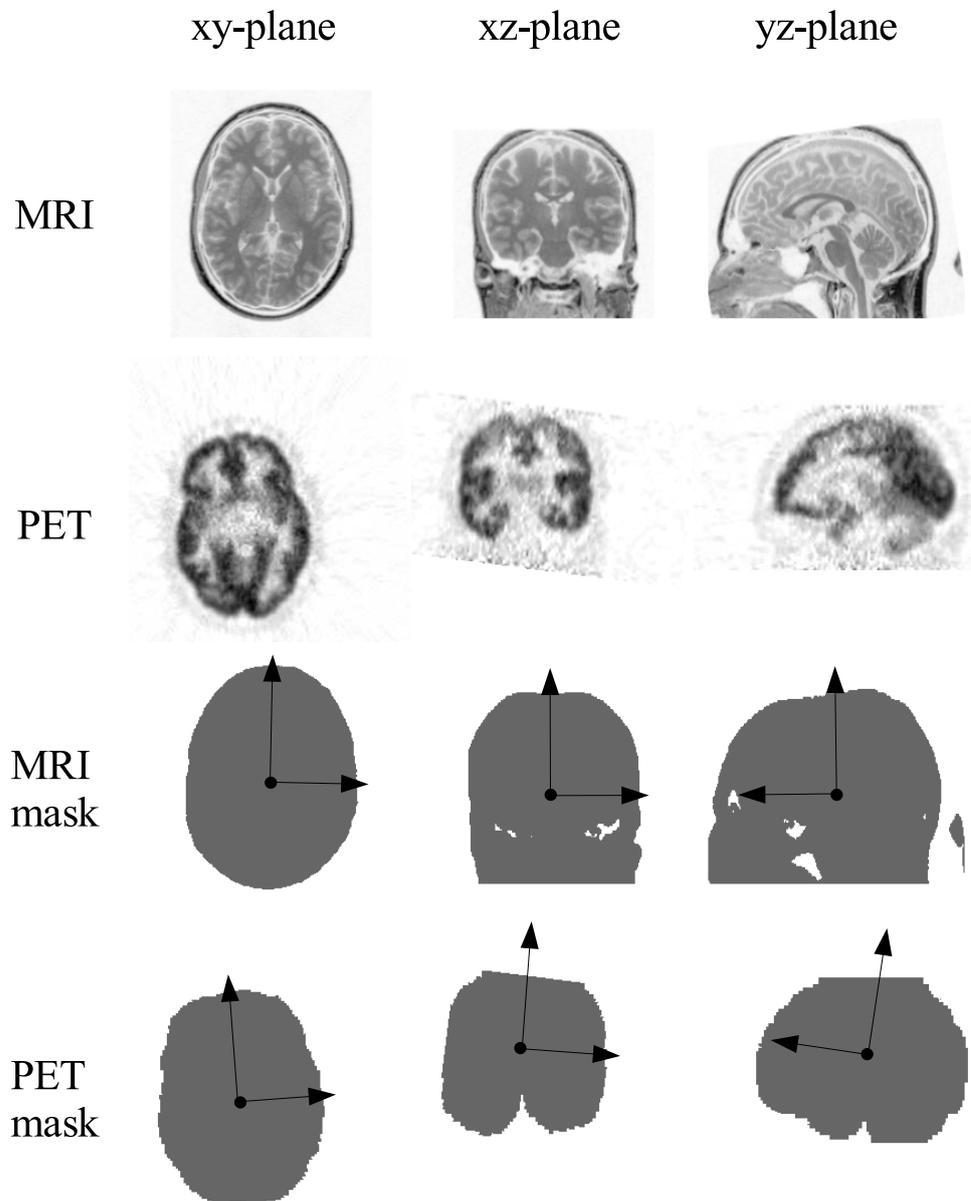


Figure 4.7: Estimation of initial displacement and orientation. The centres of mass and principal axes are computed for each image from voxels within each mask and are used for estimation of the initial transformation.

4.6.3 Corrections for PET and MR images

The described technique assumes that the registered objects are identical in shape and are not rotationally symmetrical². Masks created from a PET and a MR image do not have identical shape and therefore some additional corrections must be employed in order to achieve a good result.

Brain masks of PET and MR images are very similar in shape in the transversal (xy) plane (see Figure 4.7). This makes the estimation of the translation in x and y as well as of the rotation around the z axis quite robust and mostly precise. The largest difference between a PET and MR mask is in the inferior part (negative z -direction) where MR masks include the neck area but PET masks usually not (see the xz - and yz -planes in Figure 4.7). The consequence is that the z -coordinates of centroids are systematically shifted with respect to the true alignment position and that the estimation of orientation in the xz - and yz -plane may differ in the PET and the MR image.

We therefore apply an additional rough exhaustive search over the two rotation parameters and one translation parameter with the help of the similarity measure. The similarity measure is evaluated at a low resolution (scaling factors [8,8,2] by default) for a set of L_r different values of each rotation parameter in regular intervals around the parameters determined by the principal axis transformation up to a limiting rotation extent $\pm r_{lim}$. Analogously, the displacement in the z -axis is tested for L_z values of the z -translation parameter up to a limit translation $\pm t_{lim}$. This makes up a total of $L_r^2 \cdot L_z$ evaluations of similarity function. Suitable empirical values are for example $L_r = 7$, $L_t = 5$, $r_{lim} = 30^\circ$ and $t_{lim} = 20$ mm requiring a total of 245 evaluations and taking about 4 s. The estimation of similarity at the scale level [8,8,2] is very rough and therefore the B_1 best transformations (i.e. those yielding the best similarity values) are evaluated once again at the scale level [4,4,1]. For $B_1 = 20$ this step takes about 2 s. The best B_2 transformations (e.g. $B_2 = 5$) are then optimized using a few (e.g. 10) iterations of the simplex method at the scale level [4,4,1] and finally the best result of these optimizations is used as the initial transformation estimate for the main optimization procedure. For a common PET-MR image pair the whole procedure takes about 8-15 s.

The advantage our approach is that three of the six transformation parameters are computed purely analytically yielding a robust estimation without involving the similarity evaluation or iterative optimization. It reduces dimensionality of the problem and increases chances that the estimation of the remaining parameters succeeds and that a good initial estimate is provided

²If the object is rotationally symmetrical around some axis then the two corresponding principal axes are not uniquely defined.

for the main iterative optimization procedure.

We will further refer to the MMM algorithm extended by the segmentation procedure and by the modified principal axis transformation as MMM-PA. The robustness and computational time of MMM-PA will be compared against the basic MMM algorithm in section 4.10.

4.7 How many bins to use for computation of mutual information?

In this section we will focus on the selection of a suitable number of bins for computation of the MI-based measures. We will demonstrate that the number of histogram bins is an important parameter that may to a large extent influence smoothness of the MI function and thus robustness and accuracy of a registration. This issue requires yet more caution when applying multi-scale schemes (section 4.3). The main goal is finding of a suitable histogram size for a PET-MR registration.

Distribution of image intensities into histogram bins is effectively a re-quantization of the image. A small number of bins reduces detail in image as a larger range of image intensities is mapped into a single bin, such that a worse accuracy might be expected. For example, using 16 bins is equivalent to that image having a 4 bit intensity range. On the other hand, a large number of bins makes the histogram sparse. This may lead to a poor approximation of the joint and marginal probability distributions from the corresponding histograms which would in turn make the mutual information function less smooth with more local maxima.

Figure 4.8 shows the shape of mutual information as a function of rotational misalignment around the x and y axes for different histogram sizes. The graphs were computed from a PET and a MR image. A too small histogram size reduces detail in image and may introduce some false local minima (Fig. 4.8 top). Using of too many histogram bins seems to be even worse because it spoils the “global” smoothness of MI and introduces many local minima (Fig. 4.8 bottom). A suitable compromise needs to be found (Fig. 4.8 middle).

4.7.1 Statistical background

The question of a suitable bin width for computation of histograms has been known to statisticians for a long time, the first account probably being the Sturges’ rule in the year 1926. Since that time, several attempts have been made to cover the problem with a suitable theory and many other rules have been proposed, some theoretically well-founded, many empirical and

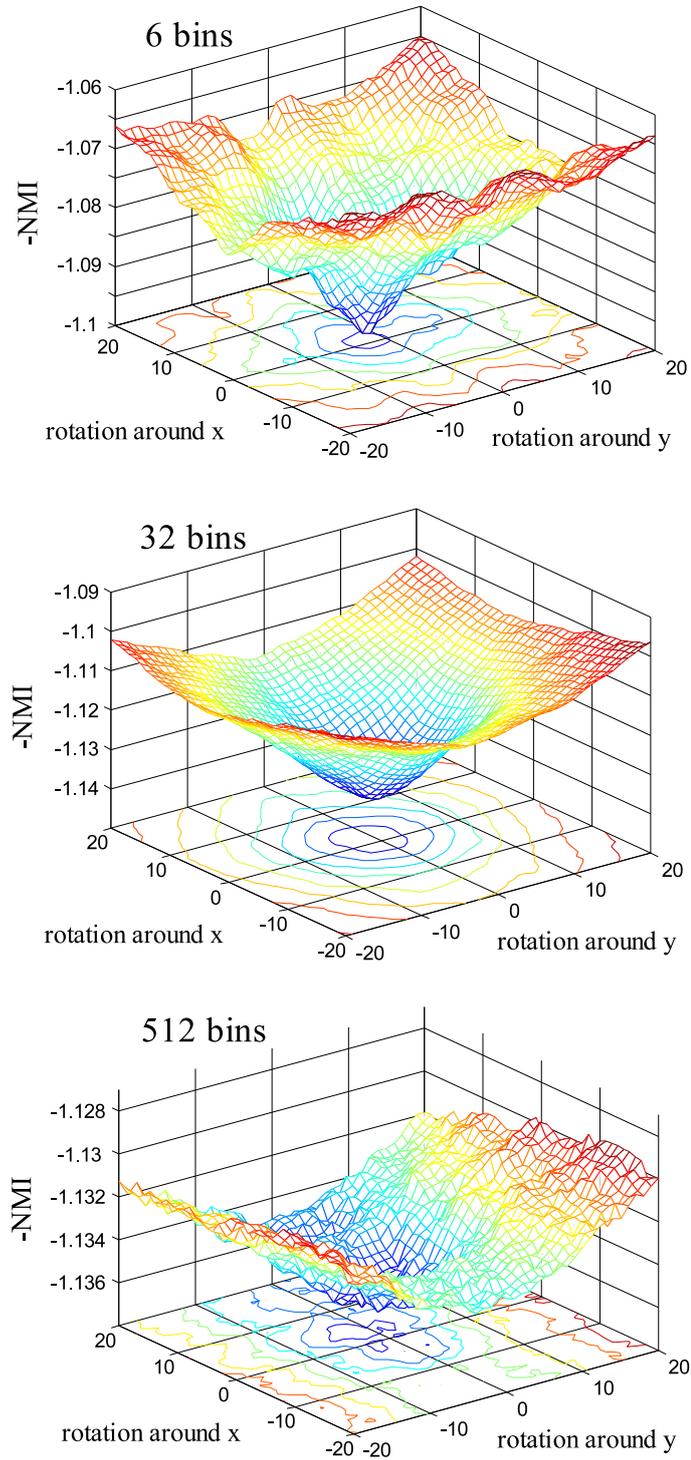


Figure 4.8: Plots of the mutual information as a function of rotation around the x and y axis for various histogram sizes.

just as many oracles. See [Wand \(1997\)](#) for a review of the most important approaches. As noted in [Keating and Scott \(1999\)](#), the potential of these rules is limited and this basic problem still remains unresolved. In practice, the choice of bin width depends on the practitioner's motives and experience and the character of the data.

Theoretically better-founded rules mostly aim at minimizing some error measure between the original density function $f(x)$ and the approximating histogram function $\hat{f}(x, w)$ where w is the bin width. It was found that, with respect to L_p norms, the optimal rate of decrease of the bin width w is $N^{-1/3}$ ([Wand, 1997](#)). The most often used error measure is the mean L_2 norm, the Mean Integrated Squared Error (MISE):

$$MISE(\hat{f}(\cdot, w), f(\cdot)) = E \int_{-\infty}^{\infty} \left(\hat{f}(x, w) - f(x) \right)^2 dx. \quad (4.13)$$

In the works of Scott, Freedman and Diaconis several rules have been proposed that try to minimize MISE (see [Wand \(1997\)](#) for a summary of these rules). The rules for the bin width w are generally of the form

$$w = \hat{C} \cdot N^{-1/3} \quad (4.14)$$

where \hat{C} is some suitable statistic. If v_{max} , resp. v_{min} are the maximum and minimum values in the image then the number of histogram bins B corresponding to the width w is:

$$B = \frac{v_{max} - v_{min}}{\hat{C}} \cdot N^{1/3} \quad (4.15)$$

A popular choice for \hat{C} is ([Keating and Scott, 1999](#))

$$\hat{C} = 3.49 \cdot \min \left(S, \frac{IQR}{1.349} \right) \quad (4.16)$$

where S is the sample standard deviation of the data and IQR is the inter-quartile range³. It has been noted in [Wand \(1997\)](#) that this choice of \hat{C} produces only a rough approximation to the MISE-optimal bin width and has no good properties for large samples or when the data is skewed or when the histogram consists of more than one mode. Unfortunately, large number of samples is exactly what we have in medical images and the data is bi-modal (background and object). The latter is the reason why the formula for \hat{C} in equation 4.16 is not suitable in our case and would produce far too large histogram size.

Histogram is not the only way how to estimate the value of mutual information. An interesting approach was presented in [Fransens et al. \(2004\)](#)

³Inter-quartile range is the 75%-quantile minus the 25%-quantile.

where kernel density estimation is used instead of histograms for computation of MI in registration of snapshots of the earth surface. On the contrary to histograms, the kernel density estimation is well theoretically founded, even with regard to multi-variate density functions like the joint entropy. This class of methods, however, cannot be easily applied here because it is computationally very demanding for sample sizes $> 10^5$ (Fransens et al., 2004; Sain, 1999).

4.7.2 Histogram size and mutual information

As there has been little research of the bin width in statistics, there has been nearly no research about the influence of bin width on the properties of MI-based functions and on the robustness and precision of a multi-modal registration. It is not clear how the bin width should be chosen in order to get a smooth MI function with large capture range on one side and on the other side to make the MI function precise enough in the vicinity of its optimum. More specifically, there is no evidence that selecting a MISE-optimal bin width is optimal with respect to the properties that one needs for a good registration.

Most works about MI-based medical image registration use a fixed number of bins without giving reasons for the choice, or the information about histogram size is missing. It is not unusual to find methods that use 256 bins for PET-MR registration which is a choice that results in too sparse histograms and gives suboptimal results, as we will see in the following tests. Surprisingly, in a recent and very comprehensive survey of MI-based registration (Pluim et al., 2003) the question of histogram size is not mentioned.

So far we have been discussing estimation of 1D marginal histograms only. Mutual information also requires estimation of a 2D distribution function. Theoretic foundations for multi-variate probability density estimation are available for methods based on the kernel density estimation (e.g. Sain, 1999; Fransens et al., 2004) that are computationally too demanding to be applicable here. Estimation of multi-variate probability density with the help of multi-dimensional histograms seems to be an unexplored area of statistics. We will therefore use the simplest generalization of the 1D case and use the same number of bins for the 1D histograms and for both dimensions of the 2D histogram (i.e. if the 1D histogram has N bins, the 2D histogram has N^2 bins). This simplifies implementation of mutual information since both marginal histograms can be efficiently computed by summation over the joint histogram.

Multi-scale registration

From the equation 4.14 it follows that a convenient value of histogram size depends on the number of samples in the image. Registration with a multi-scale approach uses substantially different amounts of samples at each level which means that the common approach, i.e. using of a fixed number of bins at all levels, needs not be appropriate. At a scale level with factors $[f_x, f_y, f_z]$, the average number of counts in the histograms is $f_x \cdot f_y \cdot f_z$ times smaller and the histogram becomes very sparse for larger subsampling factors. For example, a common PET image with 128x128x47 voxels provides 770048 samples. When using 256 bins there are on average 3000 samples per bin in the marginal histograms and about 12 samples per bin in the joint histogram. On a scale level with factors $[4, 4, 1]$, however, the average sample count in the marginal histograms is approx. 190 for the marginal histograms and only 0.7 for the joint histogram.

To the best of my knowledge, the only paper touching the topic is that of [Thurfjell et al. \(2000\)](#). They studied the influence of histogram size on the accuracy of registration of MR and simulated SPECT images in a multi-scale registration and came to the conclusion that using too many bins may significantly decrease registration accuracy. They also suggested using of different histogram sizes at each scale level.

Since there is not enough theoretical foundation to derive a suitable bin width for medical images we will experimentally examine the relation between bin width and registration quality in more detail on simulated and real PET and MR images and use the results to select a suitable number of bins for the type of multi-modal registrations considered here.

4.8 Experimental setup

In the following sections, we present several experiments that examine the relationship between the histogram size and registration quality and that validate the proposed MMM-PA method. Two types of test data are used for these evaluations:

- **Simulated, artificially misaligned PET images** derived from high-resolution MR images. The transformation needed to align them is precisely known which makes them suitable for rigorous tests of registration accuracy and robustness. These tests are suitable for a basic validation of a registration procedure but care must be taken when interpreting the results for real images.
- **Clinical PET-MR image pairs**: These images stem from clinical examinations where the alignment transformation is not known. The

quality of registration must be evaluated visually by an experienced physician or by some indirect measure, a rigorous evaluation of accuracy is not possible. These images should demonstrate that the results acquired for simulated data are in accord with the outcomes for real images.

Three quantities evaluated in these tests are the *robustness*, *accuracy* and *computational time*. Accuracy can be evaluated using error measures computed from the true transformation and the transformation computed by the registration algorithm. The error measures can be computed only for the simulated data where the true misalignment is known precisely. In most cases, the errors are either very large, indicating that the registration failed, or comparatively small for successful registrations. This allows to easily distinguish between successful and failed registrations. The term robustness stands for the rate of unsuccessful registrations to the total number of registrations.

All used PET and MR images originate from the PET Laboratory of the Max-Planck-Institute for Neurological Research in Cologne.

4.8.1 Simulated PET images

The simulated PET images were generated from a set of 5 high-resolution, high-contrast T1-weighted MR images of normal patients acquired using a 1.5 T Gyroscan Intera scanner (Philips Medical Systems, Best, The Netherlands). The image dimensions were 256 x 256 x 200 voxels with voxel size 1.0 x 1.0 x 1.0 mm.

For each MR image, a simulated PET image was created using a fully automatic procedure described below. The procedure is based on the simulation procedure used by [Kiebel et al.](#) for cross-validation of SPM and AIR ([Kiebel et al., 1997](#)). High-resolution, high-contrast MR images were used to enable robust segmentation of cortex and white matter.

1. Brain volume was extracted from MR image using the Brain Extraction Tool ([Smith, 2002](#)).
2. The extracted volume was automatically segmented into gray matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) using the segmentation tool *FAST* from the FMRIB software package ([Zhang et al., 2001](#)).
3. Voxel intensities of the MR image were scaled so that the mean intensities of GM, WM and CSF were related by a ratio of 10 : 3 : 1;
4. A random rigid-body transformation was applied and the image was resliced to a dimension of 128 x 128 x 47 voxels with voxel size 2.2 x 2.2 x 3.125 mm using tri-linear interpolation.

5. The image was smoothed by an isotropic 3D Gaussian filter with FWHM of 6 mm.
6. Additive Gaussian white noise with variance equal to 20 % of the mean intensity value of brain voxels was applied.
7. The image was smoothed again by an isotropic 3D Gaussian filter with FWHM of 2 mm.

In this way, a misaligned MR-PET pair was created with a known exact transformation needed to realign it. The simulation sequence was repeated several times with different transformation parameters for each MR image.

Registration errors were evaluated using a fixed set of 14 control points located on the boundary of the brain volume in each MR image. For each registration, these points were once transformed using the exact original transformation and once using the transformation computed by the registration procedure. The mean and maximum Euclidean distances Δ_E and Δ_{max} between the two sets of transformed points were used as measures of registration error for one patient. Due to the location of control points on the brain boundary the value of Δ_{max} approximates the maximum displacement that can be encountered within the brain volume and is suitable for distinction between successful and unsuccessful registrations. We will denote as $E\{\Delta_E\}$ the mean value of Δ_E over all patients.

4.8.2 Clinical PET and MR images

Evaluation with real data is always problematic because the precise matching transformation is not known and cannot be determined retrospectively. A standard (prospective) evaluation technique consists in utilization of fiducial markers fixed on patient's head during scanning in PET and MR scanner. The alignment transformation is then computed by means of least squares fit of the set of markers in the PET image to the set of markers in MR. The procedure is demanding for both patients and physicians because it must be assured that the markers remain in the same position during both measurements. Such procedure could not be performed at the time of this work and therefore the quality of registration had to be evaluated by indirect error measures.

Several pairs of real PET and T1-weighted MR images were used in the tests. Each pair was coregistered using the MMM algorithm, the results were checked visually and the resulting transformation parameters were recorded. These parameters served only as a reference for discrimination between successful and unsuccessful registrations and for generation of random transformation parameters where a possibly slightly imprecise alignment is not

a problem.

For each pair, a total of N misplaced PET images were created by transforming the original PET images with random transformation parameters generated around the reference transformation. Each misplaced PET image was then coregistered to the corresponding MR image by each of the evaluated methods.

As the correct alignment parameters were not known, the mean and maximum errors of each registration could not be computed. Instead, the accuracy was estimated by the standard deviation of the computed transformations in all N runs. The motivation for this choice was an assumption that the more precise a registration method is, the less dispersed will be the resulting transformations.

The standard deviations were computed with the help of the 14 control points $c_i, i = 1, \dots, 14$ that were displaced in the MR volumes in the same way as in the tests with simulated data. For a patient p , the N test runs produced N vectors of transformation parameters $\alpha_{p,t}, t = 1, \dots, N$. Every control point c_i was transformed, for each patient, using the computed parameters into locations $c'_{i,p,t}$. Mean locations (centres of mass) $\mu_{i,p}$ of all $c'_{i,p,t}, t = 1, \dots, N$ were computed and from them the standard deviations $\sigma_{i,p}$ were determined that represented the dispersion of results at a control point i of a patient p . The mean value of $\sigma_{i,p}$ over all control points and patients, $E\{\sigma_E\}$, was used for evaluation of registration accuracy.

This measure is only informative, since it may be influenced by unknown systematic errors. For example, a certain registration type might give very consistent resulting transformations with a small variance but systematically shifted from the true alignment.

4.9 Test of the influence of histogram size

The aim of this test is to examine how the registration outcome of a PET-MR registration depends on a good choice of the number of histogram bins used for computation of mutual information. It examines the relation between the number of bins, number of voxel samples, registration accuracy and robustness. It also verifies if and how the dependency on histogram size changes when performing registration with or without background masking.

We want to evaluate two basic characteristics. First, we need to know how the histogram size influences the shape of mutual information in vicinity of optimum because it directly influences *precision* of registration. Second, we want to estimate the impact of histogram size on the capture range of mutual information which influences *robustness* of registration.

For the test of precision we generate images with small initial misregistrations and evaluate how precisely the registration method aligns the images.

The test of robustness uses images with larger initial misregistrations and the evaluated property is the ratio of successful registrations where we define successful registrations as those with registration error smaller than some threshold.

Test setup

Both simulated and real data were used in this test. The initial misregistrations were generated randomly from a normal distribution with mean $\mu = 0$ and standard deviation σ :

- for the test of *precision*: $\sigma = 3^\circ$ for rotation and $\sigma = 5$ mm for translation parameters,
- for the test of *robustness*: $\sigma = 9^\circ$ for rotation and $\sigma = 15$ mm for translation parameters.

25 artificially misaligned PET-MR pairs were generated for each registration type (precision, robustness), image type (simulated, real) and every patient. MR images of 5 patients were used for generation of simulated PET images. The real PET-MR pairs originated from 4 patients and consisted of the following PET modalities: FDG, FMZ (2x) and MP4A. Together $5 \times 25 \times 2$ simulated and $4 \times 25 \times 2$ real PET images were created which gives a total of 450 misaligned PET-MR pairs.

The PET-MR pairs were coregistered using 8, 16, 32, 64, 128 and 256 histogram bins in single-scale registrations with scaling factors [1,1,1], [2,2,1] and [4,4,1]. Each registration was performed both with and without prior background removal. There were thus altogether 36 different registration setups for each PET-MR pair and a total of $36 \times 450 = 16\,200$ registrations runs.

4.9.1 Results and discussion

The Figures 4.9 and 4.10 show graphically the dependence of registration errors on the number of histogram bins. The relationship between the number of histogram bins and the ratio of unsuccessful registrations is visualized in Figures 4.11 and 4.12. We consider a registration as successful if its registration error is smaller than 5 mm. This threshold was chosen so that most registrations have either much higher or much smaller registration error.

We will first discuss the basic registration method, i.e. with no background segmentation. At all scale levels the errors are larger when the number of bins is at the low or high end of the scale. Optimal precision and robustness has been achieved with 32 - 64 bins. There is a tendency that

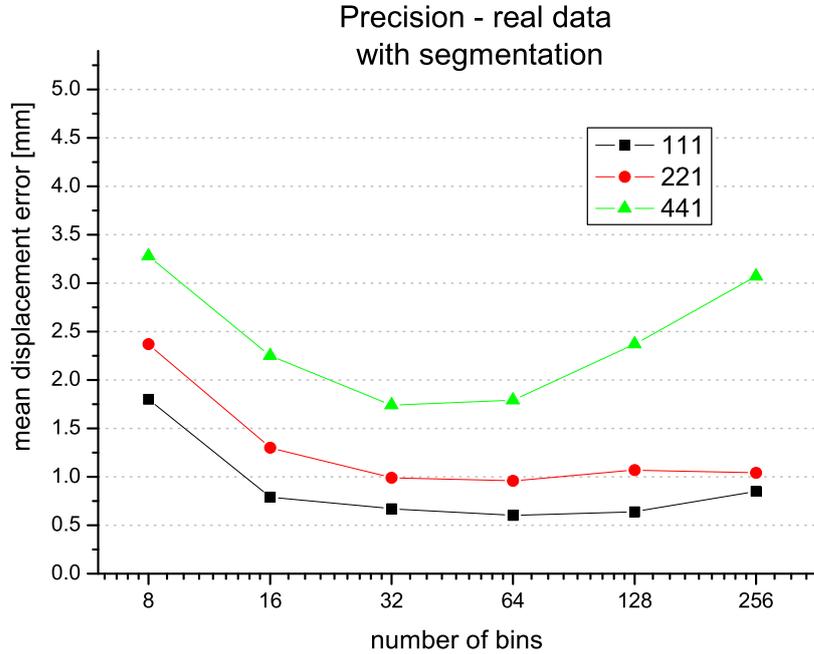
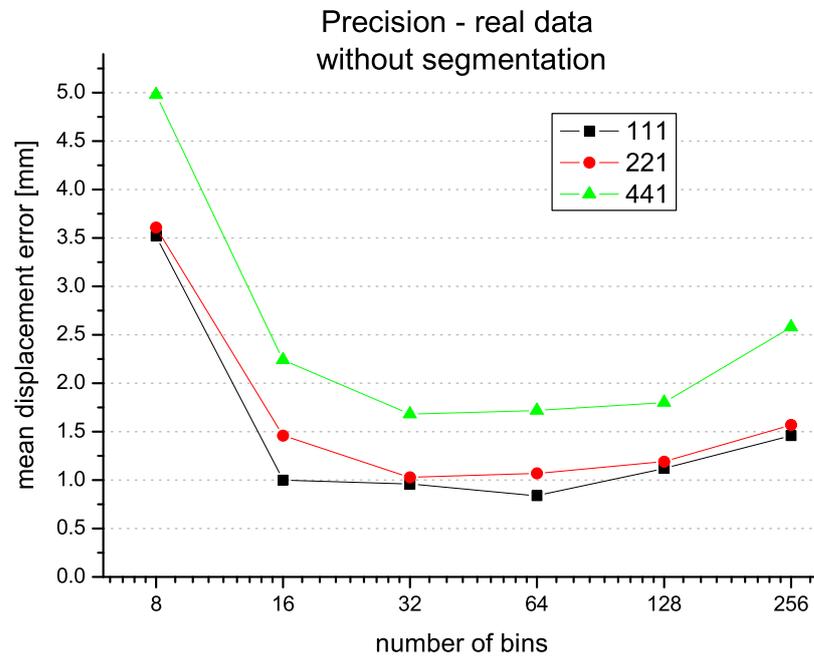


Figure 4.9: Registration errors of real PET images with and without segmentation of background.

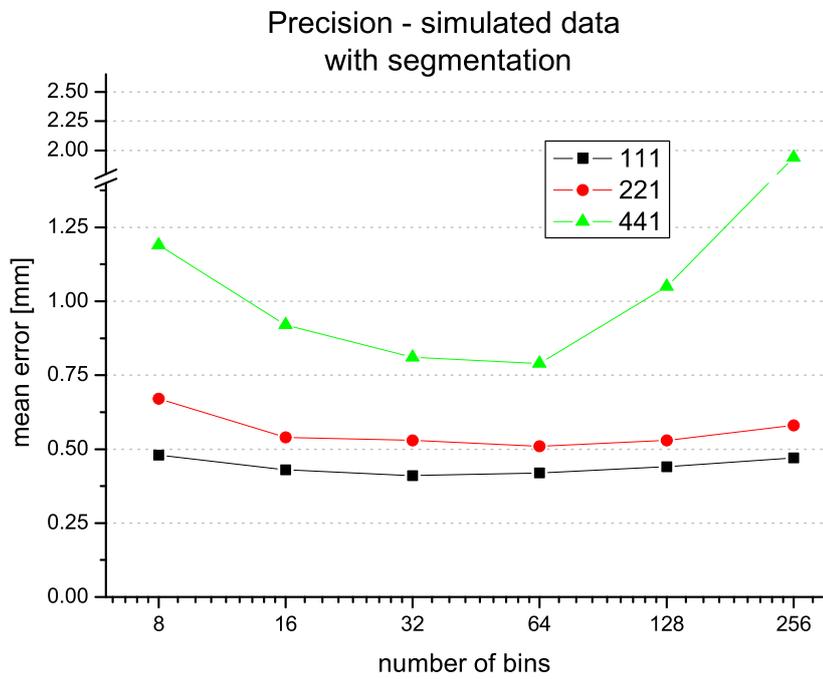
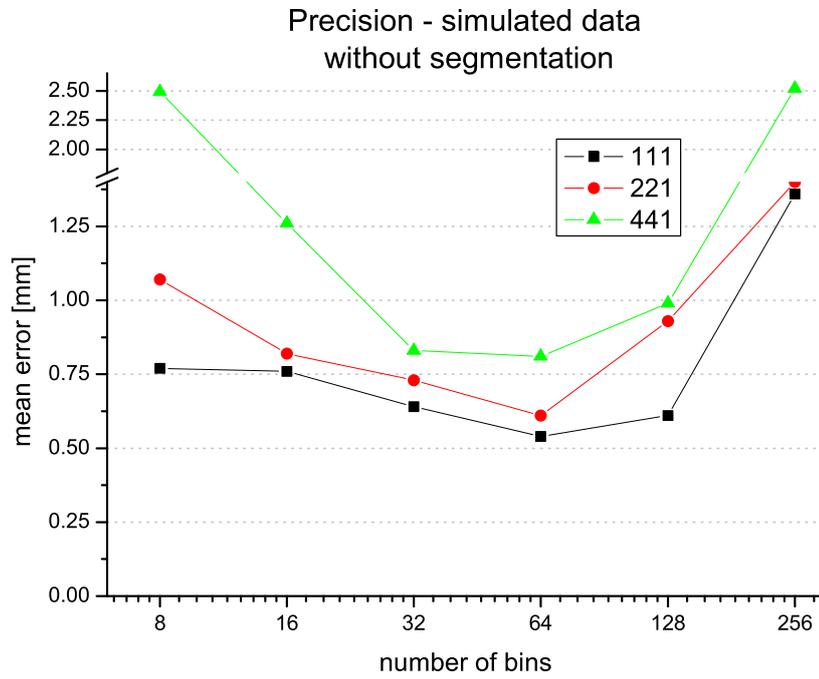


Figure 4.10: Registration errors of simulated PET images with and without segmentation of background.

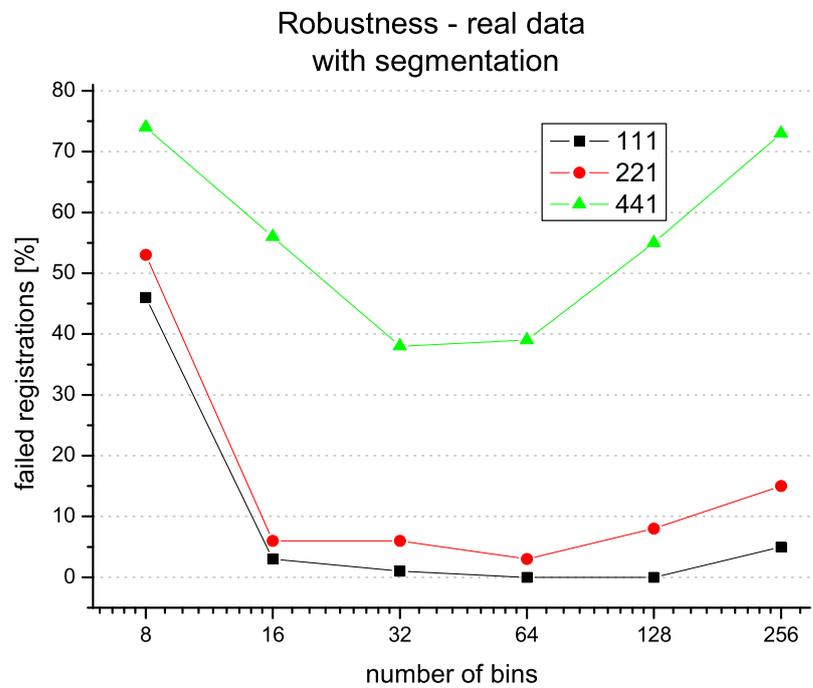
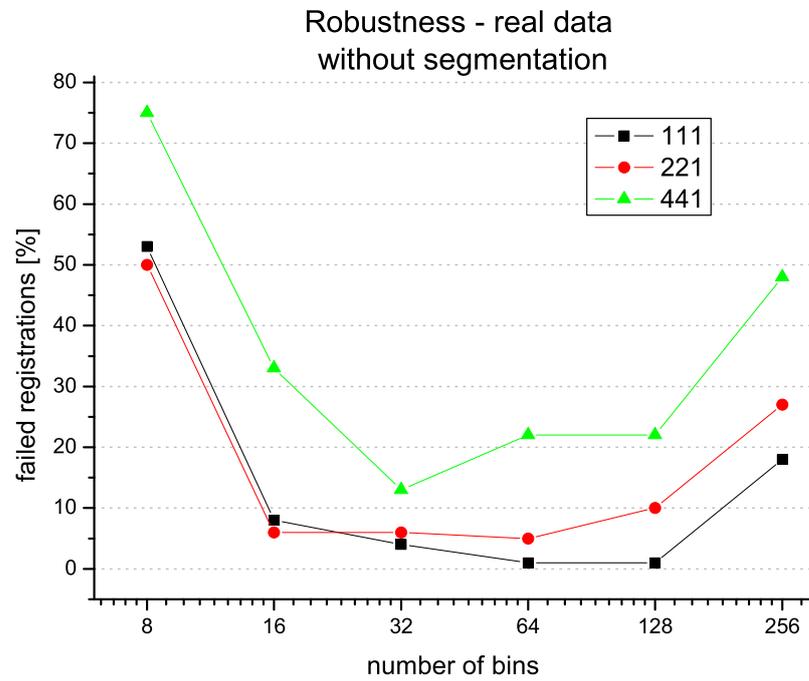


Figure 4.11: Robustness (percentage of failed registrations) for real PET images with and without segmentation of background.

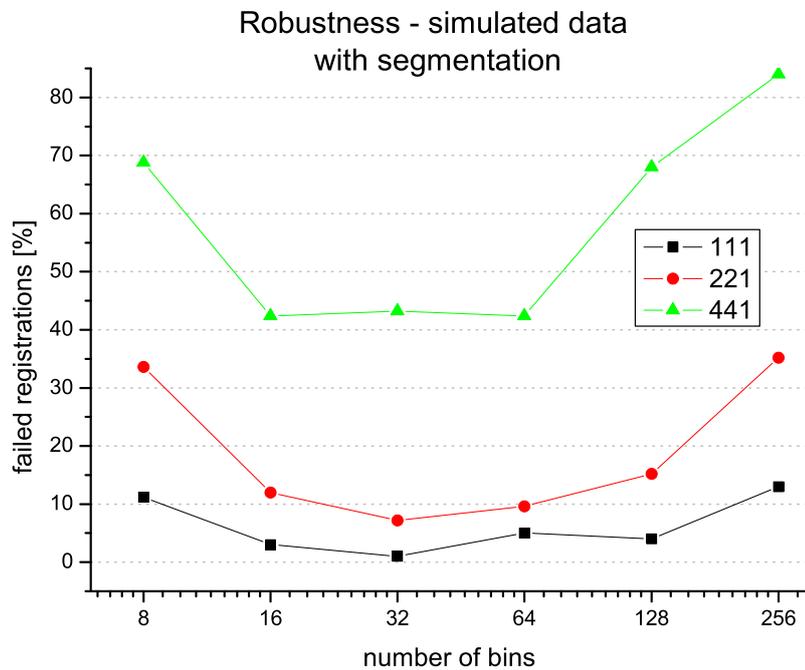
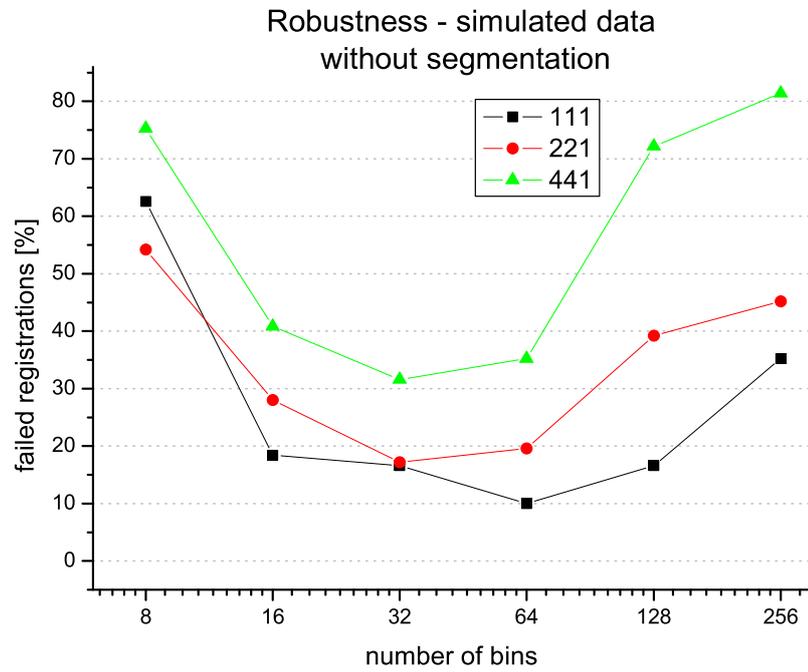


Figure 4.12: Robustness (percentage of failed registrations) for simulated PET images with and without segmentation of background.

lower resolutions prefer smaller histogram size than higher resolutions. It means that an adaptive number of bins that takes sample size into account is more appropriate than a fixed histogram size. On the other hand, it is apparently possible to achieve good accuracy and robustness even with a fixed number of bins that gives stable results at all scale levels, e.g. 32 bins.

Precision and particularly robustness gets considerably deteriorated when 8 or 256 bins are used. It is however not uncommon that authors use 256 bins for registration of PET and MR images. For example, [Maes](#) evaluated registration of PET and MR images using a very similar method to our basic MMM with a fixed histogram size of 256 bins and reported registration errors as large as 5° for rotation and 30 mm for translation ([Maes, 1998](#)).

If we compare precision of registrations with and without background segmentation we can see that removal of background improves precision at scale levels [2,2,1] and [1,1,1]. For both real and simulated images, removal of background “stabilized” the dependency on histogram size at scale levels [2,2,1] and [1,1,1] and provided a balanced precision for histogram sizes 16 - 256 (real images), resp. 8 - 256 (simulated images).

Robustness of registrations with background removal may become problematic if the initial misalignment is large and the overlap of the registered images in their starting position is small. According to the presented results this is not the case at the levels [2,2,1] and [1,1,1] that even exhibit a slightly improved robustness. At the low-resolution level [4,4,1] the smaller overlap and the small number of samples outweighed the advantage of removing of noisy background and resulted in a much higher number of failed registrations. We will see in the next section that this poses no problems if we use the global displacement estimation. It provides an estimate that is sufficiently close to the optimum to allow a robust registration at the level [4,4,1].

Looking at the differences between results of simulated and real data registrations we can see that the character of these results is very similar. This is important because the displacement error of real data registrations is just an indirect measure of accuracy whereas the error of a simulated registration is a precise measure of displacement. A direct comparison of errors of both registration types is not possible because they reflect different quantities.

An empirical choice of histogram size

On the basis of the experimental results we can select a suitable number of bins for each scale level. The optimal histogram sizes for real and simulated data, as well as for registrations with and without background segmentation are quite similar and in most tests there are no outright optima. It allows us to use the same parameters in all these cases.

The generic rule in equation 4.15 (section 4.7.1) suggests that the L_2 -optimal number of bins is dependent on the number of samples N as $N^{1/3}$. In the case of our PET-MR registration, the number of samples of the target PET image (dimension 128 x 128 x 47 voxels) is approximately $7 \cdot 10^5$ at the full resolution, about $1.9 \cdot 10^5$ at the level [2, 2, 1] and $4.8 \cdot 10^4$ at the level [4, 4, 1].

In our tests, the scale level [4, 4, 1] was the most sensitive to the number of bins and it generally achieved best results with approximately 32 bins. If we suppose that the $N^{1/3}$ dependency approximately holds for mutual information and if we require 32 bins for the [4, 4, 1] level, we can determine the number of bins at each scale level as:

$$B = 0.87 \cdot N^{1/3}$$

This equation suggests 32, 50 and 80 bins for levels [4, 4, 1], [2, 2, 1] and [1, 1, 1]. These values are in compliance with the approximate overall optimal values according to the experiments.

Summary

The most important result of this experiment is a confirmation of the hypothesis that the number of histogram bins has a large impact on registration outcome and that both too small and too high number of bins decrease registration accuracy and robustness. The rule for selecting of bin size is a rather pragmatical solution to an issue where a theoretical solution is missing. There is no prove that the relation between the number of samples and the optimal number of bins for mutual information is $N^{1/3}$. The mutual information measure is an interaction of 1D and 2D histograms and might behave in a more complex way.

The issue of histogram size is a largely unknown space in the otherwise well explored field of medical image registration based on mutual information. There are many aspects that would deserve closer examination. For example, it could be of advantage to use different histogram sizes for the marginal and the joint histogram. It could also be interesting to investigate possibilities of a variable bin width. The two modes in PET and MR image histograms (object, background) have very different characteristics and it could be of benefit to use two different bin widths, one for estimating the distribution of background voxels and one for the object voxels.

4.10 Evaluation of masking and global displacement estimation

The aim of the following test is to validate the MMM-PA method and to test if it actually increases robustness and accelerates registration. An important goal is making sure that the gain in speed is not at the cost of a lower accuracy. In addition, we examine the effect of omitting the full-resolution level (scale factors [1,1,1]).

Test setup

The test was performed with both simulated and real data. Rigid-body transformations were generated randomly from a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 20^\circ$ for the rotation parameters, $\sigma = 30$ mm for the translation in x and y and $\sigma = 20$ mm for the translation in z . The generated transformations were limited to the range of $\pm 40^\circ$ for rotation, ± 70 mm for translation in x and y and ± 50 mm for translation in z in order to make them fit into the range of expected transformation parameters and also into the image grid (see section 4.6 for details). The generated transformations were larger than usual misalignments of real images. On the other hand, such misalignments occur in practice now and then and the aim was to simulate the worst-case scenario.

4.10.1 Results and Discussion

Each PET image was co-registered to its corresponding MR image using MMM and MMM-PA. A multi-scale scheme with scale factors [4,4,1], [2,2,1] and optionally [1,1,1] was used. We denote a registration with just the first two levels as *low-resolution*, all three levels together constitute a *full-resolution* registration.

Robustness

Table 4.1 summarizes the number of failed registrations for each type of registration. For this purpose, a registration is declared as unsuccessful if the error measure Δ_{max} exceeds 3 mm for the levels [2,2,1] and [1,1,1] and 6 mm for the level [4,4,1].

The large extent of initial misalignment apparently exceeds the capability of the basic MMM method. MMM-PA performed substantially better and achieved a very low failure rate, regarding the extent of initial misalignment.

A closer look at the results revealed that nearly all unsuccessful registrations of MMM-PA were those registrations where the artificially misaligned PET image did not completely fit into the image grid. The corresponding

	real data		simulated data	
	MMM	MMM-PA	MMM	MMM-PA
total	125	125	125	125
failed at [4,4,1]	76	9	113	10
%	60.8	7.2	90.4	8.0
failed at [2,2,1]	41	6	95	6
%	32.8	4.8	76.0	4.8
failed at [1,1,1]	37	2	92	5
%	29.6	1.6	73.6	4.0

Table 4.1: Number of failed registrations for each type of registration.

PET image mask was thus partially cropped which changed the direction of the principal axes. In one such case the MMM method even performed better. There were, however, other cases where the MMM-PA registration succeeded despite a partially cropped object mask.

Omitting of the full-resolution level slightly worsened the failure rate of MMM-PA but the rate still remains very low, considering the large extent of initial misalignments.

Computational time

Table 4.2 presents average computational times and speed-up factors for each type of registration. The segmentation time includes the time necessary to threshold, erode and dilate images. The initialization time is just the time necessary for initialization of internal structures; the time needed for loading of data (a few seconds) is not included. The most time consuming part of each optimization stage is the evaluation of similarity measure. The number of these evaluations is therefore also included in the table.

As expected, the removal of background substantially accelerated the registration process. The numbers in parentheses in Table 4.2 are the speed-up factors for optimization times of MMM-PA compared to MMM and the reduction factors for the number of similarity evaluations needed to converge at each level.

The global estimation of misalignment added 14.2s on average to the overall registration time of MMM-PA. On the other hand, it provided a good starting estimate for the optimization and eliminated a number of similarity measure evaluations that would have otherwise been necessary (compare the number of evaluations for MMM vs. MMM-PA).

Recall from section 4.5.5 that the time of similarity evaluation is mainly dependent on the number of samples in the target volume. Removal of back-

	real data		simulated data	
	MMM	MMM-PA	MMM	MMM-PA
initialization [s]	2.0	2.0	2.1	2.1
segmentation [s]	-	6.3	-	6.5
global alignment [s]	-	14.2	-	8.0
optimization [4,4,1] [s]	34.7	9.6 (3.6)	28.7	7.4 (3.9)
evaluations of MI	217.9	162.2 (1.3)	194.0	188.4 (1.0)
optimization [2,2,1] [s]	96.9	27.7 (3.5)	70.8	21.5 (3.3)
evaluations of MI	153.4	113.1 (1.4)	144.4	116.9 (1.2)
optimization [1,1,1] [s]	237.5	89.3 (2.7)	218.9	59.0 (3.7)
evaluations of MI	94.9	89.2 (1.1)	99.0	90.7 (1.1)
<i>full-resolution</i>				
total time [s]	371.1	149.0 (2.5)	320.5	104.6 (3.1)
<i>low-resolution</i>				
total time [s]	133.6	59.7 (2.2)	101.6	45.6 (2.2)

Table 4.2: Mean computational times and speed-up factors.

ground reduces the image volumes to approximately 40% which implies a speed-up factor of 2.5 for a single similarity measure evaluation. The smaller number of similarity evaluations needed for convergency of MMM-PA further decreased optimization time and yielded, for the real data, a speed-up factor of 3.61 and 3.5 for the first two levels and 2.7 for the full-resolution level.

When the full-resolution level is omitted, a low-resolution MMM-PA registration takes less than one minute and is more than two times faster than the low-resolution MMM registration and more than six times faster than the full-resolution MMM.

Precision

An important question is whether the preprocessing in MMM-PA does not result in a loss of accuracy. Table 4.3 compares the mean registration errors over all successful registrations. The registration errors of MMM-PA are smaller than the errors of MMM in both the low- and full-resolution registrations. This is probably due to the elimination of noisy background that might have influenced accuracy of MMM.

The average precision of low-resolution registrations is only about 10% worse than that of the full-resolution ones. Moreover, the low-resolution registration with MMM-PA gave on average more precise results than the

$E\{\sigma_E\}$ [mm]	real data	
	MMM	MMM-PA
level [4,4,1]	1.83	1.37
level [2,2,1]	1.32	0.93
level [1,1,1]	1.20	0.70
$E\{\Delta_E\}$ [mm]	simulated data	
	MMM	MMM-PA
level [4,4,1]	0.87	0.71
level [2,2,1]	0.40	0.34
level [1,1,1]	0.34	0.24

Table 4.3: Mean registration errors over all successful registrations.

full-resolution registration with MMM. It suggests that the full-resolution step could be omitted in MMM-PA for the sake of a significant reduction of registration time.

Summary

The proposed method, MMM-PA, performed better than the basic MMM method in all three evaluated aspects: computational time, robustness, accuracy. The combination of an automatic large volume reduction with a fast analytical estimation of gross alignment resulted in an average speed-up of factor 2.5 for a full-resolution registration and 2.2 for a low-resolution registration. The speed-up factor is slightly smaller in the low-resolution case because the constant preprocessing time becomes more pronounced.

MMM-PA achieved a better accuracy than MMM at all levels. More importantly, the low-resolution MMM-PA was more precise than the full-resolution MMM. The achieved precision is undistinguishable by a visual check. For many applications it is therefore convenient to use MMM-PA without the full-resolution level and profit from the considerable acceleration.

MMM-PA should not be used when the brain object does not completely fit into the image volume. In such case the principal axes transformation may easily produce a wrong transformation estimate and be a cause of an unsuccessful registration.

Other applications

The described method concentrated on the registration of PET-MR images. The registration framework is, however, general enough to be easily extended

for other applications. For example, alignment of PET images acquired using different radiotracers can be accomplished using nearly the same parameters as for a PET-MR registration. Motion correction of a sequence of PET frames uses the SSD or CC similarity measure instead of MI, otherwise the parameters are similar. All these registration types were also implemented and successfully used in a number of applications at the PET Laboratory of the Max-Planck-Institute for Neurological Research in Cologne.

For a coarse co-registration of images from different patients, the rigid-body transformation model can be extended to an affine model. A precise matching, however, requires utilization of some kind of non-linear deformations which is the topic of the next chapter.

Non-linear registration

5.1 Introduction

In the previous chapter we were dealing with registration of images from the *same patient*. Rigid-body transformation model - rotation and translation - was sufficient and most suitable for the task. In this chapter we will consider non-linear inter-subject registration, i.e. registration of images of *different patients*, and we will focus on registration of PET images.

The most important applications of non-linear registration are:

- Alignment of an image or a set of images with an image template for the purpose of statistical analysis of both normal and abnormal functional or anatomical variability. This process is called *spatial normalization*.
- Construction of normal and disease-specific atlases and normative images.
- Alignment of a model of anatomy (e.g. atlas) with a particular image for the purpose of segmentation and interpretation.

Images of different patients may be very dissimilar due to a normal variability of human brain. It is therefore not sufficient to use just rigid-body transformations if a precise match of the shape and inner brain structures is required. In order to accommodate for this normal variability it would be possible to use affine transformations, i.e. rigid-body + scaling + shear. While this may suffice for some applications, in most cases a more local matching is required which calls for utilization of some kind of non-linear deformations. An example registration is shown in Figure 5.1 where a patient's PET image is warped to a template image. In this way it gets transformed into a standard space which enables subsequent statistical analysis and comparison.

In the following sections, we will introduce a novel algorithm for a non-linear registration of PET images and evaluate it on various types of clinical and simulated images.

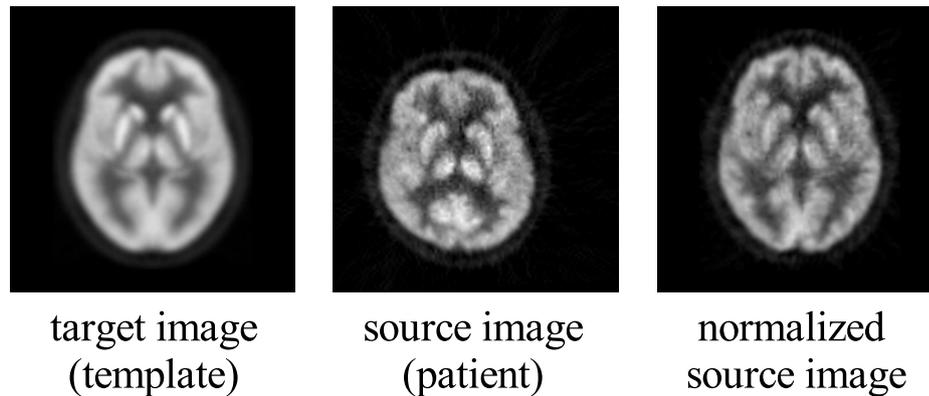


Figure 5.1: Non-linear registration of a source image (middle) to a target image (left). The registration algorithm warps the source image so that homologous features in both images correspond (right).

5.2 Rationale for a new approach

At present, there is a lot of research in the field of non-linear registration of high-quality MR images. MR scanners often produce images with in-plane pixel size of about 1 mm and provide a distinct and diversified pattern that makes possible extraction of higher-level information like edges and surfaces. Some methods for registration of MR images are to some extent able to match larger gyri and sulci in cortex (e.g. cortical flattening approaches). However, these methods are mostly not fully automatic and require, to various degree, interaction with user.

PET images on the other hand have significantly lower spatial resolution (in-plane pixel size of about 2.5 mm in better scanners) and do not have a significantly diversified pattern. PET images also contain a higher amount of noise and reconstruction artefacts. The character of the images thus does not allow reliable identification of features. For this reason most of the methods developed for high-dimensional warping of MR images would not work with PET images. Non-linear registration of PET images requires utilization of the full image contents, similarly to the rigid-body registration method described in chapter 4.

The registration task is usually regarded as an attempt to match images “as precise as possible”. However, we must consider that in the end our aim is to compare images in order to detect *differences* between them and therefore an exact match should not even be attempted. For example, when matching a lesioned brain to a normal brain it is desirable to preserve the area of lesion from being destroyed. Moreover, there exists no true one-to-one mapping between cortical structures (gyri and sulci) of two different brains (not even for healthy patients) because some structures may exist in one brain and not

in the other. Any method that tries to match brains exactly must fold the brain to “create” artificial structures and let other structures “disappear”. It is therefore not appropriate to match images beyond a certain extent and resolution.

Instead of directly registering PET to PET, we could use corresponding MR images, perform a non-linear inter-subject MR-MR registration and a rigid-body PET-MR registration for each image and transfer the non-linear transformation parameters from the MR to the PET images. There are at least two reasons why this approach is not always optimal and why it is necessary to have a procedure for a direct PET-PET registration. Firstly, MR images are not always available. Secondly, a MR-to-MR registration matches *anatomical* structures in both images. Although the functional structure in PET images highly correlates with anatomical structure, there is no evidence that they are always tightly linked. The purpose of spatial normalization of functional images is usually to maximize the sensitivity and specificity of subsequent statistical analyses of brain function. For this purpose, a direct registration of PET images might be more appropriate.

The method described in this chapter utilizes parts of the framework of the rigid-body registration method described in the chapter 4. The main difference is in the transformation model where rigid-body transformations are replaced by a non-linear transformation model. This extension brings along many difficulties. Unlike the registration of images from the same patient where the transformation model is well defined by rotations and translations, there is no clear rule as to what kind of transformations should be used to map brain image of one patient to a brain image of another patient. The non-linear registration problem is therefore ill-posed without suitable constraints:

- The transformation must be continuous and smooth.
- Topology and connectivity have to be maintained (no folding over).

In addition, an algorithm suitable for non-linear PET-PET matching must handle the specific characteristics of PET images:

- Low resolution, significant level of noise and reconstruction artefacts.
- No detailed pattern (in contrast to MR images).

The goal of the work presented in this chapter is developing of a method that:

- works well for low-resolution and noisy images,
- is robust and sufficiently precise,
- works fully automatically for a given class of images,
- *guarantees a one-to-one deformation*, i.e. it produces no folding-over or discontinuities,
- generates a deformation function that is smooth,
- is fast enough to be applicable in the time frame of common clinical applications,
- can be easily used by non-experts and if possible does not require special knowledge about the method and its parameters.

In the following text we will present a new voxel-based method that fulfils these requirements, the *Hierarchical Free-Form Block Matching* (HBM).

5.3 Hierarchical Free-Form Block Matching

Given a target image \mathcal{R} and a source image \mathcal{S} , the aim is to estimate a non-linear transformation (deformation field) that warps the image \mathcal{S} to the shape of the image \mathcal{R} .

Recall from chapter 2.3 that registration of \mathcal{S} to \mathcal{R} effectively means finding a suitable transformation that maps voxels of \mathcal{R} into the space of \mathcal{S} . The resulting resampling of \mathcal{S} is then performed via backward mapping, i.e. each voxel v of \mathcal{R} is mapped by the transformation into some point v' in \mathcal{S} , the value of v' is interpolated from the neighbouring voxels in \mathcal{S} and assigned back to v . We can thus formulate the registration task in the following way:

Find a smooth one-to-one function $\mathbf{w} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ that maps each point $\mathbf{p}_i \in \mathcal{R}$ to a point $\mathbf{q}_i \in \mathcal{S}$, $\mathbf{w}(\mathbf{p}_i) = \mathbf{q}_i$, so that the deformed image $\mathbf{w}(\mathcal{R})$ and \mathcal{S} maximize some similarity measure.

5.3.1 Similarity measure

Since we want to register images from the same modality, both the sum of squared differences (SSD) as well as the correlation coefficient (CC) described in chapter 3 can be used as a measure of similarity. We have chosen the correlation coefficient because it turned out to be a more robust estimator of

similarity of PET images than the commonly used SSD. Moreover, CC evaluates linear relationship between intensities of two images and has therefore no problem with differing global intensity magnitudes. In the case of SSD, this must be handled by estimation of a suitable scaling factor or by including an additional optimization parameter (see section 3.3). CC also has the nice property that its range of values is clearly given. Its values fall into the range of $\langle 0, 1 \rangle$ which makes a convenient regularization of the measure easier (more about regularization in section 5.3.9).

Correlation coefficient is convenient for data with diversified intensity values like PET images. It fails, however, on data that contain only a small range of intensities, e.g. binary images, and in such case the sum of squared differences is the measure of choice. The non-linear registration method presented in the following text is general enough and does not rely on a specific measure. The most suitable measure can thus be chosen depending on the intended application. All tests with PET images described further use the correlation coefficient, validation tests with binary images use the sum of squared differences. Optimization of the similarity measure is accomplished by the simplex algorithm (section 4.2) like in the multi-modality registration.

The HBM method consists of two main parts: an *affine part* that provides a rough initial alignment followed by a *non-linear part* that tries to adjust misalignments at a local level.

5.3.2 Affine part

The affine part provides an initial co-registration of images. The framework is analogous to the multi-modal registration described in chapter 4. Here, the correlation coefficient or the sum of squared differences is used instead of mutual information as the measure of similarity and the rigid-body transformation model is replaced by the affine model (chapter 2.3). The affine part also utilizes the multi-scale approach (section 4.3).

5.3.3 Non-linear part

The affine registration results in a mapping that is only roughly valid in the image volume as a whole. Different parts of image would need to be more precisely matched, with slightly different transformation parameters than those provided by the global affine transformation. This inspires to divide the image into regular blocks and to perform a local affine registration for each block separately. After these local co-registrations, resulting transformations need to be somehow interpolated among neighbouring blocks.

We will first deal with generation of a smooth warping function and then derive a subdivision schema that makes a fine local co-registration possible.

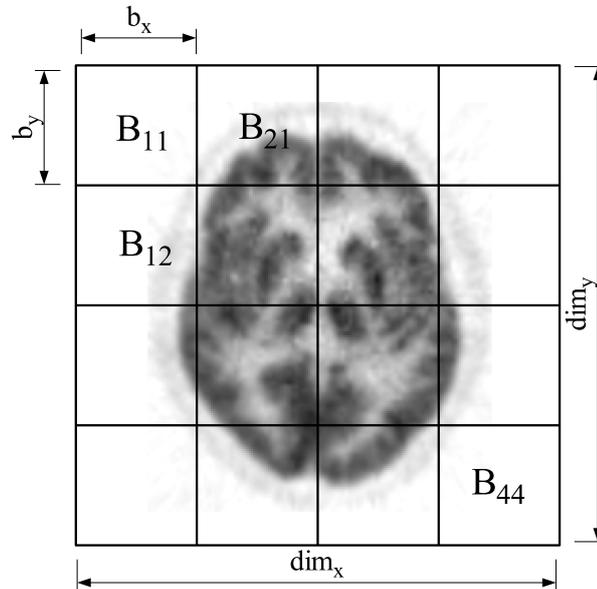


Figure 5.2: Subdivision of an image into uniform blocks (2D example).

5.3.4 Free-form Deformation

Let \mathcal{R} be a 3D image with dimensions $dim_x \times dim_y \times dim_z$ voxels that is uniformly divided into $n_x \times n_y \times n_z$ rectangular parallelepipedic equal-sized blocks B_{pqr} where $p, q, r \in \mathbb{N}$, $1 \leq p \leq n_x$, $1 \leq q \leq n_y$, $1 \leq r \leq n_z$ (Figure 5.2). Each block has size $b_x \times b_y \times b_z$ voxels where $b_i = dim_i/n_i$. Suppose without loss of generality that the image dimensions are multiples of block dimensions so that b_x , b_y and b_z are integer numbers. (If this was not the case we would pad the volume with virtual “empty” voxels, see section 5.3.8 for details.) Next, we co-register each block separately to a corresponding area in the image \mathcal{S} . Let t_{pqr} denote the affine transformation computed for the block B_{pqr} . The transformation t_{pqr} maps the block B_{pqr} to an area $t_{pqr}(B_{pqr})$ in the image \mathcal{S} .

We will now describe a convenient and efficient way to create a smooth deformation from the affine transformations computed for each block. The method is based on the Free-form Deformation algorithm (FFD) that was proposed by Sederberg and Parry as a powerful modeling tool for 3D deformable objects (Sederberg and Parry, 1986). The basic idea of FFD is to deform a volume by manipulating a 3D parallelepiped lattice of control points enclosing the volume. The displaced control points determine a deformation function that specifies a new position for each point of the enclosed volume (see Figure 5.3). The deformation function is thus fully specified by the *initial lattice* of control points, by the *transformed lattice* of displaced

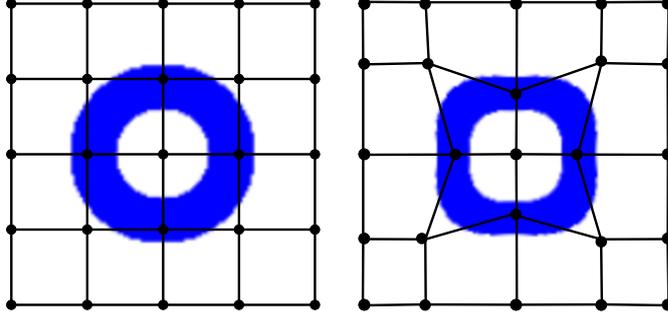


Figure 5.3: Free-form deformation of a volume: initial lattice of control points (left) and lattice of manipulated control points defining a deformed state (right).

control points and by an interpolation function that we will describe later.

5.3.5 Lattices of control points

The original FFD algorithm works with lattices of *points* but the registration algorithm described here uses *blocks*. We must therefore translate the blocks and their corresponding affine transformations to some suitable lattices.

Control points of the *initial lattice* Φ overlaid over the image \mathcal{R} can be defined in a straightforward way: They correspond to the vertices of blocks (Figure 5.4). Each control point is thus a shared vertex of at least one and at most eight blocks and each block B_{pqr} contains eight control points $\phi_{(p+l)(q+m)(r+n)}$, $(l, m, n) \in \{0, 1\}^3$.

Computation of the *transformed lattice* Φ' is illustrated in Figure 5.4. For each $(p, q, r) \in \{0, \dots, n_x\} \times \{0, \dots, n_y\} \times \{0, \dots, n_z\}$ there is an affine transformation t_{pqr} that maps blocks B_{pqr} of the image \mathcal{R} onto block $B'_{pqr} = t_{pqr}(B_{pqr})$ in the image \mathcal{S} . A control point ϕ_{pqr} that was initially shared by the ambient blocks $B_{(p-1)(q-1)(r-1)}, \dots, B_{pqr}$ is transformed to distinct locations $t_{(p-1)(q-1)(r-1)}(\phi_{pqr}), \dots, t_{pqr}(\phi_{pqr})$. We will define its resulting transformed position ϕ' as the average of these positions:

$$\phi'_{pqr} = \frac{1}{8} \sum_{(l,m,n) \in \{0,1\}^3} t_{(p-l)(q-m)(r-n)}(\phi_{pqr}) \quad (5.1)$$

For a reason that becomes clear in the following text it is necessary to enlarge both lattices by one strip of imaginary blocks and control points on each side. These outer points remain in the same positions in both lattices.

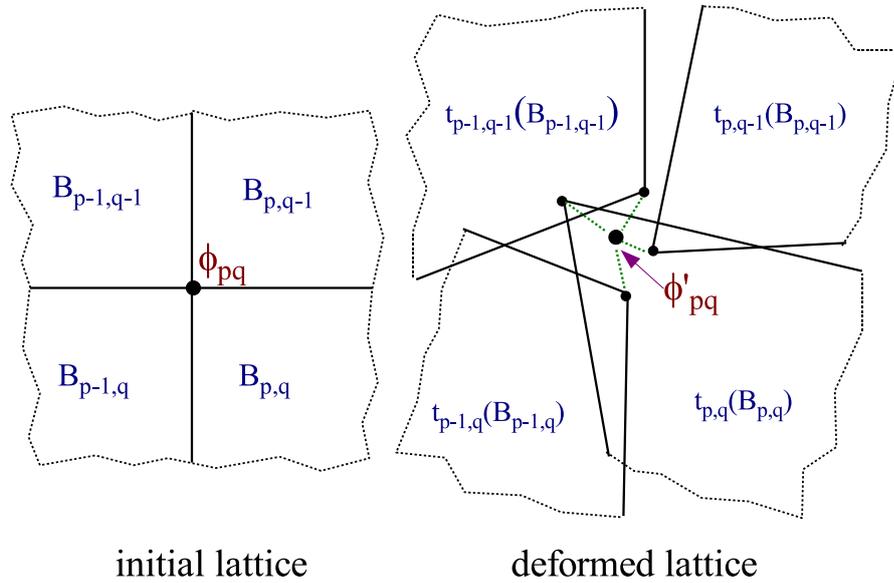


Figure 5.4: Formation of an initial and a deformed control lattice illustrated on a 2D example: Transition from blocks to control points.

5.3.6 B-spline approximation

Having defined the initial and transformed lattice of control points we can now advance to the description of the free form deformations. The transformed lattice of control points determines a deformation function that specifies a new position for each point of the transformed volume. The FFD method does not prescribe a specific type of deformation function and it is therefore possible to choose one that is suitable for the intended application. The original FFD algorithm of [Sederberg and Parry](#) used the Bernstein polynomials that resulted in a global warping function, i.e. new position of each point was computed as a weighted sum of all control points. In this work we use B-splines because they offer local control and guarantee C^2 -continuity ([Sederberg, 2003](#)). We can thus model local C^2 -continuous distortions that only affect the nearest neighbourhood of each block. Deformation of each block is influenced just by the 26 neighbouring blocks which provides computationally a more efficient solution than the Bernstein polynomials. Free form deformations with B-splines were used in a number of applications, for instance for generation of image morphing sequences ([Hsu et al., 1992](#); [Lee et al., 1995](#)) or for scattered data interpolation ([Lee et al., 1997](#)).

Consider now the deformation of voxels in a block B_{pqr} and let $\phi_{pqr} = (\phi_x, \phi_y, \phi_z)$ be a control point of the initial lattice Φ expressed in the voxel coordinates of the image \mathcal{R} . The block is assigned a set of local coordinates $s, t, u \in \mathbb{R}$ such that $0 \leq s, t, u \leq 1$. Each voxel $(i, j, k) \in B_{pqr}$ is expressed

in these local coordinates as

$$(s, t, u) = \left(\frac{i - \phi_x}{b_x}, \frac{j - \phi_y}{b_y}, \frac{k - \phi_z}{b_z} \right) \quad (5.2)$$

The deformation function $\mathbf{w} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ within the block B_{pqr} is then defined as the trivariate cubic B-spline tensor product:

$$\mathbf{w}(i, j, k) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(s) B_m(t) B_n(u) \phi'_{(p+l-1)(q+m-1)(r+n-1)} \quad (5.3)$$

where $\phi'_{...}$ are the 4^3 control points in their displaced positions in the lattice Φ' and $B_l(s)$, $B_m(t)$ and $B_n(u)$ are the uniform cubic B-spline basis functions evaluated at s , t and u , respectively. They are defined as

$$\begin{aligned} B_0(t) &= (-t^3 + 3t^2 - 3t + 1) / 6 \\ B_1(t) &= (3t^3 - 6t^2 + 4) / 6 \\ B_2(t) &= (-3t^3 + 3t^2 + 3t + 1) / 6 \\ B_3(t) &= t^3 / 6 \end{aligned}$$

Since a B-spline curve through collinear control points is itself linear, the initial configuration of control points generates the undeformed volume:

$$\mathbf{w}_{id}(i, j, k) = (i, j, k) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(s) B_m(t) B_n(u) \phi_{(p+l)(q+m)(r+n)}$$

5.3.7 One-to-one mapping

An important point in the list of desired properties of the developed registration method was the requirement to *generate a one-to-one mapping* between the target and source image. One-to-one (injective) mapping is essential for obtaining physically correct results. If a mapping function is not injective, the resulting deformed image may contain undesirable artefacts because parts of the original image get folded upon nearby parts.

Injectivity is a requirement that most of the widely used tools for non-linear registration of PET and MR images do not guarantee (e.g. SPM or AIR). Some tools partially solve the problem by using linear regularization which reduces the probability of occurrence of non-physical deformations. Although widely used, linear regularization alone does not guarantee injectivity. It is nevertheless fair to say that the concept can significantly reduce the probability of occurrence of non-injective deformations.

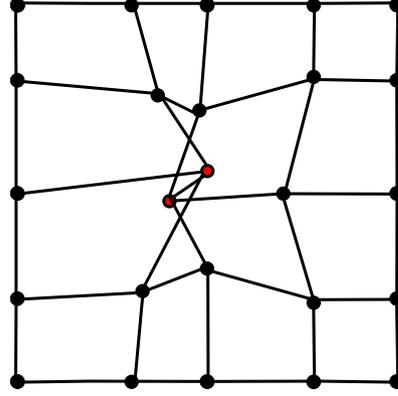


Figure 5.5: Self-intersection in a lattice of control points.

The injectivity of a warping function \mathbf{w} means that no self-intersection is introduced to the deformed volume. Self-intersection implies that at least two points in the initial volume map to a single point in deformed volume. The injectivity of the warping function \mathbf{w} can be violated in two ways. First, a *global violation* of injectivity occurs when the lattice of control points Φ contains a self-intersection (Figure 5.5). Second, the injectivity may also be violated *locally* within a block even if the lattice is not self-intersecting. An example of this case was presented in Lee et al. (1996).

Consider first the local injectivity in a single block B_{pqr} of the 3D image lattice. The deformation function \mathbf{w} within this block is influenced by the 16 control points $\phi_{(p+l)(q+m)(r+n)}$, $(l, m, n) \in \{-1, 0, 1, 2\}^3$ (and only by these). For a better readability, these points will be for the moment referred only by the indices (l, m, n) relative to the current block B_{pqr} as ϕ_{lmn} , $(l, m, n) \in \{-1, 0, 1, 2\}^3$. Let $\Delta\phi_{lmn} = \phi'_{lmn} - \phi_{lmn} = (\Delta x_{lmn}, \Delta y_{lmn}, \Delta z_{lmn})$ be the displacement of each of the 16 control points. Let $\delta_x = 1/b_x \cdot \max\{|\Delta x_{lmn}|\}$, $\delta_y = 1/b_y \cdot \max\{|\Delta y_{lmn}|\}$ and $\delta_z = 1/b_z \cdot \max\{|\Delta z_{lmn}|\}$. The factors $\delta_x, \delta_y, \delta_z$ thus express maximum displacements in units of the block size. Further let $\delta_{max} = \max\{\delta_x, \delta_y, \delta_z\}$.

Lee et al. presented a sufficient condition for the local injectivity of a 2D free-form deformation with uniform cubic B-splines (Lee et al., 1996). Choi and Lee generalized the results to 3D and derived less strict conditions for both 2D and 3D B-spline FFDs (Choi and Lee, 1999, 2000). The sufficient condition for the 3D case is summarized in the following theorem:

Theorem 1. *The deformation function \mathbf{w} given by equation 5.3 is locally injective if $\delta_{max} < D$.*

D is a constant that provides a sufficient (not a necessary) condition for the bound of the control point displacements. Its value is approximately

$D = 0.40331$. Note that theorem 1 provides a tight bound, i.e. there exists a lattice configuration that does not generate an injective function when $\delta_x = \delta_y = \delta_z = D$. For details and proof see Choi and Lee (1999, 2000).

In sum, to guarantee a local one-to-one mapping within one block it is sufficient to check during registration of each block that none of the eight vertices of the block in any of the x-, y- and z-directions moves more than D -times of the initial spacing between the control points. The check needs to be made in every evaluation of the similarity measure. The cost of the check is negligible in comparison with the time spent on evaluation of the similarity measure and it is thus very efficient.

Recall from section 5.3.5 that the resulting displacement of each control point ϕ'_{pqr} in the lattice Φ' is computed as the average of the eight transformed points $t_{(p-1)(q-1)(r-1)}(\phi_{pqr}), \dots, t_{pqr}(\phi_{pqr})$. The displacement of every one of these points is limited by the constant D , hence the average of these displacements fulfils the injective condition as well.

Global injectivity can be guaranteed by making sure that the manipulated lattice Φ' does not intersect itself in any place. This is secured if all control point displacements are less than half of the initial distance between control points in each direction. Since the theorem 1 limits displacements to approximately 0.4 times of the distance this global condition is also fulfilled.

5.3.8 Hierarchical subdivisions

In the previous sections, building blocks of a non-linear registration procedure were presented that consists of three steps:

1. Target volume \mathcal{R} is uniformly divided into rectangular blocks. Vertices of blocks compose the initial control lattice Φ .
2. Separately for each block, an affine transformation is computed that maps the block to a corresponding area in the source image \mathcal{S} , by means of optimization of a similarity measure. Transformation of each block is limited by the condition given in the Theorem 1.
3. The displaced control lattice Φ' is computed from the transformed blocks.
4. The image \mathcal{S} is deformed using the Free-Form Deformation technique with B-spline weighting functions.

The choice of fineness of subdivisions (i.e. block size) in step 1 has a great impact on the quality of registration. The smaller the blocks are, the more locally and precisely it is possible to align brain structures. Registration of

small blocks is, however, less robust due to a reduced number of used voxels and thus there is a higher risk of incorrect and unnatural transformations. On the other hand, large blocks do not allow sufficiently fine deformations. Moreover, the limit on block displacement imposed by Theorem 1 is likely to prohibit sufficiently fine alignment in any case.

To overcome these drawbacks, a multi-level hierarchical optimization technique has been designed and used as an integral part of the developed HBM algorithm. Co-registration of images \mathcal{R} and \mathcal{S} is performed in successive steps using smaller and smaller blocks which generate a hierarchy of control lattices, $\Phi_0, \Phi_1, \dots, \Phi_n$ and their corresponding displaced lattices, $\Phi'_0, \Phi'_1, \dots, \Phi'_n$. They determine a sequence of free-form deformation functions $\mathbf{w}_0, \mathbf{w}_1, \dots, \mathbf{w}_n$. Each step builds on results of the previous steps and improves the deformation on a more local level. The resulting deformation function is defined by the composite function $\mathbf{w} = \mathbf{w}_n \circ \mathbf{w}_{n-1} \circ \dots \circ \mathbf{w}_0$.

In the initial affine-only step 0, the volume of image \mathcal{R} consists of a single block B_0 with block size $\mathbf{b}_0 = (\dim_x, \dim_y, \dim_z)$. The initial lattice Φ_0 corresponding to B has control point spacing \mathbf{b}_0 . The affine transformation t_0 for this block is computed by co-registering the images \mathcal{R} and \mathcal{S} and it determines the displaced lattice Φ'_0 . Lattices Φ_0 and Φ'_0 together define the deformation function \mathbf{w}_0 .

In the next step, the block B_0 is subdivided into eight blocks with size $\mathbf{b}_1 = \mathbf{b}_0/2$ that compose the initial lattice Φ_1 . For each block, a corresponding affine transformation is determined by co-registering it to the transformed volume $\mathbf{w}_0(\mathcal{S})$. The displaced lattice Φ'_1 is computed from the transformations of all blocks and defines the deformation function \mathbf{w}_1 .

In the step k , the volume of \mathcal{R} is divided into 8^k blocks of size $\mathbf{b}_k = \mathbf{b}_0/2^k$ that constitute the initial lattice Φ_k . Each block is co-registered to the deformed intermediate image $\mathbf{w}_{k-1} \circ \dots \circ \mathbf{w}_0(\mathcal{S})$ and the displaced lattice Φ'_k is computed, defining the deformation function \mathbf{w}_k .

Subdivisions terminate at a level n when the block size drops below some selected minimum size \mathbf{b}_{min} . The resulting deformation function mapping \mathcal{R} to \mathcal{S} is given by $\mathbf{w} = \mathbf{w}_n \circ \dots \circ \mathbf{w}_0$.

Parameters of the hierarchical schema

Suitable parameters for the hierarchical schema (initial block size, minimum block size, number of levels) can be selected in various ways. The most important parameter is certainly the minimum block size that determines how fine registration it is reasonable to attempt regarding the quality and voxel resolution of the input images. It seems therefore convenient to let the minimum block size be the one input parameter and compute the remaining parameters automatically in the following way:

Let $\mathbf{b}_{min} = (b_{x,min}, b_{y,min}, b_{z,min})$ be the block size at the last and finest level. Then the size $\mathbf{b}_0 = (b_{x,0}, b_{y,0}, b_{z,0})$ of the block B_0 at the first level is defined as the tightest bounding block that contains the whole volume of \mathcal{R} and whose dimension in each x -, y - and z -axis is power of the minimum block size:

$$b_{x,0} = (b_{x,min}) \cdot 2^{l_x} \text{ where } l_x = \min \{ l \mid (b_{x,min}) \cdot 2^l > dim_x \} \quad (5.4)$$

The values l_x, l_y, l_z represent the number of subdivisions required in each axis. Let $l_m = \min \{l_x, l_y, l_z\}$. After the initial affine registration step, the block B_0 is subdivided into regular blocks with size

$$\mathbf{b}_1 = (b_{x,1}, b_{y,1}, b_{z,1}) = \left(\frac{b_{x,0}}{2^{l_x-l_m+1}}, \frac{b_{y,0}}{2^{l_y-l_m+1}}, \frac{b_{z,0}}{2^{l_z-l_m+1}} \right)$$

The first subdivision step thus equalizes the different numbers of required levels in each axis. Subsequently in every following step k , the block B_0 is subdivided uniformly in each axis into blocks with size $\mathbf{b}_k = \mathbf{b}_{k-1}/2$.

In this way, we can assure that the initial control lattices are uniform at each level which is one of the presumptions of the Free-Form Deformation technique. If the block B_0 exceeds the size of the volume \mathcal{R} then the non-overlapping parts of the block will be ignored during registration.

Computation of intermediate images

At the end of each subdivision level, a new transformed image is computed that is then used as the source image in the next level. Let \mathcal{S}_k denote the intermediate source image for level k . \mathcal{S}_k is computed as

$$\mathcal{S}_k = \mathbf{w}_{k-1} \circ \dots \circ \mathbf{w}_0(\mathcal{S}) \quad (5.5)$$

where \mathcal{S} is the original source image. A very efficient way of computing \mathcal{S}_k would be to transform it from the previous source \mathcal{S}_{k-1} , i.e. $\mathcal{S}_k = \mathbf{w}_{k-1}(\mathcal{S}_{k-1})$. This approach would, however, cause that every subsequent intermediate image would be more deteriorated since every image transformation involves interpolation. This would certainly negatively influence precision and robustness at the finest levels where a good-quality source image is desirable. Therefore, a less efficient way was chosen where every intermediate image is always computed only from the original source image and is subject to one interpolation only.

Each voxel \mathbf{v} of the volume \mathcal{S}_k is transformed to a point \mathbf{v}' in \mathcal{S} via backward mapping as $\mathbf{v}' = \mathbf{w}_0^{-1} \circ \dots \circ \mathbf{w}_{k-1}^{-1}(\mathbf{v})$ by subsequently using the lattices $\Phi_0, \Phi'_0, \dots, \Phi_{k-1}, \Phi'_{k-1}$. This transformation involves no interpolation

and is precise. The value of \mathbf{v}' is then interpolated from the neighbouring voxels in \mathcal{S} and assigned back to \mathbf{v} . The drawback is that each level requires one computation of deformation more, for each voxel, than the previous level. In other words, the time complexity increases linearly with the number of levels. This is however no big problem for the number of levels considered here (i.e. 4-6).

The computational demand also depends on the dimensions of intermediate source images. The HBM algorithm prescribes no particular dimensions but too small dimensions would obviously decrease precision of registrations in blocks. By default, the dimensions of the source image are used.

Multi-resolution optimization

The initial affine level uses a multi-resolution optimization schema as described in the section 4.3. Design of a suitable schema depends on the size and quality of input images. As demonstrated in section 4.10, omitting the full-resolution level only marginally influences precision but significantly decreases registration time. The same seems to hold for the inter-subject affine registration (see section 5.4). The HBM implementation used here for validation tests thus omits the full-resolution level and uses a two-level schema with scale factors [4, 4, 4] and [2, 2, 2].

Subsequent subdivision levels do not use multi-resolution schema and perform a full-resolution registration only. This is because the number of available voxels is much lower and a low-resolution registration would not be robust enough.

5.3.9 Linear regularization

Decreasing size of blocks on one side allows more local matching but also decreases robustness of optimization. The threshold on maximum extent of transformations presented in the section 5.3.7 guarantees injectivity and has a positive influence on robustness because it prohibits large deformations. Robustness is also significantly improved by the hierarchical schema presented in the previous section. Still, unnecessary large transformations may occur in some cases, although within limits of injectivity. For example, the optimization method might find maximum of similarity for some larger transformation, although there exists another, much smaller and better transformation that gives only slightly worse similarity. In such case it is very probable that the higher similarity for the large transformation is caused by artefacts, noise or slightly different structure in the registered images.

A common approach to avoiding such situation is the concept of linear regularization (LR). It is formulated by extending the similarity measure SM

by a second term that penalizes large deformations:

$$SM_{LR} = SM - \lambda \cdot DF \quad (5.6)$$

DF is a non-negative penalization function which equals to zero in case of no deformation. Thus, minimization of SM now becomes a task of finding a deformation that simultaneously maximizes similarity of images and minimizes the necessary deformation. In this way we can partially avoid deformations that would improve the similarity measure by a tiny amount but that would cause large and unnatural deformations. The extent of penalization is driven by the parameter λ . Linear regularization has been commonly used by many authors for non-linear registration of images (e.g. [Ashburner and Friston, 1999](#); [Hajnal et al., 2001](#); [Hellier et al., 2001](#)). The penalization function is usually *linear elastic energy*, *membrane energy* or *bending energy* ([Ashburner, 2000](#)).

Although commonly applied in renown registration packages, e.g., SPM ([Ashburner and Friston, 1999](#)), linear regularization is a debatable solution. The parameter λ joins together two quantities with different physical dimensions: an intensity measure (SM) and a geometric measure (DF). As noted in [Cachier and Ayache \(2001\)](#), the intensity similarity is related to the amount of change in intensity that must be applied to go from one image to the other, which is not uniformly proportional to the amount of motion necessary to deform one image to the other. Consequently, effect of regularization largely depends on the local variation of the similarity measure and there is no clear way of determining a suitable value for the parameter λ .

Therefore, we will make here an attempt to relate the two physically different quantities, SM and DF, using the specific properties of the HBM method. The similarity measure used in HBM is mostly the correlation coefficient with values in the range $\langle 0, 1 \rangle$, the maximum extent of transformation is limited by the condition $\delta_{max} < D$ from [Theorem 1](#). The proposed expression for a regularized similarity measure is thus:

$$SM_{LR} = CC - \left(\frac{\delta_{max}}{D} \right)^h \quad (5.7)$$

The second term is a geometric penalization function. Its values lie in the range $\langle 0, 1 \rangle$ corresponding to the range of the correlation coefficient in the first term. The exponent h determines the strength of penalization. Lower values of h cause strong penalization even for small registrations. Suitable experimentally found values are $h = 3, 4$. The advantage of this penalization function, besides its convenient range of values, is the efficiency of its computation, compared to the energy-based functions that require computation of at least first partial derivations of the similarity measure function.

In summary, the HBM method uses Theorem 1 to limit the maximum extent of transformations whereas the proposed linear regularization regulates transformations below this limit.

5.3.10 Preprocessing

Preprocessing of input images is very important in a non-linear registration. The described procedure optimizes ten thousands of parameters and is thus naturally more prone to local misregistrations and faulty estimation of some of the parameters than the global affine registration with just 12 parameters. It is therefore crucial to supply input images of a reasonable quality. Typical PET images unfortunately do not have sufficiently good quality and need to be processed prior to entering the registration procedure.

Preprocessing can also significantly improve registration speed. There are basically two potential sources of acceleration:

- **Reduction of volume:** Removing of irrelevant image areas from computation.
- **Faster convergence of optimization:** By improving image quality it is possible to acquire a smoother and more conveniently shaped similarity function and thus a faster convergence of optimization (less iterations).

The HBM method uses two kinds of preprocessing: background removal and reduction of noise.

Background removal: Prior to registration, background is automatically segmented from images using the method described in section 4.5.1.

Noise reduction: Median filtering is used to reduce random small reconstruction artefacts emerging as outliers. After that an isotropic 3D Gaussian filtering is applied to smooth the image and reduce the additive noise.

5.3.11 Summary of the algorithm

Let $register(\mathcal{R}, \mathcal{S}, B)$ be a function that co-registers images \mathcal{R} , \mathcal{S} within a block B of the image \mathcal{R} so that the resulting transformation fulfils the condition given by the Theorem 1. Let β_l denote the set of all blocks at a level l . The Algorithm 5 outlines the hierarchical registration scheme of the HBM algorithm.

Algorithm 5: Hierarchical Block Matching

Input: preprocessed images \mathcal{R} , \mathcal{S} , minimum block size \mathbf{b}_{min}
Output: deformation function \mathbf{w} that matches \mathcal{R} and \mathcal{S}
transformed image $\mathcal{S}_{\mathbf{w}}$

$l_x, l_y, l_z \leftarrow$ no. of subdivision levels in each axis (equation 5.4)
 $l_m \leftarrow \min \{l_x, l_y, l_z\}$

// block size
 $\mathbf{b} = (b_x, b_y, b_z) \leftarrow (b_{x,min}^{l_x}, b_{y,min}^{l_y}, b_{z,min}^{l_z})$

// initial intermediate image
 $\mathcal{S}_0 \leftarrow \mathcal{S}$

// loop over subdivision levels
for $l \leftarrow 0 \dots l_m$ **do**
 compose initial lattice Φ_l
 foreach block $B \in \beta_l$ **do**
 // local affine registration
 $t_B \leftarrow register(\mathcal{R}, \mathcal{S}_l, B)$
 end
 compose displaced lattice Φ'_l from all $t_B, B \in \beta_l$
 deformation \mathbf{w}_l is determined by Φ_l, Φ'_l
 $\mathcal{S}_{l+1} \leftarrow \mathbf{w}_l \circ \dots \circ \mathbf{w}_0(\mathcal{S})$
 if $l = 0$ **then**
 $\mathbf{b} \leftarrow (b_x/2^{l_x-l_m+1}, b_y/2^{l_y-l_m+1}, b_z/2^{l_z-l_m+1})$
 else
 $\mathbf{b} \leftarrow \mathbf{b}/2$
 end
end

// Results
 $\mathbf{w} \leftarrow \mathbf{w}_{l_m} \circ \dots \circ \mathbf{w}_0(\mathcal{S})$
 $\mathcal{S}_{\mathbf{w}} \leftarrow \mathcal{S}_{l_m+1}$

5.3.12 Complexity of the algorithm

The time complexity of an affine registration of two images is linearly dependent on the number of voxels N in the target image (see section 3.7). When the target image is subdivided into B blocks then the affine optimization of each blocks requires $O(N/B)$ time and the overall complexity for all B blocks remains approximately $O(N)$. The time necessary for composition of the initial and the displaced lattices is negligible.

The time for computation of the intermediate images \mathcal{S}_l increases with every new subdivision level. Computation of the image \mathcal{S}_l at a level $l - 1$ requires, for each voxel $v \in \mathcal{S}_l$, to pass through transformation lattices of levels $0 \dots l - 1$ and compute, at every level i , a B-spline interpolation of the coordinates of v using lattices Φ_i, Φ'_i . If M is the number of voxels in \mathcal{S}_l then the computation of \mathcal{S}_l has complexity $O(M \cdot l)$.

HBM with L levels has thus time complexity $O(L \cdot N) + O(L^2 \cdot M)$. The first term is the time required for optimizations and the second term corresponds to the time needed for computation of intermediate images. The quadratic nature of the second term is not a problem for the common dimensions of PET images (approx. 10^6 voxels) and for the corresponding maximal depth of subdivisions (4-5). Such registration takes about 4 min on a 1 GHz PC. It has, however, more impact when more subdivision levels are required, e.g. for high resolution HRRT-PET images. Section 5.4.3 brings more details about registration HRRT images.

The registration time can be reduced if we use smaller dimensions for the intermediate images (smaller M). This would, however, deteriorate precision of the optimization step.

5.3.13 Spatial normalization

The term spatial normalization is commonly used for transformation of an image or a set of images to some standard space, i.e. to a template brain image with a standard size and shape. Spatial normalization allows a direct comparison of corresponding structures in examined images.

The FDG16 and FMZ17 templates

For the purpose of spatial normalization with the HBM algorithm, templates for FDG-PET and FMZ-PET images were created. The FDG-PET template was created from FDG-PET images of a group of 16 healthy people (thereof 6 male). The PET images were acquired on ECAT Exact HR scanner (Wienhard et al., 1994) from CPS Innovations, Inc., Knoxville, Tennessee, USA. The mean age in the group was 59.6 ± 3.8 , median 59.5, minimum 57 and maximum 65 years. The images were first spatially normalized using the software

package SPM2 ([Ashburner and Friston \(1999\)](http://www.fil.ion.ucl.ac.uk/spm), www.fil.ion.ucl.ac.uk/spm) to the standard PET template supplied with the software. According to the information from the authors, this PET template corresponds to the space of the ICBM152 template defined by the International Consortium for Brain Mapping. The intensity in the normalized images was scaled so that the mean intensity in the brain area in each image was unity and then a mean image of these images was computed. We will refer to this template as the FDG16 template. The same procedure was applied to the group of FMZ-PET control images of 17 healthy people (thereof 10 male) and resulted in the FMZ17 template. The mean age of this group was 50.5 ± 19.6 , median 61, minimum 22, maximum 74 years and all images were acquired on the ECAT Exact HR scanner. Both templates had dimensions 91 x 109 x 91 with an isotropic voxel size of 2 mm. To fasten the spatial normalization process, binary masks corresponding to the head volume were created for both templates and used for removal of background.

5.4 Evaluation

The difficulty in evaluation of algorithms for non-linear registration is in defining a criterion for a well registered image. For real images of human brain a definition of optimal result does not exist, not even theoretically. This is simply because the variability of human brain at a local level is high and even the topology of two healthy brain images does not exactly correspond (in particular the topology of gyri and sulci). Alignment of structures can thus be achieved only to a certain level. Moreover, it is not clear what structures are desirable to be in alignment - this depends on the intended application in particular. Visual verification of results by an experienced person is therefore always necessary. It means that in the end the quality of a normalization algorithm is evaluated by experience acquired in routine use. The following sections present several validation tests of the HBM method.

5.4.1 Binary geometric images

The basic behaviour of the HBM algorithm was evaluated on several artificial binary images with various geometrical objects. The aim was to demonstrate that the proposed procedure is able to closely match images even though their shape does not allow a precise match with a C^2 -continuous deformation function. Three pairs of images were created and in each pair one image was registered to the other using HBM. In Figure 5.6, the first column contains the target images, the second column the source images and the aim was to automatically deform each source image to match the target. The images had dimensions 128 x 128 x 128 voxels. The minimum block size was set to 2 x 2 x 2 voxels resulting in 1 affine and 6 non-linear levels.

Registration quality was estimated by the ratio of mismatched voxels at each level. The ratio was computed as the number of non-overlapping voxels of the two objects divided by the total number of voxels in the target object. The results are presented in Table 5.1. In all cases the HBM method substantially decreased the initial misalignment. The affine level (SL0) accounted mainly for global differences in size. For the first pair (cube+sphere) the number of mismatched voxels was halved at the affine level whereas the improvement yielded by the other pairs was smaller. This corresponds to the differing volume ratios in each pair. The first non-linear level (SL1) did not achieve any improvement worth mentioning, in contrast to the last three levels that were able to halve the number of mismatched voxels from previous levels. This indicates that the first non-linear level could be omitted in order to decrease registration time.

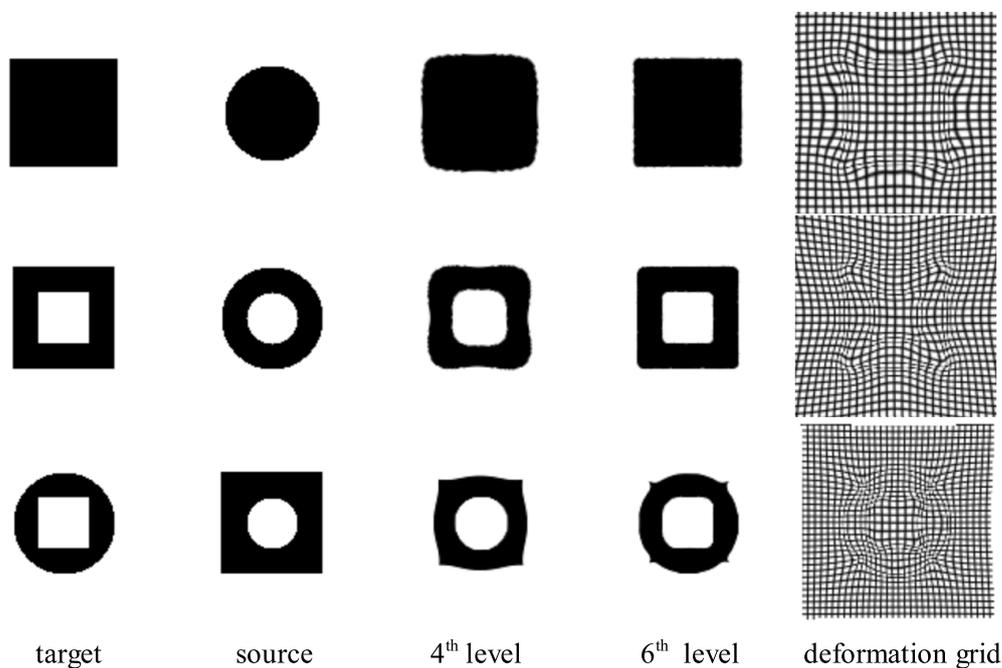


Figure 5.6: Non-linear registration of artificial binary images. Source images (second column) are matched to target images (first column). The resulting images after 6 subdivision levels are in the fourth column. The last column shows a regular grid deformed using the resulting transformation.

target image	initial position	HBM - subdivision levels					
		SL0	SL1	SL2	SL3	SL4	SL5
cube - full	64.1	30.9	30.5	25.9	15.0	4.5	0.9
cube - hollow	44.8	29.0	28.3	21.2	12.0	5.7	3.0
sphere - hollow	44.8	36.6	31.1	24.0	15.7	9.0	5.7

Table 5.1: The percentage of mismatched voxels for the initial position of registered objects and after each subdivision level. The percentage is computed as the number of non-overlapping voxels divided by the number of voxels in the target volume multiplied by 100.

5.4.2 Normalization of FDG-PET and FMZ-PET images

This section evaluates the quality of alignment achieved by the hierarchical subdivision schema of the HBM algorithm on real medical images. For this purpose, a group of 16 normal FDG-PET images (the same as the group in section 5.3.13) was spatially normalized to the FDG16 template. The images had dimensions 128 x 128 x 47 with voxel size 2.2 x 2.2 x 3.125 mm. The template had dimensions 91 x 109 x 91 with an isotropic voxel size of 2 mm.

The minimum block size was set to 4 x 4 x 4 voxels which resulted in 6 subdivision levels (SL) - 1 affine and 5 non-linear. Regularization factor 4 was used for non-linear registrations. According to the hierarchical schema, at the finest level each image was subdivided into $32^3 = 32\,768$ blocks. Therefrom approximately one third was actually used, the remaining blocks contained just background or non-brain parts. For each block, 12 affine transformation parameters were computed. The deformation on the finest level was thus modelled by approximately $1.3 \cdot 10^5$ parameters.

Experiments showed that the first SL (subdivision into 8 cubes) brought only a tiny decrease of the standard deviation and the resulting image after this level did not differ much from the affine result. The first SL was consequently skipped in all tests to reduce registration time. Note that this observation is also in agreement with the results of the previous section 5.4.1.

During every registration, intermediate images were stored after each subdivision level. Finally, the mean image and image of standard deviations for the group were computed from the resulting images at each level. The image intensities were first scaled so that the mean intensity within the brain volume was unity, where the brain volume was defined by the same binary mask as the one used for masking of the template (section 5.3.13). The mean and standard deviation images for each level were then computed voxel by voxel across all intermediate images at that level and smoothed with an isotropic gauss filter with FWHM 4 mm. The same procedure was also applied to a group of 17 normal FMZ-PET images (the same as the images in section 5.3.13) that were normalized to the space of the FMZ17 template.

The HBM procedure does not optimize at the full-resolution in the initial affine level (section 5.3.8). It would be therefore inappropriate to make comparisons just with the intermediate result of the initial affine step. Therefore, the control images were also co-registered using an affine-only procedure that corresponds to the initial affine step of HBM with an additional full-resolution optimization. We will refer to this procedure as AFF.

Figure 5.7 presents the images of standard deviations in an axial slice at selected subdivision levels for both FDG-PET and FMZ-PET images. The images demonstrate a significant overall decrease of standard deviation (σ), especially apparent in the brain rim, putamen, thalamus, etc.

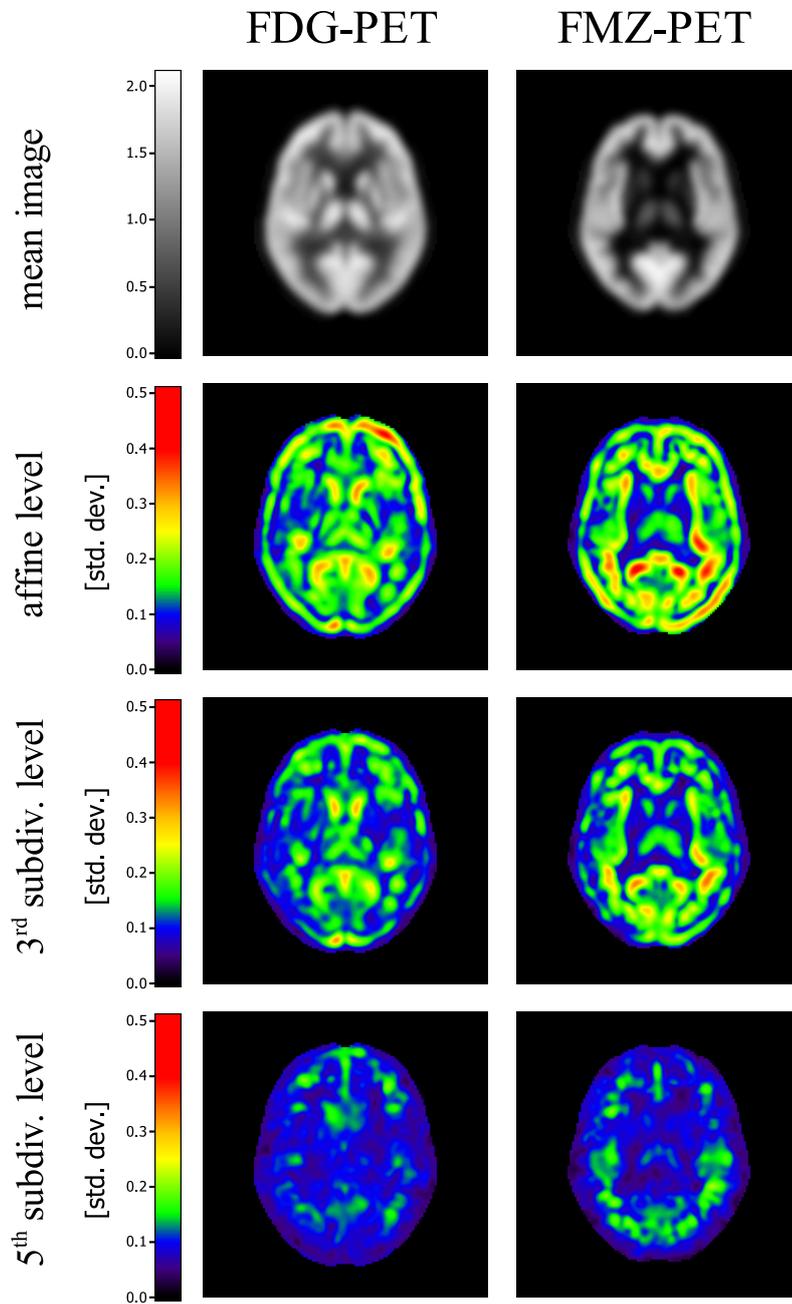


Figure 5.7: Standard deviations of the group of 16 FDG-PET images (left column) and 17 FMZ-PET images (right column) after the affine level and after the 3rd and 5th subdivision level of HBM. The first row contains the mean images of the groups at the 5th subdivision level.

Table 5.2 summarizes, for the FDG group, the average σ within the brain volume at each level. The last column presents the total decrease of σ compared to the results produced by AFF. HBM achieved on average a 38 % smaller standard deviation than AFF. For FMZ images the reduction was similar. Besides the overall decrease of σ it is important to look at smaller brain regions. The average σ values for several selected regions - basal ganglia, brain rim and cerebellum - show that the non-linear registration achieves much better alignment of smaller structures than the affine-only registration.

The graph in Figure 5.8 shows the reduction of standard deviation with the increasing number of subdivision levels. Averaged over the whole brain, each subdivision level reduces σ by about the same amount. Nevertheless, the curves of basal ganglia and cerebellum show that for some areas the largest decrease occurs not until the last two levels which demonstrates the importance of fine subdivision levels.

The average computational times are summarized in Table 5.3. The time required for computations at each subdivision level is split into the time for registration of blocks and the time needed for computation of the intermediate image \mathcal{S}_k for the next subdivision level, resp. for computation of the resulting image after the last level SL5. The first column contains times of AFF, the remaining columns correspond to subdivision levels of HBM. Notice that the time of block registration in AFF is significantly longer than in SL0 of HBM but the average standard deviations (Table 5.2) are nearly the same. This justifies omitting of the full-resolution optimization in the affine level of HBM.

After the last level, the image is warped and resampled into the dimensions of the template image which has about 17 % larger volume than the source image dimensions used for computation of intermediate images (128 x 128 x 47 source dimensions vs. 91 x 109 x 91 template dimensions). Taking this into account for the time of SL5, we can see that the transformation time increases linearly with the number of levels which is in accordance with the discussion in the section 5.3.8. Starting from the second subdivision level, the computational time of block registrations slightly decreases. The reason is that with increasing fineness of subdivisions there are more blocks that fall outside the brain and that are not registered at all. These relationships are graphically presented in Figure 5.9. The total time for all levels, 242 s, plus the time needed for initial preprocessing of images, approx. 8 s, give on the whole slightly more than 4 minutes needed for spatial normalization of one PET image.

Std. dev. [$\times 10^{-1}$]	AFF	HBM subdivision levels					Reduction [%] SL5 vs. AFF
		SL0	SL2	SL3	SL4	SL5	
brain	1.40	1.44	1.29	1.15	1.00	0.87	38 %
bas. gan.	2.80	2.92	2.73	2.54	2.01	1.14	59 %
brain rim	3.44	3.47	2.87	2.38	1.59	0.92	73 %
cereb.	2.67	2.71	2.56	2.41	1.93	1.18	56 %

Table 5.2: Average standard deviations of the FDG group within the brain volume and in selected regions (basal ganglia, brain rim, cerebellum) achieved by AFF and by each subdivision level of HBM. The last column presents the percentual decrease of standard deviation after SL5 compared to AFF.

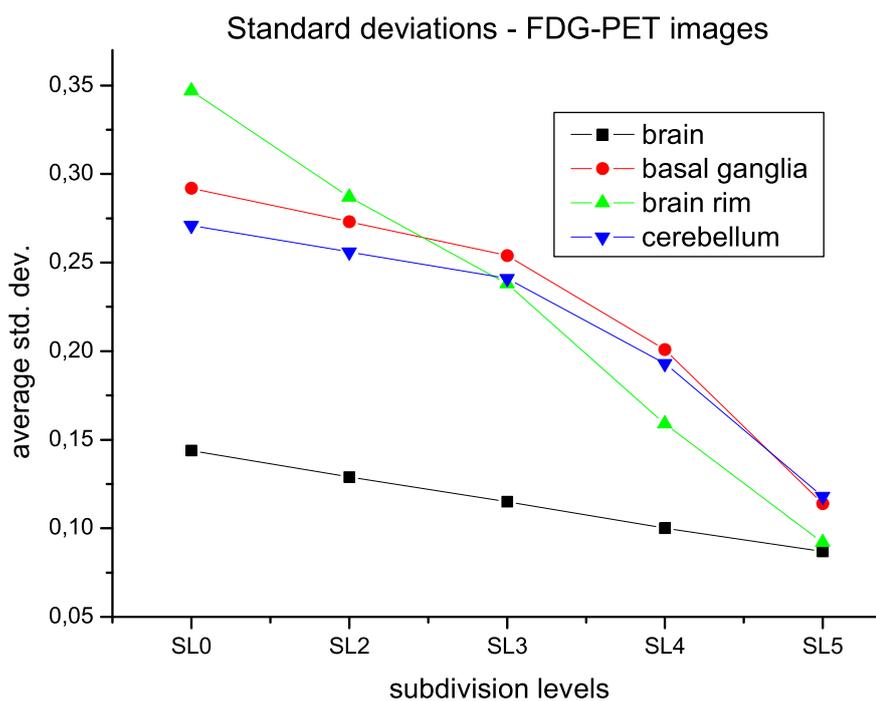


Figure 5.8: Reduction of the average standard deviation of the FDG group with the increasing number of subdivision levels.

Time [s]	AFF	HBM subdivision levels				
		SL0	SL2	SL3	SL4	SL5
optimization in blocks	72	21	48	34	32	35
transformation	2	2	7	12	18	33
time per level	74	23	55	46	50	68
total time	74	23	78	124	174	242

Table 5.3: Average computational time of each subdivision level. The computational time of a single level is the sum of the time needed for co-registration in all blocks plus the time for computation of intermediate images \mathcal{S}_k at levels SL0-SL4, resp. for computation of the resulting image after the 5th level.

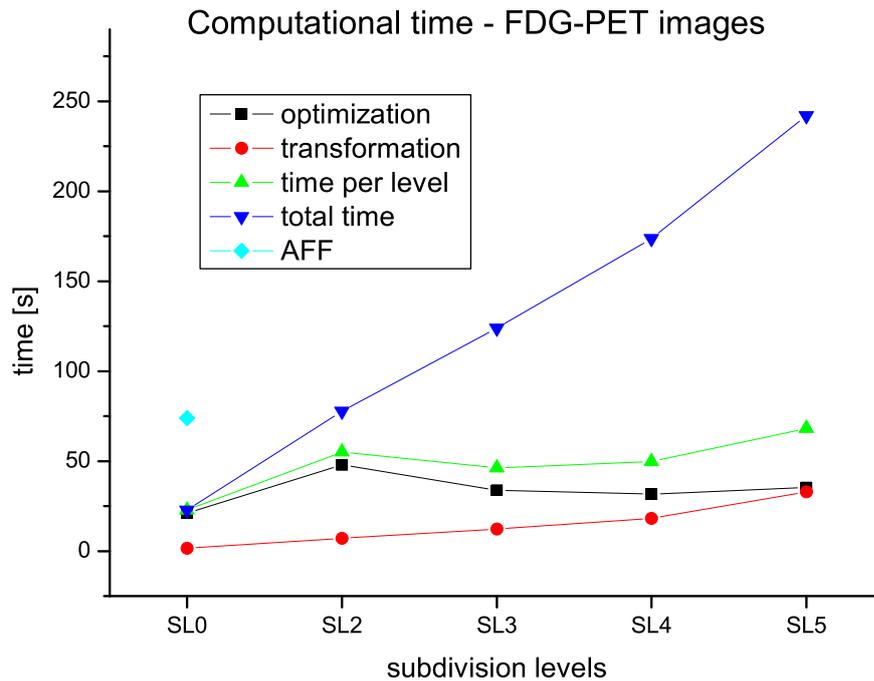


Figure 5.9: The time required for each subdivision level and the total time. The time per level is the sum of the optimization time and the warping time.

5.4.3 Normalization of high-resolution HRRT PET images

The HBM method was also evaluated on high-resolution PET images acquired on the most advanced type of PET scanner nowadays - the High-Resolution Research Tomograph (HRRT) (CPS Innovations, Inc., Knoxville, Tennessee, USA). FDG-PET images of a group of 12 healthy volunteers were used (10 male, 2 female), the mean age was 39.8 ± 13.6 , median 35.1, minimum 25 and maximum 64 years. Each volunteer acquired an injection of 370 MBq FDG and the activity was recorded for one hour. The acquired data was subsequently corrected for attenuation, random and scatter and reconstructed by OSEM3D in 3 iterations with span 3 in an image with dimensions $256 \times 256 \times 207$ voxels and voxel size of $1.22 \times 1.22 \times 1.22$ mm (Wienhard et al., 2002). The figure 5.10 shows one of the images in the group. Compared to the common PET images evaluated in the previous section, HRRT images present a considerable increase of data size - more than 17 fold. It brings along much higher demands on memory space and computational performance.

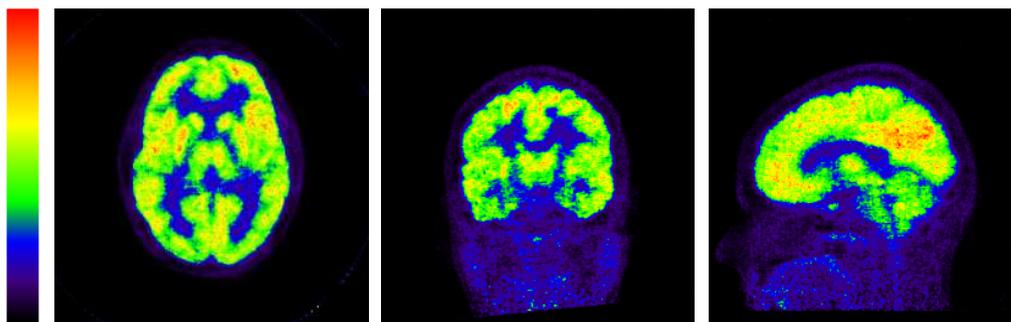


Figure 5.10: An example HRRT FDG-PET image in three orthogonal slices.

As in the previous test, the minimum block size was set to $4 \times 4 \times 4$ voxels. This resulted in 7 subdivision levels - 1 affine and 6 non-linear. Again, regularization factor 4 was used for non-linear registrations and the first non-linear level was omitted. At the finest level each image was subdivided into $64^3 = 262\,144$ blocks, from which approximately one quarter was used. The deformation on the finest level was thus modelled by approximately $4 \cdot 10^7$ parameters. In order to acquire a high-resolution template for the test, the 12 images were first spatially normalized to the FDG16 template, saved in their original dimensions and averaged, resulting in a high-resolution template called HRRT-FDG12.

The 12 HRRT PET images were then normalized to the HRRT-FDG12 template using HBM and AFF. Mean images and images of standard deviations were computed for each subdivision level as in the previous test and smoothed with an isotropic gauss filter with FWHM 4mm. Figure 5.11

presents the images of standard deviations in three orthogonal slices at selected subdivision levels.

Table 5.4 summarizes the average standard deviation σ within the brain volume at each level. The last column presents the total decrease of σ compared to the results produced by AFF. HBM achieved on average 39 % smaller standard deviation than AFF. Similar to the results of the test with FDG-PET images, the improvement in alignment quality is best demonstrated on smaller regions, a few of which are listed in the table.

The graph in Figure 5.12 shows the reduction of standard deviation with increasing number of subdivision levels. The results are very similar to the results of the previous test with common FDG-PET and FMZ-PET images. Note that standard deviation within the region of cerebellum in SL2 and SL3 is even worse than in SL0. This means that there were other brain structures in the optimized block that had had higher impact on the optimization. A fine alignment of cerebellum was not possible until finer subdivision levels.

The average computational times are summarized in Table 5.5 and presented graphically in Figure 5.13. The decrease of optimization time is slightly more pronounced here than in the case of common FDG-PET images but otherwise the time course is very similar in both cases. Due to the larger data size, the HRRT FDG-PET registration takes much longer - about 48 minutes. This is however still feasible for many clinical applications.

5.4.4 Evaluation on clinical images

The evaluations presented in the previous sections demonstrate that the HBM method is able to align a group of normal images so that the standard deviation of voxel intensity in the group is substantially reduced. This aim of the following test was to provide an informal, informative evaluation of the HBM method on clinical images with various types and extent of deterioration.

The method was used to spatially normalize 504 FDG-PET images from a large PET image database, created in cooperation with several European PET centres in the frame of the NEST-DD project. The database contains mostly images of patients with Alzheimer's and Parkinson's disease. The HBM method has also been tested on a set of 38 FMZ-PET images of patients suffering epilepsy, tumour or stroke. The results were evaluated visually. In all cases, the normalization produced correct and visually precise results within healthy tissue. It also gave correct alignment of diseased tissue in patients with Alzheimer's and Parkinson's disease as well as for patients with smaller lesions.

In several images with larger lesions the algorithm showed a tendency to diminish the lesion area. The next section deals with this issue.

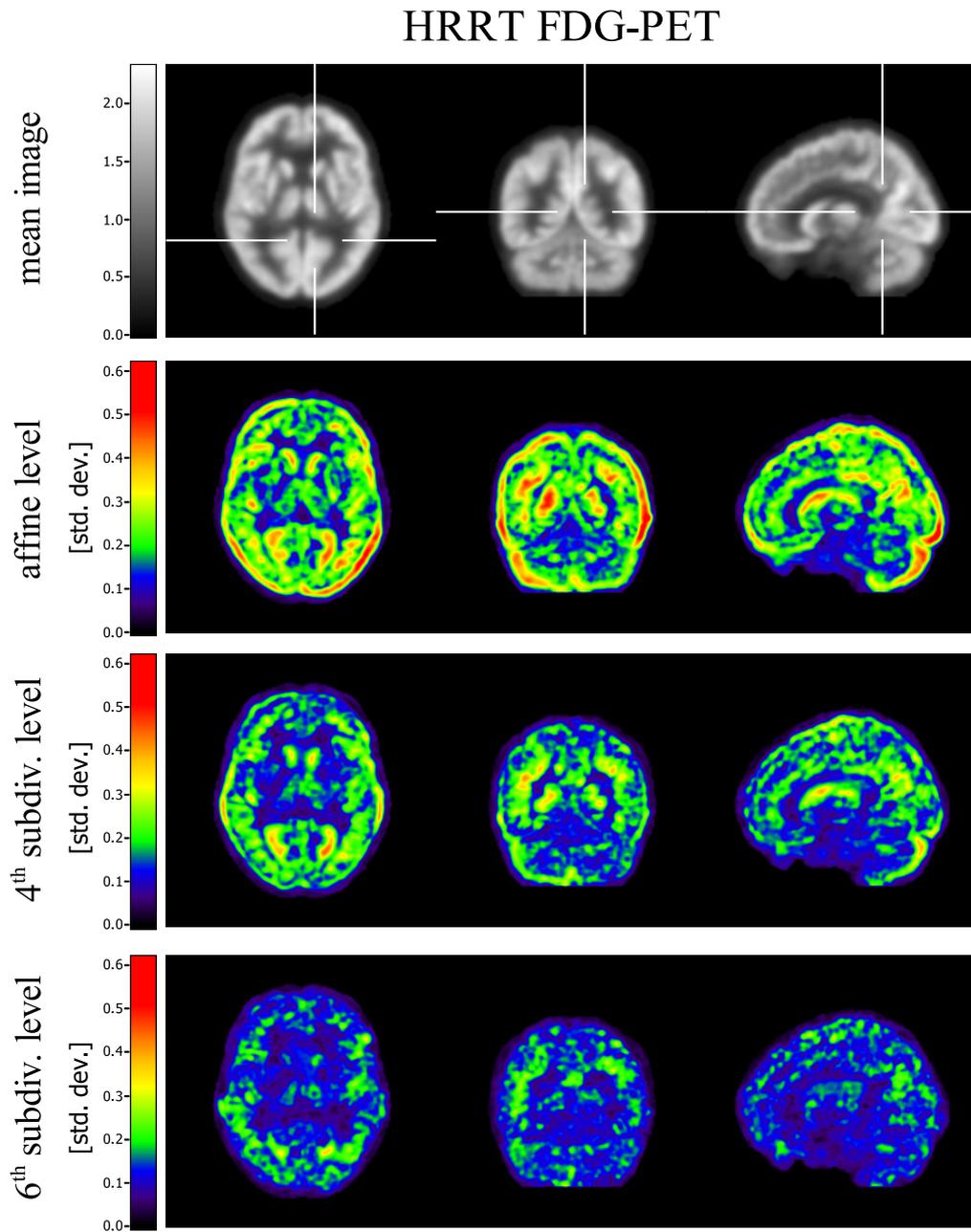


Figure 5.11: Standard deviations of the group of 12 HRRT FDG-PET images in three orthogonal slices after the affine level and after the 4th and 6th subdivision level of HBM. The first row shows the mean image of the group at the 6th subdivision level.

Std. dev. [$\times 10^{-1}$]	AFF	HBM subdivision levels						Reduction [%] SL6 vs. AFF
		SL0	SL2	SL3	SL4	SL5	SL6	
brain	1.40	1.40	1.37	1.21	1.06	0.94	0.86	39 %
bas. gan.	3.76	3.77	3.67	3.63	3.00	1.71	1.10	71 %
brain rim	4.85	4.89	3.77	2.94	2.10	1.59	1.24	74 %
cereb.	3.79	3.78	3.93	3.81	2.99	2.05	1.49	61 %

Table 5.4: Average standard deviations in the brain volume and in selected regions (basal ganglia, brain rim, cerebellum) achieved by AFF and by each subdivision level of HBM. The last column presents the procentual decrease of the average standard deviation after SL6 compared to AFF.

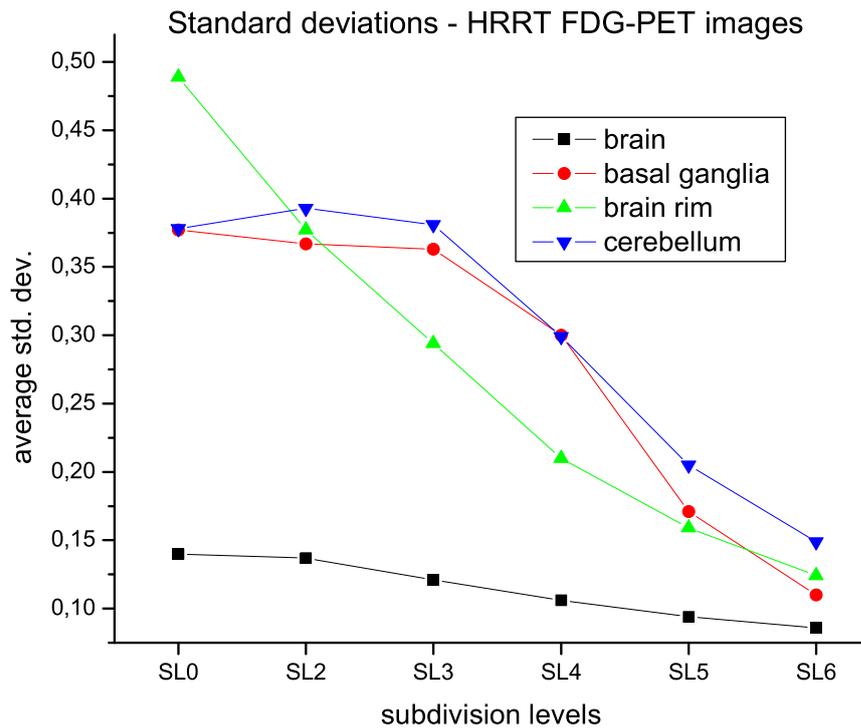


Figure 5.12: Reduction of the average standard deviation with the increasing number of subdivision levels.

Time [s]	AFF	HBM subdivision levels					
		SL0	SL2	SL3	SL4	SL5	SL6
optimization in blocks	423	127	311	267	184	137	146
transformation	28	24	128	235	339	442	548
time per level	451	152	439	502	523	578	694
total time	451	152	591	1093	1615	2194	2888

Table 5.5: Average computational time required at each subdivision level of the HRRT PET spatial normalization. The computational time of a single level is the sum of the time needed for co-registration in all blocks plus the time for computation of intermediate images \mathcal{S}_k at levels SL0-SL5, resp. for computation of the resulting image after the 6th level.

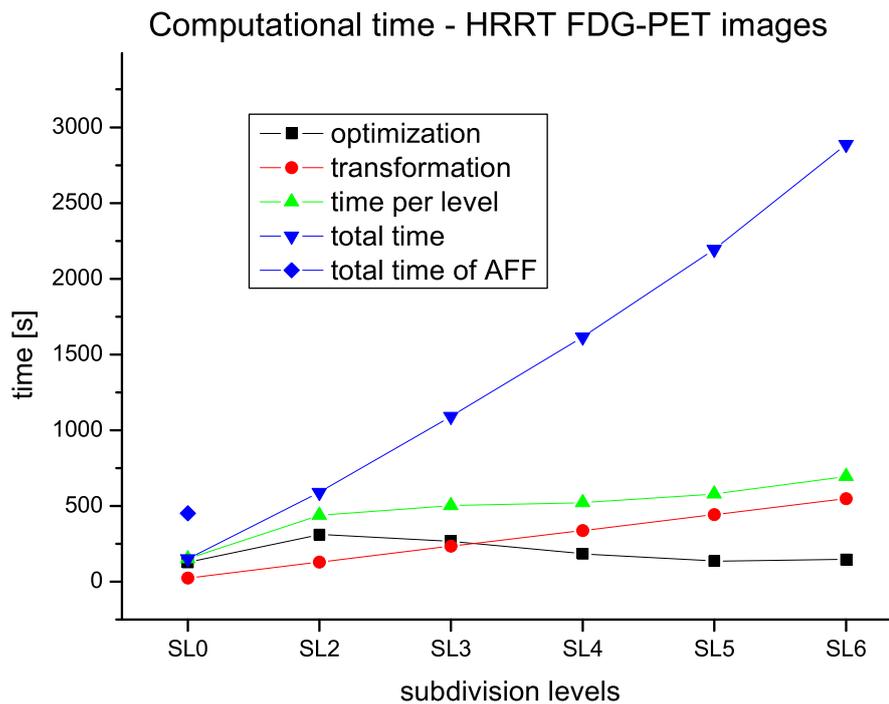


Figure 5.13: The time required at each subdivision level and the total time of the HRRT PET spatial normalization.

5.5 Spatial normalization of images with lesions

Spatial normalization of images may become problematic when input images contain pathological lesions, e.g. epileptic focus or stroke-impaired tissue. In such cases, automated normalization algorithms attempt to diminish the lesion while trying to reduce mismatch between the image and the template. For example, if the lesion comes out as an area of decreased intensity then the algorithm tries to pull surrounding areas with higher intensity values onto the lesion and “close” it (see Figure 5.14).



Figure 5.14: A transversal slice of a PET image with a lesion after affine-only spatial normalization (left) and the same image after the full non-linear normalization (middle). The lesion area in the middle image is largely reduced which negatively influences subsequent analysis. The image on the right is a regular grid transformed using the deformation of the middle image. It shows that excessive deformations take place in the area of lesion.

Several researchers addressed this issue in recent articles. [Stamatakis et al.](#) studied the influence of the degree of locality in non-linear registration in presence of lesions in SPET images ([Stamatakis et al., 2001](#)). They came to the conclusion that there are basically two possibilities to preserve the area of lesion. One option is to use affine transformations only which is currently the standard approach to this problem, although it cannot provide sufficient accuracy. The latter option is to employ non-linear deformations with some kind of lesion masking. The latter case means that a binary mask needs to be created defining the lesion area. The mask is then passed on to the normalization algorithm that excludes the area from computations. This approach was found superior to the affine-only approach ([Stamatakis et al., 2001](#)). [Brett et al.](#) studied influence of lesion masking in T1-MRI images and came to a similar conclusion ([Brett et al., 2001](#)).

In both cases, binary masks had to be created manually by a physician and this considerably increases the total processing time. It also means that the whole normalization process is no longer reproducible because different

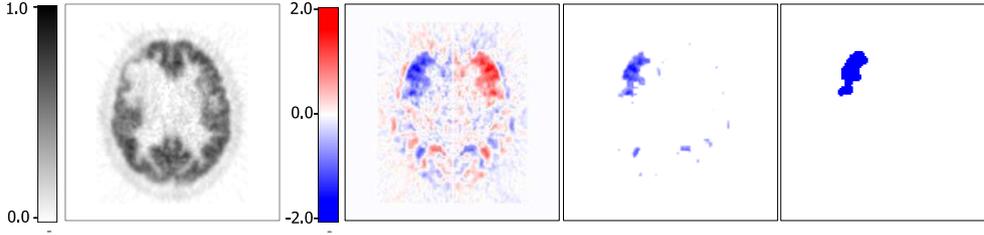


Figure 5.15: Creation of a lesion mask: An image with a lesion (a) and its mirrored image are subtracted to produce a difference image (b). The difference image is thresholded (c) and small clusters are removed by erosion and dilation (d) which yields the desired lesion mask.

physicians would create different masks for the same lesion object. To address this issue, we propose a procedure that attempts an *automatic* masking of lesions during spatial normalization of images of human brain with HBM.

5.5.1 Creation of lesion mask

In the proposed procedure we utilize the fact that human brain is largely symmetric across the sagittal plane. Focal lesions are mostly present only in one hemisphere which means that at the site of lesion there is a pronounced asymmetry.

Suppose that an image \mathcal{S} is centred within the image volume in the x -direction, i.e. it is roughly symmetric across the central yz -plane¹. Let $\bar{\mathcal{S}}$ be a *mirrored image* created by flipping the original image \mathcal{S} across the central yz -plane:

$$\forall \text{ voxel } (i, j, k) \in \mathcal{S} : \bar{\mathcal{S}}(i, j, k) = \mathcal{S}(\text{dim}_x - i - 1, j, k) \quad (5.8)$$

The idea is to subtract the image \mathcal{S} and its mirrored version $\bar{\mathcal{S}}$ which results in a difference image \mathcal{S}_d . Large values in \mathcal{S}_d correspond to highly asymmetric voxel intensities. If a lesion is presented in the image \mathcal{S} , the difference image shows very high values in one hemisphere and very low values in the other hemisphere. Suppose that the lesion in \mathcal{S} expresses itself as a region with lowered intensity. Potential lesion voxels are then the voxels with high negative values. In the next step we apply the segmentation procedure described in section 4.5.3, i.e. thresholding followed by binary erosion and dilation. The erosion and dilation use the same structuring element in this

¹All templates used in this thesis are centered in the x -direction and are roughly symmetric. In consequence, every image normalized to these templates is also approximately symmetric across the the central yz -plane.

case². Figure 5.15 illustrates this process.

We use a quantile-based thresholding which means that for some threshold p , p -% of all voxels with the largest negative values are marked as potential lesion voxels and the rest is marked as background. A suitable threshold does not need to be selected very precisely, as demonstrated in section 4.5.3. The size or diameter of lesions considered for this procedure is roughly in the order of 10-50 voxels. Smaller lesions are hardly detectable in this automatic way but on the other hand, such small lesions do not have a significant influence on the normalization, as demonstrated in Stamatakis et al. (2001). With respect to the considered lesion sizes, the choice of thresholds was tested on several lesioned images with the result that thresholds between 0.90 – 0.99 resulted in nearly the same masks. The outcome of subsequent morphological operations is also robust with respect to the choice of the size of structuring element. The aim of these operations is to remove isolated asymmetrical voxels that do not belong to the lesion area (see Fig. 5.15c). Taking into account the minimum considered lesion size and the tiny size of clusters of scattered voxels not belonging to the lesion, a spherical structuring element with a diameter of about 3 voxels can remove all, or most, non-lesion voxels while preserving the lesion voxels in the mask. Analogously to section 4.5.3, a precise selection of the size of structuring element is not crucial. This means that a single set of parameters is likely to work for a class of lesions with diverse sizes.

5.5.2 Masking within the hierarchical framework

We incorporate the described masking procedure into the hierarchical framework of HBM in the following way. First of all, we use a symmetric template centred in the x -direction which implies that the registered image becomes more symmetric with each level of HBM. The increasing symmetry then enables using of the automatic lesion masking.

The initial affine step usually does not result in a sufficiently symmetric image. On the other hand, the affine step and the very next step are negligibly influenced by presence of a lesion and there is consequently no need for masking. On the subsequent levels the influence of lesion increases and masking becomes needful for a proper alignment. We apply the masking procedure at each level starting from SL3 and always recompute the lesion mask from the current intermediate image \mathcal{S}_k . The mask is thus updated after each level and reflects all transformation changes.

Figure 5.16 shows an example of spatial normalization of a lesioned PET image. The masking procedure used the 95%-quantile for thresholding followed by erosion and dilation with a spherical structuring element with ra-

²Such morphological operation is called “opening”.

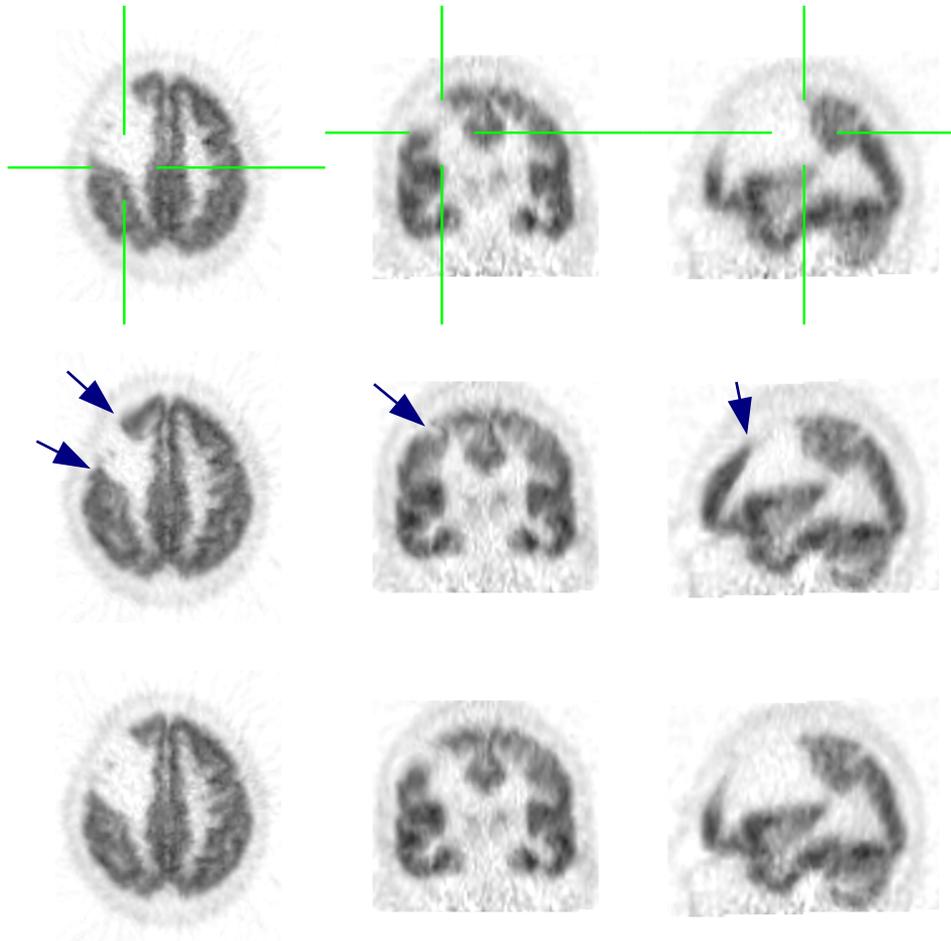


Figure 5.16: The figure shows three orthogonal slices of a PET image with lesion. In the upper row there is the original PET image realigned to a standard position using affine transformations. The middle row shows this image after a non-linear normalization *without* lesion masking. The lesion area is substantially diminished (see arrows) which is apparent mainly in the coronal slice (middle column) where the lesion was almost closed. The bottom row shows result of the same normalization where automatic masking of lesion was used. The area of lesion is mostly preserved and only its position and global shape was adapted during the normalization process.

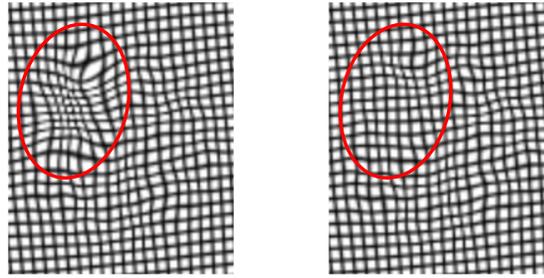


Figure 5.17: The figure shows a deformation grid resulting from spatial normalization of a lesioned PET image with automatic lesion masking (right) and without it (left). When masking of lesion is used the extent of deformation in the lesion area does not deviate from the rest of the volume, in contrast to unnaturally large deformations occurring when masking is not used.

dius 3 voxels. In contrast to the normalized image and its deformation grid showed in Figure 5.14, the lesion area in the resulting image in fig. 5.16 (c) is largely preserved. The corresponding deformation grid in the lesion area does not exhibit deformations distinctively larger than in other areas.

5.6 Discussion

We presented a novel method for non-linear registration of 3D images of human brain. The method was designed for the types of images that lack a distinctively diversified pattern, e.g. PET or SPET images. In such images a reliable identification of landmarks, segmentation of fine structures or a precise extraction of brain surface is not possible which prohibits application of algorithms capable of generating high-frequency deformations.

The number of transformation parameters effectively used for modelling of deformations by the HBM method is in the order of hundreds of thousands to millions. It enables precise alignment of functional information in PET images to a very local level. Using of more parameters would be necessary for matching of fine cortical structures but the low resolution and lack of diversified pattern in PET images do not allow such attempt. As mentioned in the introduction to this chapter, it does not need to be always appropriate to attempt registration beyond a certain resolution.

There exists a number of methods that are designed or at least applicable for non-linear registration of functional medical images. None of the commonly used methods assures one-to-one mapping of images. HBM implicitly guarantees one-to-one mapping and C^2 -continuous deformations which yields

fluid image warping. It generates visually well registered images without undesirable ripples and foldovers.

The performance of the method was evaluated on a group of normal FDG-PET and FMZ-PET images and on a group of high-resolution HRRT FDG-PET images. High robustness was confirmed by a power test on a set of 504 mostly pathologically impaired FDG-PET images and 38 FMZ-PET images.

The method was designed with respect to the requirement to have a fast method applicable for routine clinical applications. During optimization, only very efficient affine transformations are used. Afterwards, smoothness of deformation is supplied by B-spline warping. In this way it was possible to make the method sufficiently fast to be applicable in many clinical applications, even though it works with a relatively large number of transformation parameters. On a PC with 1 GHz processor a spatial normalization of a PET image with dimensions $128 \times 128 \times 47$ voxels takes about 4 minutes. The computational time increases linearly with the number of voxels in registered images. The total computational time can be divided into the time required for optimization of blocks at all levels and the time needed for warping of intermediate images after each block. The optimization time increases linearly with the number of levels whereas the time for warping increases quadratically. The quadratic behaviour is unwelcome but it does not hamper the performance for the number of levels considered here. The registration could however become more time-consuming if significantly larger images appear at some time in the future, requiring more subdivision levels for a precise local alignment. This would then require further optimization.

A difficulty arises when a non-linear registration method needs to be applied to images that contain lesions. With the intention to reduce differences between a template and the lesioned image, a normalization algorithm tends to deform the image so that the area of lesion is destroyed (“closed”). Up to now, the problem has usually been solved by limiting the transformation model to affine transformations that offer only a very limited precision or by using a lesion mask that needs to be manually created. In this chapter, we proposed an automatic method for masking of lesions and incorporated it into the HBM method. It allows precise normalization of the image volume and at the same time preserves the lesion area from being diminished.

Implementation

One of the goals of this work was creation of a robust registration tool that could be used in clinical environment. Since the beginning the development was tightly linked with a feedback from physicians who were using the software for the purpose of research as well as for medical diagnoses.

The registration methods described in this thesis were implemented in a commandline tool called MMM which runs under Windows NT/2000/XP and Linux. The software was written in C++ in a portable way. Input for MMM is an XML-based setup file that specifies input files and registration parameters. It reads and writes images in Analyze97 and ECAT7 formats.

The software development was made largely in MS Visual Studio .NET on a PC running Windows XP and partially using KDevelop and CBuilderX on SuSE Linux 9.0. The tests in chapter 4 were performed in an automated way with the help of Perl scripts. Evaluations of the tests were made in Perl, Matlab, Origin and in VINCI.

6.1 Graphical user interface

VINCI, a graphical tool for medical image analysis in Windows NT/2000/XP, has been developed at the Max-Planck Institute for Neurological Research in Cologne (MPI) since several years (Vollmar et al., 2001). It is the standard tool for medical image analysis in research and clinical applications in the PET laboratory of MPI. VINCI was designed in a modular way and is easily extensible by external plug-ins which communicate with the core engine via XML commands. To facilitate the usage of MMM, a graphical interface was created as an external plugin for VINCI. This step was a major improvement which significantly facilitated and accelerated the routine work in the PET laboratory (Cizek et al., 2004). The user interface was implemented in a way that allows easy operation without a need to understand the registration procedure and its parameters. On the other hand, many parameters can be tuned by an experienced user. Registration progress is presented on a fusion of images and updated every few seconds with the current best transformation (Figure 6.1). In this way the user can immediately see if the



Figure 6.1: Snapshot of a running registration in VINCI.

registration evolves in the right direction. In case of problems it is possible to pre-align images manually per mouse before an automatic registration. Tools are provided for visual inspection of registration results, image volumes can be resliced and zoomed synchronously in an intuitive fashion with an optional contour rendering. The registration tool has been successfully used since more than one year in a number of clinical applications.

6.2 Multi-threading

The most time consuming part of the rigid body registration (MMM) as well as the non-linear registration (HBM) is the many times repeated evaluation of similarity measure during the iterative optimization. Each evaluation of similarity measure involves transformation of all or a subset of the image voxels. All three presented measures, SSD, CC and MI, are computed from contributions of pairs of corresponding voxels in both images (chapter 3). Computation of every such contribution is independent on the remaining voxels which means that it is possible to distribute the computation to several processors. Each processor computes contributions in different part of the

image and at the end all contributions are merged.

Multi-threading has currently been implemented and tested for the MMM method. The performance was tested on a SGI machine with four 700 MHz processors running SuSE Linux 8.1 and on a PC with two 3 GHz processors running Windows 2000. On the four-processor machine an acceleration factor of 3.7 was achieved for the optimization part of a registration of large high-resolution HRRT PET and MR images and about 3.5 for a pair of common PET and MR images. On the two-processor computer the acceleration factor was about 1.8 for high-resolution and about 1.7 for low resolution images. The results demonstrate that the registration procedure is very suitable for parallelization. At the present time, when the large cost reduction of multi-processor PCs made them affordable for many clinical facilities, multi-processor registration is an attractive option for additional reduction of the time spent on a registration. Registration on multi-processor systems becomes yet more needful for registration of the new high-resolution HRRT PET images.

Conclusions

The aim of this thesis was to develop algorithms for a fast, robust and precise alignment of 3D medical images of human brain. The focus was on multi-modal registration of PET and MR images of a single patient and on non-linear matching of inter-subject PET images.

7.1 Contributions

The most important contributions of this thesis can be found in chapters 4 and 5. Both chapters are concluded with a discussion of the presented solutions. This section briefly summarizes the results.

7.1.1 Rigid-body registration

In chapter 4 we described a registration algorithm for a 3D registration of multi-modal images of the same patient based on maximization of mutual information. The general framework has been used and validated in many medical applications. The increasing performance of modern computers has already enabled routine usage of automated registration methods but there is still a considerable demand after yet faster and robuster algorithms that would facilitate and accelerate work in the tight time frame of clinical applications.

Masking of background and global displacement estimation

We proposed several modifications to the basic MMM algorithm that aimed at increasing speed of registration and that tried to make the automatic procedure more reliable.

Both input images are first efficiently segmented into background and object voxels by an automatic thresholding procedure followed by a sequence of morphological operations. Only the object voxels are then used for evaluation of similarity measure which brings more than threefold acceleration on the optimization part.

The segmentation step also provides an approximate information about

the shape, position and orientation of the brain volume in each image. We exploit this information by a modified principal axes algorithm that provides a rough estimate of mutual position and orientation of the registered images in a geometrical way, without purely relying on the similarity function. The estimated transformation is then passed on to the optimization method. This approach is efficient and provides an estimate that is more likely in the “capture range” of the optimization method.

These procedures were combined and implemented in a registration algorithm called **MMM-PA**. Tests with simulated and clinical images demonstrated that it is much more robust than the basic **MMM** method. It is, however, less suitable when one of the images is partially cropped because the registered brain shapes are then no longer similar. **MMM-PA** achieved an acceleration of factor 2.5, resp. 3.1 for registration of real, resp. simulated images and yielded a better accuracy than **MMM** for both real and simulated images. The accuracy of a low-resolution **MMM-PA** registration was even better than the accuracy of the full-resolution **MMM**. It means that for many applications we can just use the low-resolution **MMM-PA** and profit from a 6-7 times shorter computational time compared to the current standard method, the full-resolution **MMM**.

Importance of a proper histogram size

We demonstrated that the number of bins used for computation of mutual information is an important factor that influences robustness and accuracy of registration. So far, very little attention has been paid to this parameter which might be one of the reasons for some discrepancies in results presented by different authors. Tests with simulated and clinical images proved that both too large and too small number of bins significantly deteriorate accuracy and diminish the capture range. The optimum histogram size depends on the number of samples which means that multi-resolution approaches should adapt the number of bins to the number of voxels at each scale level.

The question of a suitable histogram size belongs to yet unresolved issues in statistics. We therefore selected the suitable histogram size according to the results of our experiments.

Multi-threading

The most time consuming part of a rigid-body registration is the repeated evaluation of similarity measure during iterative optimization. The computation of similarity measure can be easily parallelized: a multi-threaded implementation of **MMM** yielded an acceleration factor 3.7 for registration of high-resolution PET and MR images on a four-processor SGI machine running SuSE Linux 8.1.

Clinical validation

The MMM tool has been in routine use in the PET Laboratory of the Max-Planck-Institute for Neurological Research in Cologne for more than one year and has been successfully applied in many clinical applications.

7.1.2 Non-linear registration

We developed a new algorithm HBM for non-linear registration 3D medical images. The algorithm estimates a suitable deformation that matches brain images of different individuals. It guarantees that the generated deformation is injective which is a missing property in existing comparable algorithms and it is one of the most important contributions of this method. It assures that the topology of the deformed image remains unchanged and that there is no folding-over which would erase intensity information. Injectivity thus improves physiological consistency of deformed medical images.

Validation

The HBM method is particularly suitable for functional images, e.g. PET images. We performed a number of tests to validate the method on various types of data: simulated geometrical objects, normal FMZ-PET brain images, normal FDG-PET images, high-resolution HRRT FDG-PET images, a set of 38 FMZ-PET images of diseased brains and a large database of 508 FDG-PET images, partially with significant distortions, reconstruction artefacts or lesions. The tests with the groups of normal images demonstrated that in comparison to an affine-only registration, the proposed algorithm is able to decrease the standard deviation of corresponding voxels across each group by nearly 40% on average, in some areas even by 70%. The results of the tests with diseased brain images are very promising. There was no dataset for which the registration would fail or give a seriously inaccurate result.

Masking of lesions

Non-linear registration of images becomes more problematic when one of the registered images contains a large lesion. In such case the registration algorithm tries to diminish the lesion in order to improve similarity of the images. We presented a solution for an automatic masking of lesions during non-linear registration which exploits brain symmetry. First results show that in this way the lesion area can be preserved without hindering alignment of neighbouring structures.

7.1.3 User interface

The described registration methods have been implemented in an application with a convenient graphical user interface. The MMM registration also provides a real-time on-screen visual presentation of registration progress. The graphical implementation presents a powerful yet easy-to-use tool for both clinical and research environments.

7.2 Future directions

Precise validation of registration quality with real data Accuracy of the MMM-PA registration of clinical PET and MR images was evaluated only indirectly, with no “gold standard” transformations available for comparison. It would be interesting to evaluate the registration accuracy on a set of PET and MR images acquired with fiducial markers fastened to patient’s head during image acquisition in both PET and MR scanner. The markers would allow a precise alignment that could serve as a gold standard against which the automatic registration could be compared. Unfortunately, such procedure is very time-consuming and demanding for both patients and physicians.

Complexity of the non-linear method The time complexity of the non-linear method increases quadratically with increasing number of subdivision levels. This is caused by the necessity to recompute intermediate source image after every level using all by then computed deformation lattices. This inefficiency could be avoided if it was possible to merge each computed lattice with a refined lattice from previous level. Creation of refined lattice poses no big difficulty. It is the considerable loss of precision resulting from interpolation of transformation vectors that hindered me in going this direction. A possible solution could be in modifying the method in a way that would make interpolation of transformation vectors unnecessary. This is a task for future research.

Masking of lesions Automatic masking of lesions presented in section 5.5 is a promising technique that could improve information gained from lesioned images. However, a precise quantification of its effects on simulated and real lesions is necessary before it can be used routinely in clinical applications.

Computation of inverse deformation The injectivity property opens many new possibilities for future development. It theoretically allows creation of an inverse transformation but there is still no clear way of generating it. It could also be worth trying to develop a hybrid algorithm that would simultaneously compute deformations in both directions, i.e. from source

image to target and vice versa, and combine the results in order to achieve more precise results.

Comparison with other registration packages It would be interesting to compare the performance of proposed algorithms with other competitive methods. This is feasible for the multi-modality registration (with the help of fiducial markers) but it is very difficult for non-linear registration because an optimum deformation that matches images of two different individuals is unknown. In fact, an ultimate “optimal deformation” does not even exist since its definition may substantially vary depending on the intended application.

One option is to make the algorithm generate a random deformation and then to let it attempt to recover the initial shape. A more challenging test would be trying to recover a deformation generated with a different model than the B-spline deformation model.

References

- Alpert, N. M., Bradshaw, J. F., Kennedy, D., and Correia, J. A. (1990). The principal axes transformation - a method for image registration. *Journal of Nuclear Medicine*, 31:1717–1722.
- Ashburner, J. (2000). *Computational Neuroanatomy*. PhD Thesis, University College, London.
- Ashburner, J. and Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7:254–266.
- Bardera, A., Feixas, M., and Boada, I. (2004). Normalized similarity measures for medical image registration. In *Proceedings of SPIE 2004*, volume 5370.
- Bookstein, F. L. (1997). Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis*, 1(3):225–243.
- Brett, M., Leff, A. P., Rorden, C., and Ashburner, J. (2001). Spatial normalization of brain images with focal lesions using cost function masking. *NeuroImage*, 14(2):486–500.
- Cachier, P. and Ayache, N. (2001). How to trade off between regularization and image similarity in non-rigid registration? In *Proceedings of the 4th International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 2208, pages 1285–1286.
- Capek, M., Mroz, L., and Wegenkittl, R. (2001). Robust and fast medical registration of 3D-multi-modality data sets. In *Proceedings of Medicon 2001 - IX Mediterranean Conference on Medical and Biological Engineering and Computing*, volume 1, pages 515–518.
- Choi, Y. and Lee, S. (1999). Local injectivity conditions of 2D and 3D uniform cubic B-spline functions. In *Proceedings of the 7th pacific conference on computer graphics and applications*.
- Choi, Y. and Lee, S. (2000). Injectivity conditions of 2D and 3D uniform cubic B-spline functions. *Graphical Models*, 62:411–427.
- Cizek, J., Herholz, K., Vollmar, S., Schrader, R., Klein, J., and Heiss, W.-D. (2004). Fast and robust registration of PET and MR images of human brain. *NeuroImage*, 22:434–442.

- Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., and Marchal, G. (1995a). Automated multi-modality image registration based on information theory. In *Proceedings of Information Processing in Medical Imaging*, pages 263–274.
- Collignon, A., Vandermeulen, D., Suetens, P., and Marchal, G. (1995b). 3D multi-modality medical image registration using feature space clustering. In *Proceedings of the 1st International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, pages 195–204.
- Collins, D. L., Neelin, P., Peters, T. M., and Evans, A. C. (1994). Automatic 3-D intersubject registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*, 18(2):192–205.
- Eddy, W. F. and Young, T. K. (2000). Optimizing the resampling of registered images. In Bankman, I. N., editor, *Handbook of Medical Imaging*, book chapter 37, pages 603–612. Academic Press.
- Fowles, G. F. and Cassiday, G. L. (1993). *Analytical Mechanics*. Thomson Learning.
- Fransens, R., Strecha, C., and Gool, L. V. (2004). Multimodal and multiband image registration using mutual information. In *Proceedings of ESA-EUSC 2004: Theory and Applications of Knowledge driven Image Information Mining, with focus on Earth Observation*.
- Gonzales, R. C. and Woods, R. E. (1993). *Digital Image Processing*. Addison-Wesley Publishing Company.
- Hajnal, J., Hawkes, D., and Hill, D., editors (2001). *Medical Image Registration*. CRC Press.
- He, Y., Hamza, A. B., and Krim, H. (2003). A generalized divergence measure for robust image registration. *IEEE Transactions on Signal Processing*, 51(5):1211–1220.
- Heckbert, P. S., editor (1994). *Graphic Gems IV*. Academic Press.
- Hellier, P., Barillot, C., and Perez, P. (2001). Hierarchical estimation of a dense deformation field for 3-D robust registration. *IEEE Transactions on Medical Imaging*, 20(5):388–402.
- Herholz, K., Herscovitch, P., and Heiss, W.-D. (2004). *NeuroPET*. Springer Verlag.

- Hill, D. L. G. and Hawkes, D. J. (2000). Across-modality registration using intensity-based cost functions. In Bankman, I. N., editor, *Handbook of Medical Imaging*, book chapter 34, pages 537–553. Academic Press.
- Holmes, A. P. (1995). *Statistical issues in functional brain mapping*. PhD Thesis, Department of Statistics, University of Glasgow.
- Hsu, W. M., Hughes, J. F., and Kaufman, H. (1992). Direct manipulation of free-form deformations. In *Proceedings of SIGGRAPH'92*, volume 26, pages 177–184.
- Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17:825–841.
- Johnson, H. J. and Christensen, G. E. (2002). Consistent landmark and intensity-based image registration. *IEEE Transactions on Medical Imaging*, 21(5):450–461.
- Keating, J. P. and Scott, D. W. (1999). A primer on density estimation for the Great Home Run Race of '98. *STATS*, 25.
- Kiebel, S. J., Ashburner, J., Poline, J. B., and Friston, K. J. (1997). MRI and PET coregistration - A cross validation of statistical parametric mapping and automated image registration. *NeuroImage*, 5(4):271–279.
- Kracht, L. W., Miletic, H., Busch, S., Jacobs, A. H., Voges, J., Hoevens, M., Klein, J. C., Herholz, K., and Heiss, W.-D. (2004). Delineation of brain tumor extent with [11C]L-methionine Positron Emission Tomography: local comparison with stereotactic histopathology. *Clinical Cancer Research*, 10:7163–7170.
- Lee, S., Wolberg, G., and Shin, S. Y. (1997). Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3(3):228–244.
- Lee, S.-Y., Chwa, K.-Y., Shin, S. Y., and Wolberg, G. (1995). Image metamorphosis using snakes and free-form deformations. In *Proceedings of SIGGRAPH'95*, pages 439–448.
- Lee, S.-Y., Wolberg, G., Chwa, K.-Y., and Shin, S. Y. (1996). Image metamorphosis with scattered feature constraints. *IEEE Transactions on Visualization and Computer Graphics*, 2(4):337–354.
- Maes, F. (1998). *Segmentation and Registration of Multimodal Medical Images*. PhD Thesis, Katholieke Universiteit, Leuven, Belgium.

- Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198.
- Maes, F., Vandermeulen, D., and Suetens, P. (1999). Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4):373–386.
- Maintz, J. B. A. and Viergever, M. A. (1998). A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36.
- McInerney, T. and Terzopoulos, D. (1996). Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108.
- Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7:308–313.
- Parker, J. R. (1997). *Algorithms for Image Processing and Computer Vision*. John Wiley & Sons.
- Pietrzyk, U., Herholz, K., Fink, G., Jacobs, A., Mielke, R., Slansky, I., Wuerker, M., and Heiss, W.-D. (1994). An interactive technique for three-dimensional image registration: Validation for PET, SPECT, MRI and CT brain studies. *Journal of Nuclear Medicine*, 35:2011–2018.
- Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (2000). Interpolation artefacts in mutual information-based image registration. *Computer Vision and Image Understanding*, 77(2):211–232.
- Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (2001a). f-information measures in medical image registration. In *Proceedings of SPIE on Medical Imaging 2001: Image Processing*, volume 4322, pages 579–587.
- Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (2001b). Mutual information matching in multiresolution contexts. *Image and Vision Computing*, 19(1-2):45–52.
- Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (2003). Mutual-information-based registration of medical images: A survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*. Cambridge University Press, 2 edition.

- Sain, S. (1999). Multivariate locally adaptive density estimation. Technical report, Department of Statistical Science, Southern Methodist University.
- Sederberg, T. (2003). Computer aided geometric design (lecture notes). <http://tom.cs.byu.edu/~557/>.
- Sederberg, T. W. and Parry, S. R. (1986). Free-form deformation of solid geometric models. In *Proceedings of SIGGRAPH'86*, volume 20, pages 151–160.
- Seul, M., O’Gorman, L., and Sammon, M. J. (2001). *Practical Algorithms for Image Analysis*. Cambridge University Press.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423,623–656.
- Shoemake, K. and Duff, T. (1992). Matrix animation and polar decomposition. In *Proceedings of the conference on Graphics interface '92*, pages 258–264.
- Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, 17(3):143–155.
- Sonka, M., Hlavac, V., and Boyle, R. (1994). *Image Processing, Analysis and Machine Vision*. Chapman and Hall Computing.
- Stamatakis, E. A., Wilson, J. T. L., and Wyper, D. J. (2001). Spatial normalization of lesioned HMPAO-SPECT images. *NeuroImage*, 14(4):844–852.
- Studholme, C. (1997). *Measures of 3D Medical Image Alignment*. PhD Thesis, United Medical and Dental Schools of Guy’s and St Thomas’ Hospitals, University of London.
- Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1995). Multiresolution voxel similarity measures for MR-PET registration. In *Proceedings of Information Processing in Medical Imaging*, pages 287–298.
- Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1997). Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical Physics*, 24(1):25–35.
- Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86.

- Thevenaz, P., Blu, T., and Unser, M. (2000). Image interpolation and re-sampling. In Bankmann, I. N., editor, *Handbook of Medical Imaging*, book chapter 25, pages 393–420. Academic Press.
- Thevenaz, P. and Unser, M. (2000). Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing*, 9(12):2083–2099.
- Thompson, P. M. and Toga, A. W. (1996). A surface-based technique for warping three-dimensional images of the brain. *IEEE Transactions on Medical Imaging*, 15:1–16.
- Thurfjell, L., Lau, Y. H., Andersson, J. L. R., and Hutton, B. F. (2000). Improved efficiency for MRI-SPET registration based on mutual information. *European Journal of Nuclear Medicine*, 27(7):847–856.
- Toga, A. W. and Mazziotta, J. C. (1996). *Brain Mapping: The Methods*. Academic Press.
- Viola, P. A. (1995). *Alignment by Maximization of Mutual Information*. PhD Thesis, Massachusetts Institute of Technology.
- Viola, P. A. and Wells, W. M. (1995). Alignment by maximization of mutual information. In *Proceedings of the 5th International Conference on Computer Vision*, pages 15–23.
- Viola, P. A. and Wells, W. M. (1997). Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154.
- Vollmar, S., Michel, C., Treffert, J. T., Newport, D. F., Casey, M., Knoess, C., Wienhard, K., Liu, X., Defrise, M., and Heiss, W.-D. (2001). Heinzl-Cluster: accelerated reconstruction for FORE and OSEM3D. *Physics in Medicine and Biology*, 47(15):2651–2658.
- Wand, M. P. (1997). Data-based choice of histogram bin width. *The American Statistician*, 51(1):59–64.
- Wells, W. M., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R. (1996). Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51.
- West, J., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer, C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Suetens, P., Vandermeulen, D., vandenElsen, P. A., Napel, S., Sumanaweera, T. S., Harkness, B., Hemler, P. F., Hill, D. L. G., Hawkes,

- D. J., Studholme, C., Maintz, J. B. A., Viergever, M. A., Malandain, G., Pennec, X., Noz, M. E., Maguire, G. Q., Pollack, M., Pelizzari, C. A., Robb, R. A., Hanson, D., and Woods, R. P. (1997). Comparison and evaluation of retrospective intermodality brain image registration techniques. *Journal of Computer Assisted Tomography*, 21(4):554–566.
- Wienhard, K., Dahlbom, M., Eriksson, L., Michel, C., Bruckbauer, T., Pietrzyk, U., and Heiss, W. D. (1994). The ECAT Exact HR - performance of a new high-resolution positron scanner. *Journal of Computer Assisted Tomography*, 18(1):110–118.
- Wienhard, K., Schmand, M., Casey, M. E., Baker, K., Bao, J., Eriksson, L., Jones, W. F., Knoess, C., Lenox, M., Lercher, M., Luk, P., Michel, C., Reed, J. H., Richerzhagen, N., Treffert, J., Vollmar, S., Young, J. W., Heiss, W. D., and Nutt, R. (2002). The ECAT HRRT: Performance and first clinical application of the new high resolution research tomograph. *IEEE Transactions on Nuclear Science*, 49:104–110.
- Wolberg, G. (1990). *Digital image warping*. Wiley - IEEE Computer Society Press.
- Woods, R. P. (2000). Spatial transformation models. In Bankmann, I. N., editor, *Handbook of Medical Imaging*, book chapter 29, pages 465–490. Academic Press.
- Woods, R. P., Grafton, S. T., Watson, J. D. G., Sicotte, N. L., and Mazziotta, J. C. (1998). Automated image registration: II. Intersubject validation of linear and nonlinear models. *Journal of Computer Assisted Tomography*, 22(1):153–165.
- Zhang, Y. Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57.

Acknowledgement

My greatest thanks go to Prof. Karl Herholz and Prof. Rainer Schrader for their support and fruitful discussions that always helped in finding the right way. Thanks to Prof. Wolf-Dieter Heiss for enabling my stay at the Max-Planck Institute for Neurological Research in Cologne.

I am grateful to Johannes Klein, Stefan Vollmar, Michael Sué and Gerhard Zündorf for many fruitful discussions and help. Thanks to all fellows of the PET Laboratory for a pleasant working atmosphere.

Many thanks to my best good friend Petra Kudějová and to my parents - for everything.

Erklärung

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. R. Schrader und Prof. Dr. med. K. Herholz betreut worden.

für C. C. C.

Lebenslauf

Persönliche Daten

Name	Jiří Čížek
Geburtsdatum	14.1.1978
Geburtsort	České Budějovice, Tschechische Republik
Staatsangehörigkeit	Tschechische Republik

Schulausbildung

1984-1992	Grundschule in Velešín
1992-1996	Gymnasium Jírovцова in České Budějovice

Studium

1996-2001	Studium der Informatik an der Mathematisch-Physikalischen Fakultät der Karls-Universität in Prag
2002-2004	Promotionsstudium der Informatik an der Mathematisch-Naturwissenschaftlichen Fakultät der Universität zu Köln; Stipendiät des Max-Planck-Instituts für neurologische Forschung in Köln.

Teilpublikation

Čížek, J., Herholz, K., Vollmar, S., Schrader, R., Klein, J., and Heiss, W.-D. (2004). *Fast and robust registration of PET and MR images of human brain*. NeuroImage, 22: 434–442.