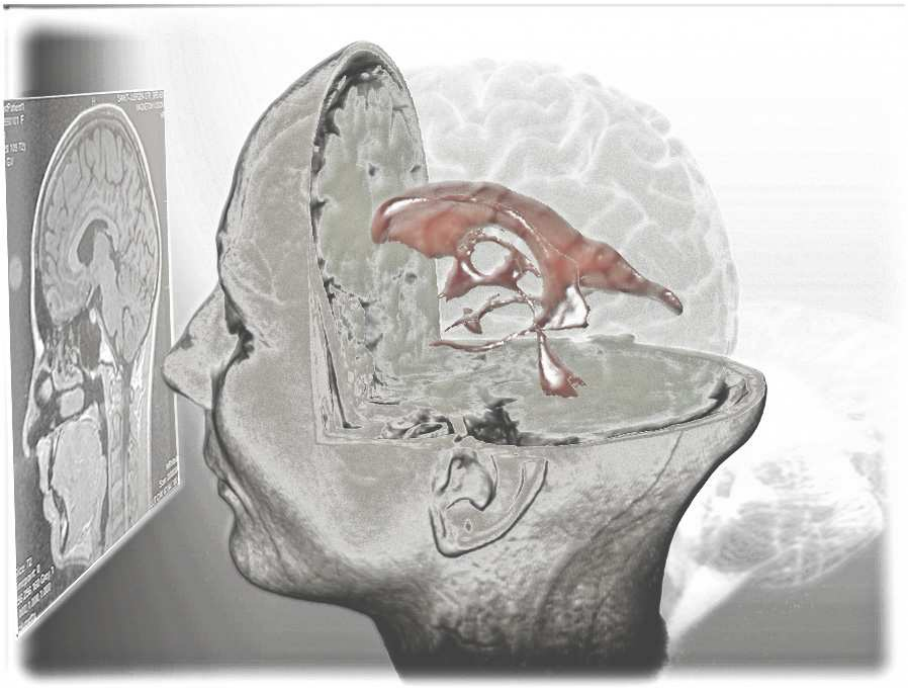
Horst Karl Hahn

Morphological Volumetry

Theory, Concepts, and Application to Quantitative Medical Imaging



MeVis – Center for Medical Diagnostic Systems and Visualization
University of Bremen
2005

Datum des Promotionskolloquiums: 11. März 2005

Gutachter: Prof. Dr. rer. nat. H.-O. Peitgen, MeVis, Bremen / D
Prof. Dr. Ir. J. H. C. Reiber, LKEB, Leiden / NL

Morphological Volumetry

Theory, Concepts, and Application to Quantitative Medical Imaging

von
Dipl.-Phys. Horst Karl Hahn
geboren am 4. September 1972 in Freiburg/Brsg.

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften
– Dr. Ing. –

vorgelegt im Fachbereich 3
Mathematik & Informatik
der Universität Bremen

Bremen, den 21. Januar 2005

Die Arbeit erfolgte bei MeVis,
Centrum für Medizinische Diagnosesysteme und Visualisierung.

The work has been conducted at MeVis,
Center for Medical Diagnostic Systems and Visualization.

Horst K. Hahn
MeVis, Universitätsallee 29, 28359 Bremen, Germany
Tel. +49-421-2189183, Email: hahn@mevis.de, http://www.mevis.de

Gewidmet
Dr. med. Klaus Hahn und
Dr. med. Dorothea Hahn,
meinen Eltern

Zusammenfassung

DIESE ARBEIT bringt quantitative Methoden auf komplexen medizinischen Bilddaten der klinischen Routine näher und überbrückt die Kluft zwischen manueller und automatischer Verarbeitung. Es werden effiziente und effektive Bildsegmentierungsmethoden beschrieben und auf bislang nicht zufriedenstellend gelöste Probleme angewendet, die relevant sind für die medizinische Diagnose und Therapiekontrolle. Die Methoden genügen allen Anforderungen, welche an einen Softwareassistenten für die quantitative klinische Bildanalyse zu stellen sind:

(a) *Anwendbarkeit* im Sinne der notwendigen Hardware und Akquisitionsprotokolle sowie einer effektiven und intuitiven Kontrolle und Korrigierbarkeit der Verarbeitung, da eine vollautomatische Lösung der gestellten Probleme nicht erwartet werden kann,

(b) *Robustheit* gegenüber Variationen hinsichtlich der Bildqualität sowie der Anatomie und Pathologie, um den Problemen der klinischen Routine standzuhalten,

(c) *Genauigkeit* aller abgeleiteten quantitativen Maße,

(d) *Sensitivität* hinsichtlich relevanter Veränderungen der untersuchten Struktur sowie

(e) *Effizienz*, um mit Standardhardware auch auf sehr großen Datensätzen anwendbar zu sein.

Der erste Teil dieser Arbeit beschreibt neuartige Methoden als Erweiterung der morphologischen Wasserscheidentransformation, die als zentrales Konzept der Mathematischen Morphologie zur Bildsegmentierung bekannt ist.

Als erstes Problem wird die Isolierung des Gehirns von umgebendem Gewebe in anatomischen Magnetresonanz-(MR)-Bildern des menschlichen Kopfes behandelt, auch Skull-Stripping genannt. Es wird eine schnelle modifizierte Wasserscheidentransformation mit linearer numerischer Komplexität beschrieben. Die Preflooding-Höhe wird als neuer Parameter eingeführt, der die Detailliertheit der Segmentierung global steuert. Dieser Parameter stellt die Basis zur Vermeidung von Übersegmentierung dar, welche als das Hauptproblem der klassischen Wasserscheidentransformation bekannt ist. Die Schlüsselidee, um

das Skull-Stripping-Problem robust zu lösen, besteht in der Anwendung einer schnellen modifizierten Wasserscheidentransformation auf die invertierten Original-Bilddaten. Im Kontrast hierzu geschieht die herkömmliche Anwendung der Wasserscheidentransformation auf ein abgeleitetes Diskontinuitätsbild. Die Qualität der vorgeschlagenen Methode wird anhand von Patienten-, Probanden- und Phantomdaten demonstriert.

Von der Theorie effektiver Mensch-Computer-Interaktion werden Richtlinien für die Konzeption interaktiver Systeme für die medizinische Bildanalyse abgeleitet. Es wird die Dualität von Interaktivität einerseits und Automatisierung andererseits diskutiert. Auf dieser Basis wird das Konzept der minimalen Interaktion vorgeschlagen, um die Möglichkeit zur Bedienerkontrolle mit der Generierung weitgehend bedienerunabhängiger quantitativer Ergebnisse zu kombinieren. minimale Interaktion wird in Bezug gesetzt zum Konzept der iterativ-interaktiven Verfeinerung, worin das System durch ein Multiskalen-Verhalten von jeder einzelnen Bedienereingabe, in Kombination mit schnellem Systemfeedback, einen Schritt näher an das gewünschte Ergebnis geführt wird.

Um den Anforderungen von minimaler Interaktion zu genügen, wird die Interaktive Wasserscheidentransformation (IWT) als Erweiterung der zuvor beschriebenen schnellen, modifizierten Wasserscheidentransformation vorgeschlagen. Die IWT stellt die erste Variante der Wasserscheidentransformation dar, welche nicht nur bei der Transformation, sondern auch während der Benutzerinteraktion schnell ist, und welche Multiskalen-Bedienerkontrolle auf großen mehrdimensionalen Datensätzen in Echtzeit ermöglicht. Iterative Verfeinerung geschieht entweder durch Setzen beliebig vieler Marker oder durch die Wahl globaler Segmentierungsparameter, beispielsweise der Preflooding-Höhe. Es wird gezeigt, dass es durch die Analyse der hierarchischen Regionen-Verschmelzungscharakteristik, welche während der IWT aufgezeichnet wird, möglich ist, die Marker und Preflooding-Höhe zur Lösung des Skull-Stripping-Problems automatisch und mit einer hohen Robustheit gegenüber Bildinhomogenität und -rauschen zu finden.

Basierend auf der IWT wird ein Konzept zur minimal-interaktiven Segmentierung von Knochenstrukturen auf computertomographischen (CT) Daten vorgeschlagen. Lösungen zweier Probleme werden beschrieben: erstens die effiziente Segmentierung der Handgelenksknochen als Voraussetzung zur Durchführung kinematischer Studien, und zweitens das Entfernen von Knochenstrukturen auf großen CT-angiographischen (CTA) Datensätzen zur vereinfachten Analyse vaskulärer Erkrankungen und Fehlbildungen, auch Bone Removal genannt. Die Herausforderung des ersten Problems entsteht durch direkten Kontakt benachbarter Knochen, inhomogene kortikale Knochendichte und hoch unregelmäßige Knochenformen. Zusätzlich zur IWT wird das Konzept der Bit-Morphologie zur recheneffizienten Generierung solider Knochenmasken vorgeschlagen. Die Herausforderung des zweiten Problems liegt in (i) der großen Vielfalt und unregelmäßigen Erscheinung kontrastmittelverstärkter Gefäßstrukturen, (ii) einer hohen intravaskulären Bildintensität, ähnlich der von Knochenstrukturen, aufgrund von Kontrastmittel, Verkalkungen, Stents, Clips usw., (iii) der beträchtlichen Größe der von Multidetektor-CT-Systemen erzeugten Original-Bilddaten und (iv) den Zeitbeschränkungen plus der Notwendigkeit einer effektiven und intuitiven, interaktiven Kontrolle, welche durch Usability-Anforderungen vorgegeben sind. Die Schlüsselidee für das Bone Removal ist die Entwicklung einer sog.

komprimierten IWT und zusätzlich eines merkmalsbasierten Klassifikators zur automatischen Knochen-Gefäß-Trennung. Nach der automatischen Segmentierung stehen die volle Interaktivität und Geschwindigkeit der IWT für die iterative Verfeinerung zur Verfügung.

Der Zweite Teil der Arbeit richtet sich auf die quantitative Bildanalyse im Allgemeinen sowie die Volumetrie bestimmter neuroanatomischer Strukturen im Besonderen. Die Morpholgische Volumetrie wird als Rahmenkonzept für die zuverlässige volumetrische Quantifizierung von komplexen dreidimensionalen Strukturen vorgeschlagen, wie sie in tomographischen medizinischen Bildern zu finden sind. Dieses Konzept kombiniert die erweiterte Wasserscheidentransformation mit einer automatischen Histogrammanalyse. Um die Anforderungen einer computergestützten Volumetrie in Radiologie und Neuroradiologie zu erkunden, wurde eine deutschlandweite Umfrage durchgeführt, deren Ergebnisse vorgestellt werden.

Eine Übersicht über die sog. Quantitative Bildgebungs-Pipeline, welche Bildakquisition und -analyse umfasst, dient als Basis für eine methodologische Untersuchung der zugehörigen Fehlerkette. Die Diskussion der verschiedenen Kategorien systematischer und statistischer Fehler liefert Richtlinien für die Evaluation entsprechender Bildanalyse-Methoden. Am Beispiel der MR-Bildgebung werden zwei wichtige und für die quantitative medizinische Bildanalyse relevante Bildcharakteristika—Pixelrauschen und Partialvolumeneffekte—diskutiert und in Bezug zu theoretischen Modellen gestellt. Die Schwierigkeiten der Volumetrie kleiner anatomischer Strukturen werden anhand von Beispielen diskutiert.

Zwei weitere Problemgruppen werden eingehend behandelt, namentlich die Volumetrie von Gehirngewebe und von cerebrospinaler Flüssigkeit (CSF). Ein Hybrid-Ansatz, welcher die IWT mit einer automatischen, modellbasierten Histogrammanalyse kombiniert, wird vorgeschlagen zur exakten Volumetrie der intracerebralen Flüssigkeitsräume, den Hirnventrikeln. Die Schwierigkeit dieser Aufgabe liegt in der hohen inter-individuellen anatomischen Variabilität, vor allem in pathologischen Fällen, und in der dünnen, lang gestreckten Form der Ventrikel, welche die volumetrische Genauigkeit behindert. Die IWT wird zur effizienten und effektiven Trennung der einzelnen Ventrikel voneinander sowie von isointensen Bildinhalten vorgeschlagen, wie etwa subarachnoidaler CSF, welche von der weiteren Auswertung ausgeschlossen werden muss. Das Histogramm-Modell ist derart konzipiert, dass es regionale Bildvariabilitäten kompensiert und sich robust dem Ausmaß von Partialvolumeneffekten anpasst. Der vorgeschlagene Ansatz führt zu einer hohen Geschwindigkeit, Flexibilität und Genauigkeit im Vergleich zu bestehenden Verfahren. Dadurch dass eine Bedienereingabe nur in Form von Punktmarkern erforderlich ist, um die interessierenden Objekte zu lokalisieren, nicht jedoch um deren Grenzen einzuzeichnen, konnte das Ziel der minimalen Interaktion mit einer hohen Inter-Observer-Konsistenz kombiniert werden. Eine besondere Herausforderung besteht bei der pädiatrischen Bildgebung durch die oft sehr dünnen Ventrikel im kindlichen Gehirn. Hierfür wird eine Erweiterung der Histogrammanalyse vorgeschlagen. Es werden Ergebnisse für eine Gruppe frühgeborener Kinder mit und ohne neonatalem Schaden der weißen Substanz vorgestellt. Darüber hinaus wird eine ultra-schnelle Methode diskutiert, welche eine reproduzierbare Volumetrie der gesamten intrakraniellen CSF innerhalb weniger Sekunden erlaubt.

Der beschriebene Hybrid-Ansatz wird erweitert und auf die Volumetrie grauer und weißer Hirnsubstanz sowie auf die Quantifizierung der sog. Brain Parenchymal Fraction angewendet. Diese Probleme stellen eine Herausforderung aufgrund des komplexen und interindividuell variablen kortikalen Faltungsmusters dar. Das vorgeschlagene Verfahren zur Gehirnvolumetrie ist konzeptionell einfacher als bestehende Methoden, welche beispielsweise auf automatischer iterativer Gewebeklassifizierung, MR-Inhomogenitätskorrektur, Markov-Random-Field-Modellierung und anatomischen Formvorlagen oder Atlanten aufbauen. Es wird gezeigt, dass die Konsistenz der vorgeschlagenen Methode vergleichbar oder besser als der zweier solcher bestehenden Methoden, EMS und SIENAX, im Hinblick auf wiederholte Untersuchungen, variable Bildauflösung, Inhomogenität und Rauschen, sowie im Vergleich zweier unterschiedlicher Tomographen ist. Zusätzlich wird die absolute Genauigkeit der drei Methoden anhand der sog. Simulated Brain Database evaluiert, wobei ein Vorteil des vorgeschlagenen Ansatzes zur Gehirnvolumetrie zutage tritt. Es wird außerdem gezeigt, dass dieser Ansatz erstmals in der Lage ist, minimale Volumenänderungen im Zusammenhang mit progressiver Hyperventilation direkt zu quantifizieren, trotz der interventionell verfügbaren sehr niedrigen Bildqualität.

Aufbauend auf den vorgeschlagenen Methoden wird die Quantifizierung des Temoralhornvolumens zur indirekten Erfassung hippocampaler und parahippocampaler Atrophie im Rahmen serieller Studien vorgeschlagen. Bestehende Ansätze zur direkten Hippocampus-Volumetrie sind entweder zeitraubend oder unzuverlässig, so dass der vorgeschlagene Ansatz als schnelle und sensitive Alternative dienen kann. Schließlich wird der Temoralhorn-Index als normalisiertes Maß zur Erfassung parahippocampaler Atrophie vorgeschlagen.

Im Rahmen von Kooperationen mit Klinik und Industrie wurden die vorgeschlagenen Methoden erfolgreich anhand einer großen Vielfalt von Patienten und Probanden sowie anhand von Softwarephantomen evaluiert. Die in dieser Arbeit vorgestellten Methoden besitzen das Potential, die medizinische Diagnose und quantitative Therapiekontrolle zu verbessern und die Objektivität medizinischer Bildanalyse zu stärken.

Summary

THIS THESIS brings quantitative methods on complex medical image data closer to clinical routine, and bridges the gap between manual and automated processing. Efficient and effective image segmentation methods are described, and applied to currently insufficiently solved problems that are relevant for medical diagnosis and therapy monitoring. The methods are designed to be suitable for software assistants that meet the requirements imposed by a quantitative analysis in clinical imaging:

(a) *Applicability* regarding the required hardware and acquisition protocols, as well as an effective and intuitive user control, since a fully automated solution of the posed problems cannot be expected,

(b) *Robustness* to variations of image quality and also of anatomy and pathology in order to face the problems of routine clinical imaging,

(c) *Precision* of all derived quantitative measures,

(d) *Sensitivity* with respect to relevant changes of the examined structures, and

(e) *Efficiency* in order to be applicable to very large data sets on standard hardware.

The first part of this thesis describes novel methods that build upon and extend the morphological watershed transform that is known as the key concept for image segmentation in Mathematical Morphology.

The first problem addressed is the isolation of the brain from surrounding tissue in anatomical magnetic resonance (MR) images of the human head, also referred to as skull stripping. A fast modified watershed transform is described with linear computational complexity. A new parameter is introduced as preflooding height that globally controls the level of segmentation detail. The preflooding height forms the basis for the avoidance of oversegmentation, which is known to be the major problem of the classical watershed transform. The key idea in order to robustly solve the skull stripping problem is to use a fast modified watershed transform on the inverted original data. This is in contrast to the traditional application of the watershed transform to a derived discontinuity image.

The performance of the proposed approach is demonstrated on patient, volunteer, and phantom data.

Guidelines for the design of interactive medical image analysis systems are derived from the theory of effective human-computer interaction. The duality of interactivity on the one hand and automation on the other hand is discussed, and a concept of minimal interaction is proposed to combine the possibility of user guidance with the ability to generate largely user independent quantitative results. This is related to the concept of iterative interactive refinement, in which each single user input in combination with fast system feedback leads the system in a multi-scale fashion one step further toward the desired result.

The Interactive Watershed Transform (IWT) is proposed as a further extension of the described fast modified watershed transform in order to meet the requirements of minimal interaction. The IWT is the first variant of the watershed transform that is fast not only at the transformation step but also during user interaction and that allows real-time multi-scale user control on large multidimensional data sets. Iterative refinement is permitted through either placing an arbitrary number of markers or selecting global segmentation parameters, such as the preflooding height. It is demonstrated that, by analyzing the hierarchical region merging characteristics as recorded during the IWT, it is possible to automatically find the markers and preflooding height required to solve the skull stripping problem with a high robustness to image nonuniformity and noise.

Based on the IWT, a concept is proposed for the minimally interactive segmentation of bone structures in Computed Tomography (CT) data sets. Solutions to two problems are described: first, the efficient segmentation of carpal bones, which is a prerequisite for quantitative kinematic studies of the wrist, and second, the removal of osseous structures from large CT Angiography (CTA) data sets, also referred to as Bone Removal, which facilitates the analysis of vascular diseases or malformations. The first is challenging due to direct contact between neighboring bones, inhomogeneous cortical bone density, and highly irregular bone shapes. In addition to the IWT, the concept of Bit Morphology is proposed to generate solid bone masks at a high computational efficiency. The second problem is additionally challenging due to (i) the high diversity and irregular appearance of contrast enhanced vascular structures, (ii) a high intravascular image intensity similar to cortical bone tissue due to contrast agent, calcifications, stents, clips, etc., (iii) the large size of the original data acquired by current multi-detector CT systems, and (iv) the time constraints plus the need for effective and intuitive interactive control imposed by usability requirements. The key idea to solve the bone removal problem is to develop a compressed IWT plus a feature-based basin classification scheme for automated separation of osseous and vascular structures. After automated segmentation, the full interactivity and speed of the IWT is available for iterative refinement.

The second part of this thesis focuses on quantitative image analysis in general, as well as on the volumetry of specific neuroanatomical structures in particular. Morphological Volumetry is proposed as a framework for the reliable volumetric quantification of complex three-dimensional structures found in tomographic medical images, combining the extended watershed transform and automated histogram analysis. In order to ex-

plore the requirements of computer assisted volumetry in radiology and neuroradiology, a nationwide survey has been conducted of which the results are presented.

An overview of the Quantitative Imaging Pipeline, comprising both image acquisition and analysis, is provided and serves as a basis for a methodological investigation of the chain of measurement errors. Various categories of systematic and statistical errors are discussed, resulting in guidelines for the evaluation of respective image analysis methods. In the example of MR imaging, two major image characteristics that are relevant for quantitative medical image analysis—pixel noise and partial volume effects—are discussed and related to theoretical models. The challenges posed by volumetry of small anatomical structures are discussed on the basis of three examples.

Two further groups of problems are addressed in detail, namely the volumetry of brain tissue and of cerebrospinal fluid (CSF). For the accurate and precise volumetry of the intracerebral fluid spaces, the cerebral ventricles, a hybrid technique is proposed combining the IWT with automated model-based histogram analysis. This task is difficult due to a high inter-individual anatomical variability, especially in pathological cases, and the thin and elongated shape of the ventricles that impairs volumetric accuracy. The IWT is proposed to efficiently and effectively separate the individual ventricles from each other and also from iso-intense image contents, such as subarachnoid CSF that has to be excluded from further analysis. The histogram model is designed such that it adapts to regional image variability and robustly accounts for varying degrees of partial volume averaging. As a result, the proposed technique is fast, flexible, and precise compared to existing techniques. Since user input is only required by point markers to locate the objects of interest, but not to delineate their boundaries, the goal of minimal interaction could be combined with a high inter-observer reliability. A specific challenge is posed by pediatric imaging with very narrow ventricles, for which an extension of the histogram analysis is proposed. Results are presented on a group of preterm children with and without neonatal white matter damage. Furthermore, an ultra-fast method is discussed allowing for a reproducible volumetry of total intracranial CSF within a few seconds.

The described hybrid approach is extended and applied to the volumetry of cerebral gray and white matter, as well as the quantification of the Brain Parenchymal Fraction, which are challenging due to the complex folding patterns of the thin cortical layer with a high inter-individual variability. The proposed brain volumetry technique is simpler than existing approaches that rely, for example, on unsupervised iterative tissue classification, MR signal inhomogeneity correction, Markov Random Field modeling, and anatomical templates or atlases. The consistency of the proposed technique is demonstrated to be comparable or superior to two such existing techniques, EMS and SIENAX, with respect to repeated examinations, variable image resolution, nonuniformity, and noise, as well as in comparison between two different scanners. Additionally, the accuracy of the three techniques is evaluated using the Simulated Brain Database showing an advantage of the proposed brain volumetry approach. Furthermore, this approach is demonstrated to be capable for the first time to directly quantify subtle volume changes associated with progressive hyperventilation in an interventional imaging setting with very low image quality.

Based on the proposed methods, the quantification of the temporal horn volume is proposed for indirectly assessing hippocampal or parahippocampal atrophy within serial studies. Existing techniques for direct hippocampus volumetry are either time consuming or unreliable, such that the proposed technique can serve as fast and sensitive alternative. Finally, the Temporal Horn Index is proposed as a normalized measure for parahippocampal atrophy.

In cooperation with clinical and industrial partners, the proposed methods were successfully evaluated on a great variety of patient and volunteer images, as well as on software phantoms. The methods have shown potential to improve medical diagnosis and quantitative therapy monitoring and to enhance objectivity in medical image analysis in various instances.

No picture tells its own story.

—E. H. Gombrich

Contents

Medicine is not philosophy.

—Hippocrates of Cos, 460–c.370 BC

The best doctor is also a philosopher.

—Galen of Pergamum, 129-c.216

CHAPTER 1

Background and Introduction

CLINICAL RADIOLOGY is currently evolving as a quantitative discipline. At the same time, film-based radiology is about to be replaced by soft copy reading. Furthermore, the number of acquired *images per case* is rapidly increasing. For example, standard MR brain imaging protocols already include far more than a hundred images and multi-detector Computed Tomography (CT) scanners produce several hundred images for routine chest examinations and more than two thousand for runoff scans that cover chest, abdomen, and extremities.

While an experienced human eye undoubtedly has superior skills in reading images qualitatively, a well-instructed computer shows its strengths in reproducibly and unremittingly quantifying specific aspects of an image. Given the increasing amount of images, the increasing specialization and detail of radiologic problems, the increasing importance of quantitative parameters, and the stagnant number of radiologists, there is a rapidly growing need for software assistants that help the radiologist read these images and derive disease or problem related quantitative measures.

Since the number of acquired images often surmounts the capacities of a radiologist reading slice by slice, a computer system is expected to help the user to efficiently handle the amount of images and to extract diagnostically relevant information with minimum user interaction. In addition, software assistants that combine information from several slices into the third dimension can solve problems that are out of reach for conventional film-based reading, which is confined to two dimensions. Such a problem is the accurate quantification of the volume of specific structures, called *volumetry*.

This thesis intends to present efficient and effective image segmentation techniques and to demonstrate how these techniques may be integrated into algorithms that solve relevant open problems of quantitative medical image analysis. The idea of the morphological watershed transform will play a central role. The presented work has been guided by the above thoughts and is closely related to the scientific fields of quantitative radiology, as well as physics, i. e. image acquisition, on the one hand, and, most importantly, image analysis on the other hand. In that sense, the following pages describe the background, against which this thesis was written.

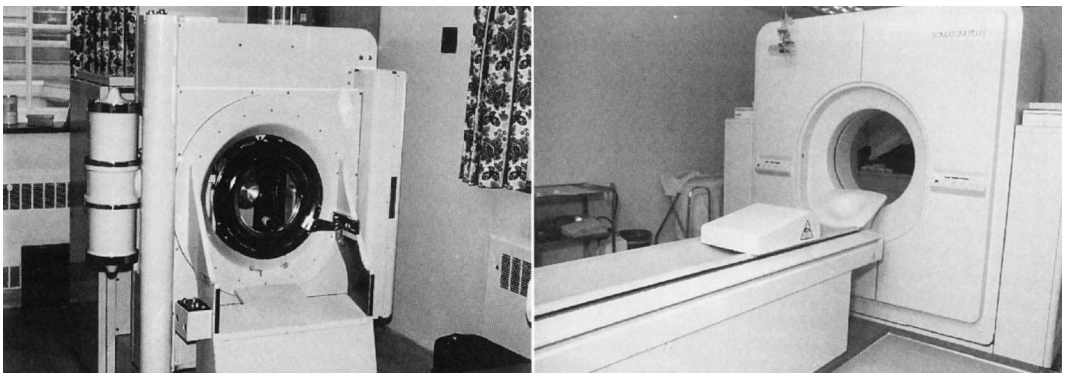


FIGURE 1.1: *left*: First generation CT scanner, about 1976. *right*: Later-generation CT scanner (SOMATOM PLUS, SIEMENS MEDICAL SOLUTIONS), about 1994 [both images from ORRISON *et al.*, 1995]. More than their exterior, the quality of the images acquired using these machines changed dramatically (cf. Fig. 1.2).

1.1 Quantitative Radiology

In March 1998[1], J. THRALL, Chief radiologist of the Massachusetts General Hospital in Boston, named three major topics for radiologic research in the early 21^{st} century as

(a) *Quantitative Imaging*,

(b) *Functional Imaging*, and

(c) *Molecular Imaging*.

While thousands of abstracts and papers are published each year on functional and molecular imaging, the first topic seems to be scientifically far less popular. This might be due to the fact that traditionally radiologists view their profession as one of recognizing pathologically relevant patterns rather than delivering quantitative measures. Another reason for this lack of scientific attraction might be that for many years radiologic methods were not well suited for quantitative measurement tools other than for measuring lengths, diameters, or angles. The usage of such tools was and has been regarded as an annoying task and often left to technicians or referring physicians [KLOSE, 2002]. The most important reason for the relatively slow progress of quantitative imaging, however, is that it is extremely difficult to control all influencing variables in order to yield reliable and significant measurements.

Consequently, *Quantitative Radiology (QR)* is not yet a common term. A query of the PUBMED database[2] yielded four matches on 16 March, 2003. While the first is dated 1976 [DEQUEKER, 1976], the most recent paper with QR in its title deals with a rather old-fashioned technique, namely the application of the classical concept of Stereology to recent problems of MR-based volumetry [ROBERTS *et al.*, 2000]. A query to GOOGLE[3], which can

[1] At the *Management in Radiology* congress held in Strasbourg, France.

[2] Search term: `"quantitative radiology"[title]`

[3] Search term: `"quantitative radiology"`

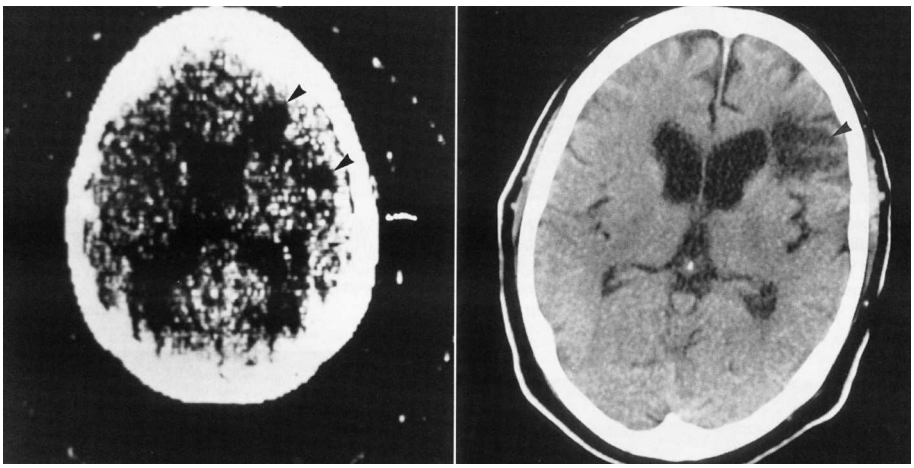


FIGURE 1.2: *left*: CT image from first-generation scanner (cf. Fig. 1.1 left), about 1976. *right*: CT image from later-generation scanner (cf. Fig. 1.1 right). In both images, focal areas of cortical and subcortical loss can be discerned (arrows), probably representing ischemic disease. [from ORRISON *et al.*, 1995]

be regarded as a modern source of knowledge [PAGE and BRIN, since 1998], on the same day yielded less than twenty-five serious web sites independently mentioning QR. None of these sites were truly dedicated to QR or even contained the term in their titles. These low numbers are surprising as today there are variety of quantitative applications based on radiologic images. We propose to use the term *Quantitative Radiology* to denominate what will be detailed on the following pages.

Knowledge, Understanding, and Interaction

"To quantify is to know [KLOSE, 2002]." This can be interpreted in two directions: firstly, that knowledge about the nature of an object's property forms the basis for its scientific quantification, and secondly, that quantification following the rules of science potentially accumulates to unbiased knowledge. However, one might like to extend the above citation. Abuse of statistical tools in the biomedical sciences teaches us that appropriate quantification requires a considerable amount of critical awareness. In a positive sense, we claim that to quantify is *to understand*.

Turning toward images, note that their visual presentation cannot be replaced by single numbers, but rather complemented. Scientific visualization provides insight into complex data, while quantitative image analysis (QIA) aims to provide reliable measures as a basis for objective comparison. In quantitative analysis of medical diagnostic images, it is unrealistic to expect a fully automated method to work in one hundred percent of all cases, solely due to the enormous natural morphological variation of the human body. Therefore, we posit that at least inspection of automatically generated results is imperative. Including user-steered correction in some percent of the cases, we also argue that to quantify means *to interact*.

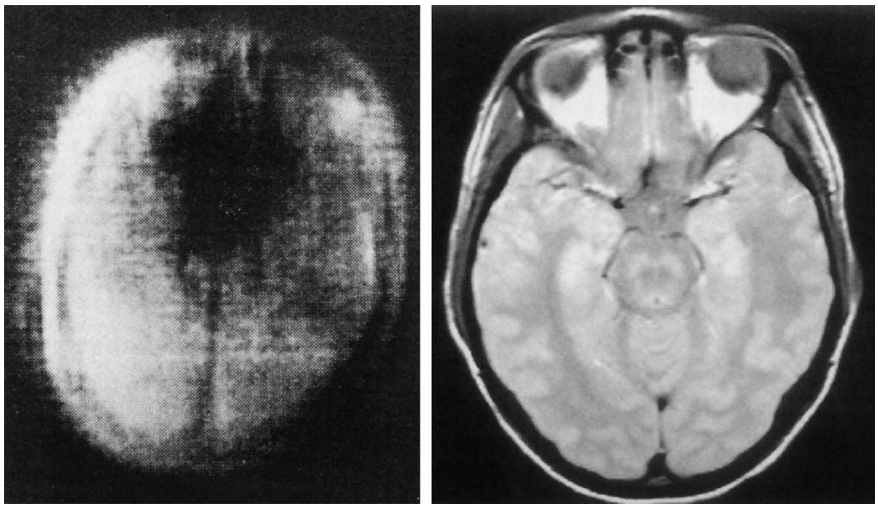


FIGURE 1.3: *left*: The first published NMR reconstructed image of a head [from HOLLAND *et al.*, 1980]. *right*: MR image of a head, about 1993. Although approximately the same weighting is used as in the left image, system and sequence improvements have resulted in a better image being produced in less time. [from ORRISON *et al.*, 1995]

Birth of Quantitative Radiology

It is difficult to identify the date of a change's beginning. In the 1960s, radiologists paid great attention to a forthcoming technique that was later named Computed Tomography (CT) or Computerized Axial Tomography. The initial motivation for CT, when developed by ALLAN M. CORMACK in Cape Town in 1958, was to achieve individual absorption rates for the planning of radiotherapy. Ten years later, GODFREY N. HOUNSFIELD at the EMI laboratories intended the same method for nondestructive material testing. HOUNSFIELD's first experiments utilized a gamma-ray source and required nine days for data acquisition of a single slice; processing the data required over two hours on a large mainframe computer. The scanning time was later reduced to nine hours by replacing the gamma-ray source with an x-ray tube. By the time the first clinical prototype CT scanner was introduced in 1972, imaging and processing times had been reduced to some minutes [HOUNSFIELD, 1973]. Computed tomography seemed to be the first technique not only to resolve the four dimensions of space and time but also to quantitatively measure properties of materials (e. g., their physical density) within the four dimensions and with an exactness that would be sufficient for clinical purposes. An outstanding example for this ability was spinal *bone densitometry* pioneered by WILLY A. KALENDER in Erlangen [KALENDER *et al.*, 1989]. Today, most radiologists associate the term *Quantitative CT (QCT)* with KALENDER'S technique and possibly to forearm bone densitometry, both facilitated through dedicated imaging systems [KLOSE, 2002, ORRISON *et al.*, 1995].

Much has happened since 1972. Remarkable technological advances have occurred since the introduction of the first clinical CT scanners. These advances have resulted in tremendous improvements in both imaging times and image resolution (Fig. 1.1 and 1.2).

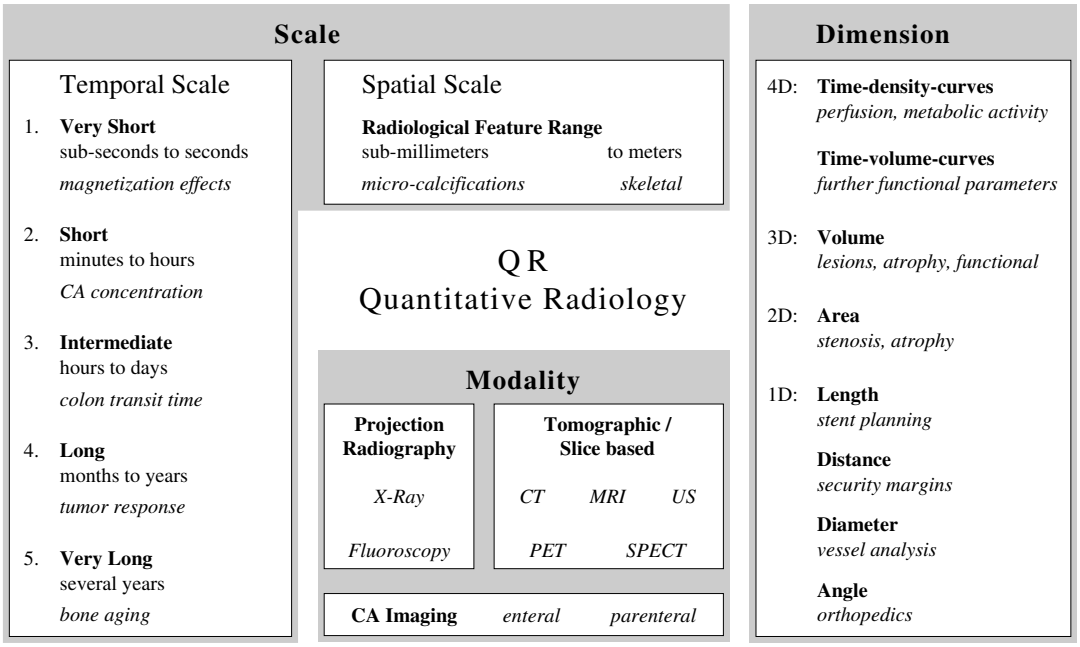


FIGURE 1.4: The field of Quantitative Radiology as defined by *Scale*, *Dimension*, and *Modality* (CA: contrast agent). Examples for different categories are given in *italics*.

The introduction of clinical magnetic resonance imaging (MRI) in the 1980s completed the transformation toward modern radiology. Since then, in addition to densitometry, a whole new world of imaging contrasts has been made available based on MRI (Fig. 1.3). The first published magnetic resonance (MR) image was by PAUL C. LAUTERBUR in 1973 [LAUTERBUR, 1973]. Yet, it took four more years until the first human MR images were published by DAMADIAN *et al.* [1977], and much longer until MRI was a suitable basis for quantitative measurements.

What we propose to refer to as Quantitative Radiology, actually had begun as various quantitative techniques in conventional radiography. Several textbooks were published in the 1960s, of which *'the Keats'* (recent edition e. g. [KEATS and SISTROM, 2001]) has probably been the most famous advocate. Since then, the field of QR has undergone dramatic developments and has diversified, now to be an integral part of various radiologic subdisciplines. Today's QR covers various questions of the form "how much?", which are summarized in Figure 1.4 according to scale and dimension of the measured quantity and to available imaging modalities [KLOSE, 2002].

Potential of QR

Before the advent of computational image analysis, the paradigms that are the subject of the present thesis were unthinkable. It is instructive to relate the different modalities on the one hand and the categories of dimension and scale on the other (cf. Fig. 1.4). Classically, one-dimensional measures are assessed by means of projection radiography. While one and two-dimensional measures are accessible from all imaging modalities, full

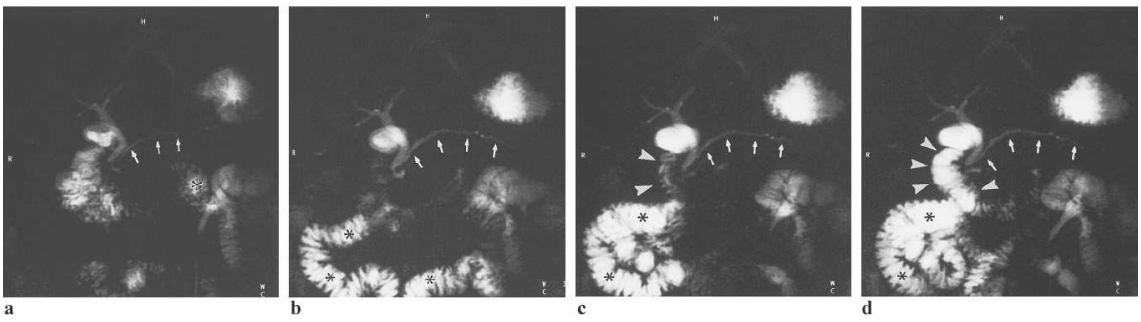


FIGURE 1.5: Example of QR: MR cholangiopancreatography images ($TR = \infty, TE = 1,100$) obtained in a 28-year-old man with suspected reduced exocrine pancreatic function (a) before stimulation with secretin and (b) 5.0, (c) 8.0, and (d) 9.5 minutes after secretin stimulation. *a—d*: After secretin stimulation, the main pancreatic duct (arrows) is consistently better visualized. A continuously increasing filling of the duodenum (arrowheads in c and d) and jejunum (*) with fluid can be seen. This fluid increase was quantified by image postprocessing under the assumption of linearity of a pixel's gray value with respect to the amount of fluid contained in the pixel. (Images courtesy of J. H. HEVERHAGEN [HEVERHAGEN *et al.*, 2001])

three-dimensional imaging is realized by tomographic techniques, such as, most commonly, CT and MRI. The following examples shall illustrate this potential.

Time-dependent characteristics, as well as functional parameters of organs and tissues are primarily examined using contrast agent (CA). Common examples are perfusion imaging using contrast enhanced CT or MRI and imaging of metabolic activity by Positron Emission Tomography (PET) (e. g., Glucose uptake using ^{18}F-FDG as a marker). Further functional scores are assessed via time-series of morphological images without CA, e. g. cardiac ejection fraction.Within neuroimaging, the term *functional imaging* primarily relates to Functional MRI (fMRI) that is facilitated by the Blood Oxygen Level Dependent (BOLD) effect that was discovered by S. OGAWA[4] in the late 1980s [OGAWA *et al.*, 1990]. While in the field of functional imaging, computational image analysis and quantification played a central role from the very beginning, this is not true for anatomical and morphological imaging where qualitative evaluation still dominates the clinical practice.

The three major slice-based imaging methods yield specific fields of application, according to their physical nature:

(a) Ultrasound (US) for sound conductance and absorption rates, as well as flow imaging (Doppler US).

(b) Computed Tomography (CT) for physical density and x-ray absorption.

(c) Magnetic Resonance Imaging (MRI) for proton density and magnetization properties of various tissues and contrast agents.

Hydrometry is a currently developing technique based on MRI. As described by HEVERHAGEN *et al.* [2001, 2002], MR hydrometry is applied, for example, to the non-invasive

[4] At the AT&T BELL laboratories, Murray Hill, NJ.

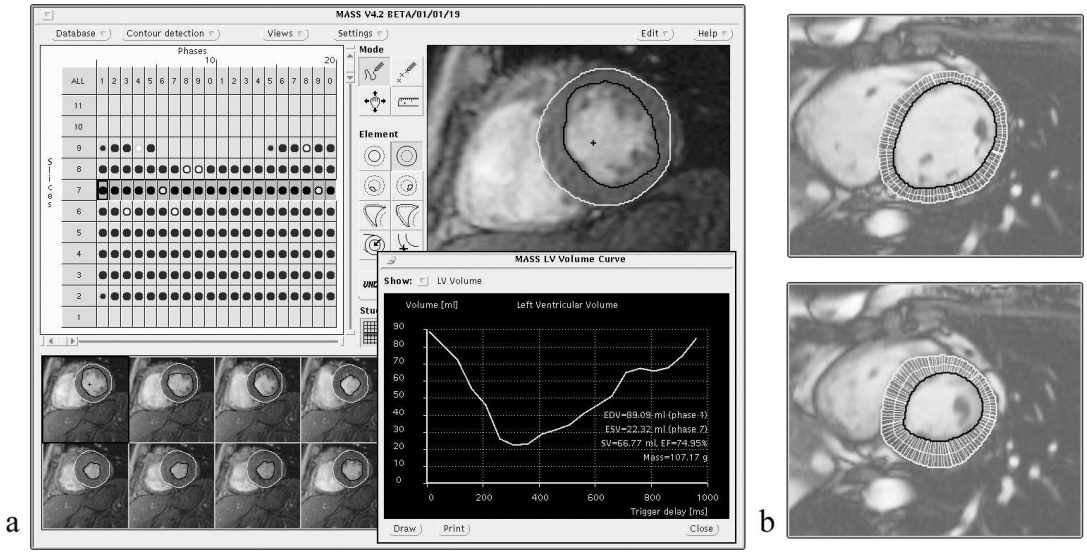


FIGURE 1.6: Left Ventricular Analysis as well-known example of Quantitative Radiology. *a*: Screen layout of the MASS software package (MEDIS, Leiden). *b*: Wall thickness measurements in an end-diastolic (top) and end-systolic image (bottom). (Images courtesy of R. J. VAN DER GEEST, University Medical Center, Leiden)

quantification of the exocrine pancreatic secretion and function. MR cholangiopancreatography images are obtained before and after stimulation with secretin. After secretin stimulation, a continuously fluid increase is observed in the images and can be quantified under the assumption of image linearity (cf. Fig. 1.5).

Established quantitative methods within the field of cardiovascular imaging include tools for Quantitative Coronary Arteriography (QCA) and Quantitative Coronary Ultrasound (QCU), as well as Quantitative Vascular Arteriography (QVA) for the analysis of peripheral arteries. These applications are supported by a large number of publications [e. g. REIBER *et al.*, 1989, 1997, DE SCHEERDER *et al.*, 1993] and also commercial products—two major relevant companies are MEDIS[5] and PIE MEDICAL IMAGING[6], both located in the Netherlands.

More recently, the classical approach of Quantitative Left Ventricular Analysis (QLV) based on ventricular angiographic images was complemented by two- and three-dimensional MRI-based image analysis methods [VAN DER GEEST *et al.*, 1997a,b, LELIEVELDT *et al.*, 2001, STALIDIS *et al.*, 2002] (cf. Fig. 1.6). The latter is an exponent of a larger group of image analysis methods that has been developed over the last decade, namely the quantitative analysis of anatomical structures using statistical shape models (cf. p. 13).

[5] `http://www.medis.nl/`
[6] `http://www.piecaas.com/`

Planimetry and Volumetry

Returning to the focus of this thesis, we ask the following question: "For which of the above mentioned quantitative applications of QR does *image segmentation* play a crucial role?" Among the categories in Figure 1.4, it is apparent that two and higher dimensional measurements require some sort of segmentation, while linear measurements, e. g. lengths and angles, can be performed manually on the original images.

For area measurements, on the one hand, object segmentation is mostly equivalent to manual drawing on single cross-sectional images; the resulting image analysis times are acceptable. Still, it must be questioned if the resulting area measurements are reliable, for example depending on the complexity of the segmented shape.

Volume measurements, on the other hand, are still a challenge to both image acquisition and analysis, and a tedious task for most radiologists, even if it is widely accepted that volumetry could substantially contribute to objective and sensitive image-based diagnosis or therapy control. The lack of efficient and problem-adapted segmentation techniques represents the main obstacle for a broad clinical application of volumetric measurements such as for tumor follow-up examinations [Schwartz *et al.*, 2000, Prasad *et al.*, 2002]. This thesis is dedicated to understanding and solving *problems* of quantitative volumetric image analysis in medicine. The introductory section of Part II presents results from an enquiry on the *requirements* of clinical volumetry (cf. pp. 85 ff.).

1.2 Physics

The process of digital image acquisition is nothing else than a physical measurement where each pixel or voxel is attributed a number being related to some aspect of real world in a complex fashion. In order to build, e. g., the *one-click volumetry* tool for the quantification of tumors and other clinically relevant lesions, however, the questions of feasibility and reliability have to be handled with care (cf. comments on evaluation criteria, p. 91). Various sources of error are associated to the physics of image acquisition and have to be identified and discussed.

Numbers and Errors

"Everything is number"—a frequently quoted doctrine attributed to Pythagoras. Already for the ancient Pythagoreans, "there was a motivating concern for *quantification* and the treatment of numbers as values (i. e., as a way of dealing with *qualities*)." From this it is only "a short conceptual step ... to the more modern concept of *measurement*" related to the "idea of a (numerical) value as something inherent in what is being measured." [Rosen, 2000, p. 65]

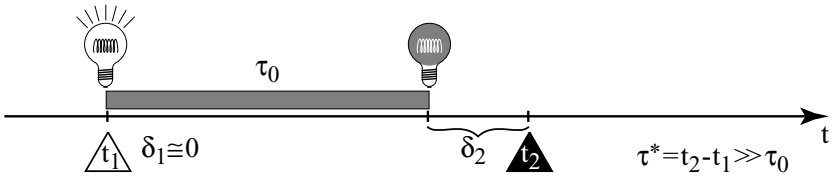
Radiologists, or medical doctors in general do not rigorously identify and analyze the specific nature and origin of measurement uncertainties. Instead, measurements based on images are commonly accepted as uncertain per se. In order to analyze this uncertainty, statistical evaluations over a series of many measurements are performed. Physics, on the other hand, would be characterized as the world of image acquisition, of radiation

protection, and of the process of quantification itself. But physics should go further than that. A standard exercise for every physicist is the estimation of measurement errors, which could be of key value when applied to clinical diagnosis. Still, most clinically oriented image analysis systems offer quantification without any sort of error estimation.

Learning from 1D—Statistical and Systematic Errors

To illustrate the difficulties of quantification, we would like to recall the difference between statistical and systematic errors. Consider a simple experiment consisting of a lamp, a timer, and a stopwatch. After the lamp was turned on, the timer ensures that the lamp will go out after a specific delay time τ_0. Let us assume that τ_0 is *exactly* the same interval for all experiments. The task of an observer is now to measure τ_0 using the stopwatch. The stopwatch's start/stop-button is pressed at times $t_i, i \in \mathbb{N}$. The resulting time of a given measurement is denoted $\tau^* = t_{2n} - t_{2n-1}, n \in \mathbb{N}$. The delays δ_i refer to the intervals between an actual lamp-on/off event and the corresponding action t_i, i. e. to the primary measurement errors.

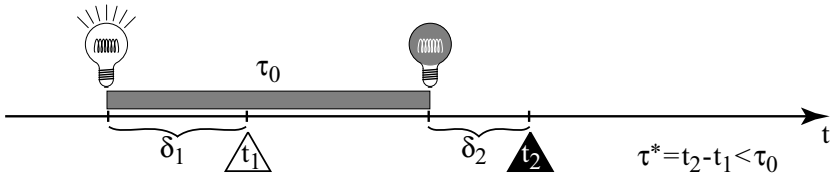
We now consider three experimental setups that are different only in the mode by which the lamp-on time is determined. In setup A, the observer manually starts both the lamp and the stopwatch at the same time t_1:



(Setup A)

The corresponding lamp-on delay can thus be regarded to be very small related to the remaining measurement uncertainties ($\delta_1 \cong 0$). In contrast, the lamp-off delay δ_2 is governed by the reaction time of the observer.

Before discussing setup A, we introduce setup B that differs from setup A in that a second person chooses the lamp-on time arbitrarily and at an unexpected moment for the observer:



(Setup B)

Here, both lamp-on delay δ_1 and lamp-off delay δ_2 are determined by respective reaction times. For both setups, the measurement time can be expressed as

$$\tau^* = t_2 - t_1 = \tau_0 + \delta_2 - \delta_1 \ . \tag{1.1}$$

Let us now talk about *statistical* and *systematic* errors. Of course, τ^* will show some variability across repeated experiments. Assuming that the experiments are statistically independent, and if the measurements are conducted in the same manner, a Gaussian is a good approximation for the resulting distribution of τ^*. The width of this distribution, as expressed by the standard deviation σ, is then an expression of the statistical errors

inherent in this specific measurement setup including the observer. Thus, statistical errors are easily revealed by means of independently repeated measurements. Comparing the setups A and B, we would find that the statistical error is greater in setup B where the total variability is composed of two variabilities at the start and end of each measurement; in setup A, only δ_2 shows a relevant variability. The (statistical) error of τ^* in setup B can be estimated using the laws of error *propagation* for the sum/difference of two (independent) variables:

$$\sigma(\tau_B^*) = \sqrt{\sigma(\delta_1)^2 + \sigma(\delta_2)^2} \quad > \quad \sigma(\delta_2) \cong \sigma(\tau_A^*) \; . \tag{1.2}$$

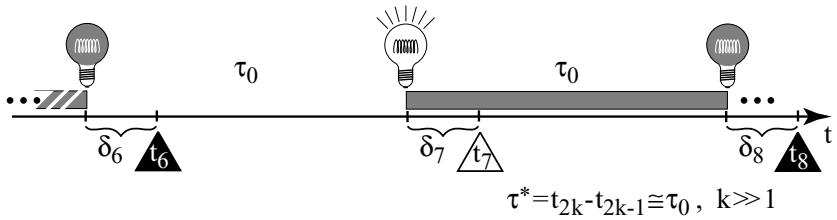
In general, if $x = f(u, v, \dots)$ and if $u, v, \dots$ are independent, the variance of x can be expressed as

$$\sigma_x^2 = \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)^2 + \dots \; . \tag{1.3}$$

More difficult to reveal, to quantify, and thus to correct are systematic errors. These errors, often referred to as *bias*, are inherent in a measurement setup and lead to situations where the mean value μ of consecutive measurements significantly rises above or falls below the underlying *true* value. In setup A, the asymmetry of δ_1 and δ_2 results in a systematic *overestimation* of τ_0.

In setup B, no systematic errors appear at first sight, since both measurements t_1 and t_2 are governed by the observer's surprise. A closer look, however, reveals that the lamp-off event is less surprising than the lamp-on event; after some experiments, the lamp-off event is anticipated due to the constant lamp-on interval τ_0. This leads to a systematic *underestimation* in setup B.

A good strategy to reveal systematic errors without available ground truth is to compare results from different measurement techniques—such as volume measured independently by MRI and by water displacement. But even if such an independent measurement is available, one should consider all factors that can jeopardize an unbiased measurement of a single method. Returning to the lamp experiment, we want to question whether it is practically feasible to unbiasedly estimate the *true* lamp-on time with a stopwatch. The only method of doing so is to *balance* the observer's surprise at lamp-on and lamp-off events. One possibility is to conduct periodically repeated measurements where all lamp-off intervals have exactly the same length as the lamp-on interval (Setup C):



(Setup C)

Still, one might argue that lamp-on events are processed faster than lamp-off events by the observer's visual system. Furthermore, the lamp's color certainly does make a difference for the processing speed within the primary visual system. We consider these two issues, however, beyond the scope of this chapter.

Standardization and Calibration

Let us now turn again to medical imaging. Images, on the one hand, are smooth mappings $\Omega \to \mathbb{R}$, and digital images, on the other hand, are sets of pixels related to a digital grid D, most often represented by discrete gray values $I : D \longrightarrow \mathbb{Z}$. While the continuous case often serves as a basis for algorithmic definitions, it is the discrete case that will receive central attention in this thesis. In many cases, it is just the sampling step leading from continuous to discrete that causes major problems—a common example is the acquisition of anisotropic volume data by most tomographic imaging protocols; here, anisotropy relates to the fact that voxel spacing is greater between slices than within.

Considering the physics of image acquisition, it is interesting to note that for CT a standardization of the intensity scale was achieved early on by the introduction of Hounsfield Units (HU). In contrast, only during the last years have there been similar efforts with respect to the MR intensity scale [e. g. Nyul and Udupa, 1999]. The reason for this lies in the facts that (a) MR offers a variety of possible image contrasts, while CT relates to a single physical property (i. e. the x-ray absorption), and (b) MRI is susceptible to electrodynamic field inhomogeneity, which are rather the rule than the exception.

At first glance, a standardized intensity scale such as HU seems to be an indispensable prerequisite for valid quantitative image analysis. Instead of intensity standardization, however, we argue that image analysis systems—also for CT data—must (semi-)automatically adapt to intensity variations.

The combination of (a) and (b) gives rise to an important topic that is frequently addressed in MR image analysis, namely the correction of image artifacts due to physical fluctuations and noise, as well as static and dynamic field inhomogeneity [Sled *et al.*, 1998, Mangin, 2000, Likar *et al.*, 2001, Arnold *et al.*, 2001]. A consequent approach to QIA is to design methods so that they are insensitive to such variations and require neither a standard intensity scale nor homogeneous image intensities. This property, also called *robustness* or *adaptivity*, is crucial for image analysis systems that are intended for widespread clinical deployment. In this sense, calibration is not an obligatory prerequisite for properly designed image analysis systems.

1.3 Image Analysis

On the whole, software systems potentially have three abilities of which we take advantage for medical imaging, namely to support and perform:

(a) *Quantification*—as opposed to qualification where humans generally are better

(b) *Derivation*—such as of parameter images from large amounts of data

(c) *Visualization*—of complex information, often in three or higher-dimensional space.

Being dedicated to the first, i. e. to performing measurements, we are projecting nature to analytical models and thus to numbers (cf. Quantitative Imaging Pipeline, Fig. 6.1). Note that a digital image already consists of numbers, however, without an analytical model

behind. These primary numbers we refer to as *data*. To derive meaningful and valuable quantitative information from these data, *models* are required that describe the image intensity associated with certain tissue types and physiological conditions, and possibly the anatomical shape and location of organs and other structures.

There have been publications of a plethora of image analysis methods for virtually every medical application. Still, working in the field, we feel that there is yet much to do.

Segmentation

Image segmentation is one of the large fields within image analysis, and aims toward identifying regions, so-called segments that have a specific *meaning* within images, for example represent an organ or an anatomically defined part thereof. Another definition of image segmentation is identifying regions that are *uniform* with respect to some parameter, such as image intensity or texture. While the latter is often used for technical reasons, the former definition should be preferred from an application point of view. We would like to make the distinction between two groups of image segmentation methods:

(a) First, methods that try to *classify* segment membership of an image element, e. g. pixel or voxel, based on its or its closest neighbors' *image intensity*. The oldest member of this group is simple thresholding. Higher level variants are the hard c-means and also the more recent fuzzy c-means clustering algorithms [Pham and Prince, 1999]. Providing a statistical approach to the same problem, the Expectation Maximization (EM) segmentation algorithms that have become quite popular during the last decade also belong to this group [Wells *et al.*, 1996, Van Leemput *et al.*, 2003]. Frequently, some sort of Markov Random Field (MRF) is incorporated into the EM approach in order to correct the class membership of misclassified isolated voxels by taking statistical neighborhood information into account. For example, such misclassifications frequently occur due to noise. The pioneering work of D. H. Laidlaw goes one step further, modeling sub-voxel histograms in order to correctly classify image elements close to object borders where partial volume effects (PVE) occur [Laidlaw, 1995, Laidlaw *et al.*, 1998].

(b) Second, a whole group of problems remains unsolved by methods that rely on local intensity information. This group can be described by *separating* specific image segments from other segments that exhibit *same or similar* image intensity characteristics (gray level, texture, etc.). Approaches to these problems usually are either contour or region based. The oldest and most basic member of this group is manual contour drawing, a technique that seems to solve every segmentation problem. To reduce interaction and time, region growing methods have been introduced that start expanding regions from initial seed points until a stopping criterion is met. Since region growing often results in jagged segmentation boundaries and to further automate the segmentation process, deformable templates were introduced, such as snakes or balloons [Kass *et al.*, 1988, Cohen and Cohen, 1993]. An extension to deformable templates is the incorporation of explicit *shape priors*, i. e. a-priori constraints on the object's shape. The *live wire algorithm* offers efficient interactive contour drawing using image-based cost functions without shape priors [Barrett and Mortensen, 1997]. Schenk *et al.* [2000] propose to overcome the problem

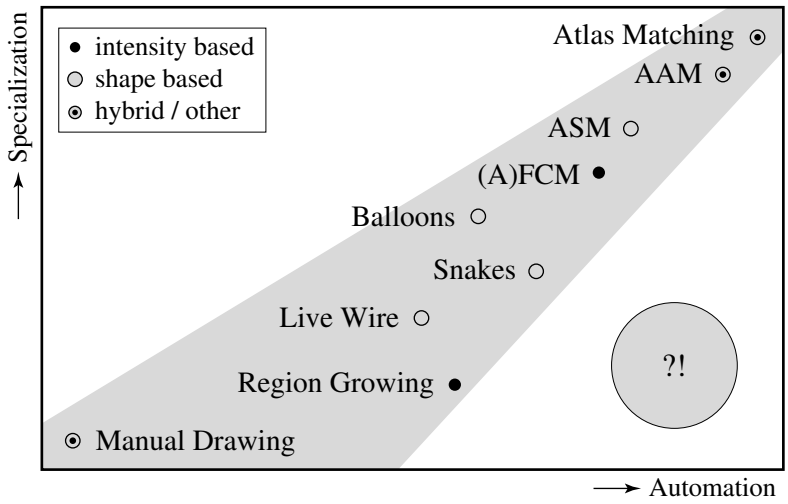


FIGURE 1.7: Qualitative chart on established image segmentation methods according to their degree of *automation* and *specialization*, which has served as motivation for this thesis. Note that these two attributes seem to be coupled—there is a void at the lower-right corner of the chart: Currently, no established segmentation method seems to exist that is highly automated but unspecialized.

of two-dimensional interaction by a combination of the live wire algorithm with shape-based interpolation resulting in significantly reduced interaction times. An automated template-based segmentation method that is neither region nor contour-based consists of matching individual images to atlases, i. e. standardized, pre-segmented, and labeled images. Active Appearance Models (AAMs) can be regarded as statistical atlases that, in contrast to Active Shape Models (ASMs), also account for image intensity [COOTES *et al.*, 1995, 2001]. A further approach that aims toward statistical description and segmentation of single objects was described by PIZER *et al.* [2001].

While the first group of methods is well described by the term *classification*, it is the second that we associate with the term *segmentation* in the narrower sense. It is common to all the above mentioned methods that the advancement of automation also involves an increasing degree of specialization. A qualitative chart of several methods is provided in Figure 1.7. It has been one of our motivations to fill the void at the lower-right corner of the chart, i. e. providing a highly automated segmentation method at a low degree of specialization. While incorporating knowledge of the shape or appearance of target objects into an algorithm leads to specialization, the direct utilization of user knowledge results in more generic methods. Efficient interaction is thus a key to broadly applicable segmentation tools. Furthermore, if interaction is well designed, a full automation for a subgroup of problems based on the generic approach comes into reach. Note that, for example, the AAMs fall somewhere in between the two strategies: Expensive interactive segmentation of a representative training set is required during a learning phase, while the knowledge acquired by this means will later serve as a basis for fully automated segmentations.

Beyond Segmentation

Even if image segmentation is an important step in the quantitative image analysis pipeline, in most cases it is not enough to answer the question on "how much" is, e. g., the width, size, or volume of the segmented object. Of course, the size of a single image element has to be taken into account. However, uncertainties remain after segmentation with respect to the object's boundaries. It is there that partial volume averaging occurs due to the limitations of digital image acquisition. Figure 1.8 illustrates these problems with a blurred and noisy image of three arrows and results in an intuitive approach to gradient and histogram-based image analysis.

One strategy to measure the width of the arrows is to analyze an intensity profile $I(x)$ perpendicular to the arrow axes, much as shown in Figure 1.8 a. Since each of the objects exhibits a uniform mean intensity, object boundaries are associated with zero-crossings of the profile's second derivative. This is equivalent to finding maxima of the profile gradient magnitude $|\nabla I(x)|$. A low-pass filter is commonly realized by convolving $I(x)$ or the gradient operator ∇ with a Gaussian of width σ. The first problem of such an approach is to find a proper filter width by which the number of gradient maxima above a certain threshold corresponds to the number of edges to be detected: For small filters—corresponding to high spatial frequencies—too many maxima are found, while for large filters neighboring maxima tend to merge (cf. Fig. 1.8 b–d). A second problem of a gradient-based measurement is that, depending on noise, sharpness, and filter width, gradient maxima are located more or less far from the true boundary location; it is non-trivial to quantitatively describe the reliability of measures derived from gradient maxima computation.

It is instructive to look at this image analysis problem from a different perspective, namely to examine the profile histogram $N(I)$ (Fig. 1.8 e) revealing the frequency of certain image intensities. If objects are represented by characteristic intensities, then on a regular grid, object sizes can be derived from the respective frequencies. For the given profile, three main peaks are observed corresponding to background (centered around $I \cong 80$), bright arrow ($I \cong 50$), and dark arrows ($I \cong 20$). Between the peaks, the histogram does not drop to zero, but rather shows plateaus, which are characteristic of the occurrence of PVE. The height of a plateau inversely corresponds, much as the width of the gradient peaks shown in Figure 1.8 d, to the spatial sharpness of the image. In contrast, the width of the histogram peaks is directly related to the amount of noise in the respective tissue (arrow) classes, which again corresponds to the proper gradient filter width, as discussed above. We will return to this figure when discussing the use of thresholds, gradients, and histograms for QIA.

1.4 Motivation

The goal of this thesis is to provide image segmentation methods that are *robust, fast, and flexible* enough to meet the requirements of a clinical setting. Furthermore, the methods are intended to serve as a basis for *reproducible* and *unbiased* volume measurements. Our

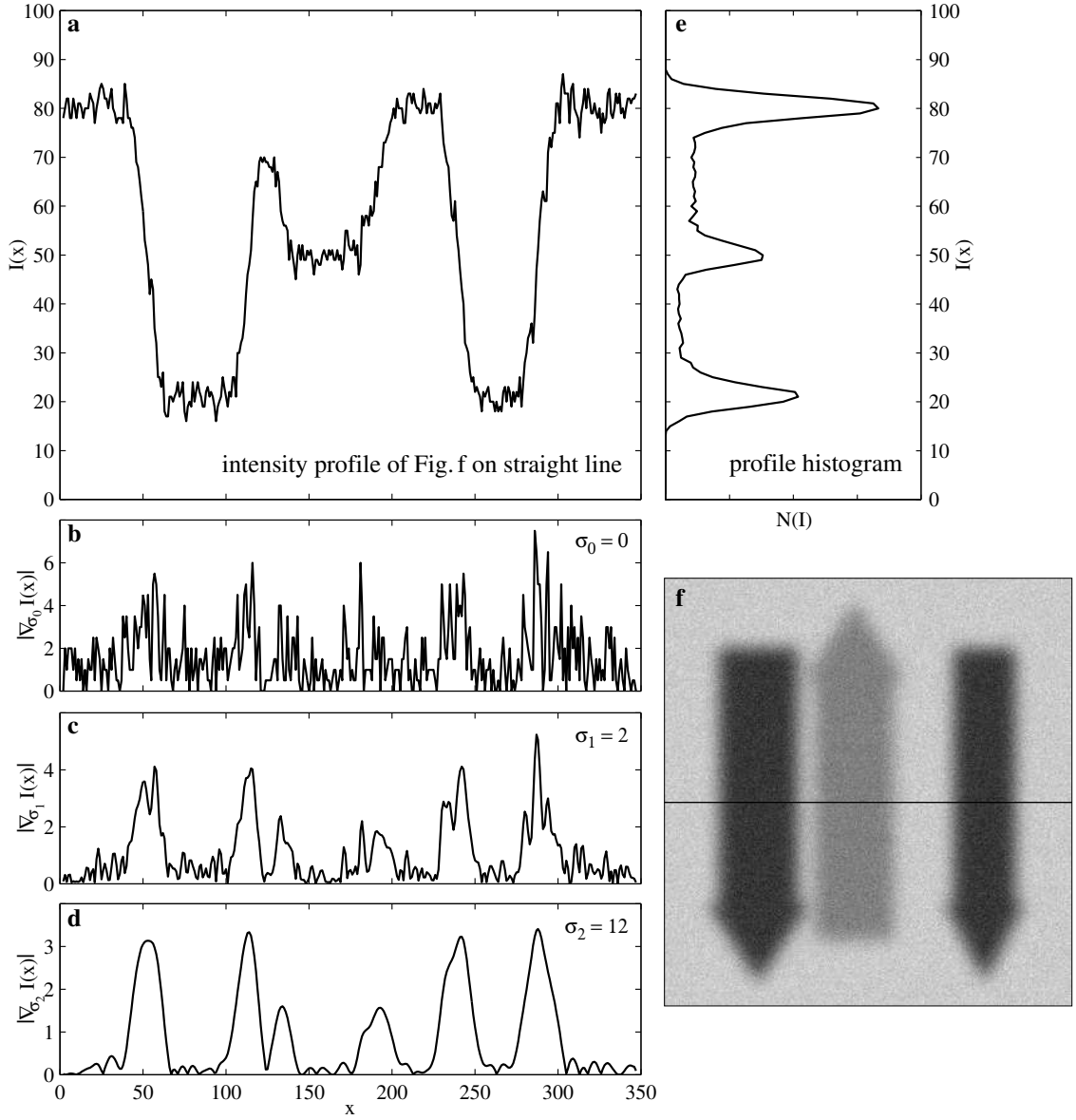


FIGURE 1.8: *Duality of Gradient- and Histogram-Based Image Analysis.* Test image for one-dimensional measurement of object size consisting of three arrows. An intensity profile perpendicular to the arrow axes yields information on the arrow widths (a). Profile gradients were computed for three different filter widths ($\sigma_i, i = 0, 1, 2$; b–d). Orthogonal to the gradients, the profile histogram is plotted (e). To generate the test image, a perfectly sharp image of the three arrows was smoothed by a Gaussian kernel and distorted by Gaussian white noise (f).

main motivation originates from the field of quantitative neuroradiology, but the concepts and methods are also applicable to other fields.

Specific Objectives

Three major objectives from the perspective of quantitative medical image analysis are summarized in Figure 1.9. If one of these is not met, the respective method will not be a good candidate for clinical deployment.

(a) Efficiency of the developed methods is crucial since (i) visual inspection is too subjective and observer dependent to yield a reliable quantification and (ii) the number of images to be quantified is too high for manual processing. In contrast to manual processing, automated and semiautomated methods can work on multidimensional data sets without being constrained to 2D representations. In addition to speed, robustness to even severe pathological variations is a prerequisite for the *applicability* of a method. Since no automated method is able to cover all of these variations, interactivity is often essential for clinical use, e. g. in terms of verification and correction of segmentation results.

(b) *Reproducibility*, which is the second main objective, relates to the precision of all derived quantitative parameters. Of course, we would like to demand both accuracy and precision of these parameters, describing their absolute and relative correctness, respectively. However, only the latter can be evaluated by systematic a-posteriori statistical analysis, since in most cases no ground truth will be available. In many cases, a relative variance of less than five percent will be sufficient to derive clinically relevant observations, such as of tumor growth. The upper limit, though, depends on the given application. For example, in order to monitor the progress of brain atrophy in neurodegenerative diseases, such as Alzheimer's disease (AD) and Multiple Sclerosis (MS), one should be able to reliably quantify relative brain volume changes of less than one percent. In contrast, accuracy, which corresponds to the absence of systematic errors, is often accessible by phantom measurements [Collins *et al.*, 1998, Blake *et al.*, 2005, Rexilius *et al.*, 2005, Schlüter *et al.*, 2005] and also by careful consideration, but not by statistical analysis.

(c) Finally, *availability* of a given image analysis method is crucial for its dissemination in clinical routine. Image analysis tools, unless provided by the image acquisition device manufacturer, should run on standard hardware, which is most commonly a Microsoft Windows or Linux PC. Even more importantly, the availability of a quantitative method can be restricted by the use of highly specialized image acquisition that is not featured by standard image acquisition devices. Therefore, throughout this thesis we rely on image acquisition techniques that correspond to what is available on standard scanners.

In order to comply with these goals, our central idea for the development of a volumetric image analysis tool is to combine (i) an automated model-based statistical analysis with (ii) flexible and effective interaction (if required) and morphological segmentation. We refer to this combined concept as Morphological Volumetry. While the first part is represented by a segmental[7] histogram analysis, the second refers to the proper

[7] *segmental*=regional with respect to specific image segments.

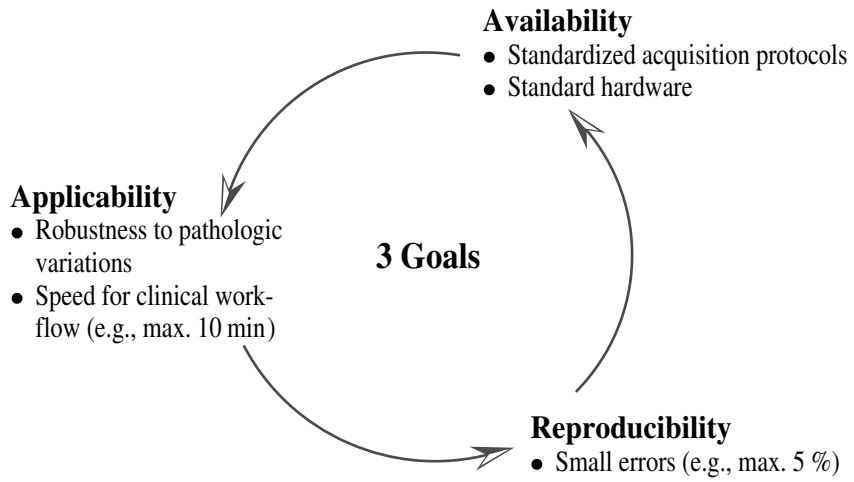


FIGURE 1.9: *Circle of Three Goals* of clinical quantitative image analysis.

segmentation step. For the development of new segmentation methods, we attempt to meet the following requirements, which we consider essential:

- ▷ To be *robust* even to considerable image artifacts of low (bias field) and high spatial frequency (pixel noise),
- ▷ To be *image-based*, i. e. to adhere as closely as possible to the original image information,
- ▷ To comprise user *interaction* techniques for efficient and intuitive control and correction of results,
- ▷ To rely on the *connectivity* of objects, but not on a-priori knowledge about shape and anatomy,
- ▷ To provide the basis for the *automatic* generation of results, which is beneficial for high-throughput applications and as initial guess before interactive refinement,
- ▷ To be fast with a close to *linear* computational complexity $O(N)$, where n is the number of image elements, especially in order to deal with future image resolutions, and
- ▷ To permit *accurate* measurements of object and tissue properties (e. g. object volumes).

Structure

In accordance to the above mentioned two-step approach of Morphological Volumetry comprising morphological segmentation plus volumetric quantification based on statistical analysis, the remainder of this thesis is divided into two main parts. The first part deals with theoretical and practical approaches toward three-dimensional segmentation that are robust without taking explicit assumptions on shape and anatomy into account, and that provide a suitable basis for the last three chapters. This second part deals with

quantification by addressing fundamental questions on the nature of volumetric measures and errors and by providing various examples, mainly from the field of neuroimaging. All examples will reveal their specific difficulty for which we will propose solutions. Each of the following seven chapters receives its motivation from the context of its application and contains its own central ideas, and is therefore readable in isolated form. Both parts are additionally preceded by their individual motivation and didactical introduction (pp. 19 ff. and 83 ff.).

For image segmentation, as well as quantification, evaluation and, where possible, validation are crucial. In some cases, we have been able to compare our results to state-of-the-art techniques from other researchers. However, in other cases, such a comparison was impossible for certain reasons—a thorough discussion of our results will try to bridge this gap. Still, in most cases the *ground truth* will remain concealed such that evaluation must be conducted with due care and attention, even if so-called 'gold standards' were available.

On the whole, this thesis is focused on finding strategies on the algorithmic level for solving image-based quantification tasks. These tasks can be divided into two questions: The first part of the thesis will answer to the question *where* the objects of interest are located in a set of images, while the second will lead to answers on *how much* of an object is contained.

It was crucial to choose the methodological framework for the first part. Initially, active shape models and its variants (cf. p. 13) were taken into consideration due to their ability to learn about an object's variability and to automatically solve according segmentation problems. However, the inter-individual anatomical variance, as well as the complexity of many of the objects, which have been in the focus of our work, were considered too large for such approach—one example is the cortical folding pattern of the human brain (cf. Fig. I.1, pp. 20 f.). Instead, we sought a framework that permits efficient three-dimensional segmentation without incorporating shape priors. Among these, we chose the *morphological watershed transform*, which forms the basis for image segmentation in Mathematical Morphology [Soille, 2003] (cf. p. 50).

Interaction, i. e. the possibility to iteratively verify, correct, or refine the results, is an essential ingredient to the first part. We developed a novel concept for interactivity, not being offered by the classical watershed transform (WT), while for each problem, a balance between user interaction and automation had to be found. Furthermore, we had to provide solutions to deal with oversegmentation, which is known to be the major shortcoming of the conventional WT [Haris *et al.*, 1998]. In contrast to the first part, the quantification step should be fully automated in most cases in order to remain observer independent. The main ingredient to the second part will be histogram analysis that permits a valid estimation of quantitative parameters, as well as their uncertainties.

PART I

ON ROBUST SEGMENTATION AND WATERSHEDS

THE FIRST PART of this thesis has been much inspired by the work of J. Serra, Director of the Centre de Morphologie Mathématique, Paris, who gave some outstanding lectures on Mathematical Morphology and its application to various challenging image analysis problems during a summer school on digital image analysis in Freiburg in October 1999.

As a first and critical step in the neurological image analysis pipeline, it is often required to solve the *skull stripping problem* on T1 weighted images. This problem is defined as removing everything but the intracranial cavity from a set of images. T1 weighted images are most common for studies of brain anatomy and pathology, offering a good contrast between different tissue types, most importantly CSF, gray matter (GM), white matter (WM), and Gadolinium enhancing tissue. The inspection of a large diversity of brain MR images from various institutions and scanners suggests that the design of highly automated image analysis methods that are able to deal with this diversity is not a trivial task (cf. Fig. I.1, pp. 20 f.).

There have been publications on skull stripping before, two well-known examples being the papers by Atkins and Mackiewich [1998] and Dale *et al.* [1999]. While the latter use a three-dimensional balloon template steered by local image intensity to approximate the intracranial cavity, the former describe a hybrid technique that builds upon nonlinear anisotropic edge-preserving diffusion filtering, automated thresholding, morphological filtering, prior spatial assumptions, and an active contour model similar to Kass' snakes [Kass *et al.*, 1988] to refine the brain contour based on local image gradients and a smoothness term.

Besides the fact that most existing techniques were computationally and/or algorithmically complex, some fundamental problems remain: On the one hand, existing methods relying on global intensity thresholds are not robust to image nonuniformity; object sizes will be underestimated in darker, and overestimated in brighter regions. In some cases, this problem can be alleviated by nonuniformity correction prior to image segmentation [e. g. Sled *et al.*, 1998]. On the other hand, methods relying on local image gradients are likely to produce erroneous results for textured objects that exhibit steep gradients at their interior or weak gradients at their edges.

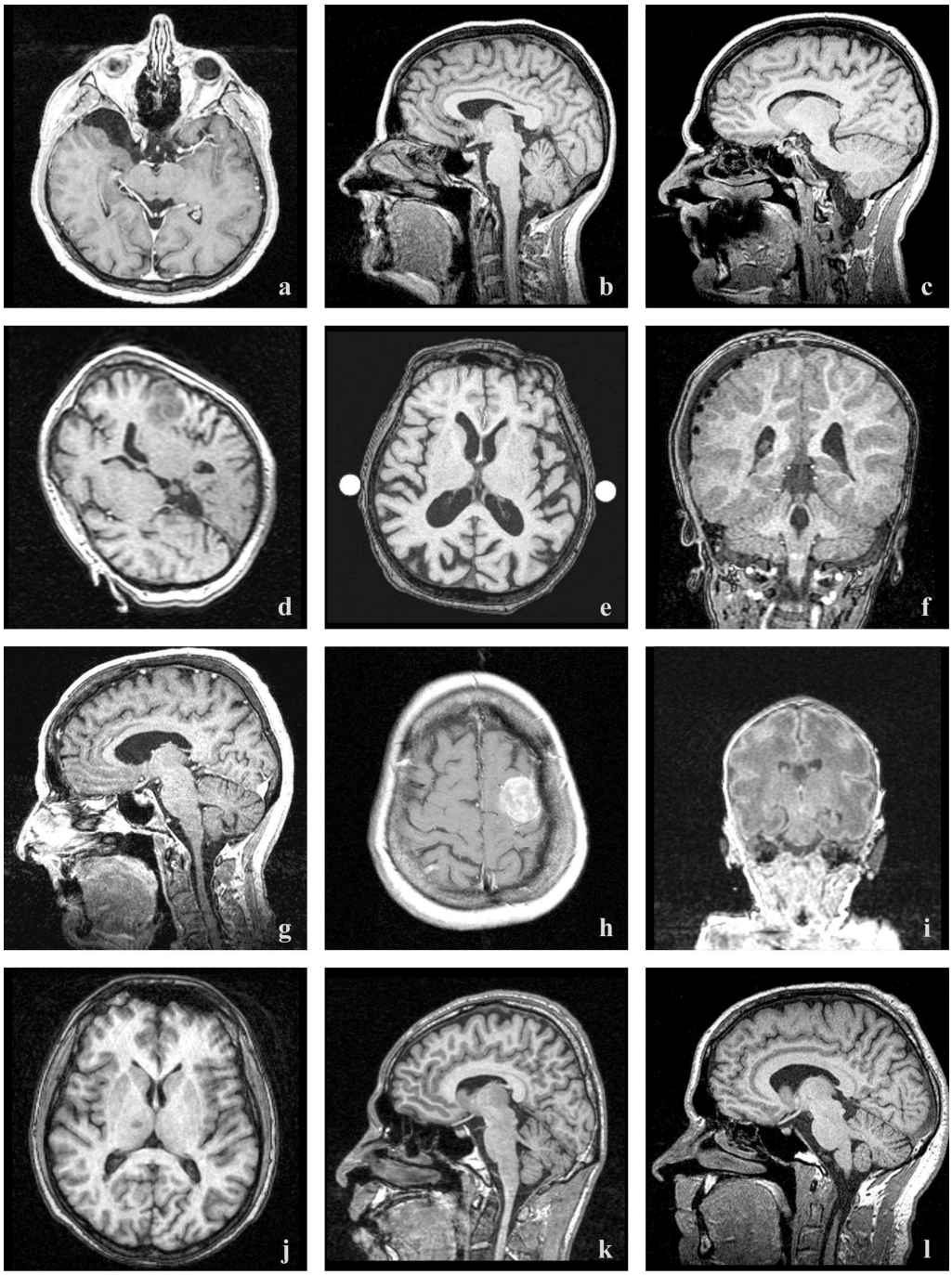


FIGURE I.1: Examples of T1 weighted brain images from various scanners and protocols. Some images (*d,o,t,u*) were acquired with an open 0.5 T system (SIGNA SP/I, GE HEALTHCARE), all other images with closed 1.5 T systems. Images courtesy of: (*a*) R. V. KUMMER, Dresden; (*b,c,g,h,l,n,r,v–x*) B. TERWEY, Bremen; (*d,o,t,u*) M. MOCHE, Leipzig; (*e,j,n*) G. SCHMID, Bochum; (*f,i*) R. ROBERTSON, Boston; (*k*) T. HOLROYD; (*p,q,s*) W. S. MILLAR, New York. (See also page 21 and Table I.1)

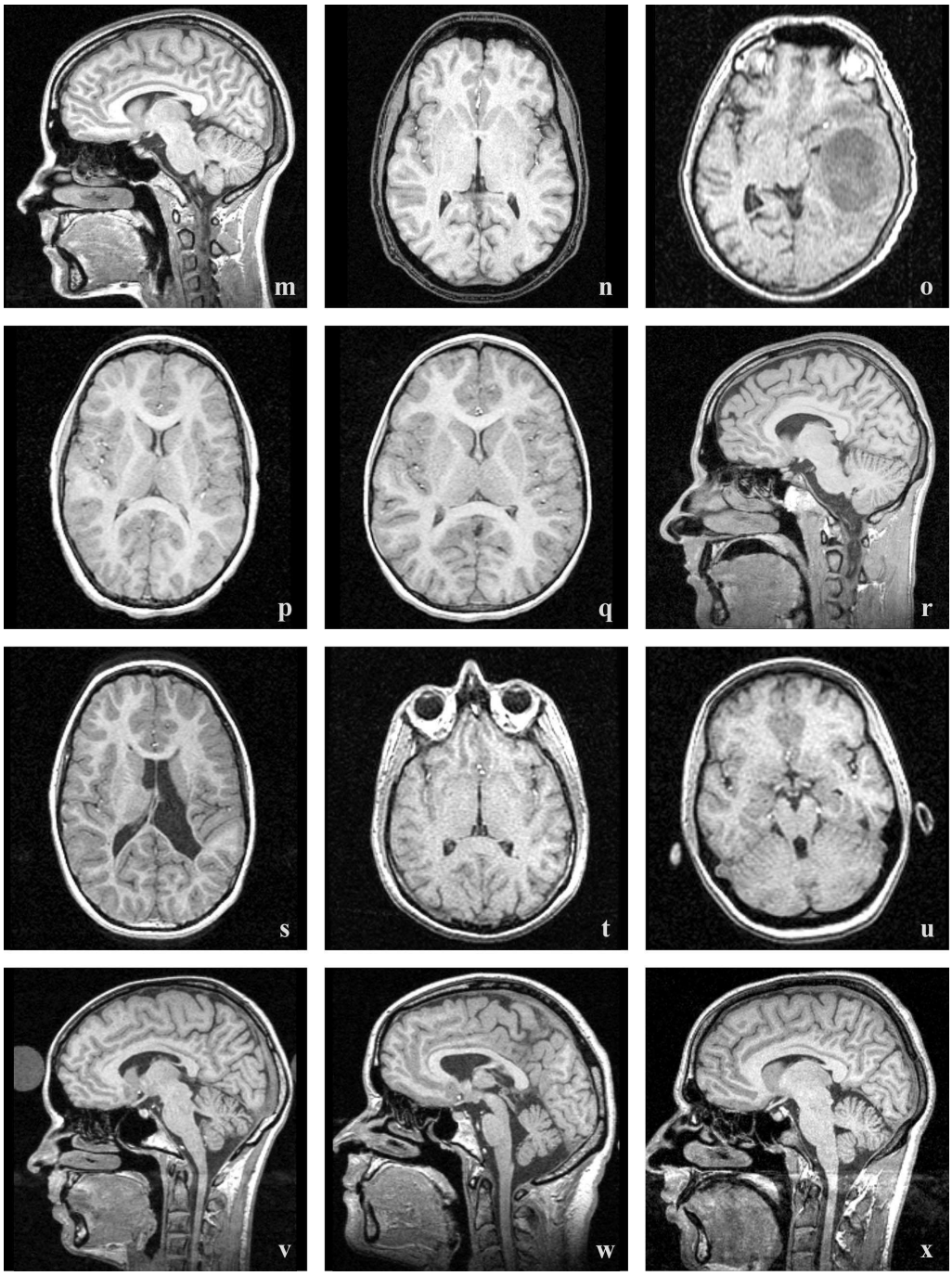


FIGURE I.1: Examples of T1 weighted brain images from various scanners and protocols (contd.). Comments: (*a*) tumor patient after surgery; (*d,g,h,o*) tumor patient before surgery; (*e,j*) MS patient; (*f*) epilepsy patient after grid implantation; (*i*) preterm baby; (*p,q,s*) follow-up at age six for preterm baby; (*b,c,k–n,r,t,u–x*) volunteer. (See also page 20 and Table I.1)

TABLE I.1: Details for Figure I.1, abbreviated sequence names, times in ms, and image resolution in mm (in-plane [x,y] × distance between slices [z]).

Fig. I.1	Sequence	Resolution	Age / Gender
a	FL3D RM, TR=32.0, TE=6.0	$0.94^2 \times 2.00$	55 y / F
b	MPR IR, TR=11.08, TE=4.3, TI=8.0	$0.50^2 \times 1.00$	34 y / F
c	MPR IR, TR=11.08, TE=4.3, TI=8.0	$0.50^2 \times 1.00$	36 y / M
d	SPGR TR=13.3, TE=2.7	$1.09^2 \times 2.00$	70 y / M
e	MPR IR, TR=14.3, TE=7.0, TI=300	$0.94^2 \times 1.30$	77 y / M
f	SPGR, TR=6.9, TE=2.6	$0.78^2 \times 1.50$	2 y / M
g	MPR IR, TR=11.08, TE=4.3, TI=8.0	$1.00^2 \times 1.00$	78 y / F
h	SE RM, TR=760, TE=20.0	$0.78^2 \times 5.00$	(cf. g)
i	SPGR, TR=35.0, TE=6.0	$0.70^2 \times 1.55$	14 days / M
j	MPR IR, TR=14.3, TE=7.0, TI=300	$0.98^2 \times 2.50$	35 y / M
k	FL3D, TR=20.0, TE=6.0	$1.00^2 \times 1.00$	M
l	MPR IR, TR=11.08, TE=4.3, TI=8.0	$0.50^2 \times 1.00$	38 y / M
m	MPR IR GR, TR=1900, TE=4.15, TI=1100	$1.00^2 \times 1.00$	26 y / F
n	MPR IR, TR=14.3, TE=7.0, TI=300	$0.94^2 \times 1.50$	29 y / M
o	SPGR TR=13.3, TE=2.7	$1.09^2 \times 2.00$	49 y / M
p	SPGR, TR=34.0, TE=5.0	$0.86^2 \times 1.50$	6y
q	SPGR, TR=34.0, TE=5.0	$0.78^2 \times 1.50$	6y
r	MPR IR GR, TR=1900, TE=4.15, TI=1100	$1.00^2 \times 1.00$	29 y / M
s	SPGR, TR=34, TE=5.0	$0.86^2 \times 1.50$	6y
t	SPGR TR=13.3, TE=2.7	$1.09^2 \times 2.00$	36 y / M
u	SPGR TR=13.3, TE=2.7	$1.09^2 \times 2.00$	36 y / F
v	MPR RM, TR=9.7, TE=4.0	$1.00^2 \times 1.00$	33 y / F
w	MPR RM, TR=9.7, TE=4.0	$1.00^2 \times 1.00$	34 y / F
x	MPR IR, TR=11.08, TE=4.3, TI=8.0	$0.98^2 \times 1.00$	28 y / M

In Freiburg, J. SERRA demonstrated how basic methods of Mathematical Morphology are able to solve complicated segmentation tasks. We sought to propose a likewise basic skull stripping method avoiding the shortcomings of deformable templates, threshold, and gradient-based approaches. The idea is to consider a three-dimensional region-based algorithm in analogy to *chocolate icing* that flows down a cake until reaching a *valley*, rather than stopping at a global or local threshold height or at pronounced gradients. When interpreting the image gray value as surface altitude of multiple cakes, the brain is to be regarded as a single connected cake that is separated from neighboring cakes (further bright intensity tissue such as muscle, fat, eyes, skin, etc.) by valleys (cf. Fig. 2.2, p. 28).

The analogous 'cake separation problem' corresponds to a two-dimensional image describing a three-dimensional topological relief, much as all of the problems discussed in Freiburg were embedded in two dimensions. Note that the sought chocolate icing algorithm is equivalent to a three-dimensional watershed transform (WT) on the original image when

interpreting the gray values as depth information, i. e. bright regions as valleys and dark regions as hills and crest lines.[8] An algorithm for a fast immersion-based WT besides its application to skull stripping lays the basis for the first part of this thesis and is described in Chapter 2. The actual implementation was inspired by a paper of MITTELHAEUSSER and KRUGGEL [1995]. The fast WT consists of two steps. First, all image elements are sorted according to their image intensity. Second, each element is processed exactly once with respect to its neighborhood in the specified order. The robustness to image noise, inhomogeneity, and resolution is evaluated on patient, volunteer, and phantom data.

Chapter 3 goes beyond specific problems and examines the theory of effective human-computer interaction, with respect to image segmentation. The duality of interactivity on the one hand and automation on the other hand is analyzed, and it is questioned to which extent user independency and interaction are compatible with each other. To derive practical guidelines for user interfaces, the time necessary to interactively solve an image analysis problem is considered as a measure of optimality. Moreover, the role of human experts, as well as the complementarity of data and knowledge are discussed.

Aiming at efficient interactive segmentation, the skull stripping algorithm is extended to a generic image segmentation tool in Chapter 4. After a formal definition of the WT, algorithmic and implementational aspects are discussed besides general problems that occur when using the WT. The Interactive Watershed Transform (IWT) is proposed, which builds upon the fast immersion-based WT described in Chapter 2, followed by a hierarchical organization of the resulting basins in a tree structure. Each local image minimum is represented as an atomic basin on the lowest hierarchy level. Sorting, processing, and tree generation are of order $O(N)$. Two mechanisms for the avoidance of oversegmentation, which is the major problem of the classical WT [HARIS *et al.*, 1998], are incorporated in the IWT, namely preflooding and markers. Both are processed close to real-time to control tree partitioning and basin merging. Furthermore, by analyzing the set of atomic basins, a concept for a full automation of the skull stripping algorithm is derived. The robustness of this automation to extreme image noise and inhomogeneity is evaluated on images from a variety of scanners.

As a further application of the IWT, Chapter 5 focuses on bone segmentation, as important for *in vivo* studies of normal and injured kinematics, and bone removal, as important for vascular diagnosis in CTA images. Separation of individual bones from CT volume data using conventional techniques is difficult due to the tightly packed and irregularly shaped anatomy. We introduce an interactive algorithm for fast and robust segmentation of bones in CT images, consisting of two steps. First, a marker driven separation of single bones is realized using the IWT. Second, solid and accurate segmentations for each bone are generated automatically using a concept that is proposed as Bit Morphology. Evaluation is conducted on ten wrist and one skull CT scans. For quantitative evaluation, all eight carpal bones from manually segmented scans of an embalmed cadaver wrist are compared through kinematic analysis to the new minimally interactive algorithm. To facilitate bone removal on very large CTA data sets, the IWT is extended by a compression step, only

[8] Later in this thesis, "the inverted original data"is used for interpreting the image gray values/intensity as depth information.

taking bright image elements into account. The efficient interactivity of the IWT is moreover combined with automatic region labeling rules for automatic separation of osseous and vascular objects.

Things are worse than most people think,
But less complicated to solve.

—D. H. D. K.

CHAPTER 2

The Skull Stripping Problem in MRI Solved by a Single 3D Watershed Transform

Abstract. *In this chapter, a robust method for the removal of non-cerebral tissue in T1 weighted magnetic resonance (MR) brain images is presented. This procedure, often referred to as skull stripping, is an important step in neuroimaging. Our novel approach consists of a single morphological operation, namely a modified three-dimensional Fast Watershed Transform (FWT)*[1] *that is perfectly suited to locate the boundaries of the brain, including the cerebellum and the spinal cord. The presented method is referred to as* Watershed Based Skull Stripping (WASS).

The main features of the WASS method lie in its simplicity and robustness. It is simple since neither preprocessing of the MRI data nor contour refinement are required. Furthermore, the skull stripping relies solely on one basic anatomical fact, i. e. the three-dimensional connectivity of white matter. As long as this feature is observed in the image data, a robust segmentation can be guaranteed independently from image orientation and slicing, even in the presence of severe intensity nonuniformity and noise. For that purpose, the watershed algorithm was modified by the concept of preflooding, which helps to prevent oversegmentation, depending on a single parameter. The automatic selection of the optimal parameter, as well as applicability are discussed based on the results of phantom and clinical brain studies.

THE SKULL STRIPPING PROBLEM is well-known and has been, e. g., studied by Kikinis *et al.* [1992], Kapur *et al.* [1996], Wells *et al.* [1996], Freeborough *et al.* [1997], Sandor and Leahy [1997], Atkins and Mackiewich [1998], Goldszal *et al.* [1998], Dale *et al.* [1999], Hojjatoleslami *et al.* [1999], and Rehm *et al.* [1999]. It is equivalent to the segmentation of the whole brain or the removal of non-cerebral tissue such as skull, scalp, veins, or meninges (cf. Fig. 2.1). Studies of brain anatomy and pathology are most commonly based on magnetic resonance imaging (MRI) due to its good soft tissue separation. Here, T1 weighted MR imaging that comprise the fastest MRI protocols available are often preferred, since they offer a good contrast between cerebral gray matter (GM) and white matter (WM), as well as between GM and cerebrospinal fluid (CSF). On T1 weighted images, CSF and cortical bone, both represented by very low image intensities, are not subject of the skull stripping problem.

[1] Note that FWT must not be confounded with Fast Wavelet Transform.

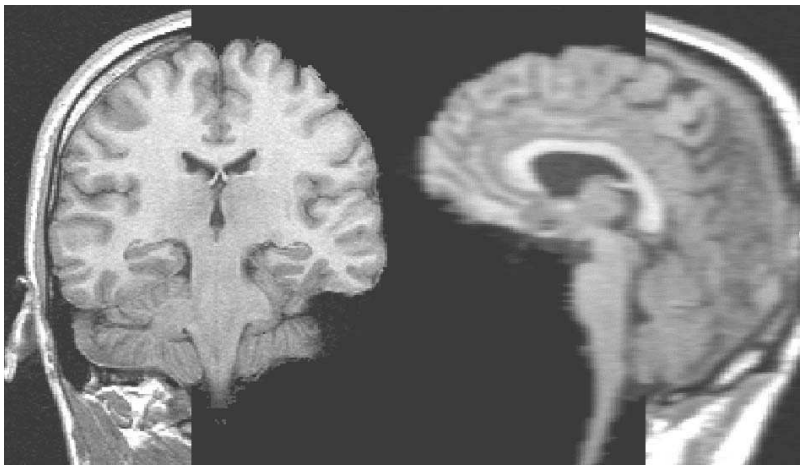


FIGURE 2.1: Fusion of a successfully skull stripped image with the original data in two projections (healthy volunteer, coronal slicing, thickness 3 mm).

Whole brain segmentation is often regarded as an essential step in a neurological image processing pipeline, either because the whole brain is the region of interest, such as in studies of Morbus Alzheimer [FREEBOROUGH *et al.*, 1997], or because the subsequently performed steps benefit from the fact that only a small set of well known tissue types is left over (i. e. WM, GM, CSF and possibly lesions and vessels), such as statistical brain tissue segmentation methods [KAPUR *et al.*, 1996, WELLS *et al.*, 1996, PHAM and PRINCE, 1999]. The latter offer a good basis for volumetric or morphometric examinations, e. g. cortex reconstructions [DALE *et al.*, 1999], and yield promising results in accounting for intensity nonuniformity [WELLS *et al.*, 1996, PHAM and PRINCE, 1999], however, their convergence is distorted by the presence of non-brain tissue [GOLDSZAL *et al.*, 1998, WELLS *et al.*, 1996, MAES *et al.*, 1999]. For brain warping techniques that are used to perform inter-subject studies, it is as well desirable to exclude all non-brain tissue from the matching process [TOGA, 1999].

This chapter is organized as follows. The following sections give a brief overview of existing methods and describe the motivation and goals of the novel approach. Then, the method is derived in detail, but postponing theoretical and mathematical considerations until Chapter 4. An evaluating section assesses both robustness and applicability. The performance, as well as the dependency on the parameter introduced below are examined. Finally, extensions and limitations of our approach are discussed.

2.1 Related Work, Motivation, and Goals

Despite the clear definition of the skull stripping problem, no standardized solution has been published yet. A good survey of the work up to 1996 is given by ATKINS and MACKIEWICH [1998]. The image processing techniques found in literature can be divided into three groups:

- ▷ *Region based.* The most common approaches sequentially apply morphological operations and manual editing. First, the white matter gray values are located using thresholding or seeded region growing, followed by a morphological opening that detaches the brain tissue from the surrounding tissue. Morphological dilation and closing are required for the segmentation to cover the whole brain without holes [Kikinis *et al.*, 1992, Hoehne and Hanson, 1992, Kapur *et al.*, 1996, Sandor and Leahy, 1997, Freeborough *et al.*, 1997, Goldszal *et al.*, 1998, Hojjatoleslami *et al.*, 1999, Lemieux *et al.*, 1999].
- ▷ *Hybrid methods.* In order to account for the shortcomings of morphological multi-step approaches, they have been combined with edge-based methods [Kapur *et al.*, 1996, Aboutanos and Dawant, 1997, Atkins and Mackiewich, 1998, Rehm *et al.*, 1999]. As mentioned above, e. g., two-dimensional contours such as snakes are applied to the morphological segmentation result in a final step [Atkins and Mackiewich, 1998]. Marais and Brady [2000] developed a highly specialized method for detecting the brain surface in sparse MRI based on an intensity model for the outer brain-bone-skin interface.
- ▷ *Template based.* More recently, some investigators succeeded in fitting a balloon-like surface to the intensity-normalized magnetic resonance (MR) data in order to separate the brain from surrounding structures [Dale *et al.*, 1999]. An elegant method implementing two coupled surfaces was published by Zeng *et al.* [1999]. Frackowiak *et al.* [1997] described another three-dimensional template, which is based on averaged volume data.

Our goal in this chapter is to develop a skull stripping procedure that is qualified to be the first step in the image processing pipeline. This means that it should be robust even to considerable radio frequency (RF) nonuniformity, as well as noise. This is not the case for most existing techniques.

Problems of existing skull stripping techniques were discussed by Holden [2001, p. 73]. Thresholding and region-growing-based approaches exhibit the problem of leaking out, as discussed by Worth *et al.* [1998], and are highly sensible to image nonuniformity. The dependency on the image uniformity remains a problem also for complex morphological operations such as proposed by Lemieux *et al.* [1999]. Deformable templates are often susceptible to image noise that has to be removed in a preprocessing step. Efficient interactive methods such as the live wire algorithm [Barrett and Mortensen, 1997, Schenk *et al.*, 2000] also provide flexible image segmentation that may be applied to the original image data. However, due to their two-dimensional nature, they are preferably applied when the number of slices is limited and when the objects are rather smooth than highly folded.

In order to be applicable both for pathological brain structures and anatomical abnormalities, the algorithm should be as *simple* as possible: First, in the sense that it consists of one (or a few) well-defined operations, such that segmentation failures can be easily understood and interaction techniques may be properly integrated. Second, in the sense that it should be based on as little assumptions and anatomical a-priori knowledge as possible.

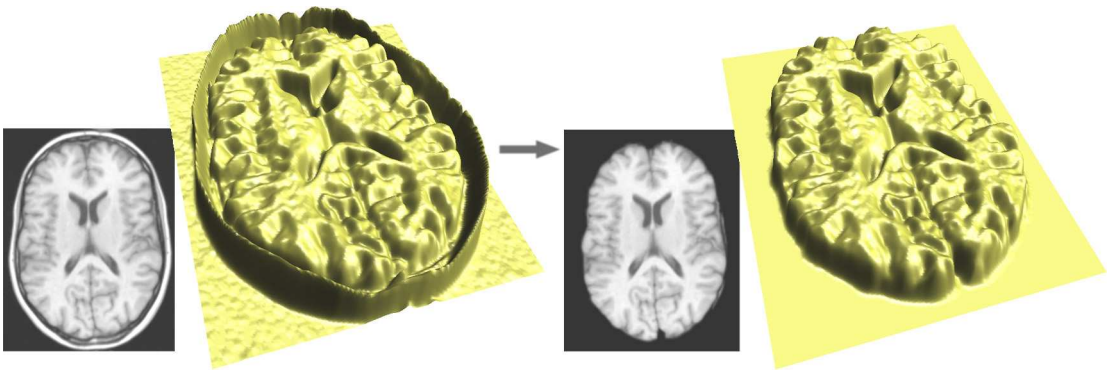


FIGURE 2.2: T1 weighted axial image (slice thickness 3 mm) of a healthy volunteer (F, 30 y). Interpretation of gray value as altitude information. *left*: Original image. *right*: Skull stripped with $h_{\mathrm{pf}} = 0.16\, I_{\max}$ and only one connected hill left, representing the brain.

We decided not to use deformable templates that require smoothness constraints, which often are not satisfied on the brain boundary, especially in its basal parts.

For clinical purposes it is desirable to provide a fast segmentation procedure that does not require much user interaction. Moreover, the computational costs should be close to proportional to the number of voxels N, to be able to deal with future imaging resolutions. Finally, motivated by an ongoing project with the UNIVERSITY HOSPITAL BOCHUM aiming for brain atrophy quantification in MS patients (cf. Ch. 8, [LUKAS *et al.*, 2004]), the segmentation, in order to serve as a basis for such a task, should be able to include most of the partial volume voxels occurring at the brain-CSF boundary.

2.2 Methods

On Connectivity

Similar to other approaches [KIKINIS *et al.*, 1992, KAPUR *et al.*, 1996, GOLDSZAL *et al.*, 1998], our basic assumption regarding brain anatomy is the connectivity of white matter.

We define two points p and q of a gray level image $I : D \longrightarrow \mathbb{R}$ to be *connected*, if a path γ inside D of adjacent image elements x exists between them that is at least as bright as the darker one of the two points:

$$I(x) \geq min\{I(p), I(q)\} \quad \forall\, x \in \gamma \quad . \tag{2.1}$$

Equation 2.1 can be easily understood when interpreting the intensity of a two-dimensional image as altitude information. For an arbitrary image slice of a three-dimensional image, the topographic interpretation of the gray values consists of several hills corresponding to bright image areas and valleys corresponding to dark image areas (Fig. 2.2 left). According to the above definition of connectivity, two regions are dis-

connected within this image slice, if they are separated by a valley, otherwise they are connected.[2]

Here, the significance of the intensity characteristics of T1 weighting for our approach becomes clearly visible: The connected WM is surrounded by darker GM and even darker CSF, and can thus be regarded as the top of a hill. WM regions that are not connected in 2D must be connected in 3D, since the WM is interconnecting all functional parts of the brain. For a robust skull stripping technique, it is hence crucial to work fully in 3D, in contrast to 2D techniques that do not account for the brain's three-dimensional connectivity.

Assuming that the brain is surrounded by CSF, and all non-brain tissue, which need to be removed, show brighter image intensities than CSF, skull stripping becomes equivalent to isolating a single hill in the four-dimensional landscape.[2] This hill then represents the whole brain including the cerebellum and the spinal cord, as long as they are connected within the image. The valley corresponding to CSF and other low intensity tissue such as bone and meninges will then define the accurate border of the segmentation result (Fig. 2.2 right).

Solution using a Single 3D Watershed Transform

We now interpret gray values as depth information, i. e. we consider a *depth image* or a gray level inverted *altitude image*. From this viewpoint, hills become basins and valleys become crest lines (Fig. 2.3). The morphological operation that partitions an image into regions, each of them corresponding to a connected basin, is the watershed transform (WT) [Serra, 1982]. Thus, all that needs to be done for a whole brain segmentation, is to perform a *single* 3D watershed transform on the inverted original data[3] and to choose the catchment basin representing the brain. For a mathematical definition of the watershed transform both for the continuous and discrete case refer to Section 4.1. In this chapter, in order to pronounce ideas and applications, we will use a more qualitative description.

Before introducing an algorithm to perform the fast watershed transform, we should take a more detailed look at our problem: WM cannot always be regarded as connected in the strict sense defined above, since its image intensity is not constant, even without the presence of RF nonuniformity and noise. Therefore, we have to weaken our criterion for connectivity. We do so by allowing the connecting path to show a lower intensity than the darker of the two connected points up to a maximum difference.

In words of the watershed transform, this is described by the concept of *preflooding*, which we introduce: Prior to the transform, each catchment basin is flooded up to a certain height above its bottom, i. e. the darkest image element, called preflooding height h_{pf} and will only be regarded as a separate region as long as it holds the water inside. Otherwise, it will be merged with the deepest neighboring basin. In this chapter, h_{pf} shall be regarded as a positive constant value.

[2] Figures 2.2 and 2.3 present didactic three-dimensional projections corresponding to a single image slice of the four-dimensional problem corresponding to a three-dimensional image volume.

[3] Recall that usually the WT is applied to the gradient image.

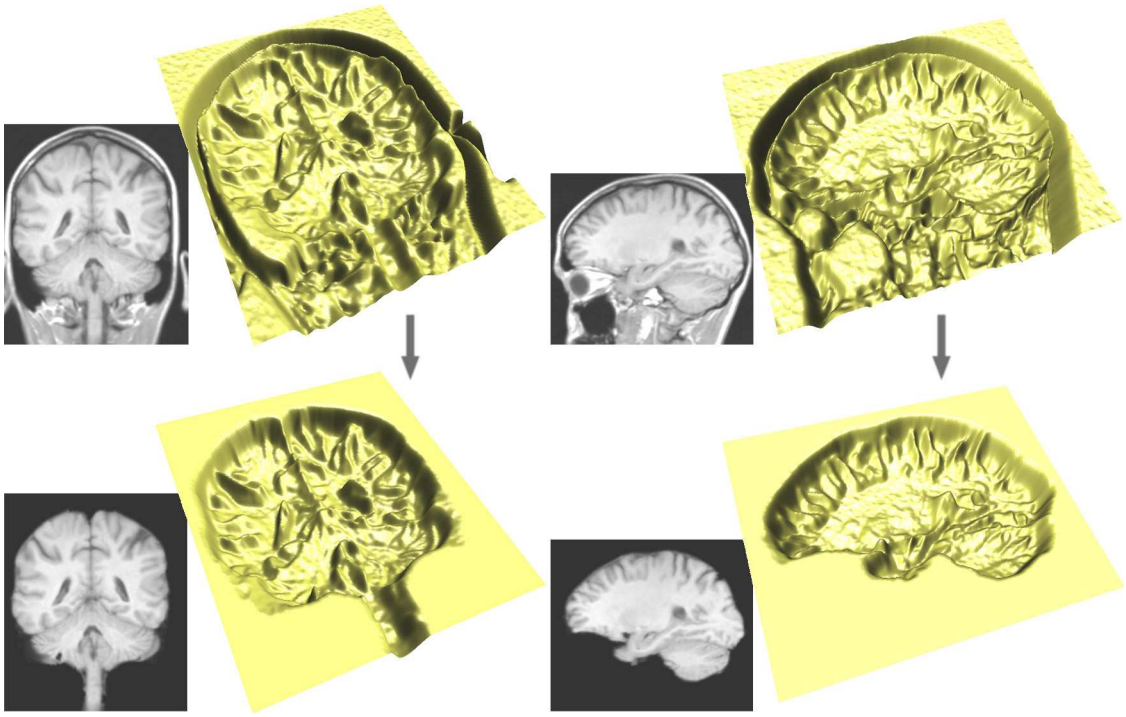


FIGURE 2.3: Same volunteer as Figure 2.2, but coronal (left) and sagittal (right) sections (both 3 mm thickness). Now, the image is interpreted as depth information, resulting in crest lines—dark regions in the original image—and basins—one of which is the brain—, which can be separated by a watershed transform. *top*: Original image. *bottom*: Single separated basin after watershed transform.

Furthermore, it is not true that the brain is fully surrounded by CSF, as assumed above. Rather, nerves such as the optical, olfactory, and auditory, connect the WM with eyes, nose, and ears. However, these connections are rather thin with respect to the image resolution. In all practical cases, this results in a decreased image intensity due to partial volume averaging, such that the above idea is not jeopardized.

A Modified Fast Watershed Algorithm

The applied fast watershed transform starts by sorting all voxels (of the gray level inverted image, Fig. 2.3) according to their intensity in an ascending order.[4] Each voxel with regard to its six direct neighbors on a regular orthogonal 3D grid is processed exactly once, until the last voxels (the highest altitudes) have been processed. If during processing, a voxel has some already processed neighbors (i. e. voxels of same or lesser intensity), *voxel-basin merging* is performed. Otherwise, a new basin is formed since after sorting, an isolated voxel must represent a local intensity minimum. If two or more neighbors have already been processed belonging to different basins, these are tested for *basin-basin merging*. This

[4] Ascending order in the gray level inverted image means that the voxels with the highest *original* image intensity are processed first.

approach is similar to the one described by MITTELHAEUSSER and KRUGGEL [1995], but simpler in that the voxel-basin merging is unconditional. We now introduce two new criteria for the merging procedures:

(a) *Voxel-basin merging*: Each voxel will be merged with the deepest neighboring basin, i. e. the basin with the lowest seed voxel.

(b) *Basin-basin merging*: All neighboring basins whose depth relative to the current voxel intensity is less or equal to the preflooding height h_{pf} will be merged with the same basin as the voxel itself.

After the transform with an appropriate preflooding height, one basin should exist that represents the cerebrum together with all parts that are connected via WM, i. e. cerebellum and spinal cord. This basin usually is the largest existing one[5], unless the field of view (FOV) was selected too large, such that image parts outside of the head can form a larger basin; in this unusual case, one could choose the basin containing the center of FOV or a manually selected basin. These rules will be used in Section 4.4 together with an automated estimation of the proper preflooding height for highly robust and fully automated skull stripping.

The number of basins is monotonically decreasing whereas its sizes are increasing with increasing h_{pf}. Figure 2.4 shows the typical characteristics of that behavior. Here, the region-based nature of our method becomes visible, providing the basis for *robustness*. The segmentation result does not change smoothly with varying h_{pf}. Instead, a staircase pattern is found with varying step size. We have found that one of these steps correspond to the desired brain region (cf. position 3 and 4 in Fig. 2.4 and triangles in Fig. 4.11), such that, depending on the image quality, a more or less broad range of proper h_{pf} values exists.

2.3 Quantitative Evaluation and Results

This section will evaluate the usefulness of the described procedure for a variety of brain images. We use clinical data from two different scanners, as well as phantom data to answer the following two questions:

(a) Is it possible to successfully segment the whole brain with our method?

(b) If yes, what is the range of proper preflooding heights?

The answer to the second is expected to depend on the image noise level and other aspects of image quality. For each image, the level of noise n is measured as the standard deviation of the high-frequency signal present in the image background. In the following, noise levels n and preflooding heights h_{pf} will be given in units relative to the maximum image intensity $I_{\max}$.

As mentioned before, in clinical MRI segmentation studies no ground truth is generally available. Our *gold standard* was defined by an expert radiologist who classified the

[5]The volume of the largest basin at a given h_{pf} value is denoted $V_{\max}$ (cf. Fig. 4.11).

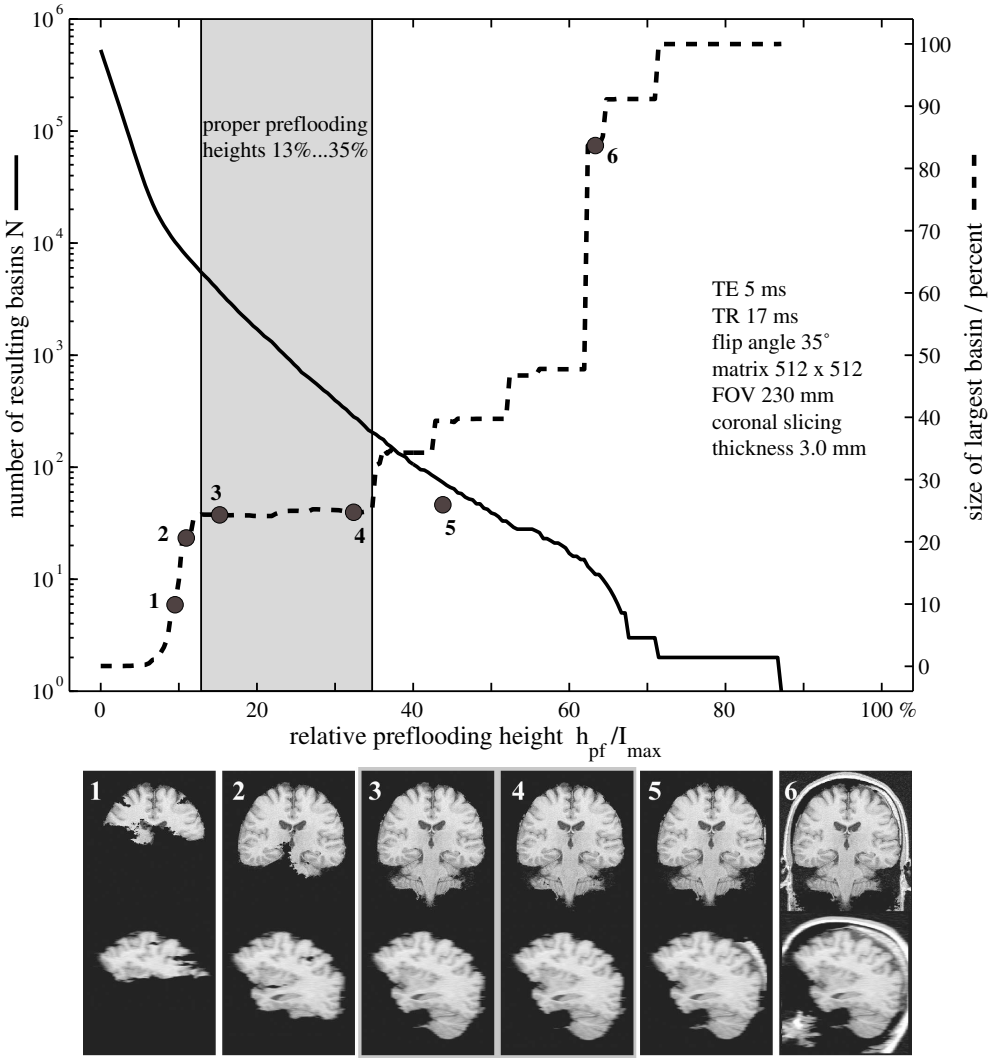


FIGURE 2.4: Algorithmic behavior of the modified watershed transform in one typical example. The number of resulting catchment basins (solid line) and size of largest basin (dotted line) are plotted as a function of preflooding height h_{pf}. For six h_{pf} values (filled circles) the result is shown in coronal and axial views. For the 5^{th} value where under-segmentation did already occur, the brain was not contained in the largest basin (filled circle below dotted line). For both the 3^{rd} and 4^{th} values, even if more than 15 % apart, a proper segmentation result is obtained (cf. Fig. 4.11 for further examples).

segmentation procedure as unsuccessful when small parts of cerebrum or cerebellum were excluded or pieces of skull, skin or eyes were included in the segmentation result. On the other hand, parts of meninges or veins that in the images often seem to be connected to gray matter were tolerated within a successful segmentation. Moreover, accuracy was assessed by comparing the overlap of a manual expert segmentation with the results of our skull stripping method. On all seven manually segmented data sets, the overlap was more than 96 %.

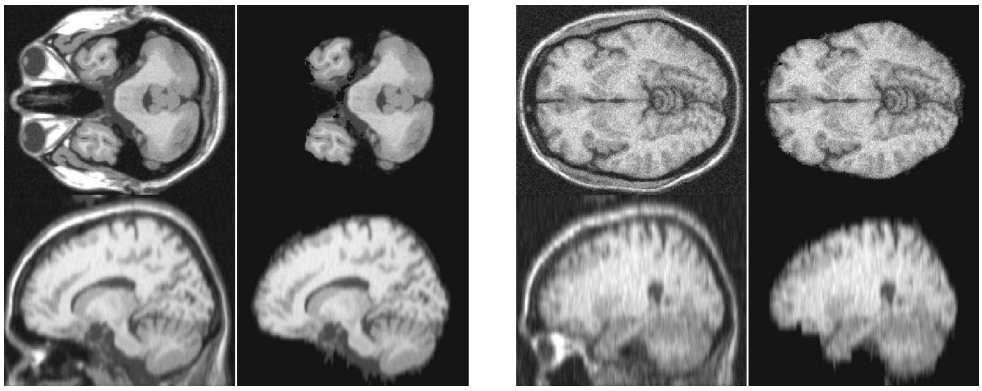


FIGURE 2.5: Two images of the MNI Simulated Brain Database and application of the skull stripping procedure, axial and sagittal view. Image simulation parameters (slice thickness / noise level / RF nonuniformity): 3 mm / 3 % / 20 % (left) and 5 mm / 9 % / 40 % (right).

Phantom Studies

To offer a reproducible data source, we extensively tested our method on the SIMULATED BRAIN DATABASE (SBD) [COCOSCO *et al.*, 1997, COLLINS *et al.*, 1998] from the Montréal Neurological Institute, McGill University (MNI)[6]. All available 90 T1 weighted data sets of the normal brain database were processed. The results were satisfactory: All data sets were successfully segmented by our method with respect to the above-mentioned criteria, even for extreme slice thickness (9 mm), noise (9 %) and RF nonuniformity (40 %). Two examples are shown in Figure 2.5. We will see later in Section 4.4 that the WASS method can also be fully automated at a robustness that extends far beyond these numbers (cf. Fig. 4.12).

Clinical Brain Studies

For further evaluation, 43 clinical brain images of healthy volunteers and patients were acquired with two MR scanners: MAGNETOM VISION PLUS 1.5 T (SIEMENS MEDICAL SOLUTIONS) and S15-ACS 1.5 T (PHILIPS MEDICAL SYSTEMS), using various 3D T1 weighted acquisition protocols (TR 8.1–17 ms, TE 4–5 ms, flip angle 12°–35°) and reconstruction parameters (no interslice gap, coronal, axial, and sagittal slicing, slice thickness 1–5 mm). In all 43 cases, the above described method succeeded in segmenting the brain. See Figure 2.6 for examples. However, in two cases the preflooding value h_{pf} for correct segmentation turned out to be unexpectedly high, as will be discussed below.

Robustness and Parameter Dependency

In order to chart the above results and to evaluate the robustness of our method, we plotted the range of proper h_{pf} values against the noise level n, both in units relative to the maximum image intensity I_{max} for 45 of the 90 processed phantom data sets (Fig. 2.7).

[6] Image download Jan 2000.

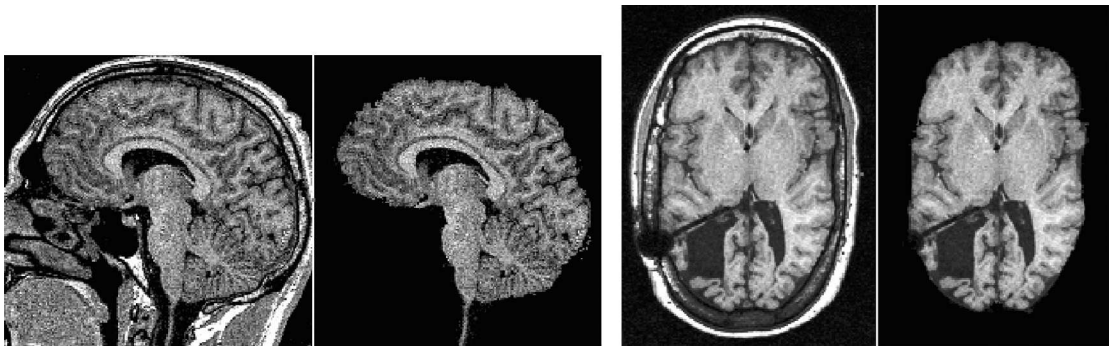


FIGURE 2.6: Clinical brain data sets and application of skull stripping procedure. Two problems are addressed; *left*: extreme image noise level (acquisition parameters: slice thickness 1 mm, TR 8.1 ms, TE 4 ms). *right*: pathological abnormality (enlarged left ventricle, with drainage).

TABLE 2.1: Quantitative evaluation of robustness in terms of the range of proper h_{pf} values on 43 clinical and 90 phantom images. The h_{pf} range is partitioned in four categories; the larger the range, the higher the robustness of a given segmentation result.

Range of proper preflooding heights $(h_{\mathrm{pf},max} - h_{\mathrm{pf},min})/I_{\max}$	num. data sets phantom	clinical
0 (failure)	0	0
< 5 %	6	2
5–10 %	25	9
> 10 %	59	32

In turned out that the proper h_{pf} values roughly scale linearly with the noise level. The dotted line represents the equation $h_{\mathrm{pf}} = 0.11\ I_{\max} + 3.5\ n$ chosen manually such that all 45 data sets are correctly segmented. This behavior makes it possible to automatically select an appropriate preflooding height. The broader the h_{pf} range, the more robust the segmentation will be to further noise and inhomogeneity (Table 2.1).

Numerical Issues

The complexity of the modified watershed transform is linear in the number of voxels N for voxel processing, and in the worst case proportional to $N \log N$ for sorting when using a Quick Sort algorithm [PRESS *et al.*, 1988–1995]. However, for discrete gray value images, a fast bucket sort algorithm [ISAAC and SINGLETON, 1956] can be deployed that is also linear in the number of voxels, such that the overall WASS algorithm is computed with linear complexity $O(N)$. On a Pentium IV, 1.7 GHz, one million voxels are processed in less than a second.

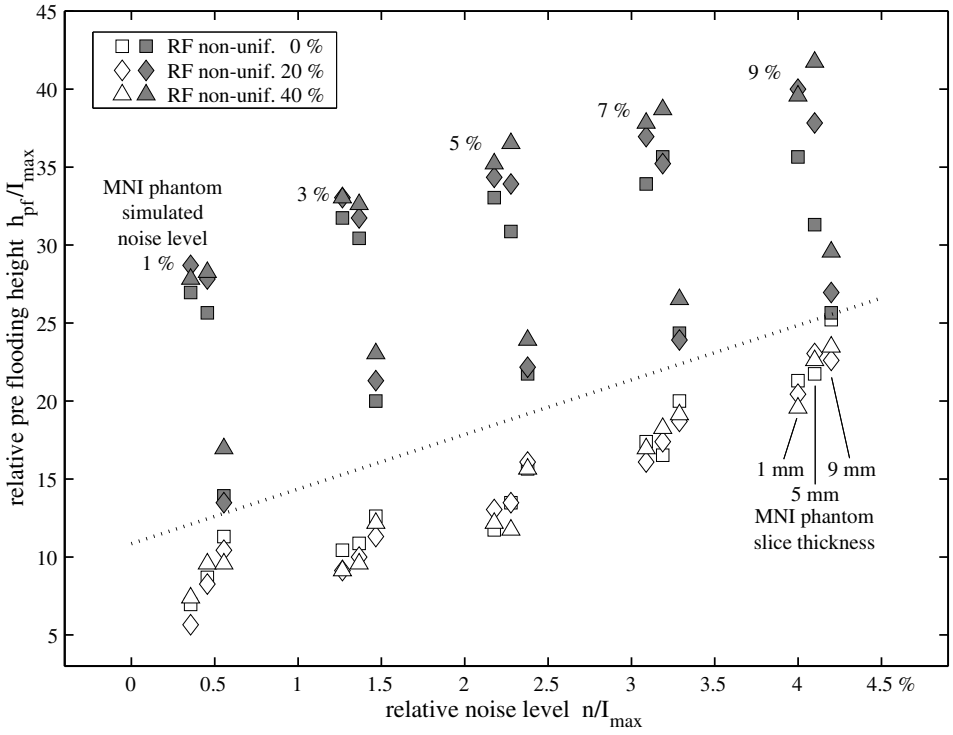


FIGURE 2.7: Range of proper preflooding heights for skull stripping (minimum: white, maximum: black markers) plotted as a function of noise level for 45 data sets of the MNI database. For each noise level, slice thickness is increasing from left to right (1 / 5 / 9 mm).

2.4 Discussion

The described WASS algorithm is able to successfully segment the whole brain in all 133 data sets, without any preprocessing. For any given image, the parameter h_{pf} can be varied over a certain range without changing the output, which is a measure for the robustness of the method. The sensitivity in comparison to manual expert segmentation is estimated to be more than 96 %. Differences (less than 4 %) are mainly located in the dark intensity region at the brain boundary, i. e. the interface between bone, meninges, and CSF. In some cases, larger parts of the superior sagittal sinus were included.

The results were satisfactory even where the range of proper preflooding heights h_{pf} has been rather narrow. This was the case in six of 90 simulated images due to a slice thickness of 9 mm, which is too thick for a good separation of eyes and frontal lobes (cf. Fig. 2.5 right), and in two of 43 clinical data sets (cf. Table 2.1). When inspecting these two images, we found some isolated bright voxels in various parts of the brain, thus increasing the required minimum h_{pf} value.

The merging rules introduced in Section 2.2 are well defined and easy to compute. They enabled the watershed algorithm to solve the skull stripping problem in all cases and lead to fast processing times. However, further improvements regarding the robustness of the method seem to be possible by modifying the merging rules. The measure for basin depth,

e. g., could be replaced by a measure that considers all gray values of a basin, not only the darkest one.

2.5 Conclusion

The modified watershed transform presented here is a powerful tool for segmenting the whole brain from MRI data sets. The application to the original MRI data rather than the gradient image is a new approach, which performs excellent with images showing an intensity characteristic comparable to the MR T1 signal. In particular, it is important that GM is at least as bright as CSF but not brighter than WM.

The 3D watershed transform with preflooding on the original data interpreted as depth image bears several advantages compared to existing skull stripping techniques, in that it is extremely robust to RF nonuniformity and noise. Moreover, it is fast and does not require any preprocessing, such as intensity normalization or denoising. The only underlying anatomical assumption is the connectivity of white matter. It is model-free in the sense that no assumptions on the smoothness or shape of the brain surface are made and no initialization of a template is required. Interaction techniques such as region selection or watershed placement can easily be incorporated.

A more sophisticated extension of this algorithm is described in Chapter 4. While in this chapter, the transform has to be performed for each choice of the parameter h_{pf}, a single transform will suffice in order to generate segmentation results for any possible h_{pf} value, and for an arbitrary number of include and exclude markers, both within less than a second. These extensions lead to a substantial increase in user friendliness, particularly in cases of varying image quality, requiring an interactive choice of h_{pf}. Chapter 4 also includes a further evaluation of robustness of this approach based on clinical and artificial images (with extreme noise and nonuniformity) as well as a full automation of the skull stripping procedure.

The presented algorithm provides the basis for a standardized brain segmentation procedure that increases reliability and reproducibility in the field of neuroimaging. It was adopted by the group of Dale, Fischl, *et al.* at the MGH NMR Center, Charlestown/MA, and integrated as a part of their software package FreeSurfer[7]. In order to generate a smooth brain surface estimate, they implemented a shrinking balloon template following the watershed-based skull stripping [Segonne *et al.*, 2004].

Publications. *Large parts of this chapter were presented at MICCAI 2000, Pittsburgh* [Hahn and Peitgen, 2000]. *As mentioned above, an extension of the skull stripping algorithm was presented together with* Segonne *et al.* [2004]. *The watershed-based skull stripping procedure is contained in a US patent application* [Hahn, 2001].

[7] FreeSurfer is available at `http://surfer.nmr.mgh.harvard.edu/`. The name of the respective software module is `mri_watershed`.

CHAPTER 3

Theory of Interaction and Automation

Abstract. *This chapter examines the theory of effective human-computer interaction, with reference to image segmentation. The field of tension is analyzed as determined by user interaction on the one hand and automation on the other hand. It is questioned to which extent user independency and interaction are compatible with each other. To derive practical guidelines for user interfaces, the time necessary to interactively solve an image analysis problem is considered as a measure of optimality. Moreover, the role of human experts, as well as the complementarity of data and knowledge are discussed.*

THERE may be nothing so practical as a good theory, but coming up with an effective theory is often difficult. By definition, a theory, taxonomy or model is an abstraction of reality and therefore must be incomplete. However, a good theory should at least be understandable, produce similar conclusions for all who use it, and help to solve specific practical problems." [SHNEIDERMAN, 1997] In our case, practical problems relate to the field of interactive systems in quantitative image analysis (QIA). Addressing their theoretical aspects is intended to help understand their general nature, as well as the nature of methods that lead to their solution. According to SHNEIDERMAN, many theories are needed to describe the multiple aspects of interactive systems.

3.1 On Effective Human-Computer Interaction

An appealing and comprehensible model of user interaction is the four-level approach that FOLEY and VAN DAM developed in the late 1970s (Fig. 3.1). "This approach is convenient for designers because its top-down nature is easy to explain, matches the software architecture, and allows for useful modularity during design. Designers are expected to move from conceptual to lexical, and to record carefully the mappings between levels." [SHNEIDERMAN, 1997, pp. 54 f.]

As an example, let us show what their model means in an interactive system that is intended to segment, visualize, and quantify the brain in volumetric MR images as described in Chapters 2, 4, and 8. The specific meanings of the four levels are described as follows:

Conceptual Level:	The user's mental model of the interactive system.
Semantic Level:	Describes the meanings of user's input and computer's output display.
Syntactic Level:	Defines how units that convey semantics are assembled into a complete instruction.
Lexical Level:	Defines the precise control mechanisms, also has to deal with device dependencies.

FIGURE 3.1: *Four-Level Approach to User Interaction* by FOLEY and VAN DAM [according to FOLEY *et al.*, 1990].

▷ *Conceptual:* An interactive system comprising two consecutive steps, where the first step is interactive segmentation, the second is interactive visualization of the segmentation result and quantitative parameters. The integration of automated steps to preset visualization parameters and to compute quantitative parameters from an actual segmentation, according to individual image statistics, is also part of the conceptual level.

▷ *Semantic:* Describes which types of information are used for input and output. Image data is displayed as gray levels on two-dimensional sections. The current state of the segmentation mask is displayed as colored overlay. For user steered segmentation, positional markers, e. g., for inclusion or exclusion of regions, are specified via a pointing device and visualized directly on the image data using appropriate colors. Parameters are grouped within a control panel. The choice of parameters that are controlled interactively, and of what is hidden from the user, belongs to the semantic level.

▷ *Syntactic:* The syntactic level of the proposed solution to the brain segmentation problem is quite simple, even if the segmentation problem itself is complex. Every user interaction represents an atomic command that is largely independent from others. The large scale syntax is already defined by the conceptual level, i. e. the order by which various tasks are performed: choice of data set, segmentation, exploration of results, and possibly iterative interactive refinement. (The latter will be motivated on p. 39.)

▷ *Lexical:* The lexical level comprises the precise mechanisms, e. g., how markers are placed and removed, or how the input device (most commonly a mouse) is translated to a 3D object transformation (rotate, zoom, etc.) at the visualization step. Two simple mechanisms are proposed for marker editing: pointing to an existing marker within a given tolerance radius will either (i) remove or (ii) select this marker, instead of creating a new one.

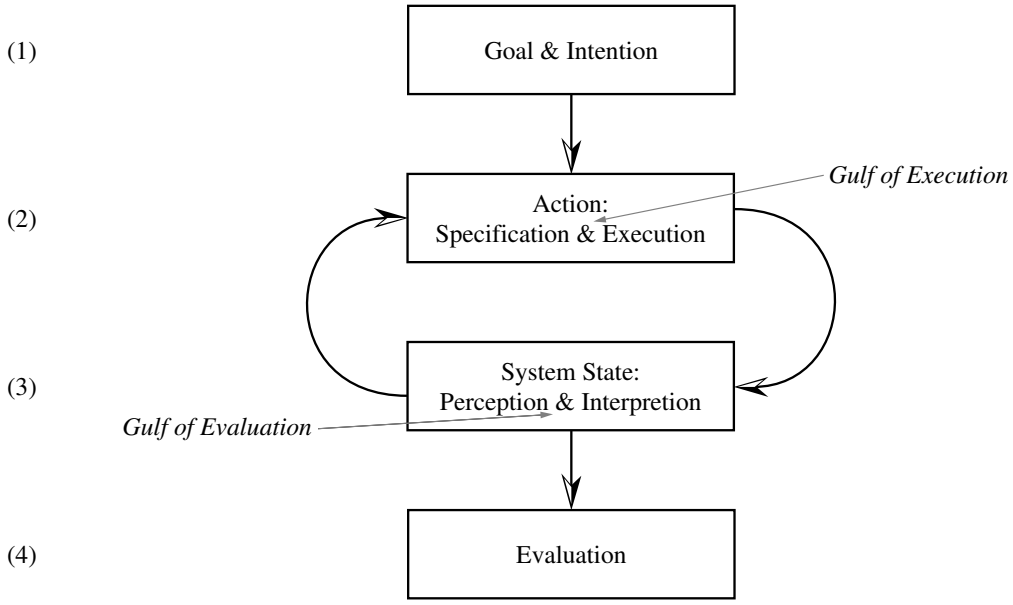


FIGURE 3.2: Cycle of action and evaluation that enables *iterative interactive refinement* (cf. text on p. 39).

In contrast to top-down models, a useful approach to forming theories of interactive systems is to describe the stages of action that users go through in trying to use a system. NORMAN [1988] offers seven stages of action as a model of human-computer interaction [SHNEIDERMAN, 1997, p. 57], which for our purpose can be reduced to these *Four Stages of Action* (stages according to NORMAN [1988] in braces):

(1) Forming the goal and intention (Norman 1–2).

(2) Specifying and executing the action (Norman 3–4).

(3) Perceiving and interpreting the system state (Norman 5–6).

(4) Evaluating the outcome (Norman 7).

It is possible to associate NORMAN's stages and FOLEY and VAN DAM'S *separation of concerns*, which describes how a user (i) forms a conceptual intention, (ii) reformulates it into the semantics of several commands, (iii) constructs the required syntax, and (iv) eventually produces the lexical token by a certain mouse click or key code [SHNEIDERMAN, 1997, p. 57].

Within the Four Stages of Action, a close relationship between stage 2 and stage 3 is desired to achieve highest efficiency of a system by enabling fast *iterative interactive refinement* of the system's state toward the goal specified by the user. NORMAN makes a contribution by describing a dynamic process placing his stages in the context of *cycles of action and evaluation*. Furthermore, the four-stages (originally formulated as seven-stages) model naturally results in the identification of the *gulf of execution*—the mismatch

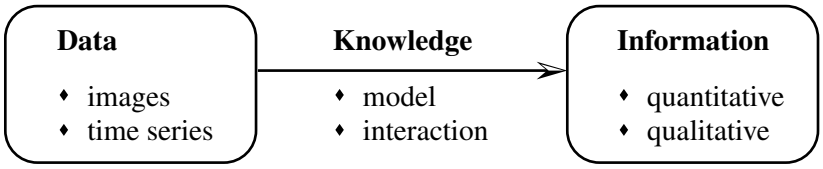


FIGURE 3.3: Diagram relating knowledge to data and derived information (cf. body text).

between the user's intentions and the perceived actions—and the *gulf of evaluation*—the mismatch between the system's representation and the user's expectations, in particular, the perception of the system's feedback after the action was carried out [SHNEIDERMAN, 1997]. This concept can be summarized schematically as shown in Figure 3.2.

3.2 On Knowledge, Information, and Models

Knowledge, on the one hand, is to be regarded as result of the scientific process. On the other hand, accumulated knowledge enters the design of models. Analytical models serve to describe specific relations and mechanisms. From such models, rules are derived on how to perform quantitative measurements and how to interpret their results. In a more qualitative way, a model can be identified with a set of assumptions, which are hypothetical simplifications of nature, created on the basis of knowledge. In addition to scientific knowledge, the implementation of analytical models requires technical knowledge, which we also call know-how. Moreover in interactive systems, case specific knowledge is contributed by expert users through interaction.

Knowledge is related to data on the one hand and to the information derived from data on the other hand. In the context of image analysis, digital images represent sets of primary information. Referring to primary information as data, images are accordingly called *data sets*. The term *information* in the narrower sense is used for the state of variables or parameters at a higher level. Many bits of data are needed to generate a single bit of information, and knowledge governs this transition from data to information. This relation is schematically presented in Figure 3.3.

In order to describe different levels of variability within the data and how these require different analytical models, we can construct a three-level pyramid where variability increases from bottom to top (cf. Fig. 3.4):

(a) Zero variability may be described by fixed models.

(b) Limited variability may be approximated by assumptions, or within a defined range of variation, automatic adaption may be feasible.

(c) Where variability and unpredictability exceeds a certain level, interaction, i. e. individual knowledge by an expert user, should guide the system to be applicable to the whole range of cases.

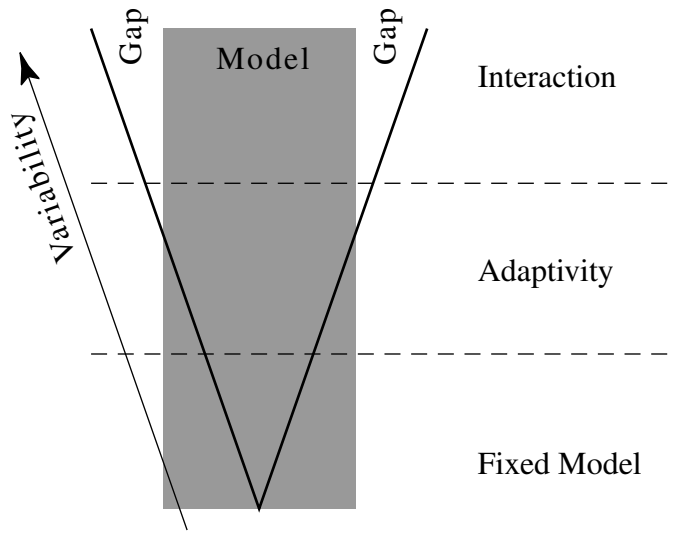


FIGURE 3.4: The *Inverse Pyramid (Large 'V') of Variable Systems* requiring fixed models, adaptive models, and interaction: The tip of the V can be fully described by a fixed parameter model. Adaptivity can cover a broader range of variability. However, where the gap between model and system becomes too large, adaptive models should be extended by interactivity.

3.3 On Interaction and Automation

The degree of automation in technical systems has been rapidly increasing as mechanical and analytical procedures became more standardized, hardware reliability increased, and software verification improved. With routine tasks, automation could be preferred, since the potential for error may be reduced. "However, I believe that there will always be a critical human role, because the real world is an *open system*; there is a nondenumerable number of unpredictable events and system failures. By contrast, computers constitute a *closed system*; there is only a denumerable number of normal and failure situations that can be accommodated in hardware and software. Human judgement is necessary for the unpredictable events." [SHNEIDERMAN, 1997, p. 83] Consequently, a balance of automation and human control has to be found.

In order to decide, which tasks within quantitative image analysis will be assigned to humans and which to machines, it is instructive to inspect the list of relative abilities of humans vs. computers/machines as presented in Figure 3.5. For example, it is unlikely that a computer will autonomously be able to design novel interactive systems that are optimized to solve new specific problems. However, a computer is able to apply a well defined set of methods to this specific problem and evaluate the results following standardized procedures. Again, in case of failure and unexpected events, humans are superior in adapting decisions and selecting alternatives.

The goal of the design of interactive systems is to provide an operator with sufficient information about the system's current state and activities, such that the operator has the capacity to perform interaction correctly. The complete system must be properly designed

Computers generally better:

- ▷ Count or measure physical quantities
- ▷ Store quantities of coded information accurately
- ▷ Make rapid and consistent responses to input signals
- ▷ Process quantitative data in prespecified ways
- ▷ Perform several activities simultaneously
- ▷ Maintain performance over extended periods of time

Humans generally better:

- ▷ Detect stimuli in noisy background
- ▷ Sense unusual and unexpected events
- ▷ Draw on experience and adapt decisions to situation
- ▷ Select alternatives if original approach fails
- ▷ Develop new solutions

FIGURE 3.5: Abilities of computers vs. humans [selected items from SHNEIDERMAN, 1997, p. 84]

and tested, not only for normal situations, but also for as wide a range of anomalous situations as can be anticipated [SHNEIDERMAN, 1997, p. 85]. As complexity of interaction increases, it is advisable to evaluate the system's most essential tasks by a representative group of operators/users.

In addition to the levels of variability (cf. p. 40), it is valuable to consider different levels of interaction. Levels of interaction most importantly refer to the gulf of execution [NORMAN, 1988], i. e. to the question on how much of a system's state can by altered by an atomic user action. The lowest interaction level provides a one-to-one correspondence between user actions and the system's variables, whereas the highest level refers to a situation where user actions potentially affect the whole system. From a usability point of view

Higher level interactive systems are commonly called *semiautomated*. From that perspective, the highest level of interaction is attained by an automated system where the user has no direct control over the process of computation but can only accept or reject the system's result without an option to modify incomplete or incorrect results. For a semiautomated system, efficiency is closely linked to the reduction of NORMAN'S gulfs of execution and evaluation. A common strategy to reduce the gulf of evaluation is (i) to provide *feedback*, and (ii) to provide *rapid* feedback. In contrast, iterative refinement is jeopardized if feedback takes too long (i. e. longer than a few seconds), even if a broad range of parameters and interaction modes if offered to the user.

In image segmentation, interaction levels are commonly identified with different levels of detail or different scales of interaction. This motivates the idea of multi-scale approaches where interaction is performed on a more or less broad range of scales, typically from coarse to fine. In contrast, for systems that work on a fixed scale or level, it is generally a great challenge to find the *right* scale or level in order to solve a particular problem. OLABARRIAGA and SMEULDERS [1997, 2001] provide a good overview on different interaction modes for two-dimensional image segmentation.

Iterative refinement is an elegant strategy to reduce the gulf of execution; repeated evaluation-execution loops are ideally performed from coarse to fine within a multi-level—or multi-scale—framework. This could result in a system that offers solutions to simple problems requiring little or no interaction, and requiring more interaction as the complexity of the problem increases. A crucial question is whether the desired correction of the system's status is possible at any time—this can be seen as the most important usability criterion in this context [Olabarriaga, 1999, Olabarriaga and Smeulders, 2001]. Still, it is questionable if the user should have interactive access and control to all details of a system. This leads to the field of tension between interaction and user independency.[1]

3.4 On Experts, User Independency[1], and Time

When designing a technical solution for a certain class of image analysis problems within QR, the most important question to answer is: *which information is used?* More specifically, for the design of a computer assistant that comprises image segmentation and quantification, we have to answer three questions:

(a) Which level—or levels—of automation and interaction to target,

(b) Which tools and inspectors to deploy for control and correction of the segmentation and quantification process, and

(c) Which prior knowledge, i. e. anatomical or functional models, to use as a basis for automation.

When these questions are answered and the objects and actions including the desired level of automation identified, one can choose from various interaction methods. Shneiderman [1997, p. 71] gives a comprehensive overview and discriminates five primary interaction styles: (i) menu selection, (ii) form fill-in, (iii) command language, (iv) natural language, and (v) direct manipulation. He identifies two opposing directions in the philosophy of interactive systems:

▷ Some designers promote the notion of autonomous, adaptive, or anthropomorphic agents that carry out the user's intents and anticipate needs.

▷ The philosophical alternative to agents and anticipation is user control, responsibility, and accomplishment (direct manipulation).

Designers who emphasize the second believe that users have a strong desire to be in control and to gain mastery over the system. Users, who seek comprehensible and predictable systems, will hesitate to use those that are autonomous, complex, or unpredictable. Conversely, users who do not consider themselves as an expert or do not want to know the details of a problem, might prefer the manipulation of agents. Still, we are much in favor of the *direct* manipulation style, as will be discussed below.

[1] We use the term *user independency* to characterize a system that generates consistent results independent of who operates the system, given that these users have the knowledge and ability to operate the system.

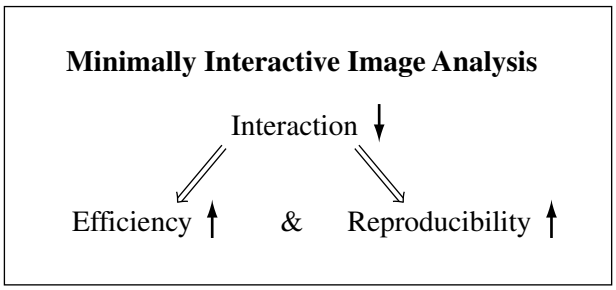


FIGURE 3.6: *Win-Win Situation of Minimally Interactive Image Analysis:* If the modes and the amount of interaction can be reduced to the very essential for a specific problem, reproducibility, and—since more images can be analyzed in a given period of time—also the efficiency of the system will be increased.

"No picture tells its own story [GOMBRICH, 1984]." In medical image analysis, as mentioned above, we often have the situation that knowledge has to be added to a set of images in order to obtain relevant information. Image gray values usually reflect a certain physical property at a given location; by comparing gray values from neighboring image elements, also contours and shapes become apparent. However, for example in CT, a gray value itself does not tell us whether it represents healthy or pathologic tissue, a bone or contrast enhancing tissue, and for bones, it does not tell us which bone it belongs to. In MRI, it cannot distinguish between two low intensity regions, such as the cerebral ventricles and subarachnoid space in T1 weighted head scans. For each piece of added knowledge or information, the designer of a system has to decide between interaction and automation.

It seems contradictory to aim at interactivity and at the same time at results that are user independent. Interactivity, on the one hand, is required to integrate individual knowledge from expert users. User independency, on the other hand, is expressed, e. g., by the demand for inter- and intra-observer reproducibility in medical image analysis. At a closer look, we note that quantitative results on a given case or image should agree between experts, such that the role of a human expert is mainly to prevent the system from failure in difficult cases. While programmed algorithms typically process each image in the same prespecified way within a range of variability and adaptability that is inherent to the algorithm, human understanding of images is mainly qualitative on the one hand and highly variable on the other hand, but can overcome the limitations of generic algorithms.

Especially in a clinical setting, but also in other domains of image analysis, interaction time is a critical factor and should be minimized. Therefore, expert users should only be asked to contribute their knowledge if required for the system to process a given data set. The system should then accept commands on various levels that are appropriate for the given problem, such that a single command has the potential to put the system into the intended state. We refer to this behavior as *minimal interaction*.

There are two strategies of increasing the efficiency of an interactive system: To minimize the number of cases where expensive interaction is required, or to implement iterative

Definition of Direct Manipulation:

- ▷ Visual representation (metaphor) of the "world of action"
  - Objects and actions are shown
  - Analogical reasoning is tapped
- ▷ Rapid, incremental, and reversible actions
- ▷ Replacement of typing with pointing and selecting
- ▷ Immediate visibility of results of actions

Major Concerns:

- ▷ Possibly increased system resources
  - Requires graphics display and pointing devices
- ▷ May be hard to program
- ▷ History and other tracing may be difficult
- ▷ Some actions may be cumbersome

FIGURE 3.7: Definition of *Direct Manipulation* as one of the five interaction styles, and major concerns associated with direct manipulation [according to SHNEIDERMAN, 1997, pp. 72 and 229].

refinement (cf. p. 39) that allows for minimally interactive processing, even in difficult cases. Of course, a combination of the two strategies is favorable. If designed properly, this results in a win-win situation where both reproducibility and efficiency are increased (cf. Fig. 3.6).

3.5 Discussion

Among the list of possible interaction styles (p. 43), we consider direct manipulation to be most appropriate for interactive image analysis. Menu selection and form fill-in do not meet the complexity, while command and natural language do not meet the complexity and topology of an image. Still, these styles do support direct manipulation in specific situations. The interaction styles discussed by OLABARRIAGA and SMEULDERS [2001] mainly refer to two-dimensional problems and are not directly applicable to the third dimension, where most problems within this thesis are embedded. For example, a two-dimensional interactive contour would correspond to an interactive balloon in three dimensions. The direct local steering of such a balloon is not a trivial task [COHEN and COHEN, 1993, KONRAD-VERSE, 2004]. Conversely, a point marker can be used to identify the interior of a specific object in arbitrary dimensions.

SHNEIDERMAN also gives a definition of direct manipulation besides major concerns that have to be taken into account when implementing a system based on direct manipulation (Fig. 3.7). By such a system, all data, system state, user input, and feedback can be integrated within an interactive rendering area. We argue that markers, which link user input and algorithmic entities with certain specified image positions, should play a central role in system design for interactive image analysis.

We are also advocating the concept of iterative refinement that has the potential to offer minimally interactive systems even for complex problems. This corresponds to hierarchical interaction designs that, according to SHNEIDERMAN [1997, pp. 31 f.], play a major role in theory and practice. In many situations, an optimal system in order

to facilitate efficient interactive refinement is accordingly represented by semiautomated multi-scale interaction. In an ideal case, both control and correction are directly linked to appropriate layers within the algorithm. Of course, the mechanisms for control and correction of the system's state must be adapted to the specific problem. Compelling cases for the efficient use of iterative refinement will be provided in Chapters 5 (bone segmentation), 7 (volumetry of the cerebral ventricles), and 8 (brain volumetry). Iterative refinement is, on the one hand, well described by NORMAN'S stages of action. On the other hand, FOLEY and VAN DAM'S model offers an intuitive approach to the hierarchical architecture of a system. Both models provide valuable contributions to the design of an interactive system.

Among the well known *eight golden rules of interface design* [SHNEIDERMAN, 1997, pp. 74 f.], two are most applicable to the challenges within this thesis, namely (i) the provision of feedback, and (ii) the reversibility of actions. To meet the user's expectations, feedback should be given not later than after two seconds. Reversal of actions may be provided, e. g., by a marker-based approach where every element of user input is visualized and may be removed or modified at a later stage of processing. Two important guidelines for data display and entry are provided by SHNEIDERMAN [1997, p. 82], the second thereof once again pronouncing the need for efficient or minimally interactive processing:

(a) *Compatibility* of data display with data entry; the format of displayed information should be clearly linked to the format of the data entry; and

(b) *Minimal input action* by users; fewer input actions mean greater operator productivity and usually fewer chances for error;

For the design of quantitative image analysis systems, it is essential to decide which parts thereof to automate. As motivated in the previous sections, full automation is not optimal as soon as failures or the need for corrections exist. If not answered by an atlas or similar automated technique, interaction should be used to answer the question of where an object is located, but preferably not to trace it's exact boundaries. In an ideal case, a single click into an object suffices to segment all its boundaries. Furthermore, all kind of statistical image analysis is to be performed automatically.

If interaction on boundaries is inevitable, active contours such as the live wire algorithm [BARRETT and MORTENSEN, 1997, SCHENK *et al.*, 2000] provide a good combination of flexibility and semiautomation. Especially for clinical applications, users desire full control over the accuracy of the system, such that deformable templates (snakes, balloons, etc.) or atlas-based segmentations often are often disadvantageous [cf. OLABARRIAGA and SMEULDERS, 2001]. Furthermore, while contours are restricted to a specific dimension, markers acting as seed points are applicable to regions of arbitrary dimension (cf. Fig. 4.13 for a four-dimensional problem).

Since markers are commonly placed within objects rather than at their boundaries, we expect that strictly marker-based segmentation, if successful, provides a better reproducibility compared to strictly contour-based interaction. In the first case, the exact boundaries are mainly governed by the image, while in the second, they largely depend on user input. Also, since edges are implicitly defined by region-based methods and vice versa,

a combination of both approaches is possible. A drawback of region-based approaches is that they generally require some sort of object connectivity. Still, if well chosen, the measure of connectivity plays a major role in strengthening the robustness of region-based segmentation.

Motivated by these considerations, we will in the following chapter extend the watershed transform (WT) to a hierarchical and minimally interactive segmentation tool. We will show that the WT is robust to both image gradients and to image nonuniformity (cf. p. 19), and that it is suitable for minimally interactive multi-scale segmentation. Special attention will be paid to the robustness and efficiency of derived algorithms.

Publications. *This chapter contains previously unpublished work.*

Alles sollte so einfach wie möglich gemacht werden,
Aber nicht einfacher.

—Albert Einstein

CHAPTER 4

Fast Watershed Transform Revisited and Automation

Abstract. *In this chapter, we introduce the Interactive Watershed Transform (IWT) for efficient segmentation of multidimensional gray scale images. The IWT builds upon a Fast Watershed Transform (FWT) in analogy to an immersion scenario*[1] *followed by a hierarchical organization of the resulting basins in a tree structure. Each local image minimum is represented as an atomic basin on the lowest hierarchy level. The FWT consists of two steps. First, all image elements are sorted according to their image intensity using a Bucket Sort algorithm. Second, each element is processed exactly once with respect to its neighborhood (e. g. 4, 6, 8 etc. direct neighbors for 2D, 3D, 4D etc. transform, respectively) in the specified order. Sorting, processing, and tree generation are of order $O(N)$, where N is the number of voxels. After computing the WT, one global parameter, the so-called preflooding height, and an arbitrary number of markers can be evaluated close to real-time in order to control tree partitioning and basin merging. The IWT combines automation and efficient interactive control in a coherent algorithm while completely avoiding oversegmentation, which is the major problem of the classical WT.*

The IWT has been successfully applied to interactively segment a large variety of objects in medical images without making assumptions on the objects' shapes (cf. Ch. 5, 7 and 8). Furthermore, the hierarchical nature of the IWT is used to automate the separation of brain from non-brain tissue, also referred to as skull stripping. Therefore, the characteristic dependency of the maximum basin size on the merging parameter is analyzed. Much as in Chapter 2, the connectivity of white matter is the only anatomical assumption for this automation. Finally, the robustness of this approach to extreme image noise and nonuniformity is evaluated.

IMAGE SEGMENTATION is a critical part of most image analysis and visualization systems for medical diagnosis, as well as therapy planning and monitoring. While fully automated segmentation algorithms have been developed for a variety of specialized tasks and types of images [e. g. Atkins and Mackiewich, 1998], interactive control and correction of those methods, as well as efficient user-steered methods rarely exist, especially for 3D and higher dimensional data [Olabarriaga and Smeulders, 2001, Pohle,

[1] In the following, we use the term *immersion-based* for algorithms in analogy to an immersion scenario.

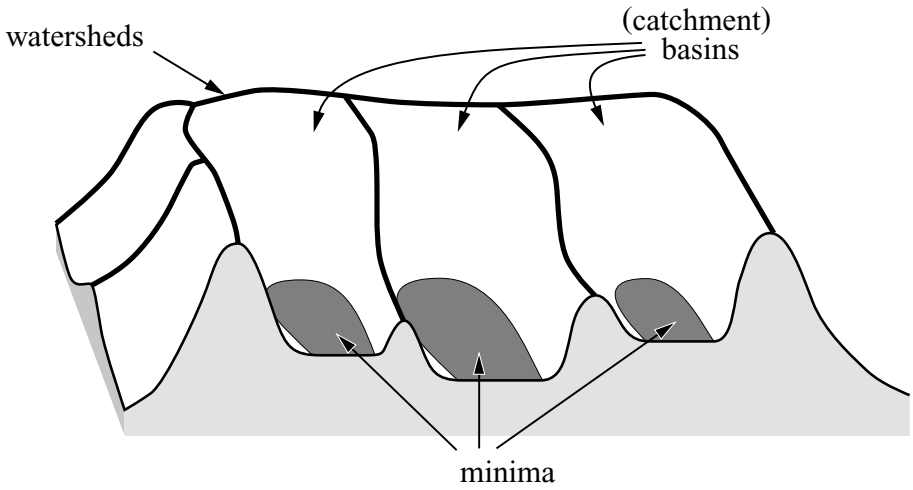


FIGURE 4.1: Topographic interpretation of a 2D image with minima, basins and watersheds [according to SOILLE, 2003, p. 269].

2004]. In clinical applications, automated methods often fail due to high anatomical and pathological variability that cannot entirely be covered by the underlying models.

The goal of this chapter is to present an algorithm for efficient segmentation of 2D, 3D, and higher dimensional gray scale images that is based solely on the image data—intensity and connectivity—and user interaction. Inspired by SERRA [1982] and SCHINDEWOLF and PEITGEN [2000], we have decided to build the algorithm upon a watershed transform (WT), which is the key concept for image segmentation in gray scale Mathematical Morphology [ROERDINK and MEIJSTER, 2000, SOILLE, 2003, p. 278]. The WT has shown broad applicability for many image segmentation problems and is known to yield robustness in extracting meaningful regions and contours. However, oversegmentation due to noise and local texture is the major problem of the WT in its original form [HARIS *et al.*, 1998]. We aim at a method that completely eliminates oversegmentation without losing the effectiveness of the WT that allows for efficient interaction. A further goal is to avoid the necessity of specific anatomical models on the one hand and any kind of preprocessing, such as noise reduction, on the other hand that are both limiting the versatility of a segmentation method. In specific cases, also fully automated segmentation shall be supported.

The chapter is organized as follows. After a reference to related work regarding the watershed transform such as its definition and algorithmic variants, we present a novel algorithm called Interactive Watershed Transform (IWT). The algorithm consists of two parts, which are described in detail. Special emphasis is put on algorithmic optimization issues. Before closing with a discussion and summary, we introduce a parametric analysis of the IWT's hierarchical basin structure. This analysis is employed to fully automatically solve the skull stripping problem, as discussed in Chapter 2.

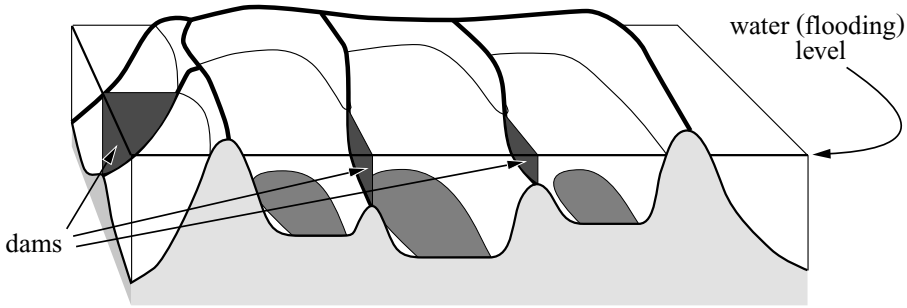


FIGURE 4.2: Interpretation of an immersion scenario on the landscape depicted in Figure 4.1. At positions, where water coming from neighboring basins would merge at progressive flooding levels, dams are constructed [according to SOILLE, 2003, p. 270].

4.1 Definitions and Related Work

Morphological Watershed Transform

Since the early 1990s, there has been a considerable amount of scientific work on the watershed transform that has originally been proposed by DIGABEL and LANTUÉJOUL [1978] as an image processing tool. In Mathematical Morphology [SERRA, 1982], an n-dimensional gray scale image is often interpreted as $(n+1)$-dimensional topographic relief where the numerical (i. e. gray) value of each image element (i. e. pixel or voxel) defines the elevation at the respective position. A topographic interpretation of a two-dimensional image is depicted in Figure 4.1. The most important notions in this context are the ones of minima, catchment basins (or simply basins), and watersheds that are separating basins from each other. Using this terminology, the WT transforms an image into a disjoint set of basins plus a set of watersheds.

A breakthrough in the computation of image watersheds was made by VINCENT and SOILLE presenting an immersion-based[1] algorithm orders of magnitude faster than previous ones, more accurate, and more flexible; their TPAMI paper [VINCENT and SOILLE, 1991] also includes a summary of earlier work, including the watershed transform on DEMs (digital elevation models), and is probably the most cited WT publication. Since then, most related publications deal with theoretical, algorithmic, parallelization [ROERDINK and MEIJSTER, 2000, MOGA, 1997], and multiresolution issues [MOGA, 1999, GAUCH, 1999], and also with the segmentation of multivalued or multispectral images [SCHEUNDERS and SIJBERS, 2002]. An excellent overview on definitions, algorithms, and parallelization strategies was recently published by ROERDINK and MEIJSTER [2000]. SCHINDEWOLF and PEITGEN [2000] propose a parallel region merging algorithm in equivalence to a watershed transform applied to the gradient image.

Our algorithm is based on the one proposed by VINCENT and SOILLE [1991], however it includes some modifications that will be described in Section 4.2. In the following, we give a possible WT definition for the continuous case, from where we will motivate and introduce the IWT.

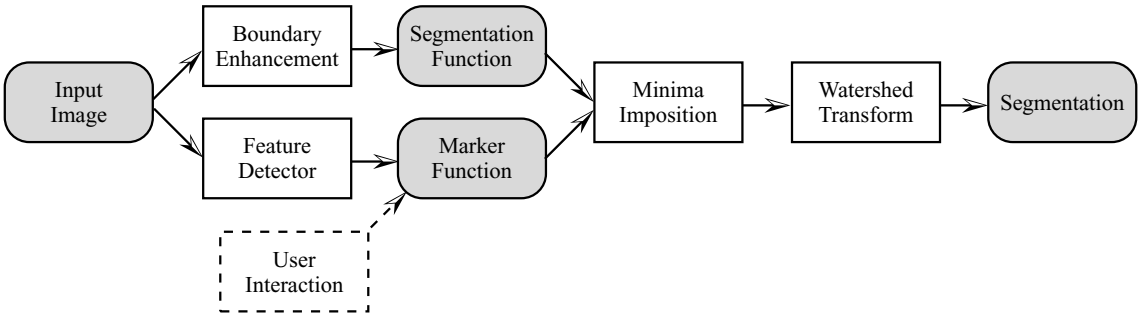


FIGURE 4.3: *Morphological paradigm for image segmentation.* Image understanding and user interaction is done at the very first stages of the process. The key to success consists in generating pertinent marker and segmentation functions. The rest of the procedure is nonparametric, the watershed transform therein playing a key role [according to SOILLE, 2003, p. 278].

Topographical Distance

For the continuous case, we restrict ourselves to the watershed definition given by MEYER [1994] based on distance functions, with different distance functions resulting in different definitions [ROERDINK and MEIJSTER, 2000]. The *topographical distance* within an image I between two points p and q in a connected domain D is defined by

$$T_I(p,q) = \inf_{\gamma} \int_{\gamma} ||\nabla I(\gamma(s))|| \, \mathrm{d}s, \tag{4.1}$$

where I is an element of the real twice continuously differentiable functions $\mathcal{C}^2(D)$, the infimum being over all smooth curves γ inside D with $\gamma(0) = p$ and $\gamma(1) = q$. The path γ with shortest distance between p and q is also the path of *steepest slope* [ROERDINK and MEIJSTER, 2000]. This directly leads to rainfalling-based watershed algorithms that simulate water following the paths of steepest slope (cf. pp.53 f.).

Definition of Watershed Transform

Let $\{\zeta_k\}_{k \in K}$ denote the (local) minima of $I \in \mathcal{C}^2(D)$ for an index set K. The watershed transform defines two types of sets, namely *basins* and the *watershed set*, depending on I. The (catchment) *basin* b_k of a minimum ζ_k is defined as the set of points $p \in D$ that are (topographically) closer to ζ_k than to any other minimum ζ_o:[2]

$$b_k = \{p \in D \,|\, \forall\, o \in K \backslash \{k\} : I(\zeta_k) + T_I(p, \zeta_k) < I(\zeta_o) + T_I(p, \zeta_o)\} \tag{4.2}$$

The watershed W_I of an image I is the set of points, which are equidistant to at least two minima and therefore do not belong to any basin:

$$W_I = D \setminus \bigcup_{k \in K} b_k \tag{4.3}$$

[2]The terms $I(\zeta_k)$ and $I(\zeta_o)$ in Eq. 4.2 are required in order to account for the different levels of the basin minima in accordance to Eq. 4.1.

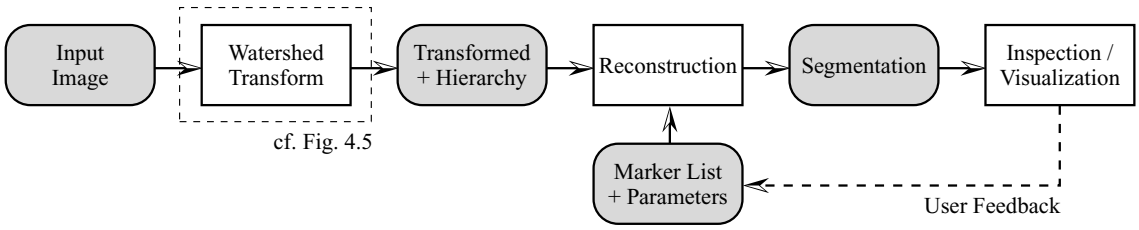


FIGURE 4.4: Diagram of main components and data flow of the IWT. In contrast to the classical *morphological paradigm for image segmentation* (cf. Fig. 4.3), where a marker function is imposed to the segmentation function *before* performing the watershed transform, user feedback in terms of markers and parameters is efficiently evaluated *after* the WT (cf. Fig. 4.5 for details of the employed WT).

Thus, the watershed transform of I is a mapping $\omega : D \longrightarrow K \cup \{w\}$ where $w \notin K$ is the watershed label, resulting in the following relations

$$\omega(p) = k \in K \quad \Leftrightarrow \quad p \in b_k \quad \text{and} \tag{4.4}$$

$$\omega(p) = w \quad \Leftrightarrow \quad p \in W_I \quad , \tag{4.5}$$

which means that basins are uniquely labeled (Eq. 4.4), and a special label w is assigned to all points of the watershed of I (Eq. 4.5) [according to ROERDINK and MEIJSTER, 2000].

Discrete Watershed Transform

Consider a digital image, i. e. a set of image elements that are defined on a regular d-dimensional grid $D \subset \mathbb{Z}^d$ with a t-dimensional function $I : D \longrightarrow \mathbb{R}^t$ assigning a value $I(p)$ to each $p \in D$. Unless otherwise stated, we assume that I is a one-dimensional function $I : D \longrightarrow \mathbb{N}$ assigning a positive integer value to each image element. Two problems arise when applying the above watershed definition to such an image. The first problem is the occurrence of plateaus, i. e. regions of constant gray value, as discussed in numerous publications [e. g. VINCENT and SOILLE, 1991, GAUCH, 1999, ROERDINK and MEIJSTER, 2000]. Here, two types of plateaus are considered, namely minimum and non-minimum plateaus. We will comment on p. 57 on the plateau problem.

The second problem, which is partly linked to the plateau problem, is the dependency of the watershed location on both the used algorithm and the grid connectivity [ROERDINK and MEIJSTER, 2000]. The two most frequently addressed families of WT algorithms are in analogy to *rainfalling* and *immersion*. In a typical rainfalling scenario, water is following the steepest slope of I until a minimum is reached. Near watersheds, the initial slope determines to which basin the respective image element belongs to.

For the immersion scenario, we imagine holes pierced in each regional minimum of the image I, which is regarded as topographic landscape [VINCENT and SOILLE, 1991]. When the landscape is slowly immersed into a lake, basins are filled progressively, starting from the minima at lowest altitude, and finally merge at the exact positions of the watershed lines. To prevent basin merging, dams are built (cf. Fig. 4.2). At the end of

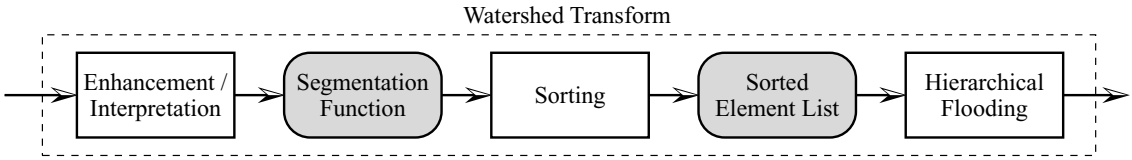


FIGURE 4.5: Detailed diagram for the watershed definition that is the first step of the IWT (cf. Fig. 4.4). As for most immersion-based[2] WT algorithms, flooding is preceded by the generation of a sorted index. Both, sorting and flooding are computed in linear time. Note that enhancement/interpretation can be omitted in many cases, where the IWT is applied to the original image data.

the immersion or flooding process, i. e. when the relief is fully immersed into or covered by water, the watershed set is defined by all image elements where dams were built. However, most of these image elements will exhibit a different slope to either side of the watershed and will therefore not be considered as part of the watershed set in a rainfalling scenario [ROERDINK and MEIJSTER, 2000]. We will provide more details on this problem in Section 4.3. Moreover, the immersion scenario often generates thick watersheds in case of non-minimum plateaus for obvious reason.

4.2 IWT—Interactive Watershed Transform

The IWT, which we introduce, is based on a Fast Watershed Transform in analogy to an immersion scenario, and hierarchically organizes the resulting basins in a tree structure [HAHN, 2001, HAHN and PEITGEN, 2003]. Each local image minimum is represented as an *atomic basin* on the lowest hierarchy level. It is only in a second step, i. e. *after* the WT, that markers and global parameters are used to control the actual segmentation result. A schematic overview of the complete algorithm is given by Figure 4.4.

In the classical paradigm for image segmentation from Mathematical Morphology, user input is only possible at the very first stage of processing as part of the marker function (according to SOILLE [2003, p. 278], cf. Fig. 4.3). In contrast, the IWT does not require user interaction before the very last stage of processing: the reconstruction step that is performed interactively close to real-time (cf. Fig. 4.4 and Algorithm 2).

We employ a WT consisting of three steps—image interpretation, sorting, and flooding (cf. Fig. 4.5). In most applications discussed in this thesis, the image is directly used as segmentation function, such that image enhancement or interpretation in a first step is omitted. In some cases, e. g. for gradient-based watersheds, the segmentation function is optionally generated from the input data. This generation is preferably nonparametric and should include some careful image interpretation. To facilitate and accelerate the flooding step, all image elements are sorted according to their image intensity before processing. In most practical cases, the number of elements exceeds the number of gray levels by some orders of magnitude, such that a *bucket sort* algorithm should provide the fastest sorting method. This procedure introduced by ISAAC and SINGLETON [1956] determines the exact frequency distribution for all gray levels, which induces the assignment of each element

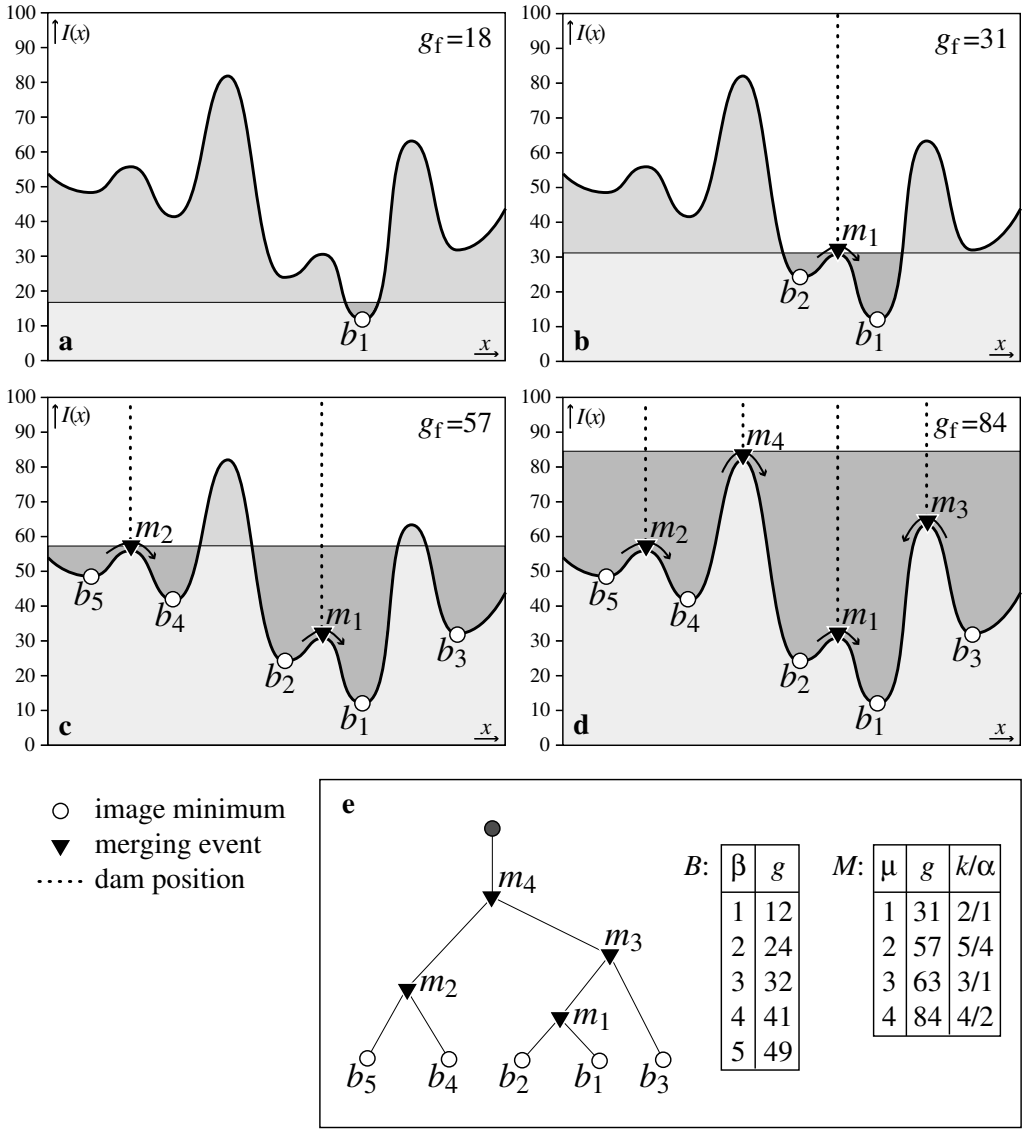


FIGURE 4.6: Illustration of *hierarchical watershed transform* as part of the IWT on a continuous 1D function $I(x)$ interpreted as a landscape. The landscape is sequentially flooded from bottom to top. *a*: At flooding height $g_f = 18$, a single basin b_1 exists. *b*: At $g_f = 31$, a second isolated minimum b_2 and the lowest separating ridge were detected. At the ridge point, a candidate for basin merging m_1 is registered and stored for future evaluation. *c*: At higher flooding heights, more and more image minima and merging procedures are registered. *d*: The flooding is complete at $g_f = 84$. *e*: All (atomic) basins and merging candidates have now been detected and are ordered in a hierarchical tree. Relevant information, i. e. the respective gray levels $g[\beta]$ and $g[\mu]$ and merging candidates $k[\mu]$ and $\alpha[\mu]$, are stored in basin table B and merging table M (cf. Algorithm 1 and Sec. 4.2 for details).

Algorithm 1 Fast hierarchical watershed transform for multidimensional digital gray scale images in analogy to an immersion scenario.

```
1: procedure HIERARCHICAL_WATERSHED_TRANSFORM
2: INPUT: input image I[p ∈ D]
3:    /* D: grid of image elements p ∈ D */
4:    /* N(p): set of image elements that are direct neighbors of p (e. g. 4, 6, or 8) */
5: OUTPUT: table of atomic basins B = {b_β}, β ∈ ℕ
6:    /* g[β]: the lowest (seed) gray value of basin b_β */
7:    /* r[β]: reference (by index) to the deepest already connected basin */
8:    /* l[β]: basin label used for the actual segmentation step (Algorithm 2) */
9: OUTPUT: table of possible merging events M = {m_μ}, μ ∈ ℕ
10:   /* k[μ]: index of basin candidate b_k that is considered for merging */
11:   /* α[μ]: index of a deeper neighbor basin b_α that b_k[μ] should be merged to */
12:   /* g[μ]: absolute gray value of the merging event m_μ */
13:
14: /* Init */
15: g_min ← min{I[p] | p ∈ D}      /* minimum gray value of input image */
16: g_max ← max{I[p] | p ∈ D}      /* maximum gray value of input image */
17: Y[p] ← 0  ∀ p ∈ D              /* clear work image Y[p] defined on D */
18: β ← 0                          /* reset basin counter */
19: μ ← 0                          /* reset counter for merging candidates */
20:
21: /* Flooding start */
22: for g_f ← g_min to g_max do    /* ascending loop over all gray values */
23:   for all p ∈ D with I[p] = g_f do /* loop over all iso-intensity elements */
24:     B_p ← {b_k ∈ B | ∃ q ∈ N(p) ∧ Y[q] = k}   /* neighboring basins */
25:     if B_p = ∅ then            /* create new basin b_β if p is isolated */
26:       β ← β + 1                /* basin counter increment */
27:       g[β] ← g_f               /* seed gray value of basin b_β */
28:       r[β] ← β                 /* self-reference for new basin b_β */
29:       Y[p] ← β                 /* element-basin assignment */
30:     else                       /* non-trivial neighborhood */
31:       /* determine index α of deepest connected basin */
32:       α ← argmin_k{BASE(k) | b_k ∈ B_p}
33:       Y[p] ← α                 /* element-basin merging */
34:       for all b_k ∈ B_p do     /* for all neighboring basins */
35:         if BASE(k) ≠ BASE(α) then /* avoid duplicate merging candidates */
36:           r[BASE(k)] ← BASE(α)  /* reference to deepest connected neighbor */
37:           /* counter increment for new candidate for basin-basin merging m_μ */
38:           μ ← μ + 1
39:           k[μ] ← k             /* register basin b_k as candidate for merging */
40:           α[μ] ← α             /* register deeper neighbor basin b_a to be merged to */
41:           g[μ] ← g_f           /* register absolute gray value of merging event */
42: end                            /* end HIERARCHICAL_WATERSHED_TRANSFORM */
```

Algorithm 1 (contd.)

```
43: /* Determination of deepest connected basin by recursive basin reference. */
44: procedure BASE(k)
45: if r[k] ≠ k then                    /* descend recursively until self-reference */
46:     k ← BASE(r[k])
47: return k                            /* end BASE */
```

to a unique cell in the sorted array [Vincent and Soille, 1991]. Still, for floating point images, a slower *quick sort* algorithm is used.

Immersion is performed after sorting. Each image element is processed exactly once in the prespecified order with respect to its neighborhood thus resulting in a very efficient algorithm. The neighborhood is determined by the connectivity of the underlying grid. In the following, we assume a regular, orthogonal, and isotropic connectivity such that the connectivity is defined by the number and arrangement of direct neighbors $N(p)$ of each element p. Most common choices are four and eight (including diagonal) direct neighbors for 2D, six and 26 direct neighbors for 3D, and eight direct neighbors for 4D grids. The exact immersion procedure is described by Algorithm 1 (pp. 56 f.). An illustration of a complete flooding process on a one-dimensional function is given by Figure 4.6.

After the transform, one global parameter, the so-called preflooding height h_{pf} (cf. Sec. 2.2, [Hahn and Peitgen, 2000]), and an arbitrary number of markers are evaluated close to real-time to control tree partitioning and basin merging. Basin merging is performed until (a) the zones of influence of different markers would be merged or (b) the preflooding height is met, as described in Algorithm 2. Some didactic examples for segmentation results are given in Figure 4.7.

Comments on Hierarchical Watershed Transform (Algorithm 1)

Since this thesis is directed toward image segmentation, we put emphasis on the decomposition of an image into labeled regions, or, in terms of the watershed transform, into catchment basins, whereas the extraction of watershed lines is not considered as an output of our algorithm. All image elements that are neighboring two or more basins are assigned to either of these basins.

Furthermore, we do not perform any kind of *SKIZ* computation (*skeleton of influence zones* [cf. Vincent and Soille, 1991, Roerdink and Meijster, 2000]), thus resulting in a (theoretically) inexact position of basin borders. To be more precise, non-minimum plateaus are not partitioned based on a distance measure but rather on the order of processing as provided by the sorting of image elements (cf. Algorithm 1, line 23). To achieve a proper SKIZ treatment, it would suffice to add an additional scanning for each threshold level [Vincent and Soille, 1991]. In most practical cases, though, the image exhibits a multitude of gray levels and is governed by noise such that image plateaus rarely exist.

The sorting step that results in an index to all image elements speeds up the whole algorithm, which is strictly linear in time. The only disadvantage is that it is memory

Algorithm 2 Reconstruction of image segments by basin labeling in dependence of markers and a global parameter (based on the output of the fast hierarchical watershed transform, cf. Algorithm 1).

```
 1: #define NOMARK 0                    /* further definitions cf. Algorithm 1 */
 2:
 3: /* Start reconstruction */
 4: procedure RECONSTRUCT_ALL
 5: for all b_k ∈ B do                  /* unmerge all basins first */
 6:    l[k] ← NOMARK                    /* unmark basin */
 7:    r[k] ← k                         /* self-reference to decouple basin */
 8: for μ ← 1 to μ_max do               /* traverse hierarchical tree from bottom to top */
 9:    COND_MERGE(μ, h_pf)              /* conditionally consider each merging candidate m_μ */
10: end                                 /* end RECONSTRUCT_ALL */
11:
12: /* Put into effect a single merging candidate m conditioned by h_pf and l */
13: procedure COND_MERGE(μ, h_pf)
14: k ← BASE(k[μ])                      /* deepest connection of basin b_k */
15: α ← BASE(α[μ])                      /* deepest connection of neighboring basin b_α */
16: if h_pf ≥ g[μ] − g[k] then          /* test basin depth against h_pf */
17:    /* test if one of b_k and b_α is unmarked or both marked identically */
18:    if l[α] = NOMARK ∨ l[k] = NOMARK ∨ l[α] = l[k] then
19:       r[k] ← α                      /* merge b_k and b_α */
20:    if l[α] = NOMARK then            /* if b_α is unmarked */
21:       l[α] ← l[k]                   /* propagate marker from b_k */
22: end                                 /* end COND_MERGE */
23:
24: /* Label lookup for computation of segmentation result for image element p */
25: procedure LABEL(p)
26: k ← Y[p]                            /* basin number lookup */
27: /* proceed directly to the deepest connected basin if unmarked */
28: if l[k] = NOMARK ∧ r[k] ≠ k then
29:    l[k] ← l[BASE(k)]                /* save label for speedup on reentry */
30: return l[k]                         /* end LABEL */
```

demanding since at least 24 bits have to be allocated for each element. (In a cubic image of 256^3 voxels, e. g., 2^{24} elements have to be addressed.) Another investment in order to achieve a fast algorithm is to allocate one additional element for the work image Y (line 17) at all borders of the grid D such that boundary checking for the computation of $N(p)$ can be omitted.

Optimization is possible by modifying the following pieces of the algorithm. Since the number of basins is not known a-priori, both the table of basins B and the table of merging candidates M should be allocated dynamically. The design of the merging table M is less critical since new elements are only stored at the end of a list, random access to elements

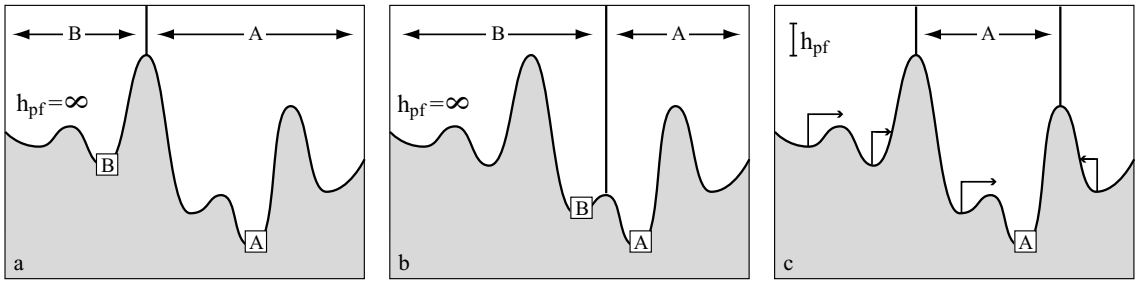


FIGURE 4.7: Three didactic examples of possible segmentations using markers and/or preflooding (cf. Fig. 4.6 for list of basins and merging events). As described in Algorithm 2, the basin labels and thus the image segments (resulting segmentation indicated by large arrows, top) are successively reconstructed based on markers (boxes near the function minima, two types: A and B) and the preflooding height h_{pf}. *a* and *b*: Markers are used to enforce watersheds between differently marked basins (vertical lines), h_{pf} being set to infinity (cf. Sec. 4.2). *c*: Alternatively, a single marker is used in combination with an appropriate preflooding height h_{pf} (ruler, top left). Two basins are merged (b_1 and b_2, cf. Fig. 4.6), whereas others remain separate (preflooding indicated by small arrows).

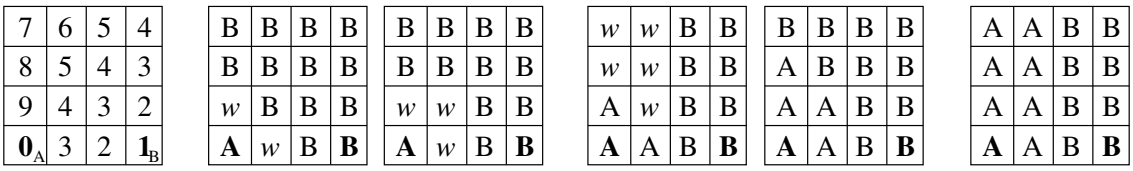
not being required. A standard template library could do a good job here. Conversely, since basin data is frequently and randomly accessed, one should carefully design the basin table B. We chose simple arrays for both tables with an initial size of 4096 elements; the size of the arrays being doubled and the contents being copied if required.

We further recommend the introduction of a single innermost loop for the determination of B_p and α (comprising lines 24 and 32). All neighboring basins b_k including their deepest connected neighbors BASE(b_k) can be, e. g., pushed to a stack, not checking for double entries that are filtered later in line 35. Then, $B_p = \varnothing$ is equivalent to an empty stack. The access to the neighborhood $N(p)$ by initializing a vector of index offsets for a respective connectivity (e. g., $[+1; -1; +n_x; -n_x]$ for 4-connectivity on a 2D grid) avoids unnecessarily expensive conditional code and obfuscating code doubling. Furthermore, the BASE(k) recursion (line 46) might be replaced by a non-recursive loop, and, if b_k is not self-referred ($r[k] \neq k$), one might store the return value in $r[k]$ for speedup on reentry.

Comments on Reconstruction of Image Segments (Algorithm 2)

In addition to the output of Algorithm 1, markers and the preflooding height h_{pf} serve as prerequisites for the reconstruction of labeled image segments. This reconstruction is intended to be computed close to real-time. Markers l^* at a position p are projected to the appropriate basin property through direct work image[3] lookup: $l[Y[p]] \leftarrow l^*$. The preflooding height usually is also controlled interactively by the user, while on each marker or parameter change the procedure RECONSTRUCT_ALL is called for fast update and feed-

[3] cf. Algorithm 1 for the introduction of the work image Y.



(a) original (b) 4-conn. (c) 8-conn. immersion (d) 4-conn. (e) 8-conn. topographical distance (f) 4-conn. FWT

FIGURE 4.8: WT on a square grid for three algorithms and two choices of connectivity (a–e according to ROERDINK and MEIJSTER [2000]). *a*: Original image with minima indicated in bold and labeled as A and B. *b*–*c*: Results according to immersion. *d*–*e*: Results according to topographical distance. *f*: Results according to the FWT algorithm introduced in Section 2.2 and also to the IWT.

back. LABEL(p) provides the current segmentation result for an image element p. Again, for speedup on reentry for the same basin, one would set $l[k]$ to a unique EXCLUDE value always if BASE(k) is unmarked (insert before line 30).

The comparison of Figures 4.6 and 4.7 reveals a seeming conflict. The merging of b_1 and b_2 in Figure 4.6 e is at a very low level in the hierarchical tree such that the segmentation depicted in Figure 4.7 b could not be obtained by only two markers. However, the trick that we use is to store the actually neighboring basins in the merging table but to think in terms of their deepest connected neighbors. When executing RECONSTRUCT_ALL for the basin and merging information of Figure 4.6 e, one can simply prove the correctness of the result. Hence, interpreting the result of Algorithm 1 as a hierarchical tree is simplistic and incorrect for specific cases.

We already commented on non-minimum plateaus. Since no costly explicit minimum or plateau handling is contained in the IWT, and since the processing of image elements is determined by the sorting order, plateaus in general will be split into several atomic basins produced by every image element that is isolated at the time of processing (cf. Algorithm 1). Nonetheless, the atomic basins belonging to a single plateau will be merged as long as the markers permit (even for $h_{\mathrm{pf}}= 0$ due to the greater-or-equal operation in COND_MERGE, Algorithm 2, line 16). Also in this case, the preflooding height can be interpreted as minimum basin depth.

4.3 Depths of the Watershed Transform

Connectivity and Algorithmic Aspects

As stated above, the results of a WT depend on both the connectivity of the underlying grid and the actual WT algorithm. While the definition based on topographical distance corresponds closely to rainfalling algorithms, immersion algorithms often are more practical in terms of complexity and processing speed. In order to understand these dependencies, it is instructive to analyze the result on a clearly arranged image, as presented in Figure 4.8. At first sight, it is surprising that in this example five out of twelve elements exist for which

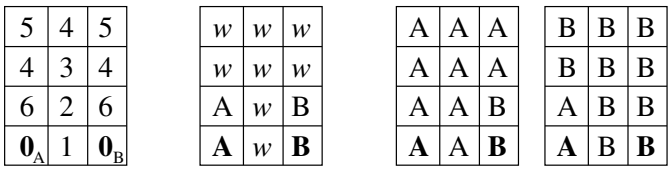


FIGURE 4.9: WT on the 4-connected square grid for two different algorithms [according to ROERDINK and MEIJSTER, 2000]. *a*: Original image with minima indicated in bold and labeled as A and B. *b*: Results according to topographical distance. Note that this result is consistent with both topographical distance and classical immersion-based algorithms for both 4 and 8-connectivity (cf. Fig. 4.8 b–e). *c–d*: Two extreme results consistent with the local condition (as defined on p. 61).

all three labels are possible (A, B, and *w*). In order to understand the difference between (b–c) and (d–e), recall that the WT definition based on topographical distance uses paths of steepest slope to determine basin partitioning, whereas immersion-based algorithms are independent of slope.

The last example (f) shows the result according to the two algorithms that were introduced in Sections 2.2 and 4.2. At watersheds, these algorithms give priority to the deepest neighboring basin. Therefore, many elements are attributed to basin A, in this example. Of course, basin depth could easily be replaced by another measure of basin priority. Note that even if produced by an immersion-based algorithm, Figure 4.8 f is actually more similar to the results based on topographical distance.

A major difference between b–e and f is that in the latter, no watershed elements are produced by definition. This is also consistent with the WT definition based on a *local condition* as introduced by A. BIENIEK *et al.* [according to ROERDINK and MEIJSTER, 2000]. Along this definition, for any image $I(D)$ without plateaus, a function ω assigning a label to each element p is called a watershed transform if the following two conditions are fulfilled:

(a) $\omega(\zeta_i) \neq \omega(\zeta_j),\ \forall\ i \neq j;$

(b) $\exists\ p' \in N^*(p)$ with $\omega(p) = \omega(p'),\ \forall\ p \in D$ with $N^*(p) \neq \varnothing.$

The first condition means that different minima ζ_i also have different labels. The subset of direct neighbors of p that is connected to p at steepest slope is denoted by $N^*(p)$; this subset is called *lower neighbors* of p [ROERDINK and MEIJSTER, 2000]. The second condition thus means that each element with lower neighbors receives its label from one of these neighbors.

The ambiguity of such a definition becomes most apparent by studying symmetrical images such as presented in Figure 4.9. Many elements that are close to the border between basins A and B can be attributed to either basin. Note that the algorithms introduced in Sections 2.2 and 4.2 could produce either of the two extreme results (c and d), according to different sorting orders. As we will see in the remainder of this thesis, however, this inaccuracy is irrelevant in most practical cases. If for some reason a symmetrical result is

required in those theoretical cases, one should perform a skeleton of influence zones (SKIZ) computation (cf. p. 4.2) or explicitly introduce a watershed label.

Treatment of Oversegmentation

The major problem of the WT in its original form is known to be oversegmentation, i. e. the tendency to produce too many basins [Haris *et al.*, 1998]. Within the concept of IWT, we provide two methods to overcome this problem. First, preflooding can be applied to merge small basins with adjacent basins before constructing watersheds, and second, basins can be marked individually in order to determine the minimum number of basins. In various instances, a combination of both has turned out to be beneficial.

The preflooding concept uses a measure of depth of a certain basin. As long as this depth is greater than the specified preflooding height, the basin will be separated from neighboring deeper basins. If not, it will be merged with the neighboring basin of highest priority. In our example, priority is equivalent to absolute basin depth. Of course, both rules—as introduced for preflooding and merging—could be altered in order to produce a different algorithmic behavior. Furthermore, in order to account for contrast variations across an image, the preflooding height can be a function of local image characteristics, such as intensity or variance. Such a function may additionally depend on one ore more parameters.

The second handle to control oversegmentation is using markers. Point markers applied to single image elements are suitable for most cases; only on very noisy data, it is appropriate to select multiple neighboring elements associated with each marker. An arbitrary number of different markers can be assigned to arbitrary elements prior to the transform for the classical WT, or prior to processing the list of merging events for the IWT. With markers, two basins are only merged if they are not marked differently. When merging a marked with an unmarked basin, the latter will receive the marker of the former. After the WT, all elements will be assigned to one of the markers—i. e. the number of different regions is determined by the number of different markers. Thus, oversegmentation is completely avoided by markers.

In the IWT, both concepts are combined in a consistent manner, such that the number and size of resulting basins depend on both: markers and the preflooding height.

Features of the Watershed Tree

The preflooding concept depends on a single parameter, which is the (global) preflooding height h_{pf}, in the standard case. As mentioned above, merging can also depend on more than one parameter, e. g. as a function of local image or basin characteristics. Since preflooding relates to the question of merging neighboring basins, it does make sense to use basin features to control the merging process. Possible basin features are gray value characteristics (such as the lowest gray value corresponding to the seed element, mean gray value, or variation thereof; for an application to bone removal in CTA data sets, cf. Color Plate C.4) as well as basin size, shape, and location with respect to some reference.

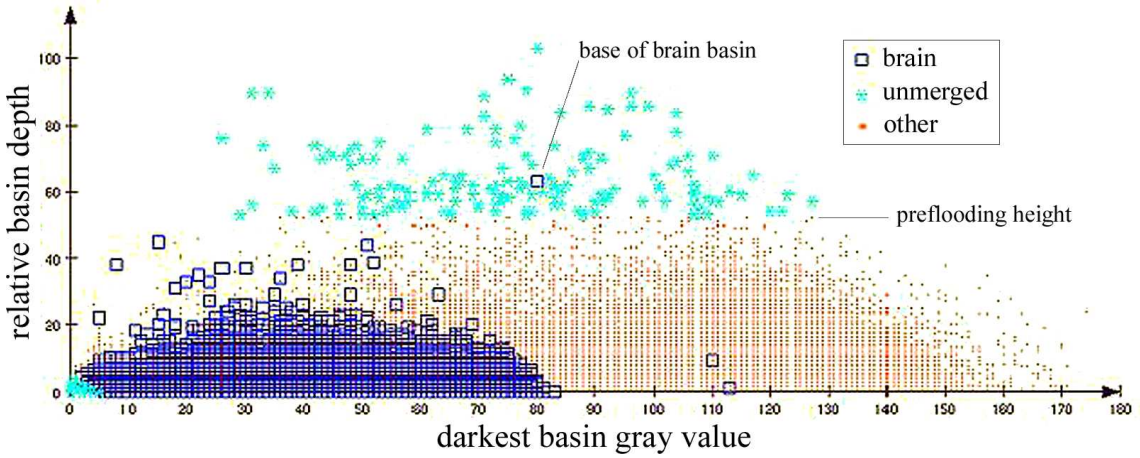


FIGURE 4.10: Parametrization of the watershed tree in a typical case when applying the IWT to a digital image. The example represents a T1 weighted brain MR image according to Figure 2.4. Three classes of atomic basins are shown: brain (merged into a single connected basin), unmerged non-brain (relative basin depth larger than h_{pf}), and other (merged with non-brain basins). The preflooding height is equivalent to a threshold on the relative basin depth, separating merged from unmerged basins. In this example, the largest emerging basin corresponds to the brain (cf. body text).

For a specific brain MR case, Figure 4.10 presents a chart on all basins—atomic and higher order—according to their seed gray value (x-axis) and relative basin depth (y-axis). In this example, the preflooding height is chosen to be a constant value. Since the relative basin depth is equivalent to the gray value difference of seed point and merging event, the preflooding height determines a line parallel to the x-axis that separates merged and unmerged basins.

From Chapter 2 we know that, for specific problems, deterministic rules for properly estimating the merging parameters are possible. For the preflooding height h_{pf}, it is obvious that the number of basins is monotonically decreasing with increasing h_{pf} whereas their sizes are monotonically increasing. It is less obvious, however, that the segmentation result does not change continuously with varying preflooding height, but—depending on the image quality—a more or less broad range of proper values exists for a given segmentation goal (cf. Sec. 2.2). If this is true, then the width of this range is a measure of robustness for the segmentation. This behavior will be used in the following section in order to provide an automation for the skull stripping algorithm.

4.4 Robust and Fully Automatic Skull Stripping

The separation of brain from non-brain tissue, in particular from T1 weighted MR images, often is a crucial step in the neuroimaging pipeline. Successful skull stripping forms the basis for various modes of brain visualization and quantification (cf. Ch. 2). For example, it facilitates preoperative visualization of the brain surface using direct volume

rendering, as well as the presentation of results from functional studies (e. g. fMRI, PET, SPECT) in the context of cortical anatomy. Moreover, skull stripping is a prerequisite for quantitative studies such as atrophy control in the course of neurodegenerative diseases, such as Alzheimer's and Parkinson's Disease, and also Multiple Sclerosis (cf. Ch. 8).

In Chapter 2, the skull stripping problem on T1 weighted MR images was presented besides an overview of existing methods and the motivation for a new method. Problems of existing methods are related mainly to computational and algorithmic complexity, to a lack of robustness in the presence of image nonuniformity and noise, as well as to a lack of anatomical flexibility. As a solution, a watershed transform of the depth image was proposed. This approach was evaluated to be successful on phantom and clinical data, a part thereof containing considerable image noise and nonuniformity.

In principle, the new method relies on the connectivity of white matter (WM) matter such that for a proper preflooding height h_{pf} (cf. p. 29), the different parts of WM are merged at an early stage of processing, whereas dams are between gray matter and extracerebral tissue (such as eyes, muscles, fat, etc.) at later stages. Dams are mostly constructed at low intensity voxels representing CSF or cranial bone.

As mentioned above, the size of resulting basins—in particular the size $V_{\max}$ of the largest basin—increases step-wise with increasing h_{pf}. In this curve, it is possible to identify a plateau that corresponds to a properly segmented brain, including cerebellum, brainstem, and spinal cord; over a certain range of h_{pf} values, the size of the largest basin does not change significantly. This characteristic behavior is well described by Figure 2.4 where both positions 3 and 4 correspond to a proper segmentation.

In Chapter 2, the desired basin had to be selected manually after the transform, whereas h_{pf} had to be chosen before the transform such that all parts of the brain are connected with each other, but disconnected from larger non-brain regions. We now seek to automate both h_{pf} and basin selection without weakening the strengths of our method, most importantly its robustness to noise, nonuniformity, and shape variation. Since in some practical cases, MR scans are only covering a part of the brain, we also tested the automated algorithm for incomplete data sets.

Analysis of $V_{\max}(h_{\mathrm{pf}})$-Curve

The basis for this automation is the characteristic $V_{\max}(h_{\mathrm{pf}})$-curve as described in Section 2.2. For low h_{pf} values, the size of the largest basin increases quickly, before passing into a step-wise increase. In order to assure that the largest basin always corresponds to the brain and not to the dark image background, we do not merge basins with gray values lower than a certain threshold (2 % of the gray value range is a suitable value). Furthermore, we assume that the basin representing the brain will not exceed a certain upper volume limit (e. g. 2.5 litres). With these modifications, the brain is reliably represented by the largest emerging basin.

The $V_{\max}(h_{\mathrm{pf}})$-curves for six different images are provided by Figure 4.11. An empirical analysis of these curves reveals that the first larger plateau corresponds to the complete brain basin. This relates to the fact that relatively small gray value variations exist within WM such that for increasing h_{pf} the brain is soon represented by a single basin. Much

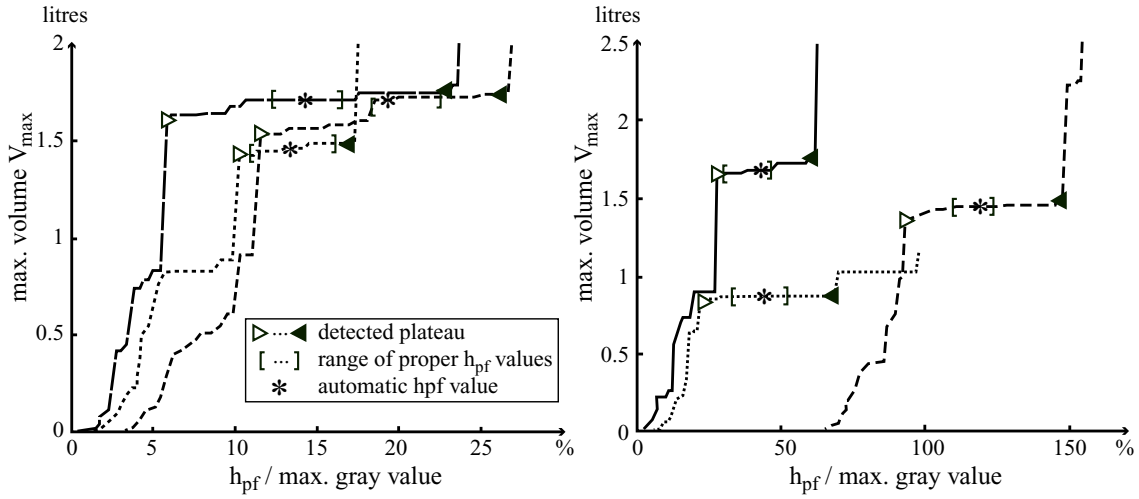


FIGURE 4.11: Characteristics of hierarchical watershed transform: Typical increase of maximum basin volume $V_{\max}$ as a function of h_{pf}. These curves are used for automatic computation of a proper h_{pf} value. *left*: Segmentations of three normal brain scans, each from a different scanner. *right*: Simulation of extreme artifacts, 50 % noise (dashed, cf. Fig. 4.12 left), 100 % ramp nonuniformity (solid, cf. Fig. 4.12 right), and an incomplete brain scan (dotted, without image). For all curves, the automatically detected plateau (triangles), the manually evaluated interval of optimal h_{pf} values (brackets), and the automatically computed h_{pf} value (asterisk) are presented. Note that h_{pf} is measured relative to the original gray value range (before adding of noise or ramp nonuniformity). (Figure according to SCHUBERT *et al.* [2002])

higher h_{pf} values are required to merge the brain with other bright regions such as skin, muscles, or eyes. To locate the first significant plateau, we use an algorithm that is robust to considerable variability of the curve. First, the starting and then the ending point of the plateau are located, whereas small steps within this plateau are tolerated. The tolerance measure relates the step size to to overall graph size. Thereafter, we choose the center of this plateau as automatic h_{pf} value. The analysis can be performed efficiently based on the hierarchical lists of basins and merging events, as produced by the IWT.

Evaluation and Results

We used clinical image data from eight different scanners with different protocols (5× SIEMENS MEDICAL SOLUTIONS and 3× GE HEALTHCARE, including one 0.5 T open MR) as well as phantom data from the SBD [COCOSCO *et al.*, 1997, COLLINS *et al.*, 1998] in order to evaluate the automatic skull stripping method. The phantom data are the same as used in Chapter 2. In each case, we manually determined the range of h_{pf} values producing an optimal segmentation result (cf. brackets in Fig. 4.11).

In non-disturbed high resolution phantom images, as well as in all phantom images up to maximum available noise (9 %), slice thickness (9 mm), and inhomogeneity (40 %), the automatic skull stripping was successful. Also for clinical images, the segmentation was successful in all cases. Successful here refers to the fact that the brain was segmented

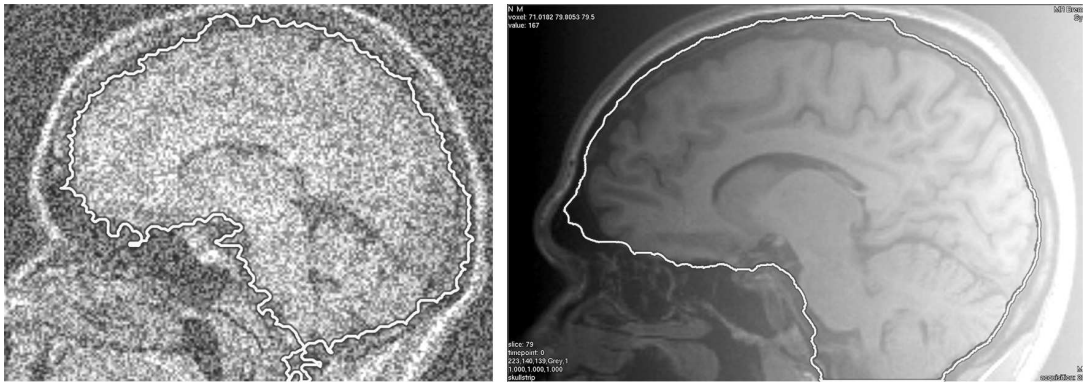


FIGURE 4.12: Segmentation results of T1 weighted MPRAGE images. *left*: Additional uniform noise with an amplitude of 50 %. *right*: Additional linear ramp parallel to x-axis with an amplitude of 100 %. (Percentages relative to the maximum gray value of the original image; MR data courtesy of B. TERWEY, Bremen; Figure from [SCHUBERT *et al.*, 2002])

as one piece and separated from scalp, fat and muscles. In less than 20 % of the cases, smaller structures were included in the segmentation or the segmentation boundary was not smooth (cf. Fig. 4.12 right).

The segmentation of twenty incomplete brain scans (unsystematically chosen cuboid parts of varying size) was successful in sixteen cases, whereas in four cases, neighboring regions (such as eyes or parts of the background) were erroneously included. However, a manual h_{pf} correction lead to successful segmentations in all cases.

Our method turned out to be extremely robust to artificially added noise (cf. Fig. 4.12 left). It produced valid brain masks even for added uniform noise with an amplitude of 30 % of the original gray value range. Even for relative noise amplitudes of 75 %, reasonable segmentations were attained, however with somewhat jagged region boundaries. Note that this artificially added noise surmounts the noise level of the SBD by a factor of greater than eight.

For artificially added linear ramp functions parallel to one of the main coordinate axes, which we used in order to simulate asymmetric image nonuniformity, the automated skull stripping procedure also showed a very high robustness (cf. Fig. 4.12 right). We could choose extreme ramp amplitudes of 200 % (relative to the original gray value range) without larger segmentation defects. Note that this artificially added nonuniformity also surmounts the SBD nonuniformity by a factor greater than five. Since the SBD is often used to evaluate brain segmentation in literature, there are currently no results known to the author that are comparable to the presented noise and nonuniformity levels.

Minor segmentation errors of the automatic skull stripping procedure are due to a limited significance of the examined graph—falsely segmented structures are mainly vessels (e. g. sagittal sinus) and other adjacent structures (e. g. dura). Since on the one hand these structures are mostly darker than white matter, and on the other hand they are close to the outer brain surface, it is promising to introduce a dependency of the merging process on either basin intensity or basin position in order to alleviate these minor problems.

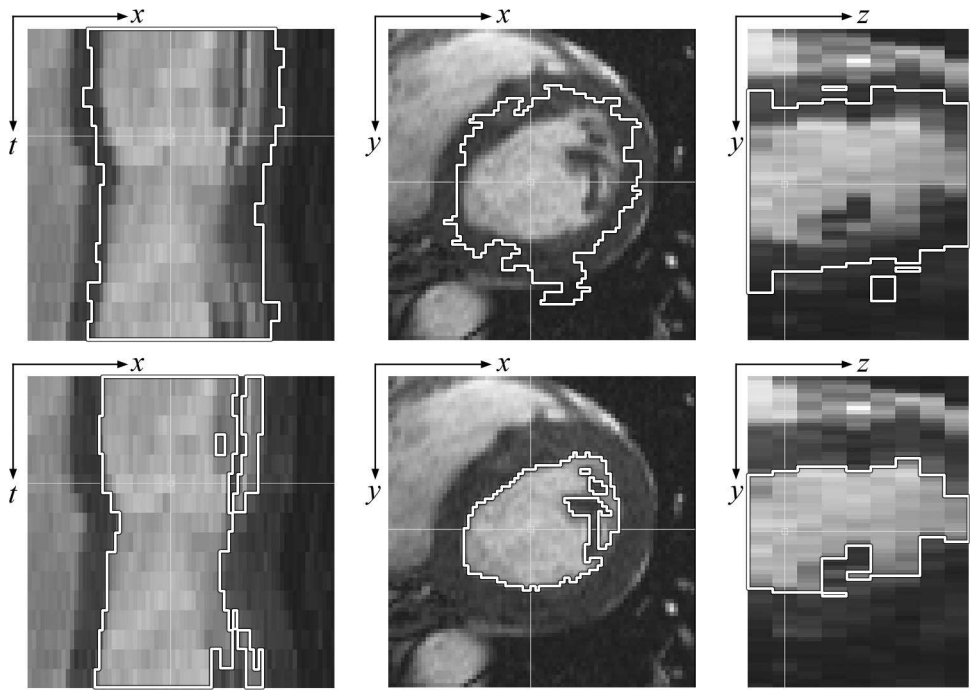


FIGURE 4.13: Application of the IWT to a four-dimensional separation problem. Three orthogonal cuts through a cardiac left ventricle cine MR data set is shown. *top*: In order to quantify the volume of the left ventricle on all phases (time points), it is separated from other bright image elements in a first step. The watershed contour is provided (bright line) after 4D IWT (8-connectivity) on the inverted original image data with a single include marker (crosshairs). *bottom*: Same as above after automatic thresholding within the selected catchment basin. (MR data courtesy of J. DEBATIN, Essen)

4.5 Discussion

Even the most advanced existing implementations of the watershed transform lack reasonable automation and efficient interactive control. The IWT combines both in a coherent and efficient algorithm while completely avoiding oversegmentation, which is the main problem of the WT. Expensive explicit multiscale computing is avoided by a hierarchical organization of all atomic regions. As a positive side effect, exact object boundaries are maintained without smoothing or other image preprocessing. The main advantage of the IWT is its interactivity—all parameters can be changed in real-time to iteratively refine the segmentation. This means that within less than a second, feedback is generated according to the cumulative user input. Compelling applications are represented by neuroanatomical segmentation and volumetry (cf. Ch. 7 and 8), by bone segmentation (cf. Ch. 5), and by vessel segmentation [BOSKAMP *et al.*, 2004], without making assumptions on the objects' shapes.

An example for 4D segmentation using the IWT is provided in Figure 4.13. The cardiac left ventricle is separated from other bright image elements in a short-axis image acquired by a clinical 1.5T MR scanner. To perform the separation on all slices and time

points, it was enough to place a single marker and to properly adjust the preflooding height, while the IWT was directly applied to the inverted original image data. For largely homogeneous images, an automatically chosen threshold can provide a good estimate of the inner ventricular border (cf. Fig. 4.13 bottom). In presence of strong nonuniformity, a secondary gradient watershed or active contour would be candidates to provide robust results.

Although the IWT is defined on gray scale images, it could easily be adapted to multi-valued images as long as a significant scalar segmentation function can be generated from the image data; possible applications are in the processing of multispectral MRI, which is often used in neuroimaging [Friedlinger *et al.*, 1999], or color images. Moreover, for specific applications, the digital grid can be generalized to any connected graph; since merely random access to all image elements and direct access to the neighbors of a given element is required [Vincent and Soille, 1991], a regular grid is no prerequisite of the algorithm. The IWT is likewise applicable to large volumes (e. g., more than 20 Mio. image elements) in 2D, 3D, and higher dimensions on standard hardware, only being limited by the working memory; 32 Bits are required per image element for both the sorted index and the work image.

However, especially for 3D and 4D applications of the IWT, not only the connectivity of the underlying grid (number of direct neighbors etc.) but also the connectivity in terms of the sampled object must be considered. For large slice thickness or large temporal distance between time points, the imaged object might change too much from one slice or one time point to the next, respectively, such that connected object components would not be properly identified. As a consequence, for 3D, the IWT performs best with an isotropic image resolution. For the above 4D example (cf. Fig. 4.13), it can be visually estimated that the object connectivity is comparable in axial and temporal direction, but much less than within the x-y-plane.

In order to circumvent the problems induced by the variety of different WT algorithms, Falcao *et al.* [2004] propose a more general transform, the so-called *image foresting transform (IFT)*, for which a unique and correct algorithm exists that can be computed by a generalization of Dijkstra's algorithm. A specific version of this algorithm fulfills the requirements of a WT.

Within the *morphological paradigm for image segmentation*, a marker function and possibly user interaction is taken into account before performing the WT [Soille, 2003, p. 278]. As stated above, the IWT does not start from markers, but evaluates the user interaction close to real-time after the transform has been performed. As a drawback, however, the atomic basins identified during the hierarchical WT are non-separable by definition. For different markers placed onto a single atomic basin we decided to give highest priority to the most recent marker, ignoring earlier ones.

Most existing multi-resolution WT techniques are faced with the problem of linking and unifying labels of regional minima at neighboring resolution levels [Moga, 1999]. For example, Letteboer *et al.* [2001] present an idea similar to our approach that is, however, far more costly and probably less robust to image variations. Unfortunately the linking between scales is not described in their paper. Our approach is different in that it works

solely on a single segmentation function, which often is the original image data, and hence does not require distinct levels of resolution. The IWT rather makes use of the hierarchy that is intrinsic to the successive flooding and basin merging process.

In many publications, including our own [Hahn and Peitgen, 2000], the SBD [Cocosco *et al.*, 1997, Collins *et al.*, 1998] is used for algorithmic evaluations. Therein, simulated noise ranges from 0 to 9 percent, inhomogeneity from 0 to 40 percent. As stated above, the noise and inhomogeneity used for our evaluation of the proposed automated skull stripping procedure (see for example Fig. 4.12) surmounts these values by a large factor. This gives a strong hint on the robustness of our skull stripping technique.

It is a shortcoming of the classical watershed transform that it is difficult to be combined with a-priori information that relates to the smoothness of the segmentation result. As demonstrated by Nguyen *et al.* [2003], in combination with a Snake algorithm, the distance-based definition of the watershed line might serve as a basis for integrating a-priori considerations about the smoothness of the segmentation contour. For three and higher-dimensional segmentation, however, balloons and deformable hyper-spheres would have to be implemented, which is difficult and time-consuming.

An alternative to generate smooth results is provided by Meyer and Vachier [2002] by proposing a viscous watershed transform. Therein, "the relief is modified so that its non viscous flooding is equivalent to the viscous flooding of the original relief. This choice allows to clearly separate the smooth procedure from the strict watershed computation and thereby to preserve the qualities and speed of the watershed transform." Grau *et al.* [2004] propose that a-priori information on the segmented objects can be directly imposed to the watershed transform by replacing the scalar relief function by a previous probability calculation. Their approach however, requires all markers to be defined before the transform, much as for the classical morphological segmentation paradigm (cf. Fig. 4.3). For the IWT, we propose to use morphological operations during iterative refinement in order to generate smooth segmentation results. Such operations are robust and can be performed at a high computational efficiency.

4.6 Conclusion

The IWT provides the basis for extremely robust, versatile, and easy-to-use segmentation algorithms. Its interactivity permits valid segmentations even on low quality images. Time consuming slice-per-slice interaction is avoided by utilizing 3D or 4D connectivity. Several powerful segmentation tools for medical imaging, predominantly CT and MRI, have already emerged. Some of them will be presented in the following chapters. Often, a combination with other image analysis techniques is favorable. For example, histogram analysis uses the output of the IWT to compute reliable quantitative parameters such as the volume of complex 3D objects. In one specific case, a reliable full automation was achieved by analyzing a characteristic graph describing the merging process. Further improvements of the automatically generated segmentation are possible by evaluating various basin features regarding intensity, shape, and position. In order to extend the IWT, future work will include the marker driven separation of atomic basins that cannot

be separated within the current algorithm, and an efficient distance driven computation of watersheds for non-minimum plateaus.

Publications. *Large parts of this chapter were presented at the SPIE Medical Imaging conference, which was held in February 2003 in San Diego* [Hahn and Peitgen, 2003]. *The automation of the skull stripping algorithm was presented with* Alex Schubert *at the BVM meeting in Leipzig 2002* [Schubert *et al.*, 2002]. *During a four-week student internship,* Alex *carried out the implementation of the graph analysis, as well as the evaluation of the automatic procedure. The Interactive Watershed Transform as well as the automated skull stripping procedure are contained in a US patent application* [Hahn, 2001].

The working body is not understandable
Without knowledge of its structure.

—Galen of Pergamum, 129-c.216

CHAPTER 5

Minimally Interactive Bone Segmentation

Abstract. *The study of normal and injured kinematics in vivo requires bone segmentation and registration. Separation of individual bones from CT volume data is difficult in case of tightly packed and irregularly shaped anatomy. In this chapter, we introduce a new interactive algorithm for fast and robust segmentation of bones in CT images, called* Minimally Interactive Bone Segmentation (MIBS)*. The algorithm consists of two steps. First, a separation of image regions each of which containing a single bone is realized using the Interactive Watershed Transform. Second, solid and accurate segmentations for each bone are generated automatically using a concept that we call Bit Morphology.*

Evaluation was conducted on one skull and ten wrist CT scans. On all data sets, satisfactory results were obtained within a few minutes. For quantitative evaluation, all eight carpal bones from manually segmented scans of an embalmed cadaver wrist were compared through kinematic analysis to the MIBS algorithm. Across all individual bones, the mean rotational and translational errors ($\pm SD$) in the manual segmentation were $0.70^\circ \pm 0.21^\circ$ *and* $0.60\,\text{mm} \pm 0.55\,\text{mm}$*, while for the proposed approach, these errors were* $0.65^\circ \pm 0.47^\circ$ *and* $0.58\,\text{mm} \pm 0.81\,\text{mm}$*, respectively. These first results suggest that the proposed method has the potential to replace manual bone segmentation and thus to dramatically reduce interaction costs.*

As an extension of the MIBS algorithm, we propose a technique for efficient bone removal on multi-detector CT data sets. Based on the IWT, two innovative parts were required, namely compression of bright image contents and automatic basin labeling, to solve this task (i) on large data sets, (ii) with a high degree of automation, (iii) fast, and (iv) with efficient user control.

NEARLY ONE THIRD of the population is significantly affected by musculoskeletal disease during their lifetime. In spite of the prevalence of joint disorders, the *in vivo* three-dimensional (3D) kinematics of our joints in normal and pathological conditions are poorly understood. For example, despite widespread clinical awareness of dynamic wrist instability, little is known about the pathoanatomy and kinematics of these conditions, hindering the design of effective treatment. The presented work aims at generating robust and reliable bone segmentation from CT images, which is a prerequisite for the study of *in vivo* kinematics [Snel *et al.*, 2000].

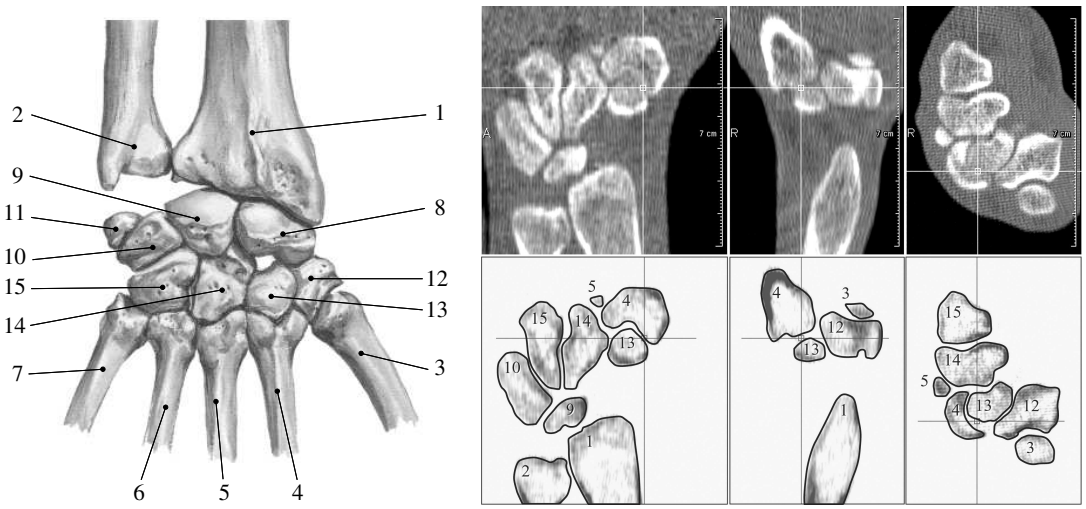


FIGURE 5.1: *left*: Drawing from an anatomical textbook [NETTER, 1989, plate 426 b] depicting the two forearm, eight carpal and five metacarpal bones. The numbers will be referred to in respective figures and text passages (1: radius, 2: ulna, 3–7: metacarpal bones, 8: scaphoid, 9: lunate, 10: triquetral, 11: pisiform, 12: trapezium, 13: trapezoid, 14: capitate, 15: hamate). *top right*: Original orthogonal sections of CT volume data of the wrist. *bottom*: Idealized result performed by manual segmentation superimposed to inverted original data.

Separation of individual bones from CT volume data is difficult in many cases. Existing automated methods are not robust nor as accurate as manual expert segmentation which, on the other hand, is tedious and time consuming, and requires considerable anatomical knowledge. To date, no general approach without shape priors is known to the author to separate and segment individual bones from high-resolution CT data within a few minutes. Actually, the problem of segmenting individual, tightly packed bones is not to find their borders—which is often done by simple thresholding—, but rather to completely segment the *correct* borders for each individual bone. In other words, separation of similar and neighboring structures plays a crucial role and is the most demanding part of the overall problem. Wrist segmentation is one of the most challenging problems in this context [SEBASTIAN *et al.*, 1998, SNEL *et al.*, 2002], while one of the major goals is the simulation of wrist motion in order to better understand its kinematics and related debilitating conditions [SNEL *et al.*, 2000, NEU *et al.*, 2000].

5.1 Related Work and Motivation

Research on bone segmentation is mainly driven by three objectives:

(a) Modeling of skeletal anatomy and kinematics [SEBASTIAN *et al.*, 1998, NEU *et al.*, 2000],

(b) Visualization thereof [WESTIN *et al.*, 1998], such as in the field of surgical planning [EVERETT *et al.*, 2000, EHRHARDT *et al.*, 2003], and

(c) Bone suppression and thus segmentation for a valid analysis of vascular systems on CT Angiographic (CTA) images [Fiebich *et al.*, 1999].

Neu *et al.* [2000] segmented the outer cortical surface of the wrist bones from CT volume images by thresholding using the 3D biomedical imaging software package Analyze^{TM}. Manual intervention was required for the separation of merged bone contours, elimination of unwanted anatomical detail, and closure of contour breaks. Once the bone contours are segmented from the image slices, custom software is used to group the contours by bone. This approach is time consuming, but yields accurate results.

Deformable models have been investigated for bone segmentation. Saxena *et al.* [2002] use a modified Snake algorithm for prosthetic finite element modeling, while Snel *et al.* [2002] use deformable triangular surfaces for 3D segmentation of the wrist. These models serve as geometric descriptors of bone contours or surfaces; they evolve under certain constraints to approximate the actual data. However, for the wrist problem, contour models often converge poorly on bone boundaries. Moreover, 3D models are computationally expensive and do not efficiently incorporate interaction for user correction of erroneous results. As an alternative and with application to hip surgery planning, Ehrhardt *et al.* [2003] have proposed atlas-based landmark recognition.

Another class of segmentation methods are those based on region growing or region merging, respectively. In these methods, initialized seeds grow by collecting statistically similar pixels. However, region growing methods tend to merge bones which are close together with respect to the image resolution. In such cases, time-consuming manual drawing is required.

A hybrid technique called *Skeletally Coupled Deformable Models (SCDM)* was proposed by Sebastian *et al.* [1998]. The technique yielded promising initial results, although its accuracy did not match the accuracy of manual segmentation. Moreover, in our experience, this method requires considerable tuning from one scan to another, rendering it less effective than manual segmentation.

Relevant general difficulties when segmenting individual bone are illustrated by a wrist CT data set in Fig. 5.1:

- ▷ Neighboring bones often are in direct contact with each other or separated by only a narrow articular space.
- ▷ At the interface between the trapezoid (no. 13) and the corresponding metacarpal (no. 4, crosshairs), the bone surface can barely be discerned visually. At a first glance, the two bones might be perceived as a single unit.
- ▷ The bone cortex intensity shows considerable inhomogeneity.
- ▷ Bones exhibit a wide variety of irregular shapes.

5.2 Morphological Single Object Segmentation

Our strategy is as follows. First, by separating the bones from each other, we reduce the *multiple bone problem* to several *single bone problems*. Once having done this, we solve the

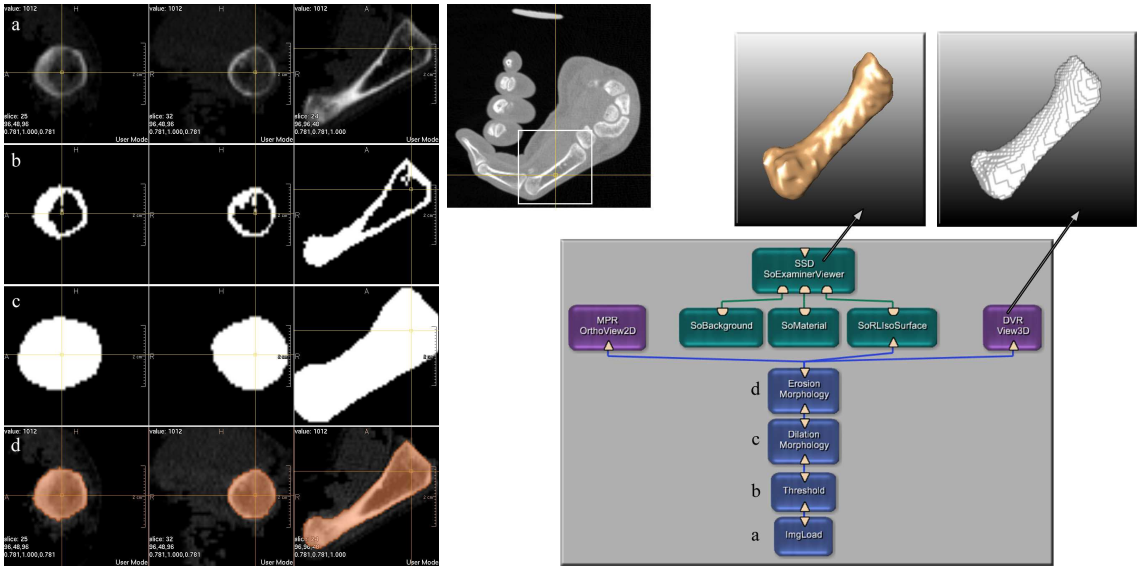


FIGURE 5.2: Segmentation of a single bone from a wrist CT data set (volunteer) in three successive steps. *a*: Original data in orthogonal sections; all other bones (cf. scout image) were excluded using a watershed transform. *b*: Results after thresholding. *c–d*: 3D morphological closing, i. e. dilation (c) and erosion (d). *right*: MeVisLab network for segmentation of a single bone (steps a–d). The result after morphological closing (c+d) is also shown in 3D as Shaded Surface Display (SSD) using marching cubes and via Direct Volume Rendering (DVR).

single bone problems all at once. To segment all single bones simultaneously in a parallel approach based on standard single-processor hardware, we propose a method called *Bit Morphology*.

Material

Images were acquired with a HIGHSPEED ADVANTAGE CT scanner (GE HEALTHCARE) of the right wrist of one volunteer and the left wrist of one cadaveric specimen. The volunteer was scanned with the wrist in neutral position at high resolution ($0.273 \times 0.273 \times 1.00\,\mathrm{mm}^3$) and with the wrist extended at low resolution ($0.781 \times 0.781 \times 1.00\,\mathrm{mm}^3$, cf. Fig. 5.2). The cadaveric specimen was used to investigate the accuracy of our method. After exposure of the individual carpal bones, the specimen was divided in two components and encased in acrylic resin. The components were scanned in four different postures and at two different pixel resolutions ($0.313 \times 0.313 \times 1.0\,\mathrm{mm}^3$ and $0.938 \times 0.938 \times 1.0\,\mathrm{mm}^3$), resulting in eight additional data sets. Furthermore, a Neurosurgery patient (F, 33 y) underwent skull CT on a VOLUME ZOOM (SIEMENS MEDICAL SOLUTIONS) at a resolution of $0.313 \times 0.313 \times 0.6\,\mathrm{mm}^3$ (0.5 mm slice thickness, 0.1 mm inter-slice gap, matrix 512, 71 slices, cf. Color Plate C.3). Before processing, the data set was resampled to a 0.6 mm isotropic grid using Sinc interpolation.

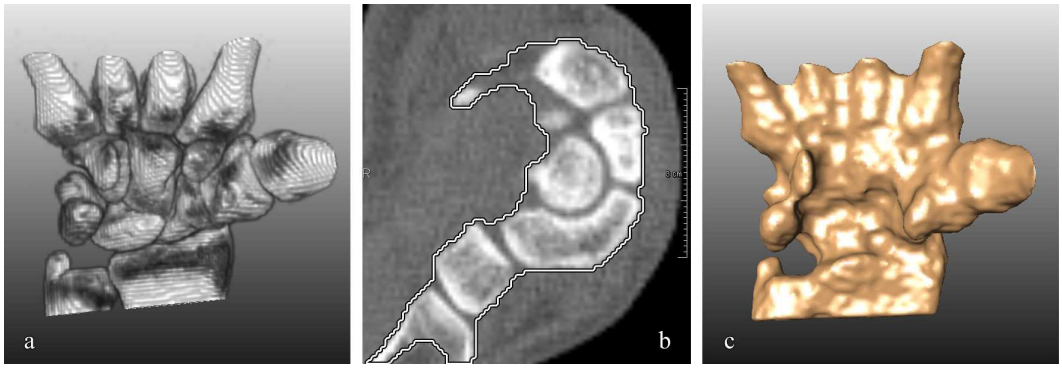


FIGURE 5.3: Separation problem occurring with neighboring bones if the same algorithm is used as in Fig. 5.2. *a*: DVR of original data. *b*: Contour overlay of erroneous segmentation result after thresholding and morphological closing. *c*: SSD of result after closing; the narrow gaps between adjacent carpal and metacarpal bones have disappeared. This problem is solved using the IWT and Bit Morphology (cf. Color Plate C.2)

Segmentation of a Single Bone

First, we would like to regard the simple case that only a single intact bone is included in a 3D CT data set (Fig. 5.2 a). Since bone cortex is most often represented by the highest intensity image elements, problems with overlapping structures do not occur; using semitransparent direct volume rendering with simple window functions for gray level and opacity, the bone surface can be easily visualized in most cases (cf. Fig. 5.3 a). However, if shape and location of a bone shall be analyzed, or if a valid triangulated representation is required, a pixel-based segmentation is a possible solution. After optionally preprocessing the data, e. g. resampling to an isotropic grid, denoising etc., a good approximation of the bone body is straightforwardly obtained using global *intensity thresholding* and 3D *morphological closing* (cf. Fig. 5.2).

5.3 Separation and Bit Morphological Closing

Bone Separation

Let us now consider a data set that comprises several bones coexisting with distances between each other, which are of the same order as the imaging resolution. In this case, applying the above straightforward approach results in an erroneous segmentation. The reason is that not only the bone marrow and other lower intensity image elements at the bone interior but also the articular spaces are filled by the morphological closing operation (cf. Fig. 5.3). Consequently, in order to be able to process each single bone in the same manner as described above, the bones first have to be separated. This separation step represents the main difficulty when trying to perform valid segmentations of individual bones in a complex arrangement of multiple bones, such as the wrist [SEBASTIAN *et al.*, 1998].

Here, the key to the bone separation problem is using a watershed transform [e. g. SOILLE, 2003], much in analogy to separating the brain from neighboring high intensity image regions (cf. Ch. 2). The original image data is interpreted as depth information; high image intensity represent valleys in an emerging four-dimensional landscape, whereas—even if only slightly darker than the bones—articular spaces are treated as crest lines separating the bone basins. Applying the WT to this four-dimensional landscape, hyperintense bone structures are separated in 3D at surfaces of lowest *peak intensity*, with peak intensity referring to the maximum intensity on a surface (cf. Color Plate C.1).

We propose a marker-based WT for efficient bone separation. In terms of connectivity, each image element will be associated with the marker to which it has the strongest connectivity. In our case, the connectivity between two image elements can be defined as the highest image intensity for which a connecting path exists that is nowhere darker than the given intensity (cf. Sec. 2.2).

As an extension to the classical watershed transform, the Interactive Watershed Transform (IWT) takes into account an arbitrary number of markers for the actual image segmentation, after executing a hierarchical watershed transform only once. This marker-based second part of the IWT, by hierarchically administering all occurring atomic basins, can be computed in well below one second (cf. Ch. 4). In an ideal case, locating a *single* marker per bone yields sufficient information for the watershed transform to simultaneously separate all individual bones (e. g., fifteen bones for the wrist including two forearm and five metacarpal bones) in the original CT image without any preprocessing.

Bit Morphology

After bone separation, the second innovative part of our method is the parallel morphological processing of the labeled image segments—i. e. basins—resulting from the watershed transform. Since morphological dilation and erosion most commonly use flat structuring elements, closing of single-object segmentation masks, such as employed in Section 5.2, can be reduced to a binary image space [cf. SOILLE, 2003, p. 116]. While the parallel processing of multiple data bits—namely 8, 16, 32, or 64 according to the data bus width—is seldomly used in digital image processing, though, we are able to independently and simultaneously perform morphological operations on up to 64 channels by interpreting the different bits in analogy to independent binary images. We call the according concept *Bit Morphology*.

More specifically, after global intensity thresholding of the original image and masking of the initially labeled bone image (IWT result, cf. Color Plate C.1 b), we project different labels to different bits of one memory word per voxel. A *bit morphological dilation* then equals a logical OR operation (compared to the MAX operation of gray scale morphological dilation, cf. [SOILLE, 2003, p. 69]); a *bit morphological erosion* equals a logical AND operation (opposed to the MIN operation of gray scale morphological erosion, cf. [SOILLE, 2003, p. 68]).

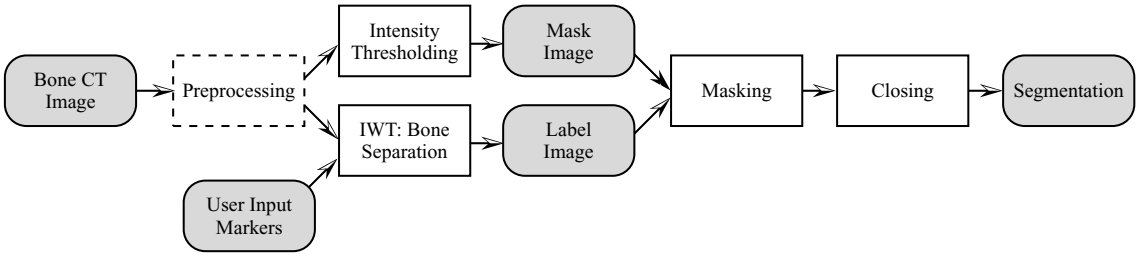


FIGURE 5.4: Diagram of main components and dataflow of MIBS algorithm. Note that in general, the methods performs well without any preprocessing. However, resampling on an isotropic grid, denoising etc. can improve the robustness and accuracy of the method. The thresholding can be either global or locally adaptive. For all examples in this chapter, global thresholding was used.

5.4 Results and Discussion

The proposed method provides an interactive approach to bone segmentation and is schematically presented by Figure 5.4. The results were inspected both as 2D color overlay and in 3D and were satisfactory in all cases, requiring a maximum of 27 markers ($min = 15$, $mean = 21.3$); two examples are shown in Color Plates C.2 and C.3. Mean processing time was 17:20 minutes ($min = 7{:}10$, $max = 31{:}10$) including interaction, bit morphological closing, storage of results, and 2D and 3D visualization.

For a quantitative analysis of segmentation accuracy, we conducted a kinematic analysis on all carpal bones in two of the cadaver CT scans. Cortical bone contours were extracted and associated manually in a complete cadaver data set using the software packages ANALYZE and MATLAB. Kinematic analysis was performed using a marker-less registration technique described by NEU *et al.* [2000] for the segmentation results of both the manual, as well as the proposed technique. Mean rotational and translational errors ($\pm SD$) in the manual segmentation were $0.70° \pm 0.21°$ and $0.60\,\text{mm} \pm 0.55\,\text{mm}$, while for the proposed approach, these errors were $0.65° \pm 0.47°$ and $0.58\,\text{mm} \pm 0.81\,\text{mm}$, respectively.

For the skull CT, separation and segmentation of individual bones worked very well. We would like to point out the quality of bone separation at the squamosal and lambdoidal sutures (cf. Color Plate C.3). These bone junctions are less than one voxel wide in many regions, and four markers per bone were sufficient for the segmentation.

As discussed in previous chapters, the classical WT is known to have a tendency toward oversegmentation [HARIS *et al.*, 1998]. In contrast, the IWT that fully operates in 3D overcomes this problem by hierarchically organizing all watershed basins before providing fast marker steered interaction. It is clear that our approach is much faster than manual contouring, but also less computationally expensive than active contour-based methods while retaining a high flexibility. On average, only one to two markers per bone are required for a valid segmentation. It is unlikely that this will be reached by two-dimensional approaches such as [SEBASTIAN *et al.*, 1998] where seeds have to placed on any bone and slice. *Bit Morphology* plays a central role in the MIBS algorithm since it facilitates the independent morphological processing of multiple neighboring objects without costly processing of single bones, one at a time.

The morphological closing can also be replaced by other closing techniques without altering the MIBS algorithm. One possible alternative is the *Rolling Ball* algorithm, which is, however, unlikely to provide faster or more accurate results than Bit Morphology. Another alternative are deformable balloon-like templates, which—after bone separation using IWT—have a good chance to converge to the correct surfaces.

Instead of a global thresholding, one could apply a locally adaptive thresholding in order to correct partial volume effects. This is likely to improve both robustness and accuracy of our method: An increased robustness is expected by a reduction of holes in the bone surfaces before morphological closing. An increased accuracy is expected since a locally adaptive equalization of partial volume effects has the potential to correctly segment both brighter and darker object borders. However, we will leave this extension of our algorithm to future research, here concentrating on the bone separation problem.

5.5 CT Bone Removal

One specific application of bone segmentation is bone removal in CTA images. We consider this problem as one of the most challenging segmentation problems existing in medical image analysis due to four reasons: (i) Large variability of vascular image patterns and contrasts, (ii) existence of high intensity structures within vessels: contrast agent, calcifications, stents, clips, etc., (iii) large variability of bone shapes and contrasts, and (iv) close vicinity or direct contact of bone and vascular structures, for example skull base, clavicle, etc. Moreover, in order to support the usability and clinical applicability of a product or software application, the most important requirements are to:

- ▷ Be applicable to very large data sets at full resolution (min. $512 \times 512 \times 1600$ voxels)[1].
- ▷ Segment and effectively remove osseous structures with a high degree of automation.
- ▷ Be fast with less than one minute computation time for as much as 1600 slices.
- ▷ Permit efficient user control and intuitive correction with full undo/redo functionality.

Existing approaches to bone removal cover automated [Fiebich *et al.*, 1999, Alyassin and Avinash, 2001, Mullick *et al.*, 2002, Suryanarayanan *et al.*, 2003] and interactive techniques [Alyassin and Avinash, 2001, Moore *et al.*, 2001, Raman *et al.*, 2002, Kang *et al.*, 2003]. Fully automated approaches vary both with respect to their speed and quality; often, only fair results are obtained [Mullick *et al.*, 2002, Suryanarayanan *et al.*, 2003]. A simple interactive technique is described by Alyassin and Avinash [2001] where the user can adjust global thresholds in order to separate vessels from osseous structures.[2] Their approach is based on fast morphological and filtering operations that are typical for bone segmentation [Fiebich *et al.*, 1999]. Interactive techniques, such as described by Moore *et al.* [2001], Raman *et al.* [2002], and Kang *et al.* [2003], often provide a higher

[1] This size has to be related to the available working memory, which is typically 500 MB for this application on current 32 bit systems, since most memory is occupied by volume rendering engines etc.

[2] Alyassin and Avinash are with the GE Research and Development Center. GE Healthcare offers a technique called AutoBone within the Advantage Windows workstation.

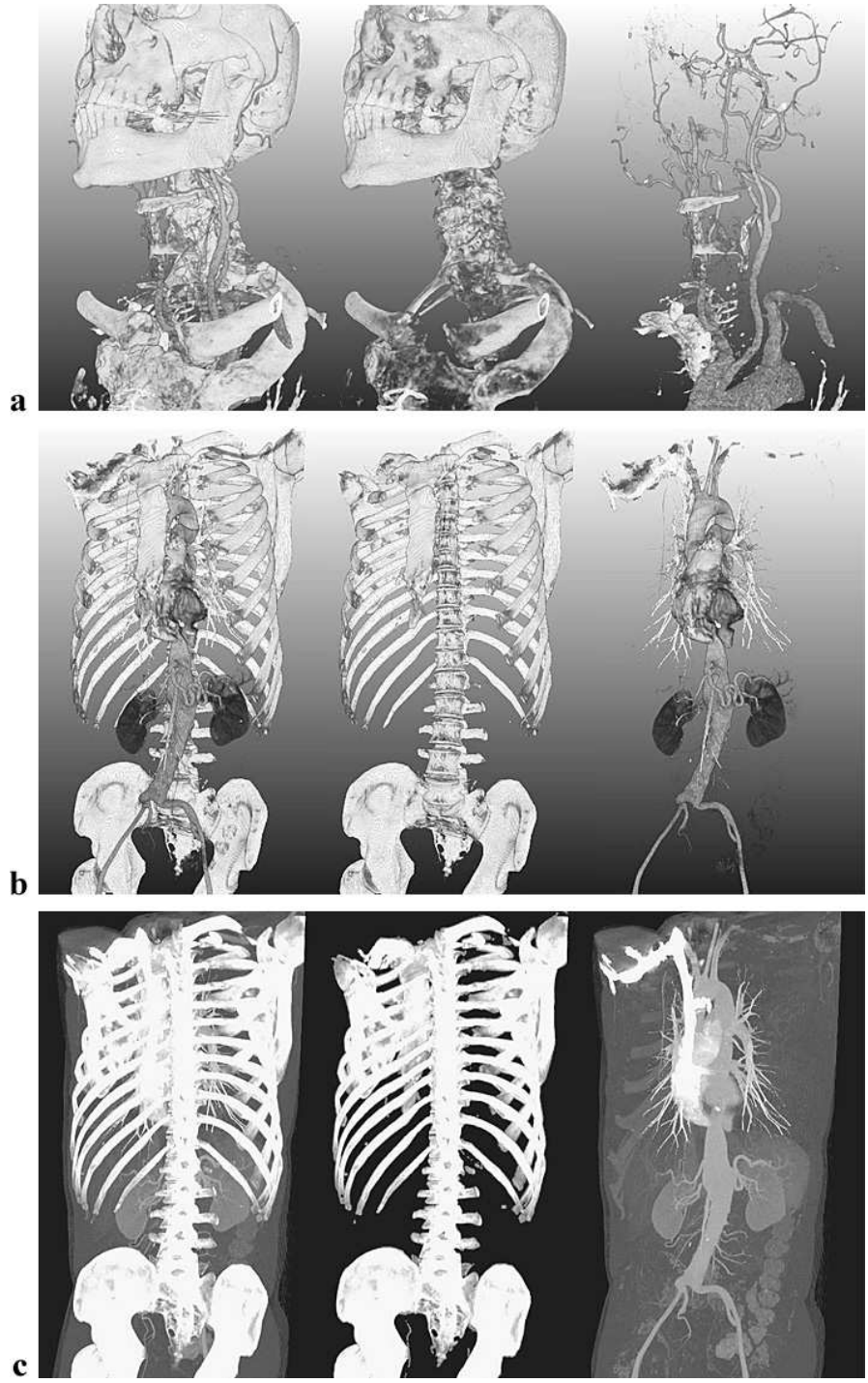


FIGURE 5.5: Two examples for CT bone removal performed by the proposed method. *left – center – right*: original data – bone segmentation – bone removal result. *a*: On head CTA images, bone removal is particularly difficult due to the close bone-vessel interface at the skull base. *b*: Thoracic and abdominal scan with ≈ 1000 slices. Problems are induced by bone-vessel contact and high concentrations of contrast agent. *c*: Same as *b*, but MIP istead of DVR rendering technique. (CT images courtesy of CLARA SOULIÉ, SIEMENS MEDICAL SOLUTIONS, Forchheim)

specificity at a lower speed. Moreover, specialized techniques provide a good segmentation quality for a restricted range of applications, such as intracranial [Moore *et al.*, 2001], knee, and femur segmentation [Kang *et al.*, 2003]. A compromise of generality and accuracy is provided by Raman *et al.* [2002] who require a single mouse-click on each bone and vessel to be separated.

Encouraged by the bone segmentation results described in Section 5.4, we present a novel bone removal technique that is based on the IWT. In order to meet the above listed requirements, two innovative parts are required, namely (i) compression of bright image contents[3] and (ii) automatic basin labeling (cf. Color Plate C.4). Fast interactive region merging is based on the hierarchical multi-scale image transform described in Chapter 4 and can be directly applied to the compressed original image data. The marker list of the IWT offers an efficient interface for full undo/redo functionality; it suffices to revert to a previous marker list in order to exactly reproduce the respective segmentation state. The proposed technique differs from existing approaches mainly in the following three points:

- ▷ The major part of the required computation is performed before any user interaction.
- ▷ Images of up to 512×512×1600 voxels at 12 Bit, corresponding to 800 MB of input data, can be processed at full resolution with approximately 400 MB of working memory.
- ▷ A multi-scale processing scheme facilitates efficient interaction with iterative segmentation refinement and automatic segmentation based on multiple local and regional image features.

More specifically, the following steps are performed successively:

(a) Histogram computation of entire input image

(b) Memory allocation in multiple blocks, each max. 50 MB in size

(c) Slice-by-slice image compression of bright image elements[3], typically above 175 HU

(d) Hierarchical multi-scale image transform, which is the first step of the IWT as described in Chapter 4 [Hahn and Peitgen, 2003]

(e) Storage of hierarchical region information for automatic and interactive segmentation

(f) Disposal of approximately 70 % of allocated memory

(g) Automated multi-scale region labeling

(h) Efficient interactive refinement of segmentation result with full undo/redo

(i) Bone mask computation on request for single slices or full volume at arbitrary resolution

(j) On-the-fly morphological dilation of segmentation result to cover partial volume region at bone surface.

[3] cf. disclaimer on p. 82.

The computational performance of the proposed bone removal technique is high with data sets of up to $512 \times 512 \times 1600$ voxels typically processed in approximately 30 seconds on a single processor Pentium IV 3 GHz; the given time is excluding histogram computation, which can be done at image loading, and interaction. All algorithmic parts are of linear complexity $O(N)$, with N being either the total number of input image elements or the number of bright image elements after compression, respectively. Additionally, more than 80 % of the total computation time is used before any user interaction, i. e. steps a–g, while less than 1 % of the total computation time is used for interactive refinement including mask computation on single slices.

Similar to the work by ALYASSIN and AVINASH, high intensity image elements are initially separated through a global threshold (step c). This threshold, however, is much lower than the one cited by ALYASSIN and AVINASH [2001] and thus provides the potential for a higher sensitivity of our method. Beyond this, specificity is warranted at later stages (steps d–h) where several regional image features are available: mean gray value (g_{mean}), maximum gray value (g_{M}), gray value standard deviation (σ), gray value coefficient of variation ($CV = \sigma / g_{\text{mean}}$), and region volume ($V$, in ml). These are used in an automatic region labeling step (cf. Color Plate C.4), and also later for efficient interactive region splitting and merging.

A specific operating mode is available for the processing of cranial CTA images posing the specific problems of large bone-vessel interfaces and closed high intensity bone structures. This operating mode is similar to the work by MOORE *et al.* [2001], but offers full efficient interactive control based on the above described bone removal technique.

Similar to the results of the MIBS algorithm described in Section 5.3, the proposed bone removal technique also offers the possibility to separate and extract individual bones or bone fragments, in most cases with a single mouse-click. This is beneficial when not all bone structure shall be removed in a volume rendering application, such as for fracture display and analysis.

5.6 Conclusion

With respect to wrist segmentation, our first quantitative results suggest that the proposed *Minimally Interactive Bone Segmentation (MIBS)* method has the potential to replace manual segmentation. From the user perspective, this method may dramatically reduce the interaction time in comparison to existing techniques while maintaining the segmentation accuracy. Still, ongoing work will include further careful validation both on cadaver and *in vivo* image data. To medical doctors, accuracy is often more important than speed, such that they will continue to do manual segmentation until they have a *proof of accuracy* from a semiautomated algorithm.

Furthermore, we will investigate, whether balloon-like templates have the potential to replace Bit Morphology at equal or superior results and performance. Finally, since no shape priors are incorporated, our method will directly be applicable to a wide range of bone structures, most relevant to other joints such as hip or ankle. Further fields of

application include preoperative visualization for surgical planning, the segmentation of all bone fragments in complicated compound fractures.

As for all CTA bone removal algorithms, current limitations exist for the proposed bone removal technique with respect to speed, memory requirements, sensitivity, specificity, and degree of automation. Due to the highly diverse appearance of both osseous and vascular structures, including calcifications, stents, clips, etc., the perfect automated technique does not exist. By implementing a consistent multi-level approach, however, the sensitivity and specificity of our method is superior to many existing approaches. Furthermore, based on a hierarchical organization of all bright image regions, efficient interaction is facilitated with a higher speed than all existing approaches known to the author. Finally, all tested images up to a size of 1750 slices could be processed at full resolution within less than a minute on standard single-processor hardware. The proposed bone removal technique has been successfully evaluated on images from a variety of different CT scanners including current 64 row detector systems. Even though the current results are satisfactory in most cases, future work will concentrate on further improving object separation at the skull base and other large bone-vessel interfaces and specificity for automatic classification of stents and large calcifications. Essential ingredients to these improvements will be an extended set of basin features and a generic *support vector* classification [Cristianini and Shaw-Taylor, 2000], which we already have successfully evaluated for this purpose.

Publications. *This chapter contains previously unpublished work. A manuscript on the MIBS wrist application is in preparation together with* Mohammad A. Upal *from Brown Medical School / Rhode Island Hospital, as well as* Georgeta E. Marai *and* David H. Laidlaw *from Computer Science Dept., Brown University.* Mohammad *conducted the kinematic analysis and evaluation,* Georgeta *helped in compiling related work, and* David *contributed to the motivation of the project. Furthermore,* David J. Crisco *provided the medical background. All four contributed with many thoughts and discussions. The application of the WT to bone segmentation as described in this chapter is comprised in a US patent filed on October 2001* [Hahn, 2001]. *First implemented in MeVisLab* [MeVis, 2004, Schenk *et al.*, 1999, Hahn *et al.*, 2003]*, the proposed bone removal technique has recently been prepared for inclusion in the* Syngo InSpace 3D *Workstation (*Siemens Medical Solutions*, cf. Color Plate C.5).* Hans Drexl *has implemented a support vector machine based classifier to improve the bone removal performance. A patent application for the bone removal algorithm is in preparation, such that details on the specific compression algorithm cannot yet be disclosed.*

PART II

ON VOLUMETRY AND HISTOGRAM ANALYSIS

IN DECEMBER 2002, A. G. SORENSEN, Director of the MGH Martinos NMR Center, Charlestown, introduced the term *NeuroVolumetrics* in order to emphasize the importance of volumetric image analysis for diagnosis and therapy of neurological diseases.[4] Likewise, the second part of this thesis stems from the question whether the Watershed Based Skull Stripping (WASS) algorithm would also be suitable for the segmentation and *volumetry* of the ventricles and other cerebral structures of interest from the same type of images as described in Chapter 2. This shall be explained below.

At first sight, the WASS approach does not seem to be applicable to the cerebral ventricles: It uses the WT on the inverted original data to locate dark areas where the brain can be separated from neighboring bright non-brain objects. Conversely, for the ventricles, the border between dark fluid filled spaces and surrounding bright brain tissue, mainly consisting of white matter, has to be found. Since the ventricular border is well contrasted in most parts, a gradient-based approach would be more appropriate, such as proposed by SCHINDEWOLF *et al.* [1999]. However, in regions such as the temporal horns, where the fluid spaces are extremely narrow, a gradient-based approach is likely to fail due to limited resolution, noise, and partial volume averaging.

On Partial Volume Effects

Partial volume effects (PVE) occur due to limited spatial resolution, when information of at least two structures is mixed within a single image element. This is particularly pronounced for narrow structures with a small volume-to-surface ratio. When the smallest object diameter is of the same order as the voxel size, the main source of volumetric error is linked to PVE. The importance of PVE in the course of volumetric image quantification has been stressed in several publications [e.g. VAN LEEMPUT *et al.*, 2003]. BALLESTER *et al.* [2000] were the first to give a direct, even if rough, estimate of confidence intervals from partial volume effects for reconstructed object surfaces, as well as volumetric measures. POKRIC *et al.* [2001] use explicit partial volume distributions within their EM

[4] At the 88th Scientific Assembly and Annual Meeting of the Radiological Society of North America (RSNA) held in Chicago, USA, between December 01 and 06, 2002.

algorithm such that intensity distributions in multimodal MR images are better explained with their help than with lack thereof. With respect to our specific problem, Wang and Doddrell [2001] state that lateral ventricular volumes would be heavily underestimated when omitting PVE.

It is crucial to note that even if all borders of an object are located, e. g. described by a binary segmentation mask, PVE still leave some open questions. It is a non-trivial task to accurately compute object volumes in the presence of PVE. Of course—assuming linearity of image acquisition—, the gray level at segmented borders can be used to estimate the partitioning of different tissues within respective image elements. For zero noise and nonuniformity, this estimation is perfect as long as only two tissues with known pure gray values are mixed in a given element.

In order to carry out such an estimation, one would ideally expand the segmentation mask in order to cover all PVE elements, however without including neighboring regions that exhibit an image intensity similar to the segmented object. With respect to their gray values, we propose to call such regions *ambiguous image contents*. Moreover, in order to attain information on the neighboring tissues' pure gray values, one would further expand the segmentation mask beyond the PVE region wherever possible.

Note that for the cerebral ventricles such approach, extending the segmentation mask to the surrounding gray and white matter but not to extracerebral CSF, can be realized by—or even is equivalent to—applying the WT on the original image data, much as conducted for solving the skull stripping problem. This idea lays the foundations for a new group of volumetric image analysis tools, as will be discussed in this second part.

On Histogram Analysis

Recalling Figure 1.8 (p. 15), we take a closer look at histogram-based quantitative image analysis and at important differences between gradient and histogram-based methods. Here, gradient-based methods are to be regarded as a representative of direct image-based methods. In order to yield significant information, the gradient function needs to be adapted to individual image characteristics, most importantly the image resolution. Absolute image intensities are ignored by the gradient function. For example, in Figure 1.8 b–d, several noise induced gradient maxima are perceptible. By only taking the gradient image into account, it remains unclear how many objects are included in the image, unless the proper gradient scale σ_i has been chosen. Furthermore, the gradient image is directly linked to the shape, topology, and dimensionality of respective objects. For a properly chosen σ, detecting object borders by maximum gradient magnitude is quite robust. However, as described above, the uncertainty induced by PVE remains.

In contrast, the image histogram contains full information on the frequency of absolute intensities, while completely ignoring the spatial context. Of course, as signal-to-noise ratio (SNR) is linked to the sampling rate, the histogram is not independent of spatial resolution. Since volumetry is directly related to the frequency of specific image contents, histogram analysis seems to be most favorable for this purpose. Based on its statistical nature, a major feature of histogram analysis is that error estimates may easily be incorporated. Theory, practical limitations, and results are described in the following chapters.

In order to deliver valid quantitative parameters, however, histogram analysis requires an appropriate model for the intensity distributions of all image contents. Regarding the histogram in Figure 1.8 e, three tissue types are clearly discernible with approximately mean gray values of 20, 50, and 80. Furthermore, the plateaus between these three peaks correspond to intermediate gray values, which are mainly effected by partial volume averaging. In Chapter 7, a simple but effective model for the intensity distribution of PVE will be introduced. A more accurate model for the PVE distribution is discussed in Section 6.2.

Other than the number of tissue types, the number of objects is not revealed by the histogram in Figure 1.8 e, as two arrows exhibit identical image characteristics. As discussed above, for an individual quantification, it is therefore necessary to separate these ambiguous objects prior to histogram analysis. The following chapters are guided by the idea that efficient interactive 3D separation of ambiguous image contents combined with automated histogram analysis can provide objective and reproducible volumetric measurements.

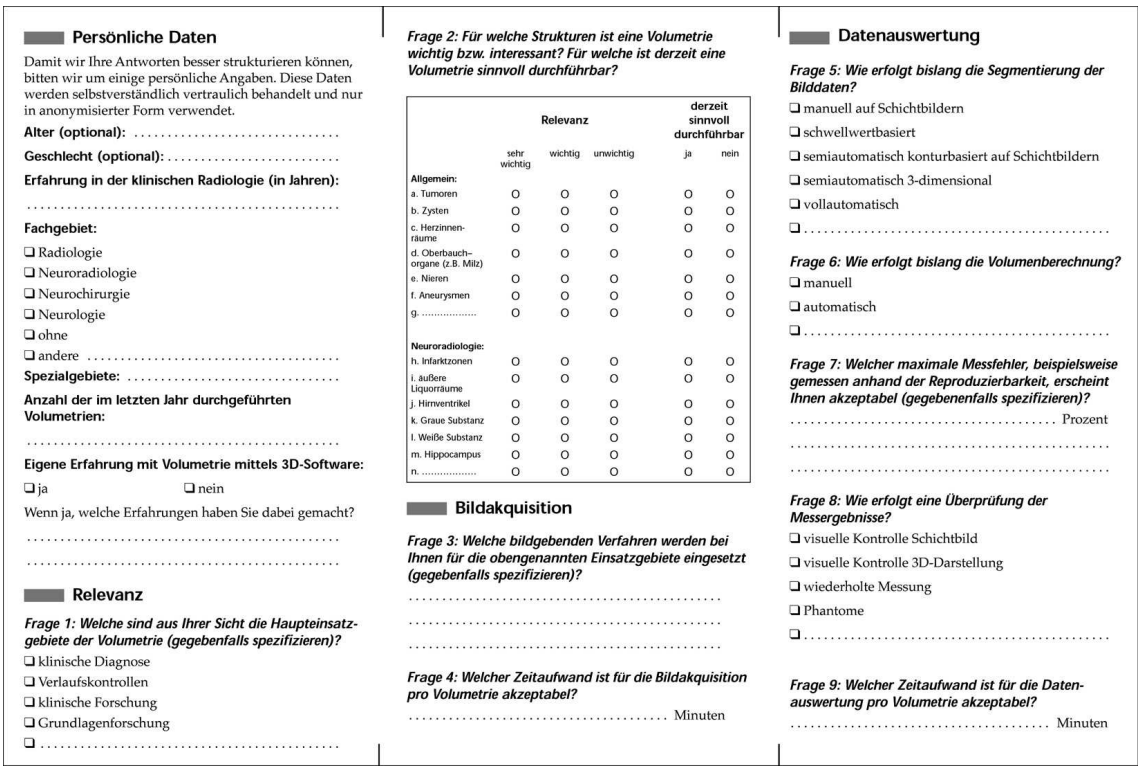
Note that, by such approach, the user specifies the objects to be included in the analysis, but not their exact borders. The latter are implicitly evaluated by a successive, automated step. For example, the user should not be able to alter the border detection—and thus the volumetric result—by means of an interactively chosen threshold value or similar tool.

Furthermore, by increasing the image resolution, most significantly the number of slices, it would be possible to simultaneously decrease the required user interaction and increase the accuracy and precision of derived parameters. This is in contrast to conventional 2D processing where the interaction costs are more or less proportional to the number of slices. Still, the precision of image-based quantitative measurements is limited by image acquisition; both SNR and image resolution, and thus PVE, are linked to the acquisition time; for practical purposes, a tradeoff has to be chosen. Therefore, a developed quantitative method must be robust to varying image resolution.

Nationwide Survey on Computer Assisted Volumetry

In cooperation with B. Terwey, Center for Magnetic Resonance Imaging, Bremen, the requirements of computer assisted clinical volume measurements were studied by polling 200 radiologists. The addresses were indiscriminately taken from a nationwide list of German radiologists and neuroradiologists with MRI facilities including university and community hospitals, as well as private practitioners. 200 questionnaires were sent in June 2000, from which 42 were returned (return rate 21 %), with varying degree of completeness. The questionnaire was designed by the author and consists of thirteen questions relating to the current practice of radiologic volumetry, to desired improvements, as well as to the relevance of volumetric measurements. The questions had to be answered by either multiple choice or free formulation. The inner page of the leaflet is shown in Figure II.1.

A preamble served to acquire some personal information on the respective radiologists. We asked for their age and sex (both optional), their experience in clinical radiology (in years), their radiologic specialization, the estimated number of conducted volumetric measurements during the last year, and their own experience with volumetry using a



Persönliche Daten

Damit wir Ihre Antworten besser strukturieren können, bitten wir um einige persönliche Angaben. Diese Daten werden selbstverständlich vertraulich behandelt und nur in anonymisierter Form verwendet.

Alter (optional): .....

Geschlecht (optional): .....

Erfahrung in der klinischen Radiologie (in Jahren):

.....

Fachgebiet:

❑ Radiologie
❑ Neuroradiologie
❑ Neurochirurgie
❑ Neurologie
❑ ohne
❑ andere .....

Spezialgebiete: .....

Anzahl der im letzten Jahr durchgeführten Volumetrien:

.....

Eigene Erfahrung mit Volumetrie mittels 3D-Software:

❑ ja ❑ nein

Wenn ja, welche Erfahrungen haben Sie dabei gemacht?

.....
.....

Relevanz

Frage 1: Welche sind aus Ihrer Sicht die Haupteinsatzgebiete der Volumetrie (gegebenfalls spezifizieren)?

❑ klinische Diagnose
❑ Verlaufskontrollen
❑ klinische Forschung
❑ Grundlagenforschung
❑ .....

Frage 2: Für welche Strukturen ist eine Volumetrie wichtig bzw. interessant? Für welche ist derzeit eine Volumetrie sinnvoll durchführbar?

	Relevanz			derzeit sinnvoll durchführbar	
	sehr wichtig	wichtig	unwichtig	ja	nein
Allgemein:					
a. Tumoren	O	O	O	O	O
b. Zysten	O	O	O	O	O
c. Herzinnenräume	O	O	O	O	O
d. Oberbauchorgane (z.B. Milz)	O	O	O	O	O
e. Nieren	O	O	O	O	O
f. Aneurysmen	O	O	O	O	O
g.	O	O	O	O	O
Neuroradiologie:					
h. Infarktzonen	O	O	O	O	O
i. äußere Liquorräume	O	O	O	O	O
j. Hirnventrikel	O	O	O	O	O
k. Graue Substanz	O	O	O	O	O
l. Weiße Substanz	O	O	O	O	O
m. Hippocampus	O	O	O	O	O
n.	O	O	O	O	O

Bildakquisition

Frage 3: Welche bildgebenden Verfahren werden bei Ihnen für die obengenannten Einsatzgebiete eingesetzt (gegebenfalls spezifizieren)?

.....
.....
.....

Frage 4: Welcher Zeitaufwand ist für die Bildakquisition pro Volumetrie akzeptabel?

..... Minuten

Datenauswertung

Frage 5: Wie erfolgt bislang die Segmentierung der Bilddaten?

❑ manuell auf Schichtbildern
❑ schwellwertbasiert
❑ semiautomatisch konturbasiert auf Schichtbildern
❑ semiautomatisch 3-dimensional
❑ vollautomatisch
❑ .....

Frage 6: Wie erfolgt bislang die Volumenberechnung?

❑ manuell
❑ automatisch
❑ .....

Frage 7: Welcher maximale Messfehler, beispielsweise gemessen anhand der Reproduzierbarkeit, erscheint Ihnen akzeptabel (gegebenenfalls spezifizieren)?

..... Prozent
.....
.....

Frage 8: Wie erfolgt eine Überprüfung der Messergebnisse?

❑ visuelle Kontrolle Schichtbild
❑ visuelle Kontrolle 3D-Darstellung
❑ wiederholte Messung
❑ Phantome
❑ .....

Frage 9: Welcher Zeitaufwand ist für die Datenauswertung pro Volumetrie akzeptabel?

..... Minuten

FIGURE II.1: German questionnaire on the requirements of computer assisted clinical volumetry, consisting of thirteen questions to be answered by either multiple choice or free formulation. (Note that only one out of two pages is shown. Layout of questionnaire and pdf document courtesy of THOMAS KLINK-TELIEPS, Bremen)

dedicated 3D software. Three women and 39 men participated in the survey. The mean age of all participating radiologists was 46.4 years ($min = 31$, $max = 63$). On average, they had 17.5 years of experience in clinical radiology ($min = 3$, $max = 33$). Among the 42 radiologists, fourteen specialized in neuroradiology, one in neurosurgery, and one in neurology. Seven already had experience with some sort of 3D volumetry software.

We present the answers for eight of thirteen questions (No. 1, 2, 4–9; cf. Fig. II.1), with results given in parentheses:

(1) "Which are the main fields of application for volumetry?"—(most important first): therapy monitoring (86 %), clinical research (48 %), clinical diagnosis (36 %), basic research (14 %), and other (12 %).

(2) "For which anatomical structures do you rate volumetry as important, for which is a volumetry currently available?"—Twelve structures were polled, six from each of the fields of general radiology and of neuroradiology. For the first part of the question, three answers were possible: very important (factor for relevance score $S_R : \times 2$), important ($\times 1$), or unimportant ($\times 0$). For the second part, two answers were possible: available (factor for availability score $S_A : \times 1$) or not available ($\times 0$).

TABLE II.1: Results according to Question no. 2 (cf. Fig. II.1): mean relevance score S_R, availability score S_A, and innovation score $S_I = S_R - S_A$ regarding the volumetry of various structures, as well as return rate n (details cf. body text). Six structures have been included in the questionnaire for each of two categories: General (G) and Neuroradiology (N). Entries are sorted according to innovation score.

Structure		S_I	n	S_R	S_A
Cerebral Ventricles	N	.82	34	1.15	.33
Cardiac Ventricles	G	.80	26	1.36	.56
Hippocampus	N	.78	33	1.10	.32
Tumors	G	.71	37	1.31	.60
Extracerebral CSF	N	.56	30	.79	.22
Aneurysms	G	.50	31	1.03	.53
Ischemic Areas	N	.48	30	.79	.31
White matter	N	.46	28	.52	.06
Gray matter	N	.42	28	.54	.12
Upper Abdominal (e. g. spleen)	G	.02	30	.59	.56
Kidneys	G	−.03	29	.36	.38
Cysts	G	−.04	32	.65	.69
Other	G/N		7		

Not all structures received answers in all questionnaires (return rate n). The answers are summarized in Table II.1 (mean values).

(4) "Which acquisition time per volumetry do you rate as acceptable?"—On average, the radiologists rated an acquisition time of eight minutes as acceptable ($min = 0$ min, $max = 20$ min).

(5) "How is image segmentation currently performed?"—Manually (83 %), thresholding (25 %), semiautomated contouring (21 %), semiautomated 3D (4 %), and fully automated (0 %).

(6) "How do you currently calculate object volumes?"—Manually (84 %), automatically (20 %), and other (4 %).

(7) "Which maximum uncertainty given in percent, e. g. in terms of reproducibility, do you rate as acceptable?"—On average, the radiologists rated an uncertainty of 6.4 % as acceptable ($min = 3$ %, $max = 12$ %).

(8) "How do you currently evaluate measurement results?"—Visual inspection on slices (55 %), repeated measurement (41 %), visual 3D (14 %), phantom measurements (9 %), and other (5 %).

(9) "Which analysis time per volumetry do you rate as acceptable?"—On average, the radiologists rated an analysis time of ten minutes as acceptable ($min = 1$ min, $max = 60$ min).

From question no. 2, two scores were directly computed as mean value of the factors (cf. above) given by the radiologists. First, the relevance score S_R ranges between zero (unimportant) and two (very important). Second, the availability score S_A ranges between zero (unavailable) and one (available). In order to derive a guideline for future development efforts in the field of computer assisted volumetry, we introduce the innovation score $S_I = S_R - S_A$, which is highest if the volumetry of a certain structure is important but currently unavailable and lowest if the tool is available but useless.

Table II.1 is sorted according to the innovation score. Thereby, three groups of structures are distinguishable according to the responding German radiologists. First, the cerebral and cardiac ventricles, the hippocampus, and tumors deserve more attention in the near future ($S_I > 0.7$). A second group would also benefit from new approaches as they become available, namely extracerebral CSF spaces, aneurysms, ischemic areas, white, and gray matter ($0.4 < S_I < 0.6$). These results suggest that the other named structures currently bear no specific motivation for the development of new volumetry tools ($S_I < 0.1$). Another structure not listed in the questionnaire was specified as important by several radiologists, namely the presurgical volumetric estimation of remaining liver volumes, e. g. in tumor surgery and—more importantly—in split liver transplantations for both donor and recipient.

It is interesting that cardiac ventricles received highest relevance scores, their return rate, however, was the lowest among all twelve structures. This could be due to the fact that cardiologists, who were not polled, might have their own imaging facilities. Conversely, it is not surprising that tumors received the highest return rate with a next to highest relevance score. The worst availability scores were received for white and gray matter volumetry at relatively low relevance scores and return rates. In conclusion, explicit white and gray matter volumetry was, at the time of enquiry, not recognized as clinically relevant topic by the participating radiologists. From recent discussions with neurologists and neuroradiologists, we assume that this has now changed.

Some of the above results are also affected by a systematic bias. Regarding a questionnaire, the two most important sources of bias are (i) who was included in the enquiry, and (ii) who did not return the questionnaire. Note that the need for a currently unavailable neuroanatomical measurement would generally be rated differently, e. g., by neurologists and neuroradiologists, assuming that the former will earlier see the need for a novel methodology. Moreover, the non-return bias probably has an effect on the relevance scores, since non-responders are also expected to have a lesser interest in a given topic.

Das, wobei unsere Berechnungen versagen,
Nennen wir Zufall.

—Albert Einstein

CHAPTER 6

Tiny World of Errors

Abstract. *In this chapter, an overview is given of the deficiencies of image acquisition in the example of magnetic resonance imaging, most importantly pixel noise and sampling effects, but also image nonuniformity and geometric distortions. These sources of error are examined from both an application and an image analysis point of view. The successive stages of image acquisition and analysis are regarded in the context of error propagation, resulting in a brief discussion of uncertainty in quantitative image analysis (QIA). It is discussed if and to which extent image analysis can compensate for diverse image deficiencies, introducing a discussion of the notions of accuracy, sensitivity, and precision.*

Volumetric measurement errors are most noticeable on objects that are small relative to image resolution. To illustrate the challenges posed by small object volumetry, a set of image analysis techniques is evaluated on three types of objects, namely a lung nodule, the pituitary gland, and Multiple Sclerosis lesions.

THIS and the following two sections are guided by *the quantitative question* of "how much?". In contrast, the classical questions to a medical image—"what?" and "where?"—are of qualitative nature. Creating answers to the quantitative question on *in vivo* images poses a fundamental problem in virtually all cases: No one can say whether the answer is correct or how far apart the answer is from the underlying truth.

6.1 Uncertainty in Quantitative Image Analysis

The *reproducibility* of a quantitative technique (cf. various types of variability in Fig. 6.1) can be tested by applying statistical methods that measure whether the technique produces the same results upon repetition. The most common example is the *standard deviation* σ_X in a set of measurements $X = \{x_1, ..., x_N\}$:

$$\sigma_X = \frac{1}{N-1}\sqrt{\sum_{i=1}^{N}(x_i - \mu_X)^2} \tag{6.1}$$

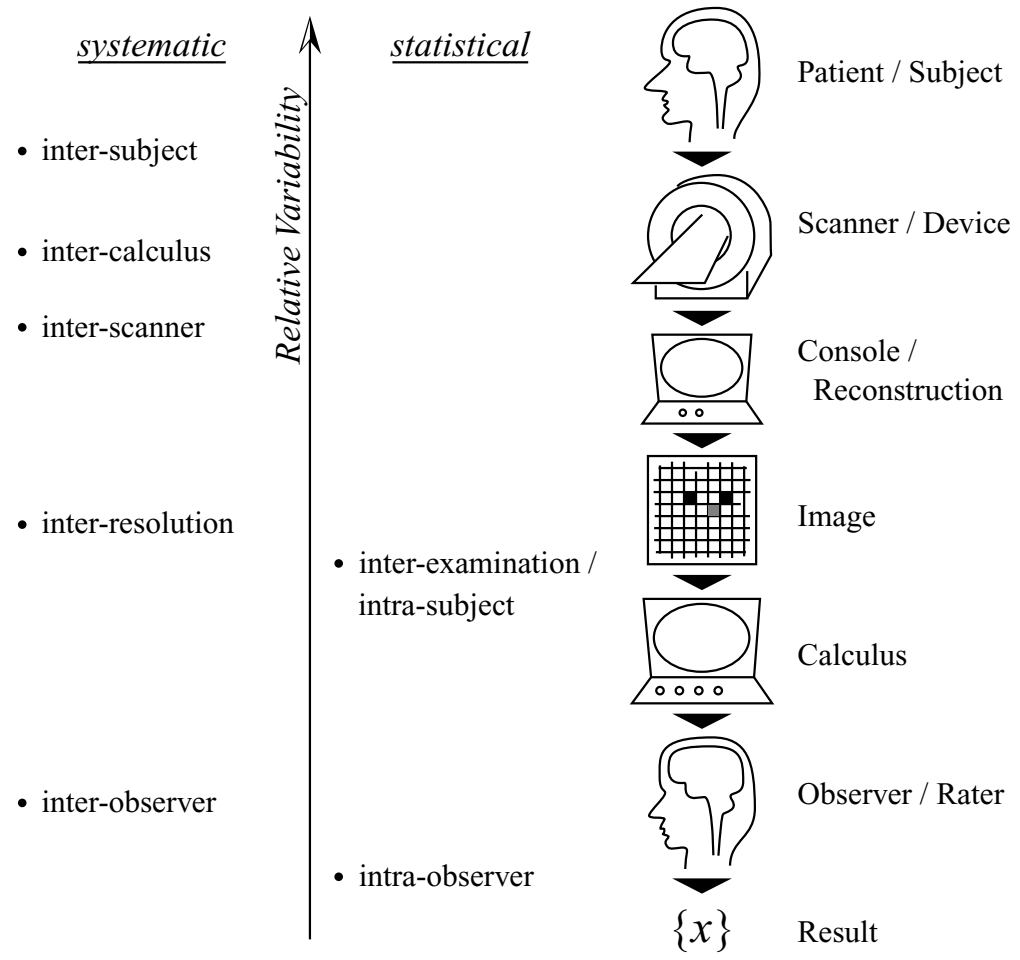


FIGURE 6.1: Chart of various types of variability as important for quantitative medical image analysis (*inter*='between', *intra*='within'), e. g. the new term *inter-calculus* indicating the difference between different image analysis methods and *intra-subject/inter-examination* indicating the difference between successive examinations of a single subject. To the right, the *Quantitative Imaging Pipeline* is shown. (Note that in the case of automated image analysis, an observer can be omitted.) It is essential to distinguish mainly statistical (center) and mainly systematic types of variability (left). The respective variability generally increases from bottom to top.

where μ_X is the mean value of X:

$$\mu_X = \frac{1}{N} \sum_{i=1}^{N} x_i \quad . \tag{6.2}$$

Instead of the standard deviation, which is also denominated SD in medical literature, a relative measure is often appropriate to assess reproducibility or variation, respectively. Based on σ and μ, we define the coefficient of variation CV as

$$CV = \frac{\sigma}{\mu} \quad . \tag{6.3}$$

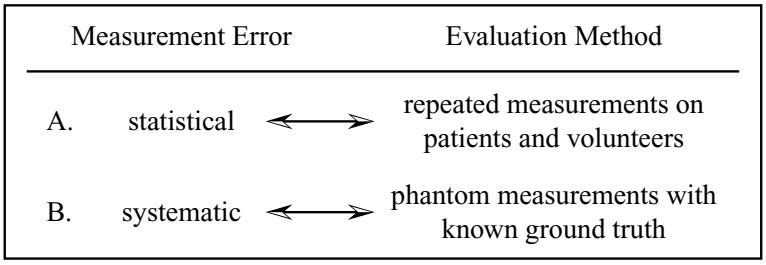


FIGURE 6.2: Appropriate methods for the evaluation of statistical and systematic errors.

In the following chapters, we will use both σ and CV to characterize the performance of proposed methods. For clinically relevant questions, a CV of approximately 5 % is often considered as acceptable (cf. answers to question 7 on p. 87). However, as will be illustrated by the following sections, a much higher reproducibility, i. e. a smaller CV, is required to differentiate stable from pathologically progressive disease in certain cases.

Conversely, much higher variability is dealt with in current oncological follow-up studies that until recently have been based on the 1979 WHO guidelines [WORLD HEALTH ORGANIZATION, 1979]. Following this classification scheme, tumor responses ranging from 50 % reduction to 25 % increase according to a *cross-sectional measurement* are classified as *stable disease* [PRASAD *et al.*, 2002]. Similarly, even the more recent Response Evaluation Criteria in Solid Tumors (RECIST) classify stable disease as less than 30 % reduction and less than 20 % increase in lesion *diameter*, with the difference that now only a one-dimensional measure is required [THERASSE *et al.*, 2000]. In terms of volume reduction, assuming a spherical lesion, 50 % reduction of cross-sectional area and 30 % reduction of diameter correspond to as much as 65 % and 66 % volume reduction, respectively, until a tumor response is no longer classified as *stable*.

Cross-sectional and diameter measurements are mostly performed manually. In order to propose an alternative to this manual processing, SCHWARTZ *et al.* [2000] have shown that in specific cases automated methods provide a higher accuracy and consistency for bidimensional tumor measurements. However, it requires a thorough evaluation and validation of current measurement techniques to narrow the above mentioned classification scheme, to become widely accepted, and thus to support early classification and detection of progressive disease. An promising example for tumor volumetry is provided in Section 6.6.

During such an evaluation it is crucial to ask whether high reproducibility suffices to prove the practical value of a measurement technique. Therefore, it is instructive to have a closer look at the various types of variability we are dealing with. At the right hand side of Figure 6.1, the *Quantitative Imaging Pipeline*, comprising image acquisition and analysis, is schematically drawn from top to bottom. To the left and center, it is listed how this pipeline affects a quantitative measurement with increasing variability from bottom to top.

In order to design appropriate evaluation protocols, it is essential to distinguish systematic and statistical elements of variability and to note that most aspects of variability are

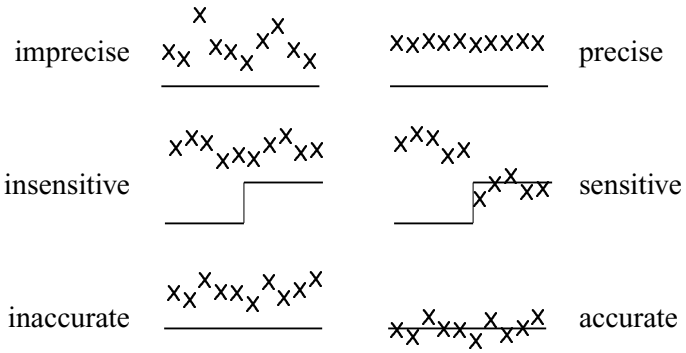


FIGURE 6.3: Schematic comparison of three different aspects of quantitative methods: precision, sensitivity, and accuracy (lines indicating the ground truth, cross symbols indicating repeated measurements). Here, precision is used in the sense of reproducibility. Sensitivity refers to the ability to show statistically significant effects. While precision and sensitivity are prerequisites for quality relative measurements, solely accuracy relates the absolute measurements to the ground truth. Note that accuracy plus precision implies sensitivity.

systematic. In a strict sense, merely inter-examination, intra-subject, and intra-observer characteristics show a mostly *statistical* pattern of variation and are thus well described by a CV or SD. Conversely, such statistical measures do not well account for the *systematic* nature of, e. g., inter-subject or inter-resolution effects.

To assess the extent of a systematic error, a reproducibility test is pointless. Rather, phantom experiments can help here (cf. Fig. 6.2) [COLLINS *et al.*, 1998, BLAKE *et al.*, 2005]. The design of an appropriate phantom for a specific question, as well as the realization and evaluation of a phantom experiment require special care and induce new pitfalls and errors [REXILIUS *et al.*, 2005, SCHLÜTER *et al.*, 2005]. As for image analysis, two types of phantoms exist, namely (i) physical phantoms that are imaged similarly to a human subject and (ii) software phantoms that include a simulation of the actual imaging procedure. Since they are less expensive and more flexible, software phantoms should be extensively used and complemented by physical phantoms only where required. Section 6.5 describes a software phantom study for MS lesion volumetry.

Restating that reproducibility is not enough from another perspective, consider a hypothetic case where the Total Brain Volume of a patient decreased by 1.5 % during the period of one year due to progressive atrophy associated with a neurodegenerative disease. Let us assume that measurements of Total Brain Volume (TBV) using a given quantitative imaging method, however, result in exactly $V_1 = V_2 = 1357\,\mathrm{ml}$ before and after this year. Let us further assume that upon repetition on either day the measurements were perfectly *precise* (cf. Fig. 6.3 for the notions of precision, sensitivity, and accuracy). In this simple example, even though reproducible, the measurements did thus not reflect the *real* volume change.

We thus conclude that besides being precise, a quantitative measurement technique must be *sensitive* to changes of the underlying truth. As motivated above, we define sensitivity as the ability to significantly reflect relative changes of a quantitative measure

(cf. Fig. 6.3 middle). However, to validate the sensitivity of a given method, one of the following is required:

- ▷ An alternative measurement method that yields reliable information on the ground truth
- ▷ A physical or software phantom, for which the ground truth is available
- ▷ Direct knowledge of the ground truth.

For example, in Chapter 8 a software phantom is used as a basis for a methodological comparison between three different approaches to brain volumetry.

It remains to consider the *absolute* values of both measurement and ground truth. While precise and sensitive measurements are sufficient for most clinical applications, there are situations that also require a high *accuracy* (cf. Fig. 6.3 bottom). Consider a longitudinal study where patients are scanned at multiple time points, but in the course of the study, the scanner is replaced by a new model offering novel acquisition hardware and protocols. If a relevant measurement is accurate, i. e. close to the ground truth on average, on both scanners, patient results along the study can be reliably compared. Accuracy is only obtained by accounting and compensating for any systematic error. Conversely, statistical errors may still occur (cf. Fig. 6.3 bottom)—and do occur in practice.

6.2 On the Physics of Uncertainty

Contemplating all different types of variability in QIA, this section asks about the reasons for these deficiencies. We focus on describing those aspects of image acquisition, which are most important for the discussion of both image analysis methodology and applications, in the example of MRI. MRI has been chosen since it forms the basis for most of the applications described in the following chapters. If not stated otherwise, our discussion refers to single coil MRI protocols commonly used for anatomical imaging. Among the characteristics of MRI, the consideration of *pixel noise* and *partial volume effects (PVE)* has proven crucial within the projects described in the remainder of this thesis. Both are artifacts generated by the specific acquisition mode and will be explained and discussed on the basis of analytical models. For an in depth discussion on the physics of image artifacts, we refer to standard literature and textbooks, such as [Haacke *et al.*, 1999].

For specific tasks, image artifacts can be seen as 'hurdles'—only some can be overcome by image analysis. Others must be handled at the acquisition stage. From an application or user point of view, the importance of image artifacts is in:

- ▷ Related errors of visually accessible information, image segmentation, and of three-dimensional visualization on the one hand
- ▷ The induction of quantitative errors on the other hand.

From a methodological point of view, their importance is related to:

- ▷ The ambiguity of gray levels

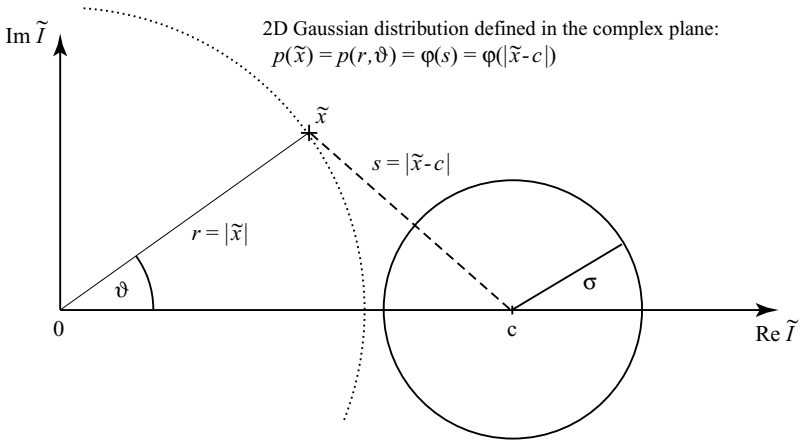


FIGURE 6.4: Schematic representation of a complex Gaussian distribution, centered at $c \in \mathbb{C}$. The probability density for this single Gaussian is a function of $s = |\tilde{x} - c|$ (dashed line). To reduce the 2D PDF $p(\tilde{x}) = p(r, \vartheta)$ to an envelope distribution $p(r)$ as a function of the magnitude $r = |\tilde{x}|$, we have to integrate $p(r, \vartheta)$ along the circle $ds = r\, d\vartheta$ (indicated by dotted line, cf. Fig. 6.5).

- ▷ The limitation of image resolution in combination with PVE that imply fuzzy object borders and image blurring
- ▷ The deformation of objects or of the image grid, respectively
- ▷ The loss of object connectivity.

Pixel Noise

Noise is often measured in terms of signal-to-noise ratio (SNR) and is the major limiting factor for high specificity[1] of image segmentation based on gray values. Often, uncorrelated white noise is assumed. It is important how amplitude and distribution of noise depend on the protocol, but also on the imaged tissues and their mean gray values. Seminal work on estimation of noise and signal in magnitude MR images was given by SIJBERS [1998].

The description of noise for MRI is well understood, but complicated by the fact that the image signal is complex, as is, in principle, the reconstructed image. In clinical anatomical MRI, however, the magnitude of the image is mostly used for diagnosis and analysis, resulting in scalar gray scale images. Consider Figure 6.4 where a complex Gaussian distribution of width $\sigma \in \mathbb{R}^+$ is schematically represented by a circle of radius σ centered at $c \in \mathbb{C}$. If the zero-centered normalized one-dimensional Gaussian probability density function (PDF) $\varphi(x)$ is given by

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{x^2}{2\,\sigma^2}}\,, \quad x \in \mathbb{R}\,, \tag{6.4}$$

we can define the $\tilde{\mu}$-centered normalized two-dimensional Gaussian $\varphi(\tilde{x})$ as

$$\varphi(\tilde{x}) = \frac{1}{2\pi\,\sigma^2}\, e^{-\frac{|\tilde{x}-\tilde{\mu}|^2}{2\,\sigma^2}}\,, \quad \tilde{x}, \tilde{\mu} \in \mathbb{C}\,. \tag{6.5}$$

[1] Here, specificity refers to a low number of false classifications of image elements.

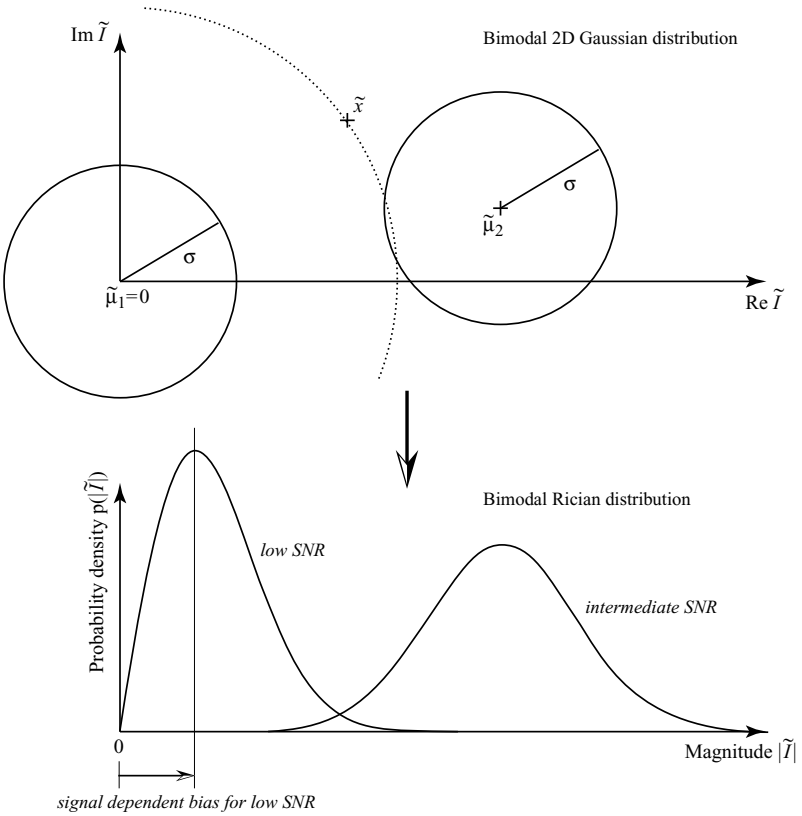


FIGURE 6.5: Qualitative derivation of RICIAN distribution from two 2D GAUSSIAN distributions with identical isotropic noise of width $\sigma \in \mathbb{R}^+$ and centered at $\tilde{\mu}_k \in \mathbb{C}, k \in \{1, 2\}$ (top). Integration along circles centered at the origin yields Rician distributions as functions of merely the complex signal's magnitude $|\tilde{I}| \in \mathbb{R}_0^+$ (bottom). Note that the left Rician distribution (low SNR) shows average and maximum values far above zero (signal dependent bias), even though the original complex Gaussian was zero-centered, and that only the right Rician distribution (intermediate SNR) exhibits a shape similar to a 1D Gaussian.

The assumption of Gaussian white noise holds for both the real and imaginary parts of the signal or image [MACOVSKI, 1996, MCVEIGH *et al.*, 1985]. In the frequency domain, where the MR signal is acquired, the noise amplitude inherent to image reconstruction is theoretically identical for all rows and can be estimated by the acquisition bandwidth.

As a consequence of the two-dimensional noise problem, the magnitude image is partly correlated, especially for low SNR. Noise of the magnitude image is best described by the so-called RICIAN NOISE [RICE, 1948, GUDBJARTSSON and PATZ, 1995, SIJBERS, 1998]. For high SNR, Rician noise is well approximated by a Gaussian, while for low SNR, not only the distribution of the magnitude image differs significantly from a Gaussian, but also the average magnitude signal is *biased*. To understand the problem of a signal dependent bias, consider two ideal tissue types with an average complex signal $\tilde{\mu}_k \in \mathbb{C}, k \in \{1, 2\}$, and Gaussian white noise of width $\sigma \in \mathbb{R}^+$ for both real and imaginary parts. Furthermore, consider the special case $\tilde{\mu}_1 = 0$, as illustrated by Figure 6.5.

The left Rician distribution ($\tilde{\mu}_1 = 0$) equals a RAYLEIGH distribution, as can be shown by integration along a circle centered at the origin with radius $r = |\tilde{x}|$:

$$p_1(r \,|\, \sigma) \;=\; \int_{\vartheta=0}^{2\pi} \varphi(r \,|\, 0, \sigma)\, r \, d\vartheta \;=\; \frac{r}{2\pi\,\sigma^2}\, e^{-\frac{1}{2}(\frac{r}{\sigma})^2} \int_{\vartheta=0}^{2\pi} d\vartheta \;=\; \frac{r}{\sigma^2}\, e^{-\frac{1}{2}(\frac{r}{\sigma})^2} \tag{6.6}$$

For a non zero-centered complex Gaussian ($\tilde{\mu}_2 \neq 0$) such as the right distribution in Figure 6.5, the integral is more complicated. The distance $s = |\tilde{x} - \tilde{\mu}_2|$ (cf. Fig. 6.4) can be calculated from c, r, and ϑ by trigonometry: $s^2 = c^2 + r^2 - 2cr\cos\vartheta$, with $c = |\tilde{\mu}_2|$ and $\vartheta = \measuredangle(\tilde{x}, \tilde{\mu}_2)$. Consequently, the Rician takes the general form

$$p_2(r \,|\, c, \sigma) \;=\; \int_{\vartheta=0}^{2\pi} \varphi(\tilde{x}(r,\vartheta) \,|\, \tilde{\mu}_2, \sigma)\, r \, d\vartheta \;=\; \frac{r}{2\pi\,\sigma^2} \int_{\vartheta=0}^{2\pi} e^{-\frac{c^2+r^2-2cr\cos\vartheta}{2\,\sigma^2}}\, d\vartheta \quad . \tag{6.7}$$

By using the zero-order modified BESSEL function of first kind $I_0(x) = \frac{1}{2\pi}\int_0^{2\pi} e^{x\cos\alpha} d\alpha$, as described by F. W. J. OLVER in [ABRAMOWITZ and STEGUN, 1972, pp. 374–377], this can be written as

$$p_2(r \,|\, c, \sigma) \;=\; \frac{r}{\sigma^2}\, I_0\Big(\frac{cr}{\sigma^2}\Big)\, e^{-\frac{c^2+r^2}{2\sigma^2}} \quad . \tag{6.8}$$

This function, which produces Figure 6.5 bottom, provides an appropriate noise model for MR magnitude images, given that the above assumptions hold. If all modeled tissues for a chosen imaging sequence, however, exhibit an intermediate SNR or better, the Gaussian model that is most often used in current literature, including our own work, is an appropriate approximation. In specific cases, we see a potential for improving qualitative and quantitative MR image analysis by employing the Rician distribution. In any case, the nature of this distribution should be kept in mind when analytically processing noisy MR images.

Point Spread Function and Partial Volume Effect

Both point spread function (PSF) and partial volume effects (PVE) are directly related to image resolution and are the major limiting factors for an accurate quantitative analysis of objects that are thin with respect to the resolution. Image resolution can be characterized by the shape and width of the PSF, separately in each spatial direction. In standard MRI, sampling is done on a limited regular grid in the frequency domain, such that the signal is multiplied by both comb function and box function [HOLDEN, 2001, pp. 220 f.]. After Fourier transform, the PSF in image space is thus composed of SINC functions. Thereby, resolution is limited by the highest acquired frequency and for 2D imaging in orthogonal direction by the RF pulse slice selection thickness, respectively. Furthermore, the PSF can be influenced, e. g., by so-called partial Fourier imaging or zero filling techniques, if applied.

For the interpretation of the image contents, it is important to note that the PSF also has a temporal component, depending on the employed 2D or 3D acquisition mode. For example, a block of slices acquired by 2D MRI sequences can store in each slice information

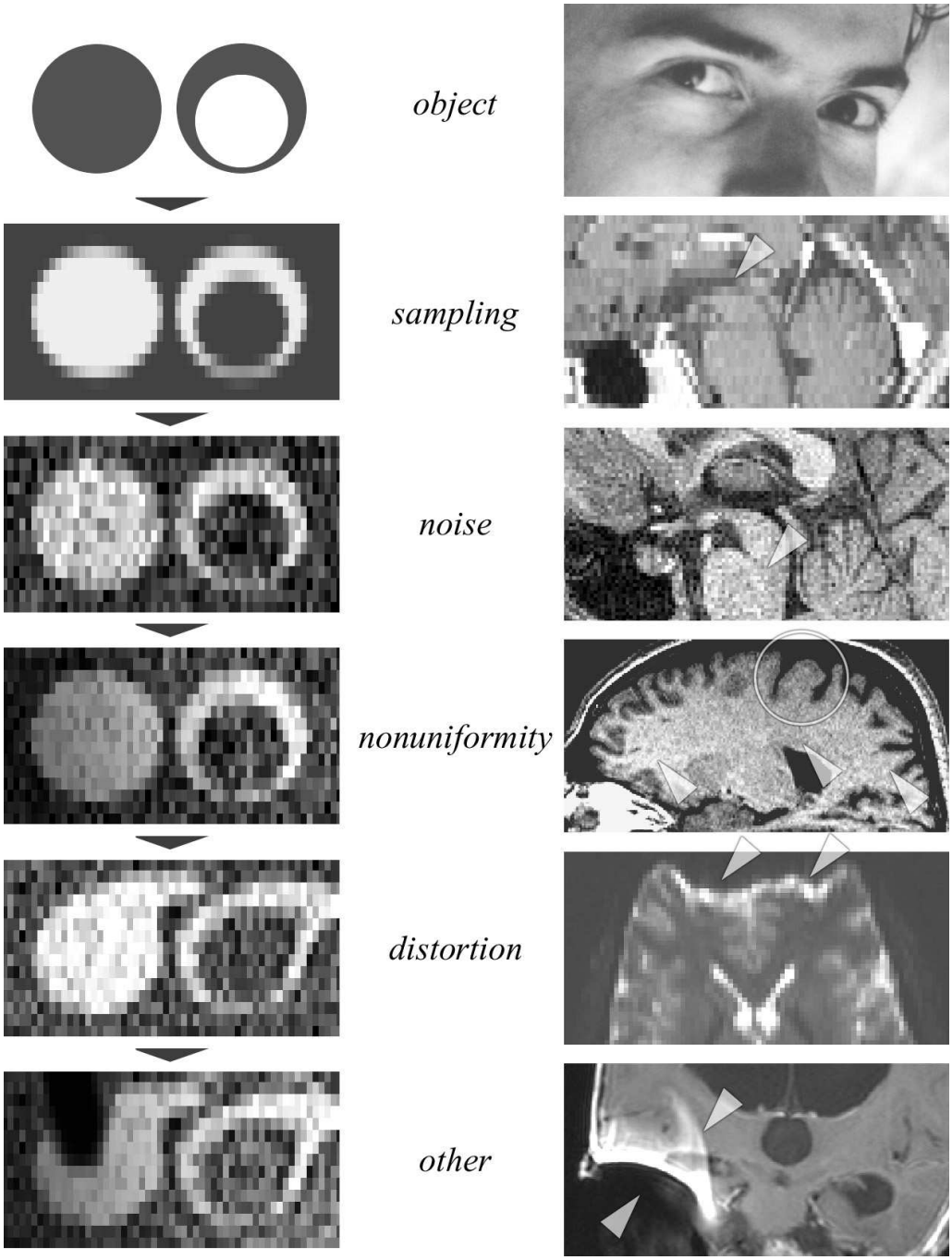


FIGURE 6.6: Schematic overview of image artifacts that are of major importance for MRI in the example of synthetic (*left*) and real images (*right*). The arrows in the right column indicate image positions where the respective artifacts become most obvious: *Sampling*: Partial volume averaging due to anisotropic voxel size at a slice thickness of 3 mm. *Noise*: Image produced by critically reducing the acquisition time at the cost of SNR. *Nonuniformity*: Varying mean intensity of WM (arrows) and varying relative contrast between GM and WM across the image related to nonuniform flip angle due to transmitting RF coil characteristics (circle). *Distortion*: Strong distortion induced by susceptibility artifacts that, e. g., occur at air–tissue interfaces in Echo Planar Imaging (EPI). *Other*: Metal artifacts in the right temporal lobe region (left in the image).

from specific non-overlapping time periods, while for 3D acquisition the contribution from the signal at specific time point to a particular voxel is more complex.

While the PSF is crucial for image generation and for voxel interpretation, PVE are of major importance from the perspective of volumetric image analysis. Consider a solid object, such as a sphere, imaged at a given three-dimensional resolution. For measuring the volume of this object, it seems obvious to count all voxels that represent the object and multiply this number by the voxel volume. However, border voxels do only partly represent the object, while also representing neighboring objects or tissue (background) for the other part (cf. Fig. 6.6, 2$^{\text{nd}}$ row, example 'sampling' for brainstem and CSF). Typically, the gray value of these partial volume voxels is found between the object's and neighboring objects' gray values. In most cases, the assumption of a linear combination—according to the linearity of the Fourier transform—of gray values according to the actual voxel partitioning holds. Limits of this assumption are induced by non-linear effects, such as saturation, or other artifacts (cf. below), and of course by SNR.

A common technique to separate an object from its background is to use a gray level threshold; simple segmentation techniques, such as most seeded region growing, and many clustering techniques also rely on some sort of threshold. The impact of the actual threshold on the measured volume becomes increasingly important as the proportion of partial volume voxels with respect to pure object voxels gets larger. This critically depends on the image resolution as illustrated by Figure 6.7 for the example of a solid sphere.[2]

It is crucial to note that providing a threshold to the user induces important observer-dependent effects on the volumetric result. In the context of QIA such as volumetry, interactive handles to a threshold, from which the quantitative result depends significantly, should thus be completely avoided.

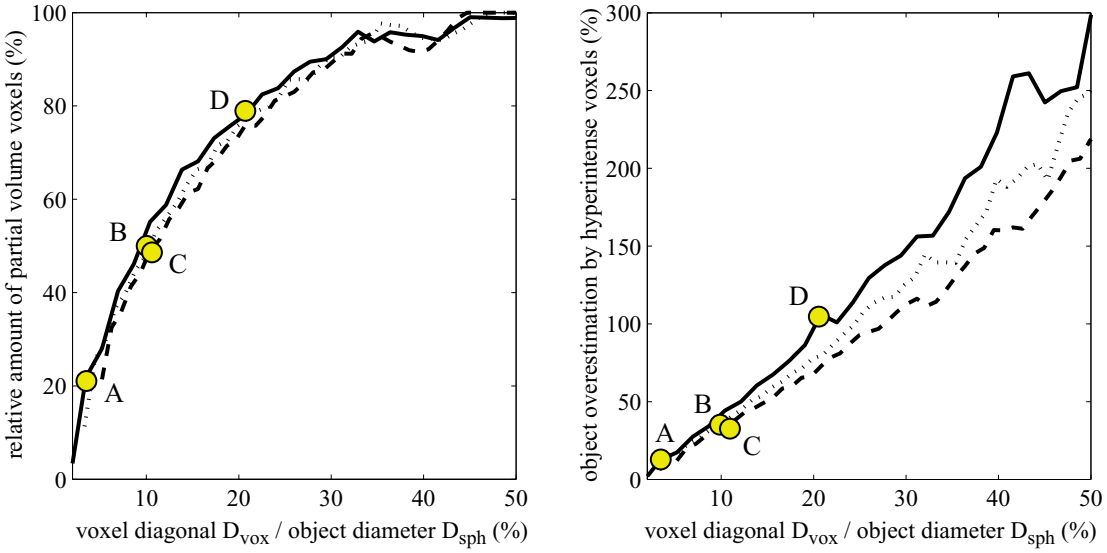
To overcome the limits of gray level thresholding, more elaborate models for the image contents are required. In the presence of noise, as mentioned in the previous section, Gaussian distributions are most often used to model pure tissue gray values. For partial volume regions, however, a Gaussian model is not appropriate since a multitude of tissue mixtures—from 0 % to 100 % for each tissue—exists, each with a different corresponding mean gray value.

A central question in this context is *which distribution* of voxel partitioning can be expected for a given image and for a certain pair of tissue types. The *uniform distribution* as proposed by Santago and Gage [1993] is mainly used in literature, including our own work. It has the advantage of conceptual and analytical simplicity by assuming that any voxel partitioning is equally probable. Still, this distribution should be used with care. Two effects are discussed here in which the uniform distribution does not provide an appropriate model for the distribution of gray values:

(a) Histogram accentuation of extreme partitions

(b) Histogram asymmetry due to thin object geometry.

[2]It has turned out as a result of preparatory studies by the author that, besides the PSF, the voxel diagonal is a good scalar measure to characterize image resolution. Note that using this measure, instead of the largest voxel extent along x, y, or z, yields a relatively close correspondence between curves for varying voxel anisotropy values in Figure 6.7.



	A	B	C	D
D_{vox} / D_{sph}:	3.5%	10.5%	11.0%	20.8%
Voxel Anisotropy:	1	3	5	1
Voxel y-extent v_y:	0.20 mm	0.95 mm	1.06 mm	1.20mm
Part. Vol. Fraction (left plot):	21.8%	51.5%	50.3%	78.2%
Overestimation (right plot):	12.8%	39.0%	36.5%	105.9%

FIGURE 6.7: Significance of partial volume averaging in the example of a sphere with an exact diameter of 10.0 mm. Perfect imaging is assumed (zero noise, box PSF). Two plots show the dependency of partial volume effects on the voxel diagonal D_{vox} (in percent relative to sphere diameter D_{sph}) for varying voxel anisotropy, i. e. ratio of longest (v_{y}) to shortest voxel extent ($v_{\mathrm{x}} = v_{\mathrm{z}}$); simulated anisotropy values are 1 (solid), 3 (dotted), and 5 (dashed lines). *left*: Fraction of partial volume voxels and all voxels containing sphere signal (hyperintense voxels). *right*: Relative overestimation of the sphere volume if all hyperintense voxels are counted. Details are provided on four examples (A–D; pure sphere and background voxels are shaded bright and medium gray, respectively). Note that despite the relatively small voxel size of 1.2 mm (isotropic, i. e. anisotropy= 1), the overestimation is more than 100 % and more than 75 % of the hyperintense voxels represent partial volume (example D). Already for extremely high resolution (0.2 mm isotropic), more than 20 % of the voxels represent partial volume (example A). The curves do not ascend monotonically due to the fact that only a single discretization was performed per voxel size, where the sphere center did not always coincide with a voxel center.

To understand the first, assume perfect cubic voxels with an ideal PSF and further assume that the distribution of voxel partitioning is best modeled by a voxel cut by a *random plane* (random support and orientation vectors). This results in a distribution that is considerably different from a uniform one, but rather shows a characteristic *U-shape* as shown by Monte Carlo simulation (cf. Color Plate C.6, red line). If the support vector of the random plane is uniformly distributed in 3D Euclidean space and if the orientation vector is uniformly distributed on a unit sphere, we introduce the term *uniform cut*.

Replacing the cube by a sphere, the distribution does not loose its characteristic U-shape, as also observed by Bello *et al.* [1998]. We have derived this distribution analytically [Jolly, 2004]. It amounts to

$$p_{\mathrm{sph}}(v) = -\frac{2\sin\left(\frac{\pi}{3} + \frac{\arccos(2v-1)}{3}\right)}{3\sqrt{1-(2v-1)^2}} \quad , \tag{6.9}$$

where $v \in [0;1]$ is the portion of the sphere that falls into a given half-space as defined by the plane. It is interesting to note that only minimal differences exist between the distributions for cube and sphere (cf. Color Plate C.6); the major difference is a stronger accentuation at the extreme partitions (below 4 % and above 96 %) for the cube distribution. We argue that for a realistic image with imperfect PSF, this accentuation should still be beyond the cube distribution.

Using two identical Gaussians with different mean values to model pure tissues (cf. Color Plate C.7 c, blue line), the *uniform distribution* of partitions yields a flattened plateau (green line). In contrast, regardless of which model we choose—sphere, cube, or something more realistic—, the effective partial volume distribution corresponding to the *uniform cut* is still symmetric, but much more accentuated at the borders (red line).

An interesting approach to overcome the uniform model is the concept of voxel histograms introduced by Laidlaw *et al.* [1998]. This approach has the potential to accurately model local partial volume distributions at a sub-voxel resolution. However, from our perspective, such a concept is limited by its computational complexity and the requirement of high SNR.

The second effect occurs if the object thickness is of the same order as the image resolution and if the object is positioned randomly on the image grid. Then, partial volume voxels will more likely represent a surplus of background. As an effect, the corresponding distribution is highly asymmetric with an accent on the small object and large background partitions ($v < 0.5$). This effect is of crucial importance, e. g., when analyzing the histogram of the cerebellum in anatomical MR images of the brain; the layers of white matter—and also of CSF—in this highly folded structure are mostly thinner than one millimeter.

Both effects, histogram accentuation of extreme partitions (*U-shape*) and histogram *asymmetry* due to thin object geometry, should be kept in mind when discussing volumetric image analysis in the following two chapters. They have implications for any statistical model used in image analysis methods such as EM or Fuzzy Clustering.

Nonuniformity

The above suggests that for given acquisition parameters a certain pure tissue is represented by an unambiguous mean intensity across the image. However, besides noise and PVE, nonuniformity (NU) is a further limiting factor when attributing gray values to specific image contents. In MRI, NU is governed by field and RF inhomogeneity, and, thus, not by a random process. The latter depend on the transmitting and receiving coils. For example, extreme NU effects are observed for surface coils. In principle, field and RF inhomogeneity can be measured by extra calibration scans and so-called shimming. Still, the imaged body individually implies further field inhomogeneity that cannot be corrected prior to the scan.

In image analysis, MRI nonuniformity is most often modeled by smooth and slowly varying multiplicative bias fields [Sled *et al.*, 1998, Mangin, 2000, Likar *et al.*, 2001, Arnold *et al.*, 2001]. Unfortunately, MRI nonuniformity is not always acting as a multiplicative bias. Instead, we also observe an inhomogeneity of relative contrast between certain tissue types. An example is the varying gray–white matter contrast that is often lower in parietal brain regions than in frontal and temporal, mainly depending on the transmitting coil characteristics (cf. Fig. 6.6, example 'nonuniformity'); with other words the relative signal drop is significantly higher for white than for gray matter in these cases.

In Chapter 8, it will be demonstrated how difficult it is to account and correct for image NU in the context of QIA. Surprisingly, the histogram analysis proposed for whole brain volumetry without an explicit NU model showed a slightly higher robustness to NU than two more complex methods that incorporate explicit NU modeling. A possible explanation is that in the image histogram, NU has a widening effect on the tissue distributions, and can in specific cases be attributed to the noise term.

Distortion

Unfortunately, geometric distortion is an important issue for specific MRI acquisition techniques. Unlike in CT, where geometric accuracy can be widely assured, MRI considerably depends on the homogeneity of the magnetic fields. An example is given by air–tissue interfaces posing problems in rapid imaging sequences; this can be observed, e. g., in Echo Planar Imaging (EPI), which is normally used for fMRI, at the inner borders of the frontal air-filled sinuses. Standard Diffusion Tensor Imaging (DTI) also uses EPI and is prone to the same strong distortions (cf. Fig. 6.6, example 'distortion'). Further details of MRI distortion are discussed by Holden [2001].

This type of distortion is potentially avoided by Spin Echo sequences where the influence of local inhomogeneity is compensated at the time of echo by refocusing. In this sense, Spin Echo are 'cleaner' than Gradient Echo sequences. Today, T2 weighted images are often acquired by Turbo Spin Echo (TSE) sequences (also called Fast Spin Echo or RARE by different device manufacturers) with good image quality at acceptable acquisition times. There were also attempts to implement DTI using TSE, e. g. by the so-called propeller imaging [Pipe, 1999].

Drawbacks of Spin Echo imaging are that acquisition times are prolonged and that the specific absorption rate (SAR) is increased due to to frequent application of RF pulses with large flip angles. In addition, techniques sensitive to changes of the relaxation time T2* require the acquisition of gradient echoes. Moreover, in BOLD imaging, the method of choice for most fMRI studies, mostly gradient EPI is employed. An alternative approach to reduce geometric distortion is represented by Stimulated Echo Acquisition Mode (STEAM) imaging, which uses gradient echos [Matthaei *et al.*, 1986].

Artifacts (and Specialized MRI)

Numerous further reasons exist for image artifacts, of which we only want to name a few. For example, the presence of metal is a problem for MRI (cf. Fig. 6.6, example 'extinction'). Not surprisingly, MRI of the jaw is complicated by tooth fillings in many cases. Another source of artifact is motion during image acquisition. Motion can be due to breath, heart beat, pulsatile blood flow, or patient movement. Most standard MRI techniques require that the imaged object remains still during acquisition, such that various types of shadows and ringing artifacts occur in case of motion.

It is interesting to note that artifacts that were first annoying have been exploited as a basis for special image acquisition techniques in various instances. One example are flow artifacts, which provide a basis for time-of-flight angiography that is able to produce a high contrast for blood vessels without the need of contrast agent. Another example are wrap-around or shadow artifacts due to incomplete sampling that are dealt with by fast parallel imaging techniques, such as SENSE (or m-SENSE) in the imaging domain and SMASH (or GRAPPA) in the frequency domain.

A third important example is Diffusion Weighted Imaging (DWI); this technique is designed such that those protons that did not move out of focus along a given magnetic field gradient direction contribute most to image formation. More specifically, repeating such acquisition for at least six independent gradient directions allows for DTI. A fourth artifact is the water–fat shift that can effect a voxel to appear dark (out of phase) or bright (in phase) without changing the physical composition of the voxel. In this case, the image linearity at object borders discussed above is undermined. The reason for this artifact is found in the fact that protons in fat and in water have a slightly different resonance frequency. More generally, this effect is exploited by so-called chemical-shift imaging (CSI), a type of MR spectroscopy that obtains spatially localized spectra [Brown *et al.*, 1982].

6.3 Implications for Image Analysis

From an image analysis point of view, it is crucial to differentiate image artifacts from real anatomical or physiological changes. For example, the growth of a tumor or progressive atrophy of brain parenchyma could be misclassified as geometric image distortion, and thus eliminated during image analysis [Holden, 2001, pp. 147 ff.]. This ambiguity represents a major difficulty when designing image analysis methodology.

Due to their various sources, it is difficult to provide valid upper bounds for image artifacts in MRI. Fortunately, in many standard situations, virtually no geometric distortion

is observed. We have scanned a large glass ball in water using 3D Gradient Echo techniques on three different scanners (MAGNETOM VISION PLUS, MAGNETOM SYMPHONY STANDARD, and MAGNETOM SYMPHONY QUANTUM, all SIEMENS MEDICAL SOLUTIONS) and found length variations between scanners less than 0.2 %, which is sufficient for most applications, but can disturb highly precise measurements.

Recalling the chart of systematic and statistical errors associated with the *Quantitative Imaging Pipeline* (cf. Fig. 6.1), it is crucial to note that only a small number of these errors are statistical or somehow random in nature and can thus be alleviated by means of proper statistical image analysis. This group of errors correspond to the variability that can be observed by repeating the exact identical experiment, assuming that the imaged subject, as well as the observing instance did not change. Furthermore, errors that occur due to physiological variation of the subject or due to variability of human observation or operation of a tool are summarized by *intra-subject* and *intra-observer* variability, respectively.

Conversely, systematic errors are generated by deterministic physical effects and are only accessible through comparison to some sort of calibration or reference. Still, we can demand that image analysis should be robust to systematic errors, such as image nonuniformity or variable resolution. However, distortion or extinction is a typical example where analysis of a single image is incapable of artifact compensation. In this sense, *optimality* of quantitative results and the underlying image analysis refers to a theoretically achievable optimum. The practical situation is yet suboptimal in many cases; in Chapter 8, we will show how different quantitative results can arise from three highly sophisticated analysis tools that serve the same purpose.

Turning toward the subject of this thesis, a segmentation technique should exhibit:

- ▷ Insensitivity to image noise and nonuniformity, possibly also to ghosting and ringing artifacts etc.
- ▷ Anatomical correctness despite the ambiguity of image contents
- ▷ Efficient interactivity to be able to integrate user knowledge,

while of a quantification method we might demand:

- ▷ Appropriate handling of PVE
- ▷ Sensitivity and specificity to relevant changes of the imaged subject
- ▷ High degree of automation and insensitivity to user variability
- ▷ Provision of error estimates.

Error Propagation

The objective estimation of measurement error propagation is a troublesome matter in QIA for three reasons:

(a) Only some error sources are of purely statistical nature (cf. Fig. 6.1).

(b) Systematic errors are mostly unaccessible by image analysis (e. g. image distortion), but rather require some sort of calibration.

(c) Human influence (e. g. patient positioning, marker placement, threshold selection, patient movement) is variable such that upper error limits cannot generally be established.

If, however, the result is calculated from a limited group of parameters, and if the variance of these parameters is known and uncorrelated, the variance of the result is calculated by Equation 1.3 (p. 10, cf. comment on p. 124). For correlated variables the mixed terms have to be appropriately added, which also requires the respective covariances to be known. This can eventually lead to a reduced variance for the result in comparison to the single variables (cf. Sec. 8.5, p. 148).

Since the world of errors in quantitative medical imaging is complex, three specific volumetry problems were chosen such that each reveals a different aspect of quantitative volumetric image analysis. The problems and the corresponding measurement errors are analyzed in Sections 6.4–6.6. Finally, Section 6.7 discusses reasons for the observed variability, as well as implications for image analysis in each specific case.

6.4 Volumetry of the Pituitary Gland

Problems and limitations of image analysis become most obvious in small objects. The first example to analyze is MR-based volumetry of the pituitary gland (also 'hypophysis').

The pituitary gland normally has a volume of less than a milliliter. Pituitary adenomas are among the most frequent intracranial tumors and are associated with high therapeutic costs due to their deep location. These benign alterations of the hypophysis often show a rapid progress and are subdivided into micro- (diameter less than 10 mm) and macroadenomas (larger than 10 mm). An indication for surgical intervention (either transsphenoidal access or frontoparietal craniotomy) is provided by fast growth or neurological deficits [Bonneville, 2000]. Especially for microadenomas, rating of hypophysis growth is impaired by the unreliability and subjectivity of manual measurements. The motivation for a precise semiautomatic volumetry is to reduce the follow-up interval for growth control after initial diagnosis and also after surgical intervention and to simultaneously decrease the error of such a rating. As an effect, decisions based on hypophysis growth can be taken earlier and more reliable, and ultimately patient care is improved.

MRI has already been established as gold standard for diagnosis of the pituitary gland [Bonneville, 2000]. It allows for differential diagnosis with respect to, e. g., suprasellar meningiomas or cystic alterations, and also allows for intraoperative monitoring [Pergolizzi Jr *et al.*, 2001]. Due to its good soft tissue differentiation, we can hope to be able to establish a semiautomatic volumetry of the pituitary gland based on MRI.

In cooperation with Joachim Böttcher, Jena, a dedicated modular image analysis system was established consisting of four successive steps (cf. Fig. 6.8):

(a) Interactive selection of an axis-parallel cuboid volume of interest

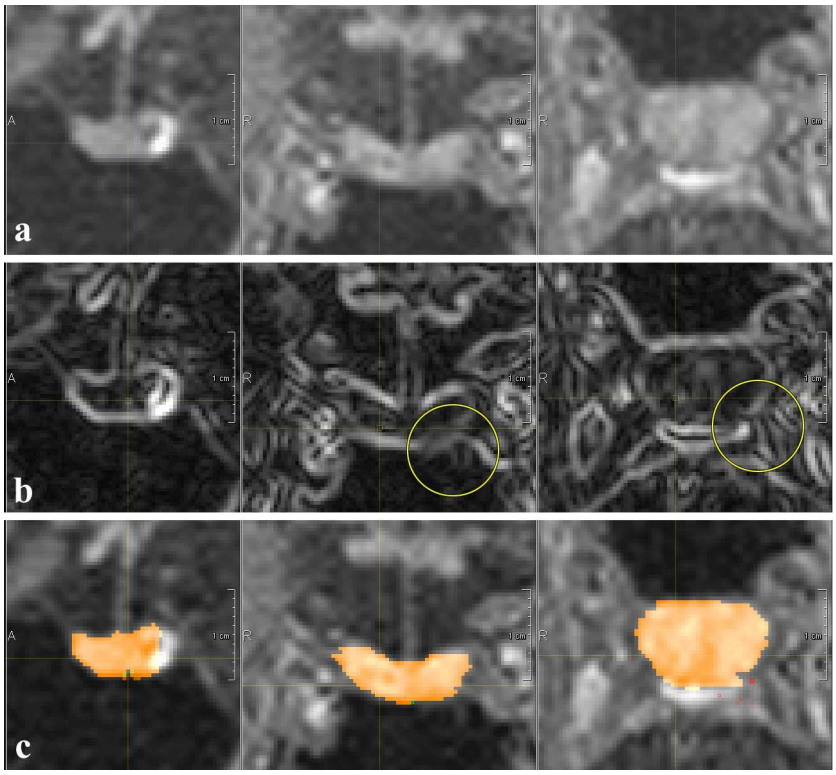


FIGURE 6.8: Segmentation and volumetry of the pituitary gland using the IWT applied to a derived gradient image (so-called *gradient watershed*). The original image is resampled to a finer isotropic grid with 0.4 mm voxel size, and is shown in three orthogonal sections (left to right): sagittal, coronal, and axial (*a*). Subsequently, a 3D gradient filter is used for edge enhancement (*b*). Finally, the IWT is used on the gradient image to interactively locate the pituitary gland (*c*). Interaction with include and exclude markers is required at positions where the object border is not well defined in the gradient image (circles). The volume is calculated by voxel counting. (MR image courtesy of J. BÖTTCHER, Jena)

(b) Resampling to an isotropic resolution (e. g. 0.4 mm) using a three-lobed LANCZOS filter[3]

(c) Local image gradient computation

(d) Segmentation using the Interactive Watershed Transform (IWT) applied to the gradient image.

While the gradient image is used as a basis for segmentation (cf. Fig. 6.8 b), the actual binary segmentation mask is displayed as an overlay to the resampled original data during interaction (cf. Fig. 6.8 c). Our approach is similar to the work by LETTEBOER *et al.* [2001] where the image gradient was introduced as *dissimilarity measure*. The difference,

[3] n-lobed LANCZOS filter kernel function: $w(t) = \mathrm{sinc}(t) * \mathrm{sinc}(t/n)$ for $|t| < n$, and $w(t) = 0$ else; with $\mathrm{sinc}(x) = \sin(x\pi)/(x\pi)$ for $x \neq 0$, and $\mathrm{sinc}(x) = 1$ else.

TABLE 6.1: Inter-examination and intra-observer results for volumetry of pituitary gland. One observer (Obs_A) has performed volumetry (analysis dates 2nd row) using the interactive gradient watershed technique (cf. Fig. 6.8) on four independent image acquisitions (acquisition dates left column, *yy/mm/dd*) of a female volunteer (35 y) without pituitary gland pathology. Mean values and standard deviations (SD) besides coefficients of variation (CV) are calculated within (intra-observer, right column) and across acquisitions (inter-examination, bottom rows). Moreover, a second observer (Obs_B) has performed the same task once on the same data (second column).

	Obs_B	Obs_A				Obs_A
	02/09/17	*02/09/17*	*03/04/07*	*03/04/11*	*03/04/11*	Mean ± SD (CV)
99/08/31	0.66	0.65	0.67	0.65	0.66	0.65 ± 0.01 (1.4 %)
99/09/09	0.65	0.65	0.61	0.59	0.64	0.62 ± 0.03 (4.2 %)
99/11/18	0.65	0.63	0.66	0.63	0.65	0.64 ± 0.02 (2.9 %)
01/09/04	0.71	0.72	0.69	0.66	0.69	0.69 ± 0.03 (3.7 %)
Mean	0.67	0.66	0.66	0.63	0.66	
± SD	± 0.03	± 0.04	± 0.03	± 0.03	± 0.02	
(CV)	(3.9 %)	(6.5 %)	(4.9 %)	(4.7 %)	(3.2 %)	

however, is that we use the IWT in order to provide instantaneous feedback to marker placement.

To achieve reproducible hypophysis volumes, it is important to define whether the fatty sellar tissue should be in- or excluded at the segmentation step. Even though our experiments show a slightly higher reliability of the measurements when including this region, an interactive separation using exclude markers is possible (cf. Fig. 6.8 c, bright tissue posterior to the hypophysis).

We have evaluated intra-observer, inter-examination, and inter-resolution characteristics for our technique on two volunteers without pituitary gland pathology. The first volunteer (F, 35 y) was scanned four times on different days, while the fourth scan was about two year later than the first three; inter-examination and intra-observer results are summarized in Table 6.1; in addition, results are provided from a second observer. The second volunteer (F, 27 y) was scanned five times on the same day; inter-examination and inter-resolution results are summarized in Table 6.2. The results are discussed in Section 6.7.

6.5 Software Phantom for Focal White Matter Lesion Volumetry

Another class of small objects, which is important in imaging neurodegenerative or inflammatory diseases such as MS, are white matter lesions. In MS, these lesions typically occur at paraventricular locations, but also throughout the cerebrospinal system. When first diagnosed, they have a diameter of a few millimeters with an elliptical or elongated shape (cf. Fig. 6.9 right). If the borders of a compact lesion can be distinguished from the surrounding normal appearing tissue, we speak of a focal white matter lesion. Since

TABLE 6.2: Inter-examination and inter-resolution results for volumetry of pituitary gland. One observer has performed volumetry (cf. Fig. 6.8) on five independent image acquisitions (rows) of a healthy female volunteer (27 y), original resolution 1 mm isotropic. Each image was resampled axially at three further resolutions (1.8, 3.0, and 4.8 mm). Mean values and standard deviations (SD) besides coefficients of variation (CV) are calculated across resolutions (right column) and within resolution across examinations (bottom rows).

	1.0 mm	*1.8 mm*	*3.0 mm*	*4.8 mm*	Mean ± SD (CV)
Scan 1	0.71	0.68	0.57	0.69	0.66 ± 0.06 (9.2 %)
Scan 2	0.81	0.73	0.73	0.66	0.73 ± 0.06 (8.1 %)
Scan 3	0.69	0.78	0.64	0.65	0.69 ± 0.06 (9.0 %)
Scan 4	0.74	0.77	0.84	0.83	0.79 ± 0.05 (5.8 %)
Scan 5	0.72	0.74	0.67	0.65	0.70 ± 0.04 (6.4 %)
Mean	0.73	0.74	0.69	0.70	
±SD	± 0.04	± 0.04	± 0.10	± 0.07	
(CV)	(6.1 %)	(5.4 %)	(14.5 %)	(10.7 %)	

a few years, quantification thereof is an obligatory part of all larger MS drug trials where T2 hyperintense[4] lesions are counted and a total lesion load volume is measured. Further parameters refer to T1 hypointense (so-called black holes) and T1 Gadolinium enhancing lesions.

From a methodological point of view, we want to question here, how accurate and how precise the volume of focal white matter lesions can be quantified. Note that typical MR imaging protocols used for this purpose in clinical routine but also in clinical studies use a slice thickness of three to five millimeters, which is in the same order of magnitude as the typical lesion extent. From Figure 6.7 left we can derive that for a large hyperintense white matter lesion with a diameter of nine millimeters—even under perfect imaging conditions—virtually all hyperintense voxels are partial volume voxels, i. e. represent both lesion and normal appearing tissue with unknown partitioning ratio.

Under perfect (infinite) SNR conditions and assuming linear partial volume gray value interpolation, the exact partitioning ratio of a given voxel and thus also the volume of a lesion, embedded in homogeneous background, could be computed if the gray value corresponding to a pure lesion voxel would be known. This is not the case, however, in most practical cases, since on the one hand very few or no pure lesion voxels exist in small and intermediate size lesions according to the above statement and on the other hand SNR is limited.

Conversely, the problem of lesion volumetry in current clinical practice and studies is of a different nature—it is performed by manual or semiautomatic tracing plus voxel counting. Human observers tend to mark most abnormal appearing (e. g. hyperintense) voxels as lesion. After voxel counting, all these voxels, no matter if partial volume or pure, are counted as pure lesion voxels. As a result, lesion volumes are heavily overestimated by

[4] *T2 hyperintense*: appears bright on T2 weighted MR images compared to normal tissue.

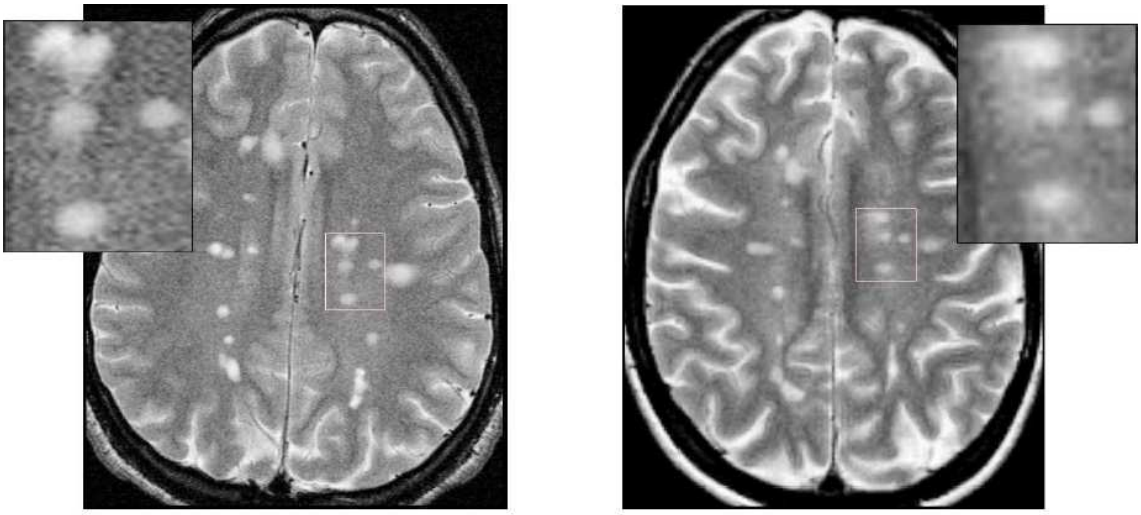


FIGURE 6.9: Examples of software phantom and similar patient data with MS lesions: T2 weighted MR images with hyperintense focal lesions. The software phantom is based on a high-resolution image from a healthy volunteer. Lesions were simulated and projected onto this image (left). Thus, the volume of the simulated lesions is known exactly (ground truth). We used such data to evaluate accuracy and reproducibility of two different volumetry methods (cf. Fig. 6.10). (Figure courtesy of JAN REXILIUS, MeVis, Bremen)

such a technique depending on the relative amount of partial volume voxels. The graph of object overestimation by hyperintense voxels shown in Figure 6.7 right can serve as an upper limit, e. g. as much as 100 % overestimation corresponds to a voxel diagonal as small as a fifth of the smallest object diameter. This corresponds to a voxel-object ratio of 1:5 (i. e. 20 %), while it is important to note that most focal white matter lesion images exhibit a higher voxel-object ratio than 20 %, resulting in an even higher percentage of partial volume voxels and thus potential overestimation.

For relative or follow-up examinations, volume overestimation would not be much of a problem, if only reproducible (cf. precision vs. accuracy, p. 92 and Fig. 6.3). It turns out, however, that manual tracing of small lesions on repeated image acquisitions is very unlikely to be reproducible, as will be discussed below.

In order to investigate these effects, REXILIUS *et al.* [2003] has implemented a software phantom for the simulation of focal white matter lesions with exactly known ground truth, *ground truth* referring both to the partial volume ratios for all voxels and to the lesion volume. The software phantom provides a realistic image model comprising:

- ▷ Three different focal lesion shapes
- ▷ Six different lesion sizes
- ▷ Sampling effects, i. e. partial volume averaging
- ▷ Variable lesion contrast
- ▷ Variable pixel noise level.

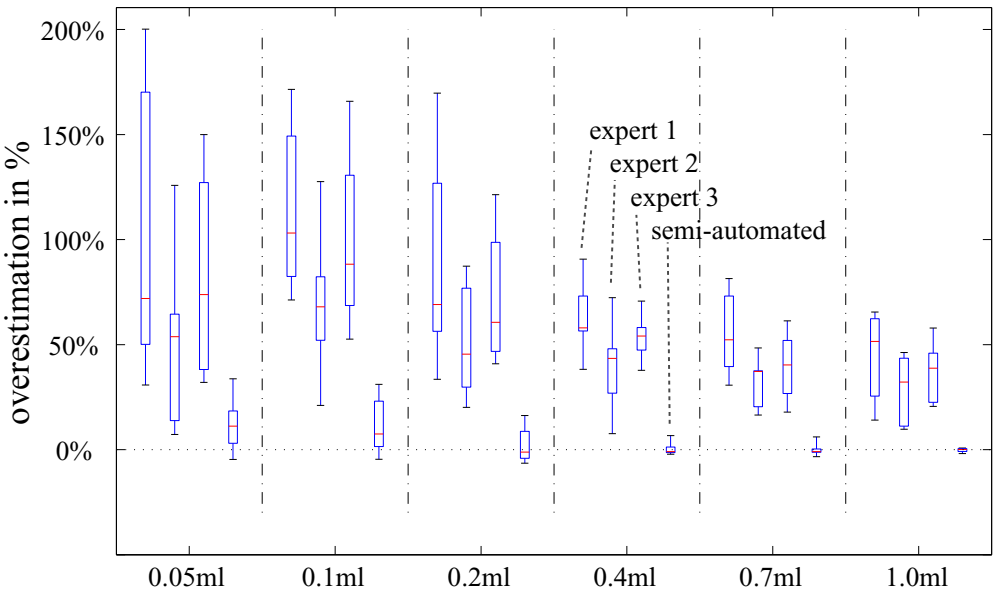


FIGURE 6.10: Comparison of partial volume modeling vs. expert manual volumetry based on realistic simulated MS lesion images (cf. Fig. 6.9 left) at six different volumes (x-axis). Each boxplot represents three independently repeated simulations of three lesion shapes, i. e. nine measurements. The y-axis shows the overestimation of the respective lesions in percent; e. g. 0 % means that the true volume (dotted horizontal line) and 100 % means that the double of true volume was measured. Each boxplot indicates 0, 25, 50, 75, and 100 percentiles. It is obvious that all three experts have systematically overestimated the true volume, and that their precision was much lower than by using the semiautomated partial volume modeling technique. (Figure courtesy of JAN REXILIUS, MeVis, Bremen)

In the same paper, an alternative volumetry technique is proposed based on the watershed transform and on a model-based regional histogram analysis; this technique is similar to the ones presented in the following chapters and is thus not detailed here. REXILIUS compared this semiautomatic partial volume analysis technique to manual outlining (and voxel counting) performed by three expert observers. The results are summarized by Figure 6.10. It is obvious that manual were worse than semiautomatic results in terms of accuracy, as well as precision. For four of six lesion sizes (0.2, 0.4, 0.7, and 1.0 ml), the worst semiautomated was even better than the best manual result.

It is instructive to separately consider the differences in terms accuracy on the one hand and precision on the other hand. Comparing the semiautomated to the manual results, precision ("how close are the results to each other for identical lesion size?") was improved by approximately a factor of four, while accuracy ("how close are the results to the ground truth?") was improved by more than a factor of ten on average. We will discuss in Section 6.7 the reasons for these differences, as well as their implications for other semiautomated volumetry techniques.

6.6 Volumetry of a Lung Nodule

One of the most important volumetry applications in clinical radiology is the quantification of tumor burden (cf. Table II.1, p. 87). The precision of such a measurement is crucial to decide early whether a tumor has grown or not. We have commented above on the current clinical practice where linear or bi-linear measurements are performed in combination with decision rules that are more than 10 years old. Modern computer-assisted volumetry has not yet been generally accepted for clinical tumor assessment. A valid method to measure the volume of a broad range of tumors at an accuracy of a few percent would have a major impact on clinical practice. In many cases, decision of tumor response or non-response to a medical treatment could be taken in a small fraction of current decision periods, i. e. a few weeks instead of several months (cf. comments on RECIST, p. 6.1).

Tumor volumetry is one of the most important applications of QIA, but also one of the most difficult. Problems arise from the fact that in general the tumor itself is not a homogenous structure, but rather comprises various compartments as, e. g., characterized by contrast enhancement or spectroscopic imaging. Furthermore, in many cases the image does not show the real tumor borders; the tumor extends beyond the visible borders as it is known, e. g., from infiltrating Gliomas. It is a currently incompletely solved problem and a challenge for image *acquisition* research to generate image contrasts that are sensitive to tumor infiltration. Conversely, image *analysis* can only aim to characterize and quantify what an image actually contains.

Being beyond the scope of this thesis to provide a general framework for accurate and precise volumetry of all classes of tumors, we want to comment on two specific cases. Hepatic metastases and lung nodules are among the most frequent found secondary tumor manifestations in both male and female patients. A standard non-contrast thin-slice CT scan of a round liver metastasis is shown in Color Plate C.8 where the lesion appears slightly hypointense at the mid-posterior tip of the right hepatic lobe. How difficult it is to segment this rather simple object can be appreciated by comparing the histograms of image samples inside (red) and outside (green) of the lesion (*region of interest*): The two distributions are heavily overlapping such that direct gray value-based segmentation methods will likely fail. With other words, the SNR is insufficient for a reliable segmentation without using prior knowledge, e. g. on the lesion shape. In such cases, we propose to trade SNR for spatial resolution and to successively use non-linear filtering techniques to result in a preprocessed image with sufficient SNR. This is detailed for two alternative preprocessing pipelines in Color Plate C.8.

Lung nodules[5] are one of the rare cases where one can hope to measure real tumor volumes due to the largely homogeneous objects and the high SNR. Here, we want to investigate the performance of a histogram-based quantification method on a single exemplary case (cf. Color Plate C.9). A lung nodule was imaged twice, base line-scan and follow-up fifteen weeks later. Both sets of images were acquired on a LightSpeed Plus, GE Healthcare at 1.25 mm slice thickness and 0.5 mm reconstruction interval. We use an image

[5]The Lung Image Database Consortium (LIDC) as an NIC/NIH funded initiative currently sets up a large database containing lung CT image data, clinical findings, and image analysis results.

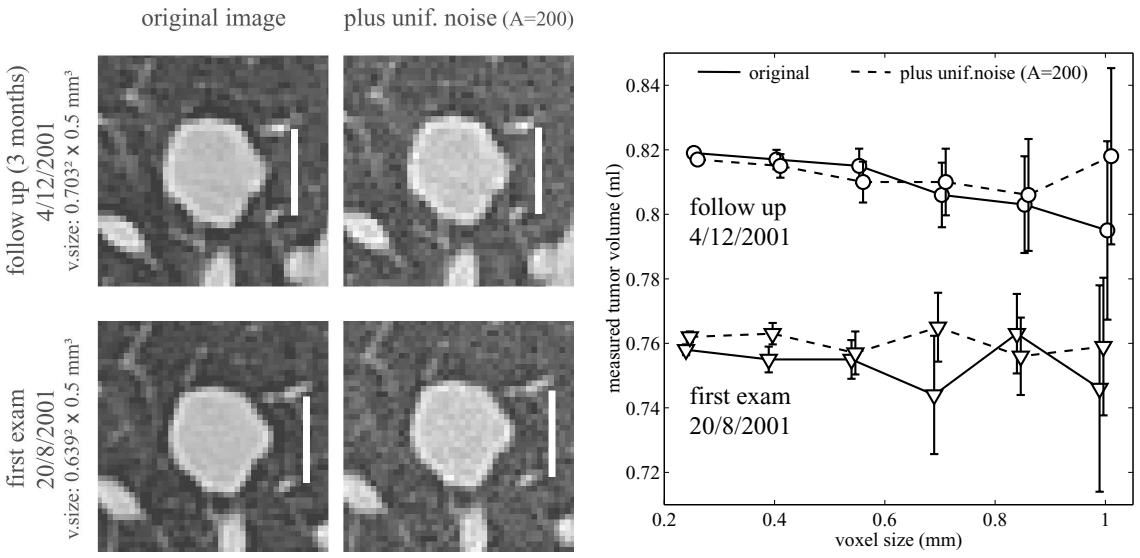


FIGURE 6.11: For two CT data sets, initial exam and follow-up of a lung nodule (courtesy of JON WIENER, Boca Raton), a series of volumes is computed, with and without initially added noise. For each of the four images (left; white bars indicate 10.0 mm), the series are generated by resampling at an isotropic resolution (ordinate of plot, right); volume computation is independently performed on all 24 images using IWT plus subsequent automatic histogram analysis. For each of the measurements, an error estimation is performed as part of histogram analysis (error bars).

analysis methodology identical to the one used above for semiautomatic MS lesion volumetry and also for cerebral ventricular volumetry in the next chapter where more details are provided. The nodule was attached to an intermediate and two smaller vessels, which were removed by adding exclude markers during IWT segmentation (cf. Color Plate C.9 b).

In order to test robustness and validity of the volumetric results, we have resampled both cases, original examination and follow-up after three months, at six different isotropic resolutions using a three-lobed LANCZOS filter (cf. footnote on p. 105). Note that the respective voxel sizes of 0.25, 0.40, 0.55, 0.70, 0.85, and 1.00 mm roughly correspond to the range used to generate the four examples shown in Figure 6.7 bottom and that also the object size is comparable to the above simulated sphere (compare Figs. 6.7 and 6.11 left). Furthermore, we have added uniform noise to both images with an amplitude of 200 HU before resampling, thus resulting in a total of four series (with and without noise; original and follow-up) of each six images (resolution).

The results from all 24 independently performed volume measurements are shown in Figure 6.11 right. All measured volumes from the follow-up scan are clearly larger than from the baseline scan. The error bars indicate the error estimation performed as part of histogram analysis, while the estimation is based on object and background noise, and on the relative amount of partial volume effects (cf. p. 124).

6.7 Discussion of Examples

Pituitary Gland

Each of the three above examples approaches the problems of small-object volumetry from a different viewpoint. The hypophysis illustrates how difficult it is to segment an object of variable shape and size when the model for the image contents is missing or incomplete. With other words:

- ▷ The hypophysis' neighboring tissue is not uniform.
- ▷ The hypophysis itself including the fatty sellar tissue at its posterior border is inhomogeneous.
- ▷ Its boundary, e. g. laterally, cannot be completely discerned in the images.

For image analysis, this implies that a valid histogram model for the object boundary can hardly be found, and that either a deformable template or user interaction are mandatory. We have experimented with a set of deformable segmentation models, but without success, most probably due to the thin and highly variable shape of the pituitary gland, as well as to its inhomogeneous border and neighboring structures. With the IWT applied to the gradient image we have implemented a method taking edge information plus user interaction into account. The interaction is such that a single mouse click placing an include marker inside of the object[6] can already lead to a valid segmentation, and that successively placed markers can be used to refine the result. In this sense, we can speak of minimal interaction, as introduced on page 44. Compared to manual tracing, this approach leads to much lower interaction times and to a higher reproducibility, with the drawback that atomic basins identified by the IWT could not be separated by the user.

Comparing the results from two volunteers, it is interesting that inter-examination variance was not lower for intra-day repeated acquisition (Table 6.2) than for a series of acquisitions on different days on another volunteer (Table 6.1). This could be explained by differences in image quality or by individual anatomy. From our experience with many volunteer images, the latter is most probably the case; in some instances, a definition of the lateral hypophysis boundaries was quite difficult using our method due to close neighborhood of iso-intense tissue. Nonetheless, even in those cases, the inter-examination variance was within an acceptable range (SD $\approx$ 0.1 ml).

Considering the results at decreasing resolution, i. e. increasing axial slice thickness, a worsening is found at 3.0 mm and 4.8 mm compared to 1.0 mm and 1.8 mm. For higher slice thickness, a systematic but non-significant underestimation was observed (cf. Table 6.2, bottom lines). The stable results at 1.8 mm are remarkable since downsampling by almost a factor of two was performed along the shortest object axis; in cranio-caudal direction the pituitary gland usually measures only five millimeters or less (cf. Fig. 6.8).

[6] Assuming that a preselected ROI fully contains the object to be segmented, and thus the ROI border never touches the object, we can fill the ROI border automatically with exclude markers, such that a single include marker can a produce meaningful segmentation.

Manual Lesion Tracing

The software phantom study conducted by Rexilius has revealed major shortcomings of manual tracing in particular and of volumetry that is based on voxel counting in general. We have to question whether the high variability of manual expert results is solely due to subjective or to objective reasons. *Subjective* refers to human unreliability or variability, while *objective* refers to more fundamental reasons that are related to image acquisition. The answer is that both classes of reasons play a major role here.

One subjective reason refers to the fact that different raters tend to mark lesions differently—they will cover the partial volume region more or less completely. A further subjective reason is that it is virtually impossible to perfectly reproduce a manual tracing, such that repeated measurements, even on the same images, will never be the same. Another subjective reason has already been mentioned above, i. e. that manual tracing systematically leads to heavily overestimated volumes, due to the fact that human raters tend to mark most of abnormally appearing (e. g. hyperintense) voxels. But even an expert, who tried to compensate for partial volume effects during manual tracing (expert 2), did overestimate the true lesion volume in all instances.

The major objective reason is related to the fact that repeated scans[7] will never produce exactly the same images. A patient positioning difference of only one millimeter will largely redistribute the partial volume voxels, such that more or less abnormally appearing voxels and more or less pure tissue voxels exist, leading to different volume measurements by manual tracing. Further objective aspects are related to image noise etc.

Using a semiautomatic voxel counting technique such as seeded region growing in a reproducible manner, one could argue that the precision on the same images could be very high, yielding similar or same results in a repeated analysis. Such a technique seems to be able to overcome all of the above listed subjective reasons for variability. Its usage could yield perfect intra- and inter-observer characteristics. Applied to repeated acquisitions of the same patient, however, even if performed on the same day and scanner, such a technique is likely to produce variable results in a similar fashion as manual tracing due to the objective aspects still remain. In cases of problematic image quality it is even possible that manual tracing performs better than region growing or other auto-contouring methods, if not properly designed.

From the above, we want to question the value of intra- and inter-observer studies in order to prove the quality of a quantitative image analysis technique. Even if necessary preconditions, such studies often found in medical research and conducted on the same images are not sufficient to prove the value of a methodology. If not at least the effects of an inter-examination experiment is known, we do not know anything about the practical value of a quantitative method. We thus argue for not accepting a quantitative image analysis tool for clinical use before its inter-examination reliability is known. We further plead for inter-protocol (resolution etc.) and inter-scanner tests in order to be prepared for reliability in longitudinal studies even if the scanner or protocol was changed in the course of the study.

[7] For the phantom study, positioning differences were simulated by axially shifting the lesion position.

It is instructive to compare the theoretical sphere example in Figure 6.7 to the manual tracing results. For the MS lesion phantom with a slice thickness of 3 mm and an in-plane voxel size of approximately 0.5 mm, we have a voxel diagonal of approximately 3.1 mm. Assuming spherical object shape, for the six phantom volumes of 0.05 to 1.0 ml, we have object diameters[8] of 4.6, 5.8, 7.3, 9.1, 11.0, and 12.4 mm, resulting in a voxel-object ratio of 67, 53, 42, 34, 28, and 25 %, respectively.

From these numbers and from the right plot in Figure 6.7 we can estimate the maximum volume overestimation[9] to approximately 300, 240, 190, 155, 125, and 110 %, respectively. A comparison to the results from manual tracing in Figure 6.10 yields that:

- ▷ These maximum overestimation values are larger than the actual results in all cases and could therefore be used as practical upper limit.
- ▷ An overestimation of 50 % or more of these values does frequently occur.

Lung Nodule

For volumetric tumor follow-up, sensitivity and precision of the employed volumetry method are crucial. From the above lung nodule example, we have learned that:

- ▷ Significant systematic errors due to limited resolution or pixel noise can be avoided by adaptive histogram analysis including an explicit model for partial volume effects.
- ▷ Within such a histogram analysis, it is possible to perform an error estimation based on the individual image characteristics.
- ▷ Attached vessels can be efficiently removed from the tumor using the IWT.
- ▷ Feasibility and stability of the results can be improved by resampling to a finer grid.
- ▷ A voxel-object ratio of 10 % already induces a significant measurement uncertainty using the given method.

The reliability of the computed results is supported by the fact that all twelve results for each of the two original images are mostly consistent within the estimated error bounds. The variability between different resolutions is mostly due to uncertainty in separating the attached vessels from the nodule. This could be improved by employing a geometric or morphological model for this separation step, as for example proposed by Kuhnigk *et al.* [2004]. Further uncertainty arises from the limited resolution in the resampled images, especially for larger voxel size.

It is interesting to note that results across resolution are slightly more consistent for the image series with additional noise than without (cf. Fig. 6.11 right). This is probably due to the fact that the separation of nodule and adjacent vessel is easier in a more structured or noisy image, since more local image minima are present in this situation and the WT associates each minimum with an atomic basin. Otherwise stated, the number of local image minima controls the level of detail accessible with the WT.

[8] Computed by the basic equation $D_{\text{sph}} = (\frac{6 \cdot V}{\pi})^{\frac{1}{3}}$.

[9] The object overestimation by hyperintense voxels is related to the voxel-object ratio through a factor of ≈ 4.5 (cf. Fig. 6.7 right).

Note that by visual inspection of the slices of largest cross-sectional area (cf. Fig. 6.11 left) it is impossible to tell a volume difference between baseline and follow-up scans. From what we know about the geometric accuracy of modern CT images, however, we can be almost certain to have measured a real effect of a slightly progressive disease in the provided example, even if small. From the results (cf. Fig. 6.11 right), we estimate the tumor growth within the interval of fifteen weeks to approximately $+0.05 \pm 0.01$ ml, corresponding to 6 % of tumor volume.

Further studies are planned to investigate the performance of the presented volumetry method for a larger sample of lung nodules and other focal lesions, as well as in comparison to other state-of-the-art techniques. Therein, it will also be important to systematically modify all relevant image parameters such as noise, nonuniformity, etc. This type of study is described for the problem of MR-based whole brain volumetry in Chapter 8.

Publications. *This chapter contains previously unpublished work excepting the following. The MS Lesion Phantom was implemented by* Jan Rexilius *and presented at MICCAI 2003, Toronto* [Rexilius *et al.*, 2003]. *The manuscript has recently been revised and extended for a journal publication* [Rexilius *et al.*, 2005]. *The project "MR Volumetry of the Normal Pituitary Gland" was presented by* Joachim Böttcher, *Jena, at ECR 2003, Vienna, and is currently being prepared for publication. The glass ball phantom to test inter-scanner geometric reliability was built by* Volker Diehl, *Bremen, and presented at DRK (German Radiological Society Meeting) 2002, Wiesbaden. Figures on theoretical voxel partitioning distributions of uniform cuts are from the Diploma Thesis of* Benoît Jolly, *Vienne, France, for which he has worked with the author for six months in 2004* [Jolly, 2004].

Pain is useless to the pained.

—Galen of Pergamum, 129-c.216

CHAPTER 7

Volumetry of Cerebrospinal Fluid

Abstract. *We present a new semiautomatic approach to segment, quantify, and visualize the intra- and extracerebral fluid spaces. The method is unique in combining three key features that are indispensable for routine clinical use: (i) Interaction times are less than three minutes for the analysis of more than 100 slices. (ii) The segmentation robustly yields anatomically correct results in the presence of image artifacts. (iii) Inter- and intra-observer, as well as scan-rescan characteristics show coefficients of variation well below two percent on average for the lateral ventricles. The method is based on thin-slice T1 weighted magnetic resonance (MR) imaging as available from most current scanners, the marker-based Interactive Watershed Transform (IWT), and automatic histogram analysis. Furthermore, an ultra-fast method is discussed for the volumetry of intracranial, i. e. ventricular plus subarachnoid fluid using T2 weighted MR projection imaging.*

Evaluation was performed on patients, healthy volunteers, and phantoms. In particular, we evaluated the new methodology for ventricular volumetry on 34 clinical volumetric MR data sets from non-sedated children (age 6–8 y) with a history of prematurity and low birth weight (< 1500 g) obtained during a prospective study. The methodology, with adaptation to small ventricular size, was capable of evaluating all 34 of the pediatric data sets for cerebral ventricular volume. The method was robust for normal and pathological anatomy, reproducible, fast, and equally applicable to children and adults.

THE STUDY of cerebral ventricles has occupied the medical community for more than 2000 years. To those who first described the brain at the time of HIPPOCRATES in 300 B.C., its most interesting features were the small, fluid-filled spaces at its center. Five hundred years later, GALEN OF PERGAMUM (129–c.216) wrote that it was in these ventricles that the animal spirits and the rational soul reside. GALEN distinguished three ventricles: One at the front of the brain, divided in two, one in the center, and one at the rear, of which a lesion had the most disturbing effects [CHANGEUX, 1985]. The Greek and, later, the Arabic traditions continued to assign brain function to the ventricles for some 1,300 years following GALEN. LEONARDO DA VINCI (1452–1519) was credited with making the first wax cast of the cerebral ventricles in the 15th century, reported in his Quaderni d'Anatomia [according to GRANT *et al.*, 1987].

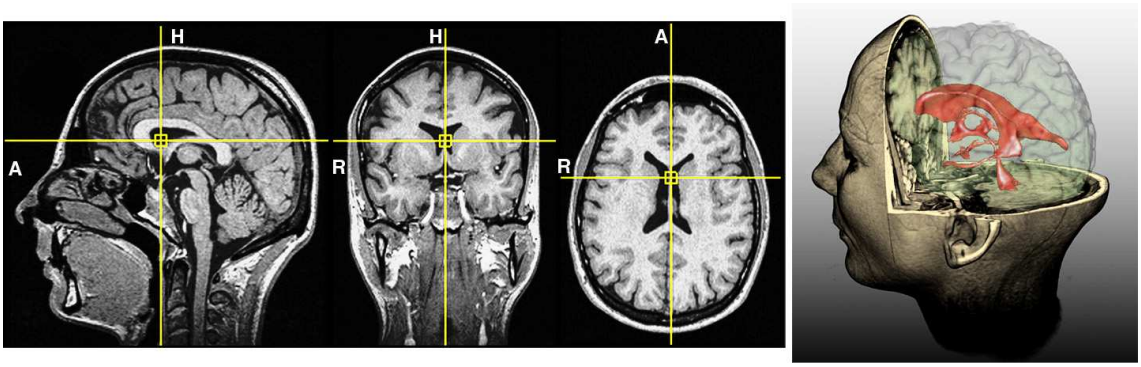


FIGURE 7.1: Healthy volunteer (35 yo woman). *left*: Three orthogonal views of isotropic T1 weighted MRI data, slice thickness 1.0 mm. *right*: Hybrid rendering of ventricular segmentation result fused with brain segmentation and clipped original data.

As belief in animal spirits died, so too did the ventricular hypothesis. However, the interest in the cerebral ventricles remained. The ability to quickly and accurately determine cerebral ventricular volume has significant diagnostic potential in the fields of neurology and neurosurgery. Possible uses include the evaluation and monitoring of brain injury (e. g., neonatal White Matter Damage (WMD), neurodegenerative disease) and the assessment of neurosurgical procedural outcomes, e. g., in patients with shunted hydrocephalus. Neonatal WMD, also referred to as Periventricular Leukomalacia (PVL), is reported to be the neuropathological basis of spastic cerebral palsy, cognitive impairment, and occasionally seizure disorder in older children [INDER *et al.*, 1999, MELHEM *et al.*, 2000]. Premature low birth weight ($\leq$ 1500 g) infants, who are at greatest risk for developing PVL [INDER *et al.*, 1999, PETERSON *et al.*, 2000], frequently undergo measurable changes in ventricular volume.

7.1 Background and Motivation

Gross injury to the neonatal periventricular white matter is commonly imaged using head ultrasound, a convenient imaging tool in the neonatal intensive care unit. By such, WMD and intraventricular hemorrhage (IVH) can be determined. Though, greater anatomical resolution and superior sensitivity to cerebral soft tissue injury is obtained through the use of MRI. Routine MRI can demonstrate abnormal signal intensity and volume loss in many premature infants with PVL, both at the time of injury and years later in adolescence [INDER *et al.*, 1999]. Late manifestations of PVL include white matter volume loss, gliosis, and cavitations with secondary enlargement of the lateral ventricles and thinning of the corpus callosum [MELHEM *et al.*, 2000]. Evaluation of these MR images for PVL by pediatric neuroradiologists and neurologists is generally a subjective process, based on years of experience.

Related Work

Quantitative measurements on CSF spaces started five decades ago. Early estimates of cerebral ventricular volumes were derived from post-mortem anatomical cast studies in the 1950s [according to Grant *et al.*, 1987]. By 1970, various linear measures have been used, but these often did not correlate well with changes in volume [Bull, 1961, Pentlow *et al.*, 1978, van der Knaap *et al.*, 1992, Kikinis *et al.*, 1992]. In the 1970s, mathematical models of the cerebrospinal fluid system have been developed to help clarifying the intracranial fluid kinetics [Marmarou *et al.*, 1978]. With the advent of CT, considerable interest in quantifying cerebral ventricular size shifted to this new imaging modality.

Through a series of publications in 1978 [Pentlow *et al.*, 1978, Thaler *et al.*, 1978, Sager *et al.*, 1978, Brassow and Baumann, 1978, Hacker and Artmann, 1978], development of modern volumetric methods based on tomographic data began. Partial volume averaging was identified as one of the central problems [Pentlow *et al.*, 1978, Thaler *et al.*, 1978]. Slice-based thresholding with voxel counting [Brassow and Baumann, 1978], bimodal histogram analysis [Hacker and Artmann, 1978], and linear interpolation have been the common techniques [Pentlow *et al.*, 1978, Sager *et al.*, 1978]. First steps toward a statistical model for CSF volumetry were done by Thaler *et al.* [1978], using compound Dirichlet-Gaussian probability distributions. First attempts to automatically identify the cerebral ventricular areas and calculate the ventricular volume were described as early as 1980, using algorithms of pattern recognition based on an anatomical model [Bull, 1961].

In the late 1980s, first MRI-based evaluations of CSF volumes were reported. In 1987, an elegant method was proposed by Grant *et al.* [1987] to derive separate measurements for total intracranial and ventricular CSF volumes from two MR projection images. Whereas overall volumes can be reliably estimated by such a method, systematic errors occur for the ventricular CSF spaces due to anatomical misinterpretation of periventricular CSF spaces. Furthermore, contra-lateral differences are not manageable directly. This technique was used by Teasdale *et al.* [1988], who showed the dependency of intracranial CSF volumes on inhalation of CO_2, hyperventilation and on the menstrual cycle. However, we did not find evidence that these results have been reproduced by other researchers.

MRI-based approaches can be divided in two groups, i. e. based on either single- or multispectral image acquisition. Kohn *et al.* [1991], Kikinis *et al.* [1992] and Friedlinger *et al.* [1999] acquired bispectral images as the basis for a two-dimensional histogram analysis. Although with higher acquisition costs, these approaches yielded results comparable to single-sequence methods. There, T2 [van der Knaap *et al.*, 1992, Tsunoda *et al.*, 2000] and T1 weighted images [Johnson *et al.*, 1993, Saeed *et al.*, 1998, Worth *et al.*, 1998] were acquired, mostly processed by manual region of interest (ROI) tracing. DeCarli *et al.* [1992] performed a model-based histogram analysis very similar to ours. However, logarithmic normal distributions were used that we did not find to account well for PVE.

Clinical Application

Due to rapid advances in computer hardware and software technology over the last decade, computer-based morphometric analysis of the adult and pediatric brain has become pos-

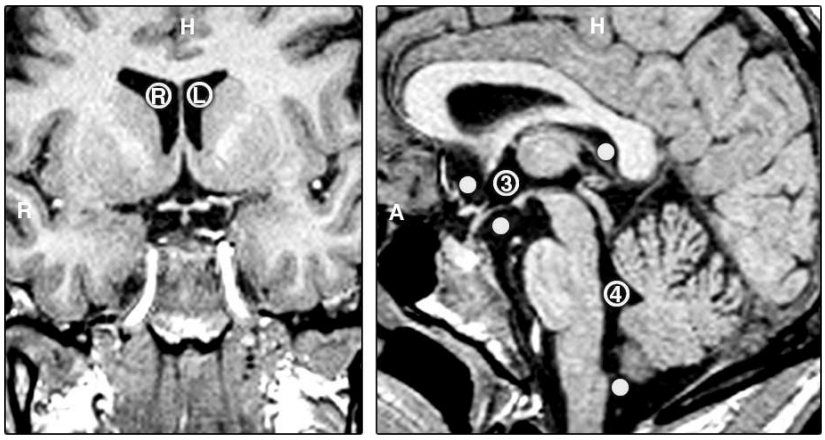


FIGURE 7.2: Typical marker positions for segmentation of all four ventricles. Five different markers are available for ventricle labeling (R, L, 3, and 4) and region exclusion (dots). The images represent a VOI interactively selected from the data depicted in Figure 7.1.

sible. Since the first ventricular measurements in adults based on MRI in 1986 [CONDON *et al.*, 1986], many techniques were developed based on acquiring large thin section data sets of the brain, and analyzing abnormalities in total and regional anatomical volume [HUPPI *et al.*, 1998]. Previously unrecognized changes in the brain structure of premature infants have been detected, including reduced volumes of cerebral cortical gray matter, basal ganglia, amygdala, and hippocampus [INDER *et al.*, 1999, PETERSON *et al.*, 2000].

Generally, these methods are labor-intensive and require an experienced team to process and analyze the morphometric data on dedicated workstations with expensive or proprietary software packages [INDER *et al.*, 1999, MELHEM *et al.*, 2000, PETERSON *et al.*, 2000]. Most of the techniques are not practical in a clinical time frame, where medical assessment and decision-making processes are frequently expected by referring clinicians within 24 hours or less. Manual or two-dimensional methods, such as described by TSUNODA *et al.* [2000] and SCHIERLITZ *et al.* [2001], often result in long interaction times or poor reproducibility, while fully automated methods do not perform well where anatomical or pathological deformation is too large. Moreover, normal pediatric cerebral anatomy shows narrow lateral ventricles of 5–10 cm length, often containing roughly 1 ml of fluid. For the volumetry of small or elongated objects, systematic errors are frequently related to an erroneous or biased processing of partial volume voxels. PENTLOW *et al.* [1978] were the first to describe the importance of partial volume effects with respect to ventricular volumetry from computed tomography.

Goals

The objective of this chapter is to efficiently segment and visualize the intra- and extracerebral fluid spaces based on MRI and, most importantly, to reproducibly quantify their volumes. No system known to the author is available for routine clinical use, due to the anatomical complexity of both the cerebral ventricular system and the subarachnoid

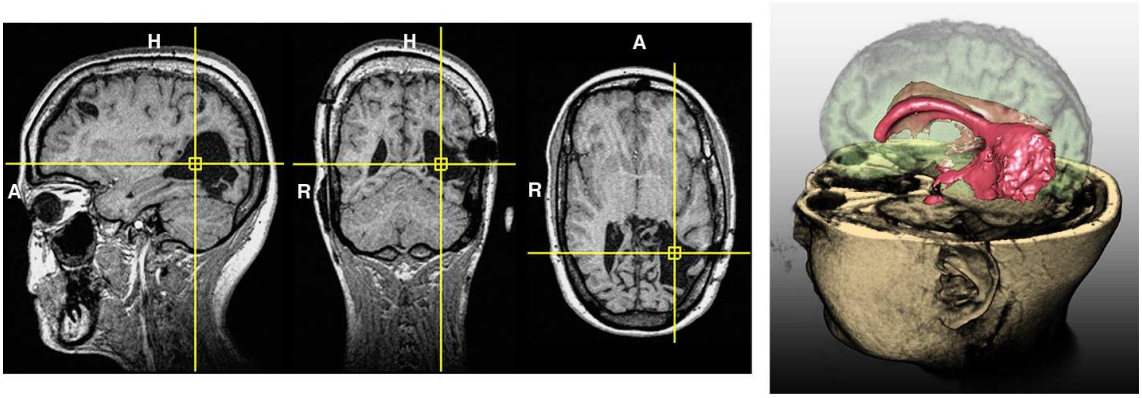


FIGURE 7.3: Patient (29 yo woman), postoperative data, total ventricular volume: 54.6 (0.6) ml.

spaces, as well as the lack of suitable and fast segmentation procedures. In order to be applicable in a clinical environment, developed methods should be

(a) *Fast*—requiring less than ten minutes for image analysis,

(b) *Flexible*—working robustly for normal and abnormal anatomy, and

(c) *Reproducible*—with less than, e. g., five percent variation.

We describe an objective methodology that incorporates interactive 3D segmentation based on a modified morphological watershed transform (cf. Ch. 4, [HAHN and PEITGEN, 2000, 2003]) and a fully automated volumetric histogram analysis [HAHN *et al.*, 2004b], with application to routinely acquired T1 weighted volumetric brain MR image data sets, e. g. from young children with PVL. This novel method is evaluated with respect to its robustness, reproducibility, and speed, conditions that are necessary for medical relevance in the clinical setting.

7.2 Segmentation of the Cerebral Ventricles

The proposed volumetry method can be divided into three major successive steps. Figure 7.4 shows a systematic overview of the image processing and analysis pipeline. More generally speaking, the three major steps are:

(a) Image acquisition with maximum contrast between the object and adjacent structures

(b) Fast and intuitive image segmentation

(c) Histogram-based volume assessment.

Minor steps are volume of interest (VOI) selection and resampling.

Isotropic 3D data for Figures 7.1 and 7.3 were acquired on a 1.5 T MAGNETOM VISION (SIEMENS MEDICAL SOLUTIONS): Flash 3D, T1 weighted, TR 9.7 ms, TE 4.0 ms, flip angle

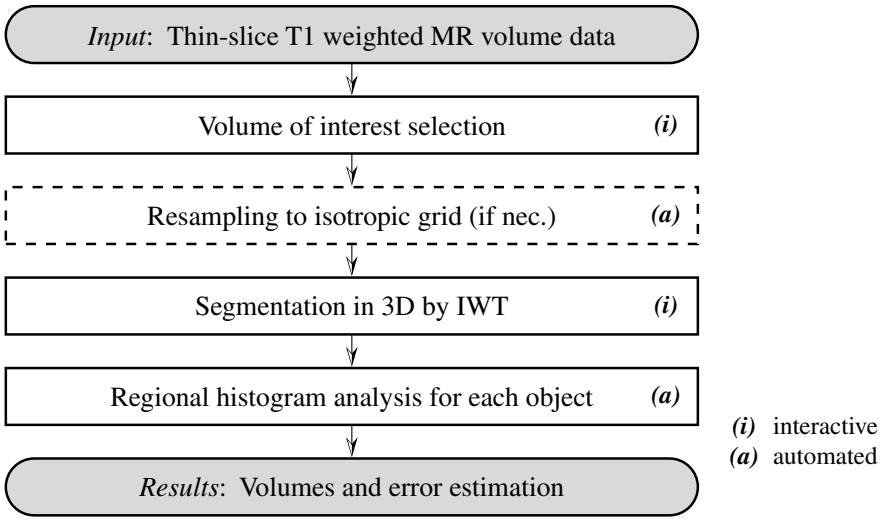


FIGURE 7.4: Schematic presentation of the semiautomated image analysis pipeline as proposed for cerebral ventricular volumetry. Two interactive and one or two automated steps are required to quantify ventricular volumes from the MR data (cf. body text; for details on the IWT, cf. Ch. 4).

20°, sagittal, FOV 256 mm, matrix 256×256, 160 slices, slice thickness 1.0 mm, acquisition time approximately eight minutes (e. g. Fig. 7.1). The thin slices provide the basis for fast 3D segmentation procedures and for minimizing partial volume effects. However, we have found that the described volumetric measurements are quite stable up to a slice thickness of 3 mm at arbitrary slice orientation (cf. Color Plate C.11).

This broadly available MR technique is combined with novel image postprocessing and quantification methods. A 3D semiautomatic modified watershed algorithm has proven to be useful for removing extracerebral tissue from T1 weighted head MRI data (cf. Ch. 2). For ventricular segmentation, the algorithm has been extended by an unrestricted number of point markers that can be interactively located to accurately define the ventricular anatomy (cf. Fig. 7.3). After specifying a volume of interest (VOI, compare Figures 7.1 and 7.3), a few markers that are defined by a single mouse click serve as initial information for the segmentation algorithm that automatically tracks the ventricular borders in 3D (cf. Color Plate C.10).

The four cerebral ventricles are segmented simultaneously by using four different markers and colors (cf. Fig. 7.2 and C.10). A 2D color overlay and an interactive 3D volume rendering serve as control tools and are updated on demand. Processing times are less than one second on a Pentium III, 1.6 GHz, for a typical region of interest (1 million voxels). In normal cases, the complete semiautomatic segmentation procedure including user interaction takes two minutes on average for all 100–200 slices. On page 125, we report on some modifications required for difficult cases.

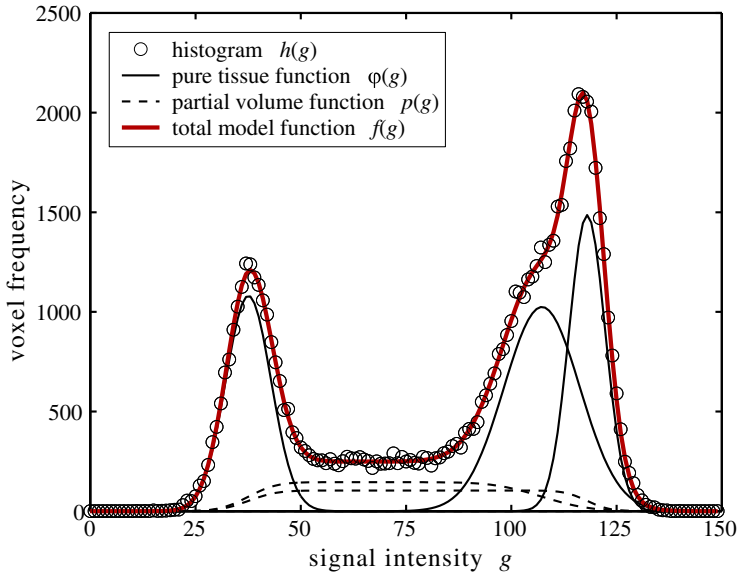


FIGURE 7.5: Model-based histogram analysis for automatic quantification of CSF volumes (cf. Fig. 7.1). The model consists of three independent Gaussians (solid lines, three parameters each: A_α, μ_α, and σ_α) and the two relevant corresponding partial volume functions (dashed lines, one additional parameter: $A_{\alpha\beta}$, cf. Eq. 7.1).

7.3 Volumetric Histogram Analysis

A final entirely automatic step robustly accounts for image noise, nonuniformity, and partial volume averaging effects that occur at object borders (cf. Color Plate C.11 for resolution effects). The volume of the four segmented ventricles is computed from the respective regional image histograms within less than one second. Similar to other researchers [THALER *et al.*, 1978, DECARLI *et al.*, 1992, WORTH *et al.*, 1998] we modeled these histograms as a trimodal Gaussian distribution $f(x)$ (cf. Fig. 7.5). However, we introduced the new terms $p^{\mathrm{pv}}_{\alpha\beta}(x)$ dedicated to model partial volume distributions, to which we refer to as *Mixed Gaussians*. Currently, only mixtures of two tissue types are considered.

Note that in MRI, Gaussians are only an appropriate model for a given tissue type with constant mean gray value and a given noise level for high SNR. Conversely, for low SNR, so-called Rician noise gives a better approximation to the data and avoids a biased model (cf. discussion of Rician noise on pp. 95 f.). In the cases discussed in this and the following chapter, though, we have found SNR high enough for the Gaussian approximation in all tissues of interest.

The histogram model corresponds to the assumptions formulated by Santago and Gage [SANTAGO and GAGE, 1993], i. e. symmetric and uniform PVE from each pair of two tissues. The symmetry assumption is approximately fulfilled as long as the imaged objects are significantly larger than the voxel size. Uniformity, however, is a theoretically imprecise assumption. Rather, the partial volume distributions are accentuated for asymmetric

partitioning. Both effects are discussed in Section 6.2. For the purposes discussed here, the uniform assumption has turned out to generate fair and stable results.

Note that our analysis is based on the fact that the watershed transform applied to the original T1 weighted MRI data (interpreted as altitude information) yields a region containing most of the partial-volume regions at the ventricular borders (interface of CSF with subcortical GM and WM) [Hahn *et al.*, 2001a, Hahn and Peitgen, 2000]. Using these assumptions for each pair α and β, we can approximate the corresponding probability densities (cf. Fig. 7.5, dashed lines) for the one-dimensional case by

$$p^{\mathrm{pv}}_{\alpha\beta}(x) \;=\; \frac{\Phi_\alpha(x) \;-\; \Phi_\beta(x)}{\mu_\beta \;-\; \mu_\alpha} \quad \text{and} \quad \Phi_\alpha(x) = \int_{x'=-\infty}^{x} \varphi_\alpha(x')\, dx' \;, \tag{7.1}$$

where x is a gray value and $\varphi_\alpha(x)$ is a normalized Gaussian centered at μ_α with width σ_α, corresponding to pure tissue α. Note that with given parameters for φ_α and φ_β, $p^{\mathrm{pv}}_{\alpha\beta}$ is non-parametric. With other words, for each Mixed Gaussian, only the amplitude $A_{\alpha\beta}$ is required as additional parameter. The complete model function

$$f(x) \;=\; \sum_{\alpha} A_\alpha \cdot \varphi_\alpha(x) \;+\; \sum_{\alpha,\beta \,|\, \mu_\alpha < \mu_\beta} A_{\alpha\beta} \cdot p^{\mathrm{pv}}_{\alpha\beta}(x) \tag{7.2}$$

is fitted to the regional histogram $h(x)$ by minimizing least-square deviations via a modified Levenberg-Marquardt method. Modifications were required to limit the parameter ranges in order to improve the robustness of the standard optimization strategy. For example, negative amplitudes were not allowed, and the order of peaks ($\mu_{\mathrm{csf}} < \mu_{\mathrm{gm}} < \mu_{\mathrm{wm}}$) was fixed. Good initial parameter values are derived from the two most prominent histogram peaks. Another option for the fitting would be to use a robust Simplex method, such as described by Kallrath and Linnell [1987].

Once the fit converges, quantifying the corresponding CSF volume is straightforward: Remaining differences between $h(x)$ and $f(x)$ are attributed to pure and mixed classes according to their probability values at x (multiplied with amplitudes A_α and $A_{\alpha\beta}$). For volume computation, the partial volume classes are added to equal parts to the corresponding tissue types. The histogram analysis has also been described by Hahn *et al.* [2004b].

Error Estimation

The volumetric uncertainty $\sigma_{V_{\mathrm{csf}}}$ is provided to the user for each ventricle, being estimated individually using error propagation. The term $\sigma_{V_{\mathrm{csf}}}$ is related to image noise, contrast, and resolution. In the regional histograms, image noise and contrast are reflected by peak widths and peak distance, respectively; image resolution directly relates to the occurrence of partial volume effects and is thus reflected by the height of the plateau between CSF and gray matter / white matter peaks. The same histogram analysis and error estimation has been used to produce Figure 6.11.

TABLE 7.1: Evaluation of total reproducibility on a healthy volunteer (M, 38 y) who underwent five separate MRI scans with varying head positions and 30 min rest between acquisitions (V_1–V_5); volumes in ml. †: V_6 acquired two months later; ‡: significantly enlarged left ventricle.

	V_1	V_2	V_3	V_4	V_5	mean ± SD	CV	V_6†
right	17.96	17.83	17.73	18.01	17.83	17.87 ± 0.11	0.63 %	17.84
left	14.37	14.26	14.06	14.37	14.23	14.26 ± 0.13	0.88 %	14.59‡
3rd	2.22	1.99	2.07	2.17	2.00	2.09 ± 0.10	4.80 %	2.01
4th	2.64	2.75	2.60	2.74	2.78	2.70 ± 0.08	2.82 %	2.84
total	37.16	36.95	36.42	37.25	36.98	36.95 ± 0.32	0.87 %	37.33
lateral	32.33	32.12	31.79	32.39	32.12	32.15 ± 0.24	0.74 %	32.48

Extension for Narrow Ventricles

The histogram analysis was extended for very narrow ventricles ($n = 15$ cases; cf. Figs. 7.9 a, 7.12 a–c) that typically occur in pediatric neuroimaging. In these images, partial volume effects play a more dominant role than we presumed when developing the methods for quantitative neuroimaging on adults [HAHN *et al.*, 2001a]. The extension of the histogram analysis automatically rejects the model fit as soon as the CSF volume drops below a certain threshold; it requests the user to specify at least ten anatomical positions that are completely surrounded by CSF. The mean gray value corresponding to these positions then defines the center of the Gaussian that represents CSF (μ_{csf}, cf. Fig. 7.8 a).

7.4 Ultra-Fast Volumetry of Intracranial CSF

Independent measurements are performed for intracranial and intracerebral fluid spaces. A hybrid approach was designed to combine the advantages of two different acquisition protocols. Figure 7.6 shows a systematic overview. All image data for the combined study was acquired on a MAGNETOM VISION PLUS (SIEMENS MEDICAL SOLUTIONS) at 1.5 T.

For volumetry of entire intracranial CSF volumes one projection data set is acquired [similar to GRANT *et al.*, 1987]: RARE, heavily T2 weighted, TR ∞, TE 1100 ms, flip angle 150°, bandwidth 156 MHz, sagittal, FOV 230 mm, matrix 256 × 240, 1 acquisition, slice thickness 160.0 mm, single-channel head coil, acquisition time 2.8 seconds. Line and point markers are used to interactively define the VOI (Fig. 7.7 a and b). After automatic background subtraction, the gray value is supposed to be directly proportional to the fluid volume represented by a certain image element. Additional noise and contributions from other tissues can be neglected due to its much shorter T2 values.

The sagittal T2 weighted projection data set provides the required information to reliably quantify intracranial CSF volumes and can be acquired within a few seconds. Reproducibility for intra- and extracerebral fluid volumes was evaluated by repeated independent measurements on the same patients and volunteers (Table 2). Image postprocessing was performed by two independent expert radiologists. No significant difference of measured

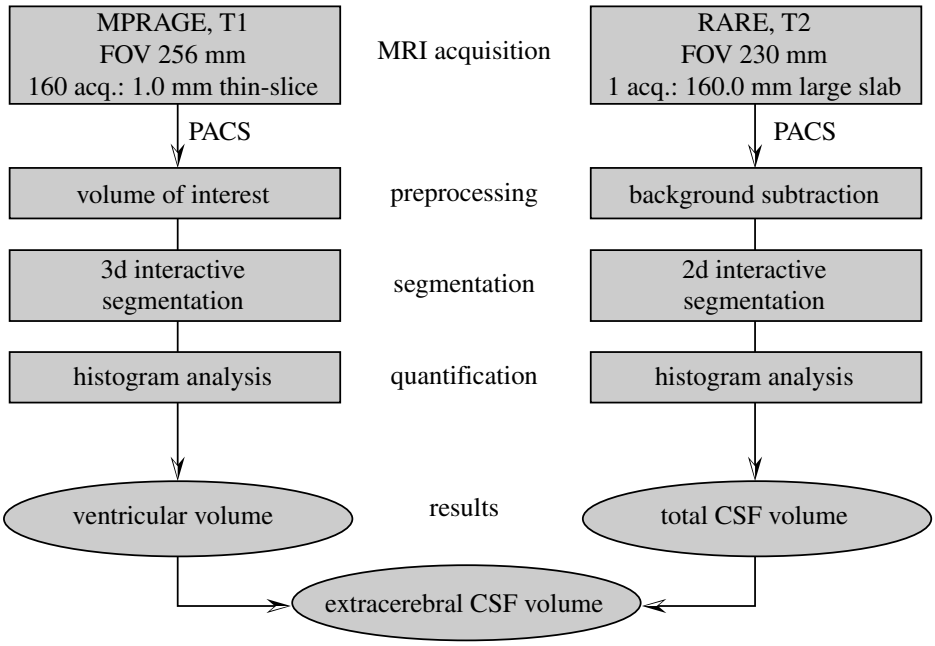


FIGURE 7.6: Dual image processing and analysis pipeline. Two independent data sets are acquired. Postprocessing is performed in three successive steps: (i) preprocessing to reduce the amount of data and to account for background noise, (ii) interactive segmentation to accurately define the anatomical structures of interest, and (iii) automatic histogram analysis.

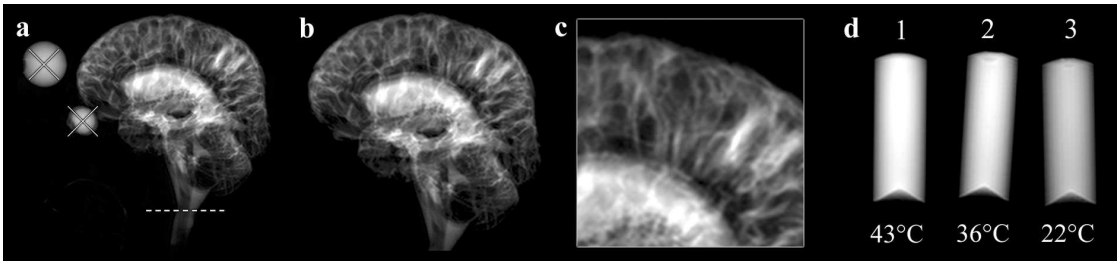


FIGURE 7.7: Volumetry of intracranial fluid volume. *a*: T2 weighted projection data with spherical phantom. Point and line markers are interactively used to define the ROI. *b*: ROI after background subtraction. *c*: Detail of b. *d*: Evaluation of temperature influence.

TABLE 7.2: Total intracranial CSF volumes: Evaluation of total reproducibility on one volunteer (28 yo man) that independently underwent MRI nine times (acquisition A_1–A_9), varying head positions, 20 sec rest between acquisitions; volumes and standard deviation (SD) in ml; cf. Fig. 7.7 a.

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	Mean (SD)	SD / mean
146.4	147.5	146.9	148.0	145.2	143.8	146.6	144.3	145.5	146.0 (1.4)	0.97 %

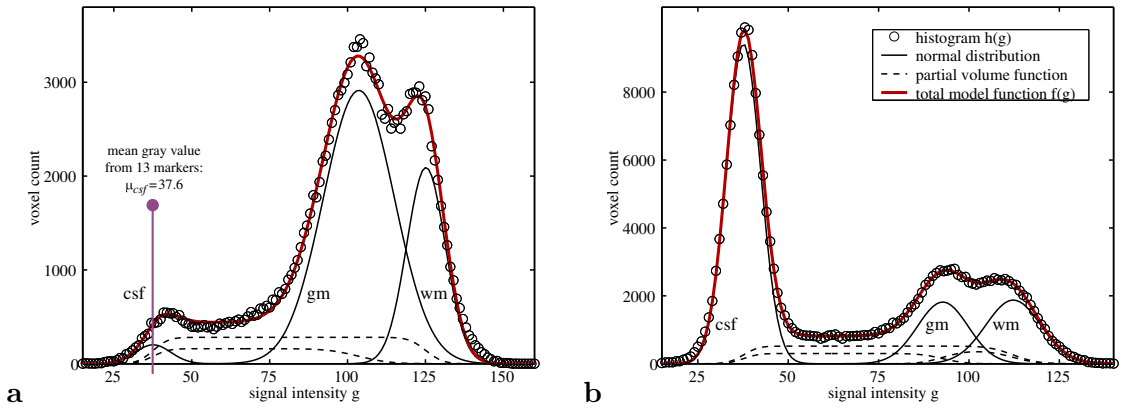


FIGURE 7.8: Model-based histogram analysis for automatic quantification of tissue volumes; for the over-inclusive segmentation resulting from the watershed transform, three tissue types are modeled using three Gaussians (solid lines) plus three Mixed Gaussians (broken lines, cf. body text). *a*: marker-based specification of μ_{csf} for histogram with sparse CSF representation (patient #255; histogram for all four ventricles). *b*: fully automated histogram analysis (patient #058; histogram for left ventricle). Compare these two extreme cases with Fig. 7.5.

cerebral fluid volume could be observed for varying head positions. Absolute calibration was performed by water phantoms (cf. Fig. 7.7 a). To our knowledge, water and CSF bear almost identical T1 relaxation times at identical temperatures ($\approx$ 3000 ms).

Previous results for the measurement of T1, as a means of temperature monitoring, have demonstrated a linear behavior with a coefficient of approximately 1.5 % per °C [DECARLI *et al.*, 1992, JOHNSON *et al.*, 1993]. Similarly, T2 relaxation time will increase with temperature, besides other parameters connected with temperature, e. g., chemical shift effects. In order to estimate the degree of influence, a series of 50 ml syringes filled with pure water at various temperatures was examined using the protocol described above. Resulting images demonstrated a significant increase of the signal by roughly 0.4 % per °C at body temperature (Fig. 7.7 d). Further studies are intended to accurately quantify this behavior in the range of 35 to 45 °C. Including calibration inaccuracy and flow artifacts, entire CSF volumes are quantified with variations of less than 4 %.

7.5 Application to Pediatric Neuroimaging

Subjects and Methods

Thirty-four children (mean age: 6.3 years, range: 6–7 years) with a history of prematurity (mean gestational age: 27.6 wks, range: 23.6–31.4 wks) and very low birth weight (mean BW: 1029 g, range: 527–1480 g) were evaluated with volumetric MR imaging. All had previously received extensive examinations at birth, including systematized ultrasonographic brain scans at postnatal days 1–3, 7–10, and 21. These children were a selected sub-sample from a cohort of 597 surviving children of very low birth weight ($\leq$ 1500 g) that were recruited by a large prospective epidemiologic study of the neurological, neuropsychological,

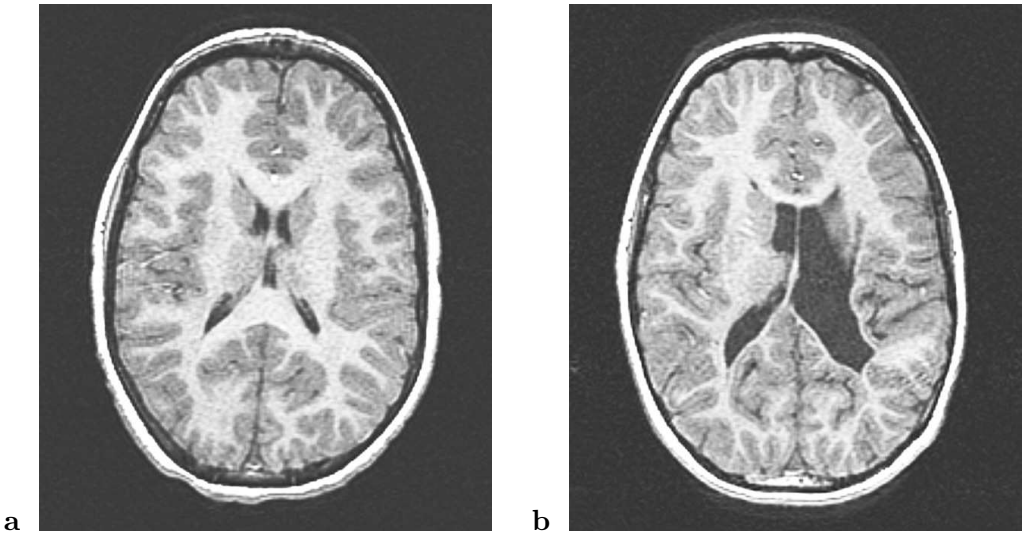


FIGURE 7.9: Original data for two preterm children at age six. *a*: patient #255; *b*: patient #058.

and neuroimaging outcomes at age of 6–7 years in order to examine the effects of early brain damage on selective aspects of cognitive development. The study hypothesized that cranial MRI obtained for this sub-sample would show reduced white matter volume and ventricular enlargement that would correlate with the presence and location of neonatal WMD as documented by ultrasound. A robust and reproducible methodology was sought to evaluate these MR data sets for volumetric changes in the ventricular system of the brain.

The proposed method to evaluate these MR volumetric brain data sets for changes in ventricular volume utilizes a combination of fast 3D marker-based segmentation and histogram analysis. The corresponding image analysis pipeline is presented schematically in Fig. 7.4.

T1 weighted anatomical images were obtained in the axial plane on a 1.5 Tesla SIGNA SP/I (GE HEALTHCARE) using a 3D SPGR (spoiled gradient recalled) sequence (1.5 mm thickness, 0.78 mm pixel spacing, no inter-slice gap, TR 34 ms, TE 5 ms, 256 256 matrix, 124 slices, acquisition time approximately 12 min). The acquired images were anonymized by the removal of DICOM (standard for Digital Imaging and Communications in Medicine) header information. All patient information including patient name, medical record number, and accession number were deleted and replaced with null fields. Each image data set was assigned a random number for correlation with the patient's clinical data set at a later date. Evaluation of all patient data was performed under a research protocol reviewed and approved by the Columbia University's institutional review board.

Results

Using the above extension for narrow ventricles, all pediatric brain ventricular systems including the smallest ventricles were reliably quantified in all 34 data sets. For 32 thereof,

TABLE 7.3: Results of ventricular volumetry for patient #255 (cf. Fig. 7.12 a–c); volumes for four objects (L, R, 3rd, and 4th ventricle), three observers (A–C), and three repeated measurements (V_1–V_3), as well as intra-observer mean, standard deviation (SD), and coefficient of variation (CV).

	Observer	V_1	V_2	V_3	Mean ± SD	(CV)
L	A	1.54 ml	1.62 ml	1.70 ml	1.62 ± 0.08 ml	(4.94 %)
	B	1.60 ml	1.72 ml	1.60 ml	1.64 ± 0.07 ml	(4.22 %)
	C	1.67 ml	2.04 ml	2.13 ml	1.95 ± 0.24 ml	(12.52 %)
R	A	1.53 ml	1.69 ml	1.76 ml	1.66 ± 0.12 ml	(7.10 %)
	B	1.67 ml	1.74 ml	1.70 ml	1.70 ± 0.04 ml	(2.06 %)
	C	1.77 ml	2.13 ml	2.19 ml	2.03 ± 0.23 ml	(11.19 %)
3rd	A	0.72 ml	0.69 ml	0.86 ml	0.76 ± 0.09 ml	(11.99 %)
	B	0.65 ml	0.67 ml	0.64 ml	0.65 ± 0.02 ml	(2.34 %)
	C	1.00 ml	0.67 ml	0.86 ml	0.84 ± 0.17 ml	(19.64 %)
4th	A	1.64 ml	1.62 ml	1.57 ml	1.61 ± 0.04 ml	(2.24 %)
	B	1.66 ml	1.66 ml	1.59 ml	1.64 ± 0.04 ml	(2.47 %)
	C	1.65 ml	1.53 ml	1.67 ml	1.62 ± 0.08 ml	(4.68 %)

clinical results have been available. Figure 7.10 shows an overview of the findings for subgroups. Whereas for GM and WM volumes only small relative inter-group differences were found (results not shown), significant differences were revealed for ventricular volumes when related to total brain volume (cf. Fig. 7.10 a and c). Furthermore, a significantly reduced IQ was observed for the WMD group (cf. Fig. 7.10 b). The increased IQ for the IVH group is not significant due to the small group size ($n = 2$).

We evaluated intra- and inter-observer variation for three experts when quantifying ventricular volume, and recorded mean values and standard deviations for all measurements (Tables 7.3 and 7.4). Intra-observer characteristics were computed by subtracting intra-observer mean values from all measurements, while inter-observer characteristics were computed from the intra-observer mean values.

Intra- and inter-observer coefficients of variation for medium and large lateral ventricular volumes were found to be less than 1 % (Table 7.6, patient #058). Larger coefficients of variation were observed for 3rd and 4th ventricles due to small object sizes and the frequently imprecise delineations in MRI. While the intra-observer standard deviations were comparable for small and medium lateral ventricular volumes, coefficients of variation were higher for small volumes (cf. Table 7.6). In one case, high inter-observer variations were recorded for the fourth ventricle due to a disagreement of the caudal limit of this object (Table 7.6, patient #058).

The original images were resampled to an isotropic grid of 0.7 mm size in order to provide sufficient information to the segmentation algorithm and histogram analysis on small ventricles. A finer grid did not change the results nor improve the stability of the method. Furthermore, we tested the above modification of the histogram analysis regarding systematic errors ($n = 19$ cases). Placing additional markers resulted in an

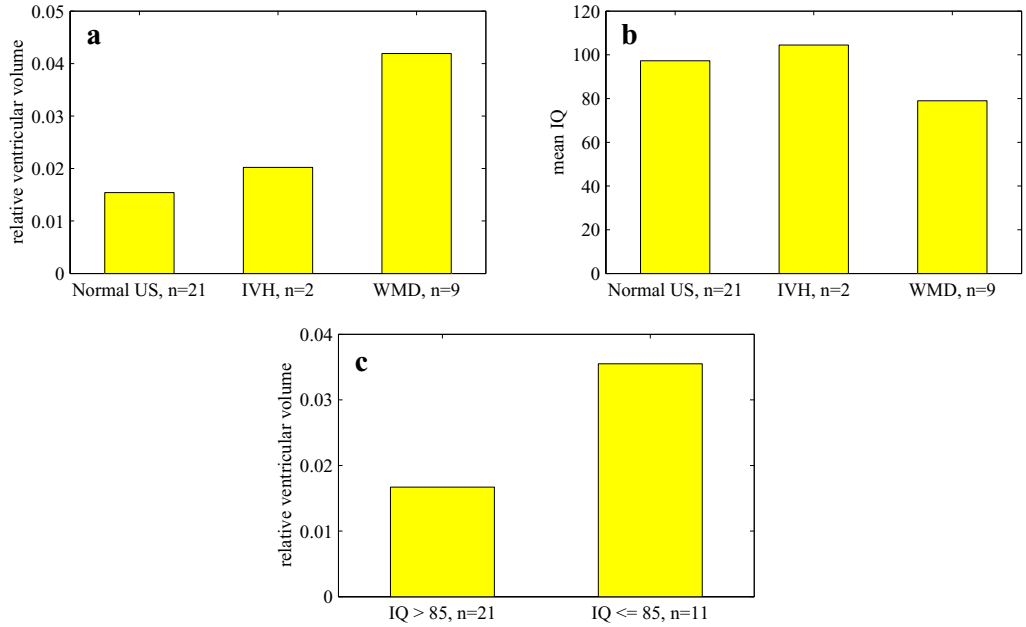


FIGURE 7.10: Group results for 32 patients at age 6—8. *a*: Ventricular volume as a proportion of total brain volume by neonatal US group. The MRI results are based on the presented method. *b*: Mean IQ by neonatal US group. *c*: Ventricular volume as a proportion of total brain volume among children with IQ > 85 vs. IQ ≤ 85.

TABLE 7.4: Results of ventricular volumetry for patient #058 (cf. Table 7.3 and Fig. 7.12 d–f).

	Observer	V_1	V_2	V_3	Mean ± SD	(CV)
L	A	47.87 ml	47.85 ml	47.91 ml	47.88 ± 0.03 ml	(0.06 %)
	B	47.86 ml	47.79 ml	48.08 ml	47.91 ± 0.15 ml	(0.32 %)
	C	48.03 ml	47.97 ml	47.61 ml	47.87 ± 0.23 ml	(0.47 %)
R	A	9.97 ml	9.95 ml	10.00 ml	9.97 ± 0.03 ml	(0.25 %)
	B	9.99 ml	9.93 ml	9.91 ml	9.94 ± 0.04 ml	(0.42 %)
	C	9.93 ml	9.88 ml	10.09 ml	9.97 ± 0.11 ml	(1.10 %)
3rd	A	2.14 ml	2.16 ml	1.80 ml	2.03 ± 0.20 ml	(9.95 %)
	B	2.19 ml	2.21 ml	2.01 ml	2.14 ± 0.11 ml	(5.16 %)
	C	2.13 ml	2.39 ml	1.73 ml	2.08 ± 0.33 ml	(15.96 %)
4th	A	1.02 ml	0.96 ml	0.93 ml	0.97 ± 0.05 ml	(4.72 %)
	B	1.70 ml	1.52 ml	1.53 ml	1.58 ± 0.10 ml	(6.39 %)
	C	1.55 ml	1.68 ml	1.57 ml	1.60 ± 0.07 ml	(4.38 %)

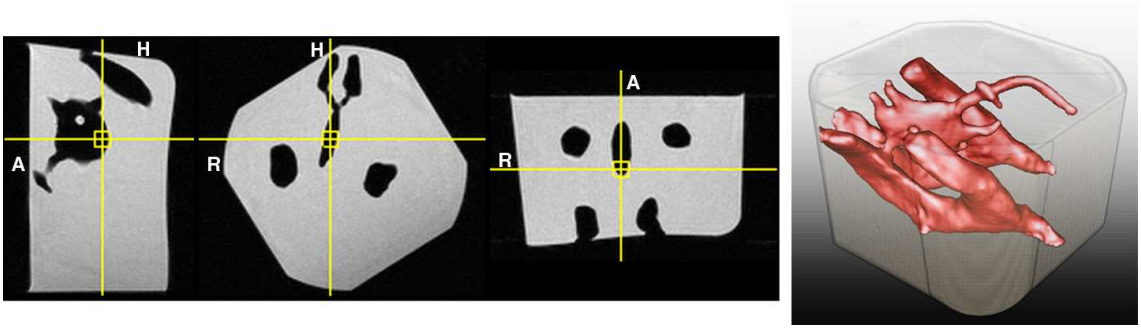


FIGURE 7.11: MR image of paraffin phantom in contrast agent. Same protocol as Figure 7.1 right. Hybrid rendering of segmentation result fused with original data.

TABLE 7.5: Evaluation of total reproducibility on ventricle shaped paraffin phantom in contrast agent (cf. Fig. 7.11), imaged using T1 weighted MRI, six independent acquisitions (V_1–V_6), independent object orientation and positioning, volumes in ml.

V_1	V_2	V_3	V_4	V_5	V_6	mean ± SD	CV
60.62	60.76	60.78	60.91	61.12	61.13	60.89 ± 0.22	0.36 %

increased mean interaction time (approximately three minutes), but did not produce biased results within the given variation. Nonetheless, by calculating one histogram parameter directly from the marker positions, an additional source for inter-observer variations was created and requires special care by the user (cf. Table 7.6, patient #255, left and right ventricle).

Physical Phantom Evaluation

In order to evaluate systematic errors of the overall method, repeated MR acquisitions on a ventricle shaped paraffin phantom were carried out (cf. Fig. 7.11) which yielded a relative inter-examination variation of 0.4 % for the total volume (Mean ± SD: 60.89 ± 0.22 ml, cf. Table 7.5); this is in accordance to the true phantom volume (61.0 ml, measured by water displacement with an uncertainty of 0.5 ml), indicating a high accuracy of the method.

7.6 Discussion

Two examples demonstrate the challenges of volumetric imaging in the pediatric age group and represent typical results of the current methodology. At the time of MR imaging, patient #255 was a 6-year-old female from a bilingual family with a history of delayed language milestones. Neurological evaluation revealed poor coordination and balance, but no evidence of cerebral palsy. Neuropsychological evaluation found poor language skills, borderline IQ, and significant behavioral problems. Direct volume rendering of

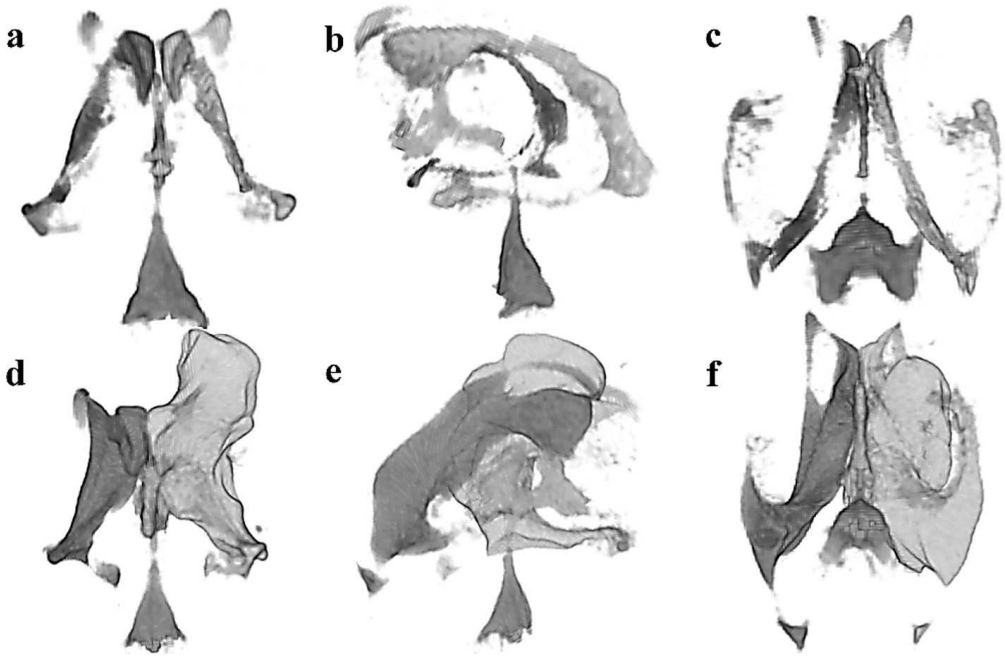


FIGURE 7.12: Direct volume rendering of segmentation result for two patients. *a–c*: patient #255; in this case, additional markers were required to specify the mean CSF gray value (cf. Fig. 7.8 a). *d–f*: patient #058, predominant on the left hemisphere.

the ventricular system demonstrates thin symmetrical lateral ventricles (Fig. 7.12 a–c). Axial volumetric SPGR imaging demonstrates similar findings (Fig. 7.9 a) and shows no significant ventricular enlargement. Volume analysis quantifies the relative symmetry of volume between the lateral ventricles (1.7 ml, left; 1.8 ml, right; cf. Tables 7.3 and 7.6).

Patient #058 was a 7-year-old male from a bilingual family with a history of speech and motor deficits, as well as attention deficit disorder / hyperactivity disorder. Neurological and neuropsychological evaluations revealed a right hemiparesis and a borderline IQ with poor non-verbal skills, but no significant behavioral abnormalities. Direct volume rendering of the lateral ventricles shows an enlarged left lateral ventricle (Fig. 7.12 d–f). An axial T1 weighted SPGR image through the level of the lateral ventricles (Fig. 7.9 b) demonstrates similar asymmetric enlargement. Volume analysis quantifies the almost five-fold enlargement of the left lateral ventricle compared to the right (47.9 ml, left; 10.0 ml, right; cf. Tables 7.4 and 7.6).

The above examples demonstrate the feasibility of rapidly quantifying ventricular volume in the pediatric brain. The methodology is robust enough to handle situations where significant ventricular asymmetry is present (Fig. 7.9 b) as well as to handle small essentially normal ventricular volume (Fig. 7.9 a). Reproducibility is extremely high with short analysis times (less than five minutes), factors that are conducive for use in the clinical setting. In addition to intra- and inter-observer characteristics, a preliminary study by [HAHN *et al.*, 2001a] evaluated the inter-examination reliability (i. e., the reproducibility between independent image acquisitions) of our method in adult volunteers; inter-examination co-

TABLE 7.6: Mean volumes, as well as intra- and inter-observer characteristics (standard deviation and coefficient of variation) derived from the results presented in Tables 7.3 and 7.4.

		Mean	Intra-observer		Inter-observer	
Pat. #255	L	1.74 ml	± 0.13 ml	(7.66 %)	± 0.18 ml	(10.55 %)
	R	1.80 ml	± 0.13 ml	(7.18 %)	± 0.20 ml	(11.25 %)
	3rd	0.75 ml	± 0.09 ml	(12.61 %)	± 0.10 ml	(12.66 %)
	4th	1.62 ml	± 0.05 ml	(2.87 %)	± 0.01 ml	(0.86 %)
Pat. #058	L	47.89 ml	± 0.14 ml	(0.29 %)	± 0.02 ml	(0.04 %)
	R	9.96 ml	± 0.06 ml	(0.60 %)	± 0.02 ml	(0.16 %)
	3rd	2.08 ml	± 0.20 ml	(9.70 %)	± 0.05 ml	(2.48 %)
	4th	1.38 ml	± 0.07 ml	(4.74 %)	± 0.36 ml	(25.93 %)

efficients of variation were determined to be less than 1 % for the lateral ventricles, which appears to be a significant improvement in comparison to existing techniques [TSUNODA *et al.*, 2000, KOHN *et al.*, 1991, KIKINIS *et al.*, 1992].

7.7 Conclusion

We presented a new semiautomatic approach to cerebral ventricular volume analysis that reliably quantifies pediatric ventricular anatomy and, at the same time, reduces image analysis and interaction times when compared to manual or semiautomatic slice-based evaluation. User-induced errors are minimized by placing markers inside the objects instead of tracing object borders interactively. Within the software assistant implemented under MeVisLab [MEVIS, 2004], orthogonal color overlays in combination with direct volume rendering serve as visual control tools during the segmentation. The 3D segmentation procedure is applicable to both normal and pathological anatomy, and does not require image morphing or an a-priori anatomical model or atlas.

Precise measurements are achieved from commonly available high-resolution T1 weighted MR images at reasonable imaging times. Image fusion and multispectral image analysis, as required by other methodologies [KOHN *et al.*, 1991, KIKINIS *et al.*, 1992, FRIEDLINGER *et al.*, 1999] are not necessary. A fully automated histogram analysis robustly accounts for image noise, nonuniformity, and partial volume effects. Moreover, asymmetry in size and volume of the lateral ventricles are directly quantified without additional user interaction. Combining short interaction times, broad applicability, and high reproducibility, the presented method meets the requirements posed by imaging and workflow conditions in the clinical setting.

Recently, the present methodology for ventricular volumetry was applied to the MRI data from a two-week-old preterm neonate. After employing a modified histogram model, also the brain volume of the baby could be quantified. Difficulties imposed by the small neonatal head size were resolved through the use of a finer resampling grid; the comparably poor image quality did not pose any problem.

Future work includes a comprehensive analysis of the clinical and morphometric data for the group of 34 children, which will also address the relationship between neonatal WMD, birth weight, gestational age, and cerebral volume. Furthermore, we will integrate the described methods into a clinically applicable software assistant comprising the interactive segmentation module plus multiple histogram models in combination with efficient 3D visualization. Cerebral ventricular volume has the potential to become an important parameter in quantitative neurological diagnosis. However, no accepted methodology for routine clinical use exists to date. We sought a robust, reproducible, and fast technique to evaluate cerebral ventricular volume in young children on a real-time basis. We described a novel volumetric methodology to segment and visualize intracerebral fluid spaces and to quantify ventricular volumes. Clinical brain ventricular volume calculations, such as in non-sedated children, can be performed using routine MR imaging, efficient three-dimensional segmentation, and automatic histogram analysis with results that are robust and reproducible.

Publications. *Large parts of this chapter are contained in an article with contributions from* W. S. Millar*, from the Dept. of Radiology, College of Physicians & Surgeons, Columbia University, New York, NY;* O. Klinghammer*, MeVis;* M. S. Durkin*, Dept. of Population Health Sciences, University of Wisconsin Medical School, Madison, WI, USA; and* P. K. Tulipano*, Dept. of Biomedical Informatics, Columbia University, New York, NY, USA* [Hahn *et al.*, 2004b]*. The fast method for intracranial CSF volumetry was presented at HBM 2001 meeting, Brighton, UK, and the combined method at CARS 2001, Berlin* [Hahn *et al.*, 2001a]*. Initial results of the cerebral ventricular volumetry were presented at MICCAI 2001, Utrecht* [Hahn *et al.*, 2001b]*.*

Vorhersagen sind sehr schwierig,
Vor allem wenn sie die Zukunft betreffen.

—Nils Bohr

Ich denke niemals an die Zukunft.
Sie kommt früh genug.

—Albert Einstein

CHAPTER 8

Brain Volumetry in Three and Four Dimensions

Abstract. *We evaluate the accuracy and precision of three different techniques for measuring brain volumes based on MRI. We compare two established software packages that offer an automated image analysis, EMS and SIENAX, and a third method, which we present. The latter is based on the Interactive Watershed Transform and a model-based histogram analysis. All methods are evaluated with respect to noise, image inhomogeneity, and resolution, as well as inter-examination and inter-scanner characteristics on 66 phantom and volunteer images. Furthermore, we evaluate the N3 nonuniformity correction for improving robustness and reproducibility. Despite the conceptual similarity of SIENAX and EMS, important differences are revealed. Finally, the volumetric accuracy of the methods is investigated using the ground truth of the BrainWeb phantom.*

We further propose a fast and robust method to obtain the Temporal Horn Index (THI) as an indirect but sensitive regional measure for hippocampal and parahippocampal atrophy, based on MRI. The THI is defined as the temporal horn volume to lateral ventricular volume ratio. The proposed method relies on efficient 3D interactive segmentation and a fully automated histogram analysis. It provides consistent THI measurements within a few minutes even for extremely small temporal horns of less than 0.1 ml. The THI obtained by volumetric MRI analysis is sensitive to hippocampal and parahippocampal atrophy and is expected to provide an early marker for pathologic changes associated with Alzheimer's and Parkinson's disease.

Finally, the proposed brain volumetry approach is demonstrated to be capable for the first time to directly quantify subtle volume changes associated with progressive hyperventilation in an interventional imaging setting with very low image quality.

VARIOUS INDICATIONS exist for brain volume measurements. Major fields of application are diagnosis, disease monitoring, and evaluation of potential treatments in MS [Rudick *et al.*, 1999, De Stefano *et al.*, 2003, Lukas *et al.*, 2004] and neurodegenerative diseases, most importantly AD [Fox *et al.*, 1996, Brunetti *et al.*, 2000]. Rudick *et al.* [1999] propose the Brain Parenchymal Fraction (BPF), which they define as the ratio of brain parenchymal volume to the total volume within the brain surface contour, as a marker for destructive pathologic processes in relapsing MS patients. De Stefano *et al.* [2003] found substantial cortical GM volume loss in MS. They propose

that neocortical GM pathology may occur early in the course of both relapsing-remitting and primary progressive forms of the disease and contribute significantly to neurologic impairment. In addition to a process that is secondary to WM inflammation, they also assume an independent neurodegenerative process, which mainly affects GM and raises the need for robust measures to independently quantify WM and GM volumes.

8.1 How Accurate is Brain Volumetry?

Fox *et al.* [1996] report a mean brain atrophy progression of approximately one percent (12.3 ml) per year in the AD group compared to less than 0.1 percent (0.3 ml) in the control group. This would mean that the relative precision of brain volume measurements must be approximately 0.5 percent in order to significantly measure atrophy within one year. More recently, Fox *et al.* [2000] found a mean rate of brain atrophy in AD patients of 2.37 % and in the control group of 0.41 %, which is larger than the preceding results by a factor two to four. Brunetti *et al.* [2000] showed highly significant reductions in AD patients for both WM and GM volume fractions and conclude that GM and WM atrophy quantification could complement neuropsychological tests for the assessment of disease severity in AD, possibly having an impact on therapeutic decisions.

In addition to global measurements, a regional analysis is important, for example to derive separate volumes for brain lobes, deep gray matter, hippocampus, cerebellum, etc. This importance has been emphasized, e. g., by Fox *et al.* [1996] and Andreasen *et al.* [1996]. Fox *et al.* [1996] visualize regional atrophy by the differences between aligned and normalized serial scans, while computing tissue loss that takes into account PVE by the integral of the change over the whole brain. Andreasen *et al.* [1996] derive volume measurements for twelve standardized regions based on the Talairach coordinate system.

Here, we concentrate on whole brain characteristics, for total brain volume ($\mathrm{TBV} = \mathrm{WM} + \mathrm{GM}$), BPF ($= \mathrm{TBV} \,/\, (\mathrm{TBV} + \mathrm{CSF})$), as well as GM and WM volumes. These measures are expected to remain important parameters in clinical imaging and within clinical trials [Rudick *et al.*, 1999, De Stefano *et al.*, 2003, Brunetti *et al.*, 2000]. Our objectives are:

(a) To evaluate the accuracy and precision of different methodologies for whole brain volumetry on a software phantom with known ground truth and on real MRI data. We aim toward contributing to a methodological comparison that we find poorly developed for image analysis in general and for brain volumetry in particular.

(b) To investigate the importance of image nonuniformity, noise, and resolution in brain volumetry and the possibility of improving results using nonuniformity correction.

(c) To propose a novel technique which is simple, fast, robust, and accurate. Robustness will be assessed with respect to nonuniformity, noise, and resolution, while the major criterion is reproducibility in terms of inter-examination (scan-rescan) and also inter-scanner characteristics.

8.2 Methodological Evaluation

For the evaluation of brain volumetry, we used phantom, volunteer, and patient data. The phantom data was obtained from the BrainWeb database [Collins *et al.*, 1998] at various noise (3 %, 5 %, 7 %, 9 %) and nonuniformity levels (0 %, 20 %, 40 %) as well as axial resolutions (1 mm, 3 mm, 5 mm; cf. Fig. 8.6). The in-plane resolution of the phantom is $(1.0\,\text{mm})^2$ throughout. The exact volumetric ground truth is known a-priori and provided on the web site; the values for GM and WM are 902.9 ml and 680.8 ml, respectively, if glial matter (6.0 ml) is counted as WM; TBV sums up to 1583.7 ml.

To evaluate scan-rescan reproducibility, we used data from three healthy volunteers (subjects 1–3), which were scanned each five times on the same day with independent repositioning and head rotation. Two of the subjects have also been repeatedly scanned twice on two different scanners on another day, such that inter-scanner characteristics are available. The two devices and acquisition protocols were: (A) Magnetom Symphony (Siemens Medical Solutions), T1 MPR 3D, TR = 1900 ms, TE = 4.15 ms, TI = 1100 ms, and (B) Magnetom Vision Plus (Siemens Medical Solutions), T1 MPR 3D, TR = 9.7 ms, TE = 4.0 ms, TI = 300 ms. We used an isotropic resolution of $(1.0\,\text{mm})^3$ for all volunteer images with an acquisition time of approximately nine minutes. Inter-examination images were acquired on scanner A for subjects 1 (M, 39 y) and 2 (F, 34 y) and on scanner B for subject 3 (F, 27 y). To limit scan times, we resampled the image data from subject 3 (all five independent acquisitions) at four different axial resolutions (1.8 mm, 3.0 mm, 4.8 mm, 7.2 mm) using a three-lobed Lanczos filter (cf. footnote on p. 105).

Various techniques exist to address brain volumetry based on MRI. We have evaluated two of them, which have reached a certain popularity, namely the software packages SIENAX [Smith *et al.*, 2002] and EMS [Van Leemput *et al.*, 1999]. Both are well documented and available on the internet for research purposes [Smith *et al.*, 2002, Van Leemput *et al.*, 1999]. Furthermore, we have evaluated a novel method that builds upon the image analysis platform MeVisLab [MeVis, 2004], referred to as MeVisLab Brain Volumetry (MBV) (cf. Color Plate C.12). We briefly describe the concepts of the three methods.

SIENAX is a fully automated command-line tool and part of FMRIB's FSL library (Steven M. Smith, University of Oxford, UK) [Smith *et al.*, 2002]. It is based on a hidden MRF model and an associated iterative EM algorithm for estimating tissue intensity parameters and spatial intensity bias field. Before this, the images are registered to a standard space brain image where a mask is used to remove non-brain tissue. SIENAX explicitly estimates PVE by evaluating the tissue intensity model on a particular voxel's neighborhood.

EMS was developed by Koen Van Leemput at the Medical Image Computing Group at KU Leuven, Belgium [Van Leemput *et al.*, 1999, 2003], and builds upon the SPM package (Wellcome Department of Imaging Neuroscience, University College London, UK). Much like SIENAX, EMS relies on unsupervised EM tissue classification, corrects for MR signal inhomogeneity, and incorporates contextual information through MRF modeling. Instead

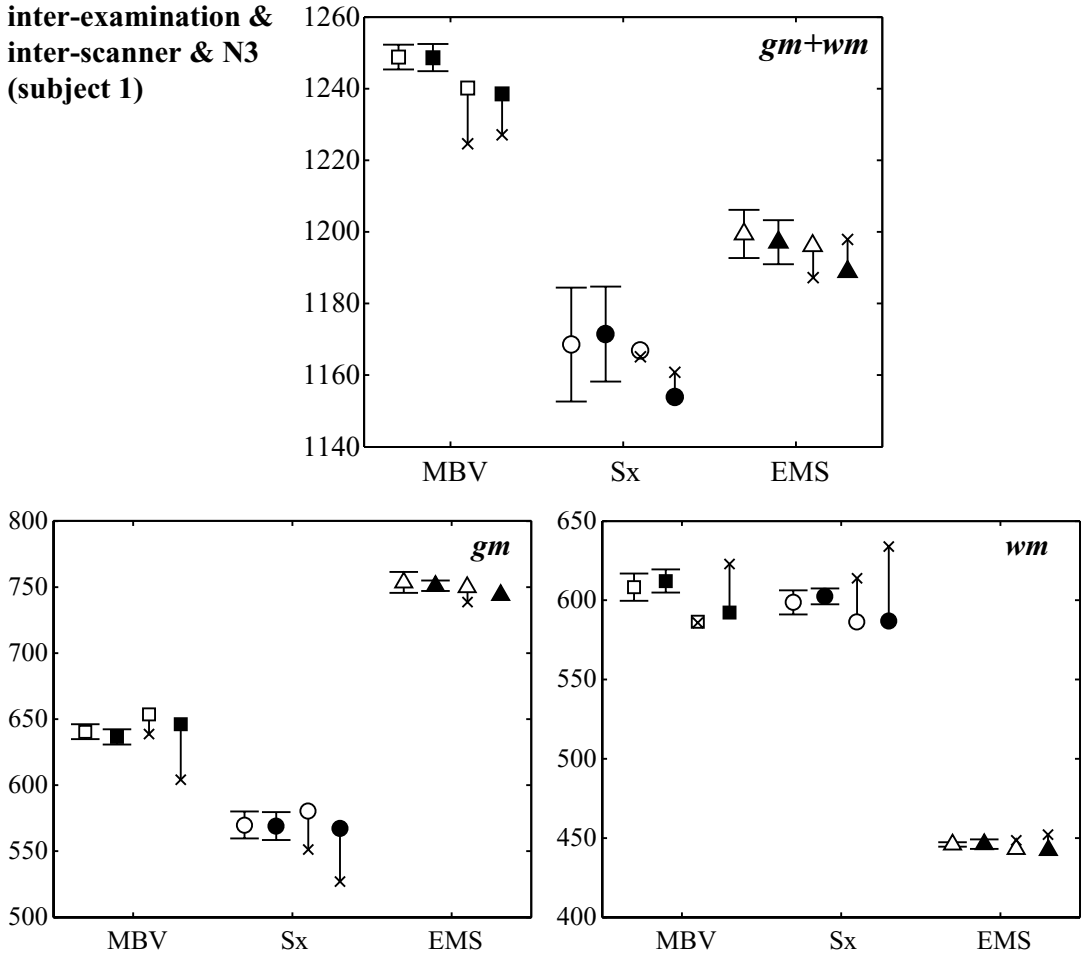


FIGURE 8.1: Inter-examination and inter-scanner characteristics (subject 1) of total brain (***gm+wm***), GM, and WM volumes for three methods (cf. Fig. 8.2–8.4): MBV (□/■), SIENAX (Sx, ○/●), and EMS (△/▲). Error bars indicate single SD. Candle plots show inter-scanner differences (□/■/○/●/△/▲ for scanner A, × for scanner B), each symbol representing the mean of two independent acquisitions. Empty and filled symbols represent the results before and after N3 nonuniformity correction, respectively.

of using a brain mask, a digital brain atlas containing prior expectations about the spatial location of tissue classes is used after affine registration based on mutual information to initialize the algorithm.

MBV is simpler than the two other methods in that it does not comprise EM classification, MRF or nonuniformity modeling. Rather, MBV relies on skull stripping and histogram analysis only. More precisely, the following four subsequent steps are performed (cf. Color Plate C.12): (*i*) Interactive definition of a cuboid ROI on three orthogonal and synchronized 2D views. (*ii*) Automatic resampling to an isotropic grid (spacing $0.9\,\mathrm{mm}^3$) using a Mitchell filter for x, y, and z directions. (*iii*) Skull stripping using the marker-

driven, three-dimensional IWT[1], which was described in detail by Hahn and Peitgen [Hahn and Peitgen, 2003] and in Chapters 2 and 4. (*iv*) Automatic histogram analysis for the 3D region defined by the IWT result as described in Section 7.3 [Hahn *et al.*, 2004b].

The analysis is now based on a model consisting of four instead of three Gaussian distributions for pure tissue types (WM, GM, CSF, and bone/air), as well as dedicated partial volume distributions for mixed tissue types. The histogram model corresponds to the assumptions formulated by Santago and Gage [Santago and Gage, 1993], i. e. uniform PVE from each pair of two tissues, and is fitted to the histogram by minimizing least square deviations. The fourth class to cover air and bone tissue is required since the IWT when applied to the original data (interpreted as depth information) includes most of the partial volume regions at the outer brain surface (GM–CSF interface) and extracerebral CSF (CSF–bone interface) (cf. Ch. 2, [Hahn and Peitgen, 2000]).

Finally, we investigated the N3 method by Sled *et al.* [1998] in order to improve the volumetric results in the presence of a low-frequency image bias field caused by RF inhomogeneity and typical for MR images. N3 employs an iterative approach to estimate both the multiplicative bias field and the true intensity distribution. It requires only two parameters to be selected, one controlling the smoothness of the estimated nonuniformity, the other controlling the tradeoff between accuracy and convergence rate.

Results

Several results were acquired on a total of 66 images (phantom: 10, volunteer: 43, N3 corrected: 13). We computed inter-examination mean and standard deviations (SD) of TBV, BPF, GM, and WM volumes for the three subjects (error bars in Figs. 8.1–8.4). For subjects 1 and 2, we also assessed inter-scanner variability (candle plots in Figs. 8.1, 8.2, and 8.4). For the inter-scanner images and for all images of subject 1, we applied N3 nonuniformity correction using the parameters suggested by Sled *et al.* [1998] (filled symbols in Figs. 8.1–8.4). Figure 8.5 shows the dependency of inter-examination mean values and variations on the axial image resolution. Figure 8.6 comprises measured volumes and ground truth for ten phantom images (varying noise, nonuniformity, and resolution) in a combined graph. The characteristics for TBV and BPF over all subjects are summarized in Table 8.1.

Since it is an interactive technique, a second observer used MBV to analyze the five original images of subject 3 (diamonds in Fig. 8.3). We found inter-observer differences (mean $\pm$ SD) for TBV, GM, and WM volumes to be $+4.5 \pm 4.9$ ml, $+3.6 \pm 3.4$ ml, and $+1.0 \pm 1.6$ ml, respectively. From our experience with MBV, we did not observe significant variations between observers [Lukas *et al.*, 2004], so that we only chose the subject for inter-observer test that required most interaction.

We used all methods without modification or – except the N3 correction – image preprocessing (resampling, noise reduction, etc.). SIENAX (FSL 3.1, July 2003) was operated under Linux on a Dual 2.8 GHz Xeon with processing times of approximately eleven

[1]The robustness of skull stripping based on a modified watershed transform has been shown first by Hahn and Peitgen [Hahn and Peitgen, 2000].

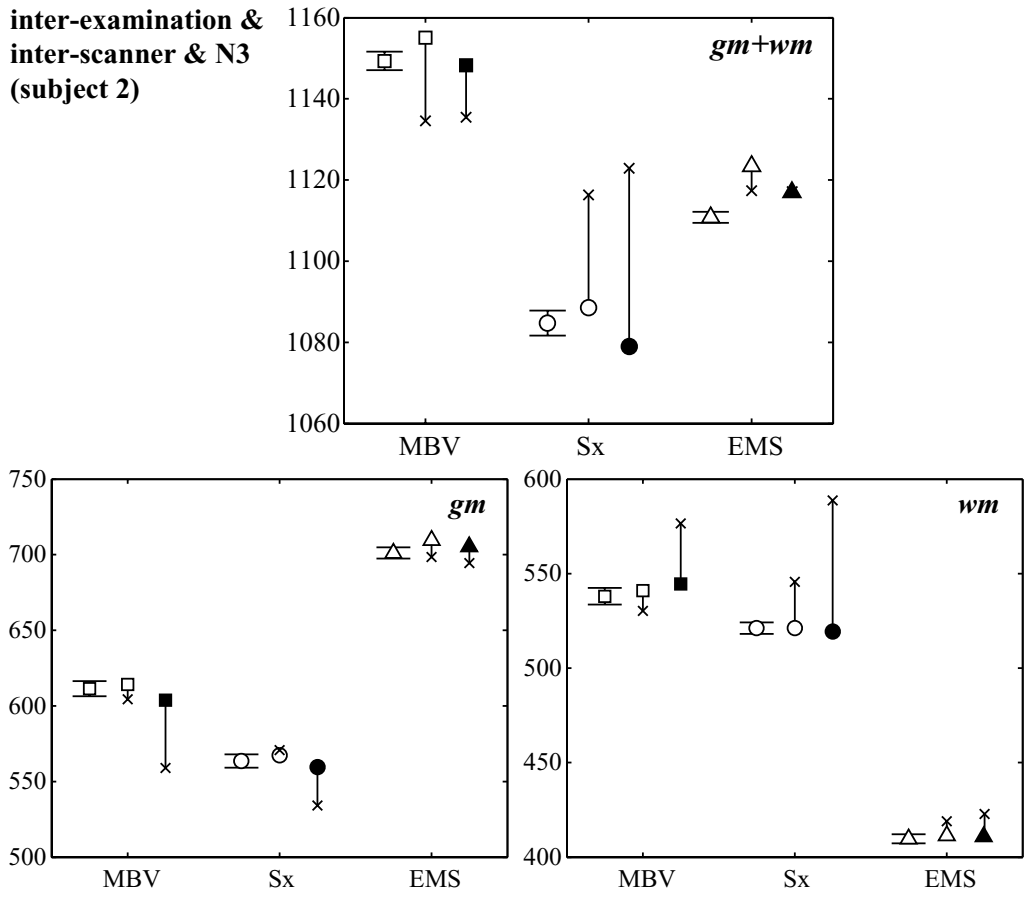


FIGURE 8.2: Inter-examination and inter-scanner characteristics (subject 2) of total brain (***gm+wm***), GM, and WM volumes for three methods (cf. Fig. 8.1 and 8.3): MBV (□/■), SIENAX (Sx, ○/●), and EMS (△/▲). Error bars indicate single SD. Candle plots show inter-scanner differences (□/■/○/●/△/▲ for scanner A, × for scanner B), each symbol representing the mean of two independent acquisitions. Empty and filled symbols represent the results before and after N3 nonuniformity correction, respectively.

minutes per case. For our evaluation, we used the given brain volumes before normalization. EMS and MBV were operated under Windows 2000 on a 1.7 GHz Pentium III. For EMS, we used SPM99 and MATLAB 6.1 with processing times of approximately 26 minutes per case (15 min registration plus 11 min segmentation), including modeling of MRF and fourth-order bias field. EMS volumes were calculated as weighted sum of tissue probability maps. For convergence, we had to manually provide the EMS registration with good initial parameter settings (e. g., 20° frontal rotation was a good starting point in many cases). For MBV, steps *i*, *ii*, plus *iv* required less than one minute, while the interactive skull stripping (step *iii*) required approximately 1–4 minutes for marker placement and visual inspection of the actual segmentation, resulting in an overall analysis time of less than five minutes. For all image processing and evaluation, we paid attention to three important aspects:

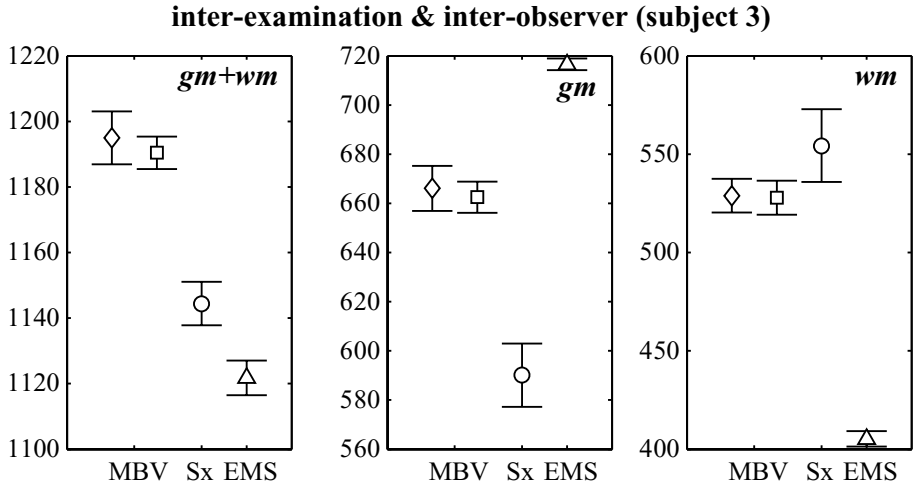


FIGURE 8.3: Inter-examination characteristics (subject 3) of total brain (***gm+wm***), GM, and WM volumes for three methods (cf. Fig. 8.1 and 8.2): MBV (◊/□), SIENAX (Sx, ○), and EMS (△). Error bars indicate single SD. For MBV, results from two observers are provided (□ for observer 1, ◊ for observer 2).

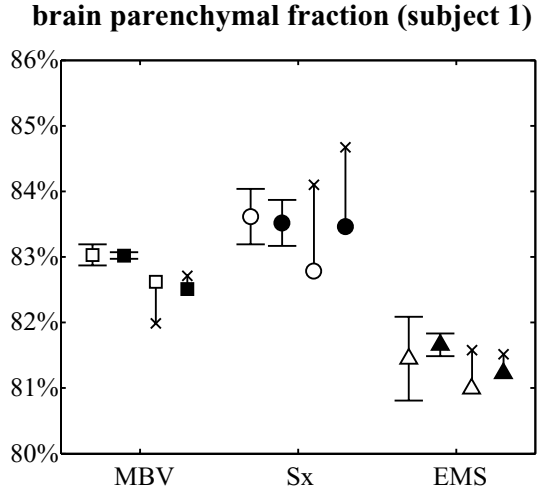


FIGURE 8.4: Inter-examination characteristics (subject 1) of Brain Parenchymal Fraction (BPF) for three methods (cf. Fig. 8.1): MBV (□/■), SIENAX (Sx, ○/●), and EMS (△/▲). Error bars indicate single SD. Candle plots show inter-scanner differences (□/■/○/●/△/▲ for scanner A, × for scanner B), each symbol representing the mean of two independent acquisitions. Empty and filled symbols represent the results before and after N3 nonuniformity correction, respectively.

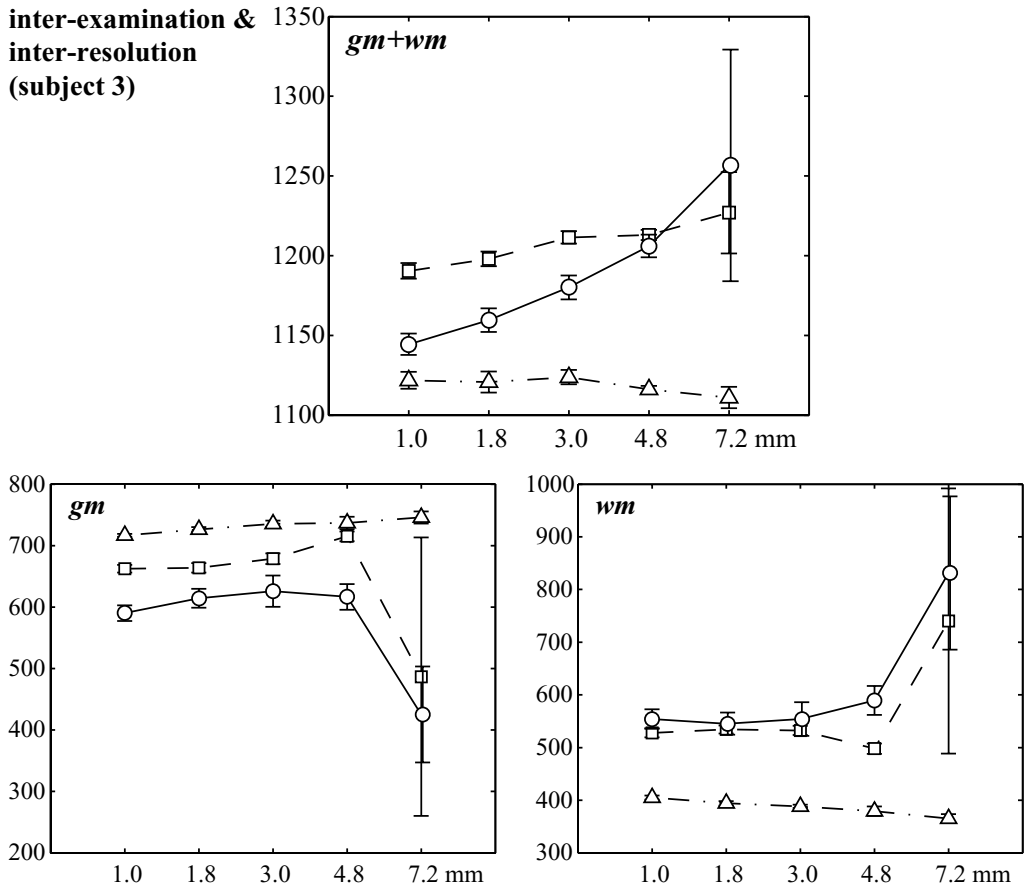


FIGURE 8.5: Inter-examination characteristics for five different axial resolutions (1.0–7.2 mm, subject 3) of total brain (***gm+wm***), GM, and WM volumes for three methods (cf. Fig. 8.3): MBV (□), SIENAX (○), and EMS (△). Each element is computed from five independently acquired and resampled images. Error bars indicate single SD.

(a) Our own method and its parameters were not altered after the first data set was analyzed.

(b) The comparative evaluation of all three methods was conducted exactly once by a person (BENOÎT JOLLY), which has not been involved in any method development.

(c) For the phantom, the operator was blinded with respect to the real tissue volumes.

8.3 Discussion I

Within our study, each of the three methods revealed advantages and disadvantages. Even though similar in concept, we found considerable differences between SIENAX and EMS. The inter-examination characteristics were slightly better for EMS and MBV than for SIENAX, while EMS performed best for GM and WM (cf. Figs. 8.1–8.3) and MBV for

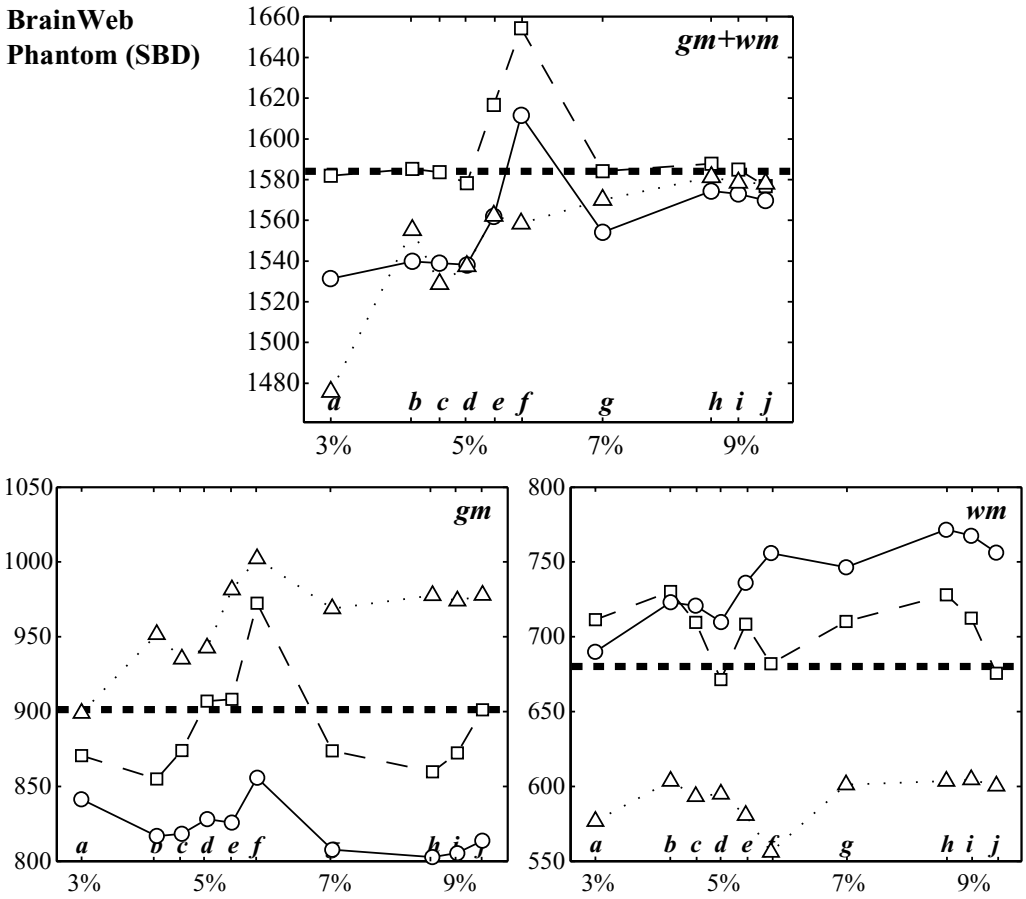


FIGURE 8.6: Volumetric results on ten phantom images (BrainWeb, SBD, [COLLINS *et al.*, 1998]) with known ground truth (bold dashed lines) for total brain, GM, and WM volumes for three methods: MBV (□), SIENAX (○), and EMS (△). Simulated noise levels are indicated on x-axis. Nonuniformity levels are 0 % (b,h), 40 % (d,j), and 20 % (others). Axial slice thickness is 3 mm (e), 5 mm (f), and 1 mm (others).

TBV and BPF (cf. Table 8.1 left). In total, including inter-resolution (Fig. 8.5) and inter-scanner results (cf. Table 8.1 right), EMS was the most robust method.

One important question, to which a phantom can provide answers to a certain extent, is which method gets closest to the ground truth. Despite its stability, EMS yielded the highest GM volumes and lowest WM volumes on all images, which is consistent with a GM overestimation and WM underestimation compared to the phantom ground truth (Fig. 8.6). Conversely, SIENAX consistently underestimated GM and overestimated WM volumes compared to the ground truth, and yielded lowest GM volumes on all images. WM volumes were similar for SIENAX and MBV in some cases (Figs. 8.1 and 8.2). On the phantom data, MBV was closest to the ground truth on average and yielded values between the other two methods in most cases (cf. Fig. 8.6). We also analyzed images of AD patients and found similar results (cf. Color Plate C.13).

TABLE 8.1: Overall characteristics for TBV and BPF, from left to right: mean inter-examination SD (n=20 images at 1 mm slice thickness: S1, S2, S3, S1+N3); effect of N3 (n=13 images, mean and SD of pair-wise differences); effect of N3 on inter-scanner differences (n=16 images, mean difference $\langle V_A - V_B \rangle$, without → with N3).

	inter-examination SD		N3 effect		inter-scanner N3 effect	
	TBV (ml)	BPF (%)	TBV (ml)	BPF (%)	TBV (ml)	BPF (%)
MBV	3.6	0.17	-0.8 ± 4.2	0.00 ± 0.41	$18.0 \rightarrow 12.2$	$0.56 \rightarrow 0.11$
Sx	9.7	0.30	-2.0 ± 10.9	0.22 ± 0.52	$14.8 \rightarrow 25.4$	$1.17 \rightarrow 1.61$
EMS	4.9	0.33	-1.4 ± 9.6	0.09 ± 0.46	$8.3 \rightarrow 4.0$	$0.34 \rightarrow 0.16$

The systematic behavior of EMS could be influenced by the atlas-based prior tissue probabilities; in Color Plate C.13 right, a comparably broad GM layer can be discerned. The GM underestimation of SIENAX could be caused by a too rigorous brain masking, which removes some of the partial volume voxels at the pial brain surface (cf. Color Plate C.13 center).

BPF is a normalized measure of brain atrophy, which can be used in cross-group studies. When computing the mean inter-examination SD for BPF, MBV performed slightly better than the other methods (cf. Fig. 8.4 and Table 8.1 left).

Pair-wise comparison of results with and without N3 revealed the least changes of TBV and BPF for the MBV method, showing its high robustness to image nonuniformity (Table 8.1 middle). Note that this is despite the fact that MBV is the only method that comprises neither nonuniformity correction nor spatial (MRF) regularization, such that we expected it to benefit most from the N3 method. Moreover, N3 did not significantly reduce inter-scanner differences. It yielded some improvements for TBV and BPF, but only for EMS and MBV (Table 8.1 right); GM and WM differences were mostly worsened by N3 (cf. Figs. 8.1 and 8.2).

We still investigate further techniques to correct for image nonuniformity, even if we find that our method is largely independent of such artifacts. A promising approach is the white matter method as proposed by sled [SLED, 1997], which can be beneficial for brain volumetry in comparison to techniques that are not specific to brain images. Also promising is the M4 method, which is based on the minimization of entropy [LIKAR *et al.*, 2001].

Conclusion I

In conclusion, EMS was most robust for very large slice thickness and between scanners, and with best inter-examination characteristics for WM and GM. SIENAX has a major advantage in its full and robust automation. MBV showed best inter-examination characteristics for TBV and BPF, and was least influenced by N3. Despite its interactivity, MBV was the fastest of the three methods, but only if a human operator is available. Together with A. SCHUBERT, we therefore investigated a full automation of MBV's remaining in-

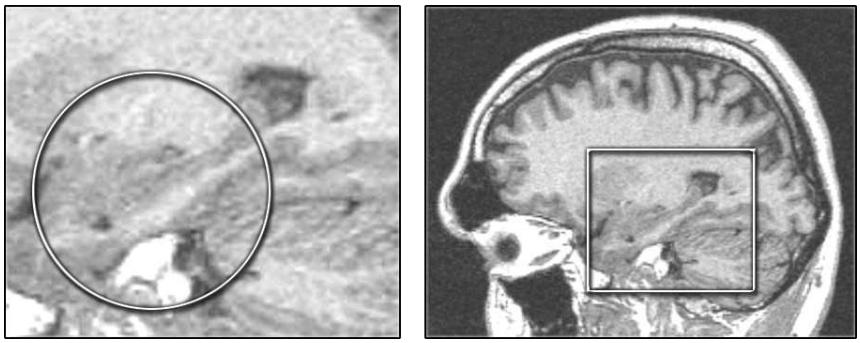


FIGURE 8.7: Illustration of the hippocampus segmentation problem. To the left, a close-up of the MR image to the right is shown (box). In many cases, the hippocampal borders are not well defined (circle).

teractive steps [SCHUBERT *et al.*, 2002] (cf. Sec. 4.4). Still, we see an advantage in MBV's possibility to interactively control and refine the brain mask and that it does not depend on image registration or anatomical template.

Future work includes a more extensive investigation of these and other methods on a variety of patient images. In particular, the systematic behavior with respect to image resolution (cf. Figs. 8.5 and 8.6 e/f) requires further research on a larger set of MR images. Other methods include SIENA (also FSL) that is designed for longitudinal brain volume measurements. One issue that is important for clinical use was not explicitly addressed in this chapter, namely sensitivity. While reproducibility is a measure for the robustness of a method, small changes, e. g. associated with GM atrophy, could remain undetected by a highly reproducible method. This needs to be addressed in future studies.

8.4 Fast and Robust Quantification of Parahippocampal Atrophy via Temporal Horn Index

After discussing the problems of global volumetry of cerebral gray and white matter, we address a specific problem, which is of interest to many Neurologists and Psychiatrists. Diverse indications exist for the quantification of hippocampal volumes. One of the most important is to provide a quantitative marker in monitoring of pathologic changes associated with dementia in Alzheimer's (AD) and also Parkinson's diseases [JACK *et al.*, 2003, SILBERT *et al.*, 2003]. According to SILBERT *et al.* [2003], cerebral ventricular and hippocampal volumes are sensitive to the accumulation of cortical neurofibrillary tangles and senile plaque, which are a key to judging the neuropathology of AD.

Currently, no fast and sensitive method for direct hippocampus volumetry is available for clinical use. To our knowledge, current segmentation methods are either time consuming or unreliable. Direct hippocampus segmentation is a challenging task, since parts of its border are poorly differentiated in clinical images ([SHEN *et al.*, 2002], cf. Fig. 8.7). Purely manual processing with total interaction times of about one hour per data set is frequent. Even highly advanced segmentation methods based on deformable shape models

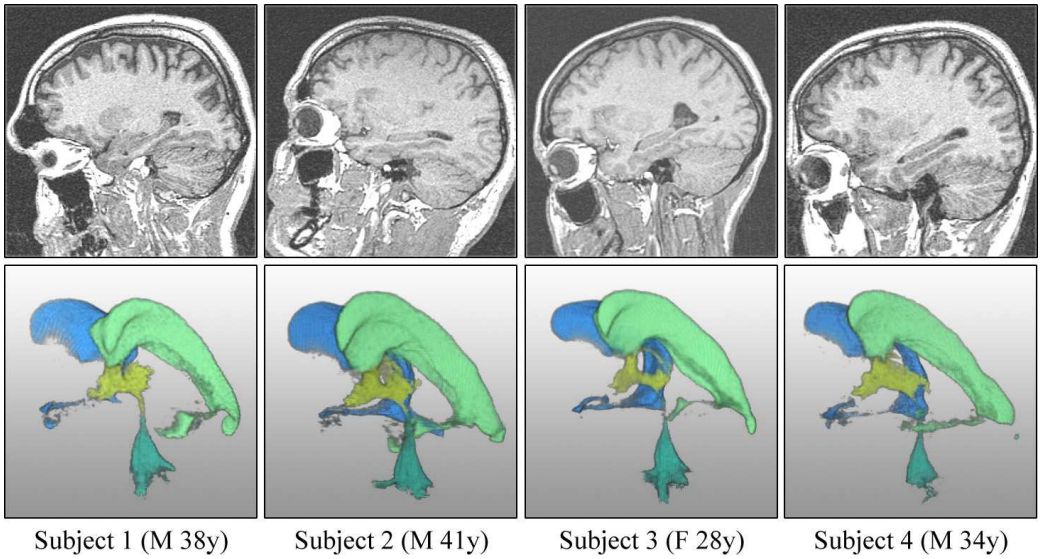


FIGURE 8.8: Material used for the evaluation of the proposed method. T1 weighted MR images from four volunteers (top); five acquisitions each. Using a software assistant based on the MeVisLab platform [MEVIS, 2004], the volumes of the cerebral ventricles were measured (bottom).

and atlases are reported to require at least 10–30 minutes per data set for manual landmark placement [SHEN *et al.*, 2002, HOGAN *et al.*, 2000] with a performance comparable to manual segmentation. In addition, systematic volumetric errors are often related to partial volume effects, which are crucial for any thin and elongated object.

Since the temporal horns of the cerebral ventricles are adjacent to the hippocampal formation, we propose to use the temporal horn volume (THV) as an indirect and sensitive regional measure for hippocampal and parahippocampal atrophy. Furthermore, we propose to consider THI in order to obtain a specific and normalized measure. Moreover, special emphasis was placed on both reproducibility and speed of the proposed image analysis method.

Methods and Material

In four healthy volunteers (age 28–41 y, cf. Fig. 8.8), 20 thin-slice T1 weighted MRI data sets (MAGNETOM SYMPHONY QUANTUM, SIEMENS MEDICAL SOLUTIONS, 1.5 T, MPRAGE, 256×256 matrix, 1.0 mm isotropic voxel size, 160 contiguous sagittal sections, acquisition time approximately 9 min) were acquired, five acquisitions on the same day each. The volunteers were repositioned in the scanner between acquisitions.

Segmentation of lateral ventricles including frontal horn, trigone, posterior horn, and temporal horn was performed applying the IWT to the original images. The IWT, which was described previously [HAHN and PEITGEN, 2003], fully works in 3D; only a few landmarks (approximately two to six) suffice to segment both lateral ventricles. This segmentation was performed twice for every data set, once with and once without the temporal horns, resulting in two volumes for each side, which are denominated LVV (lateral ventric-

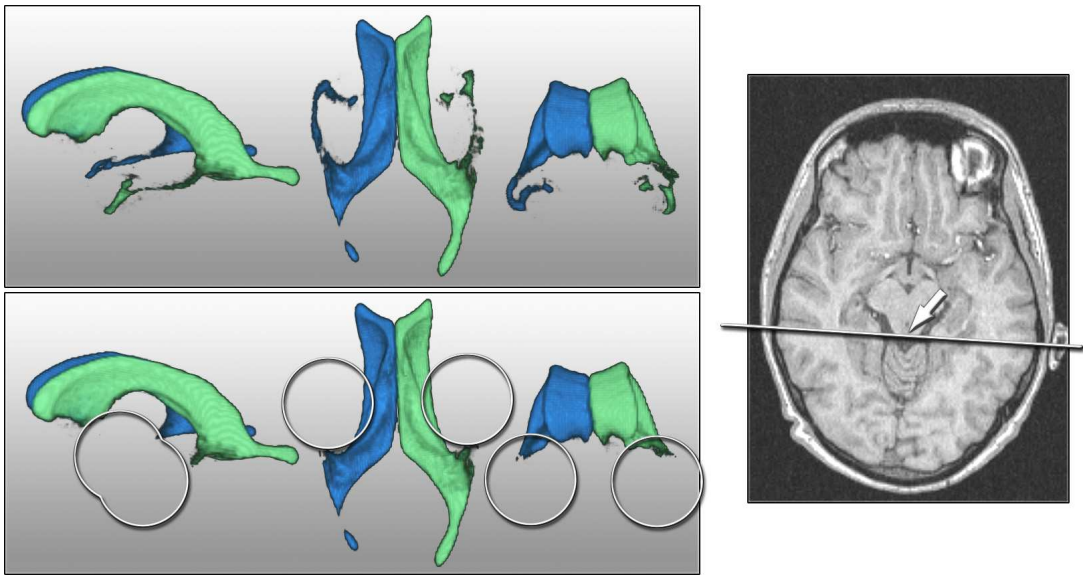


FIGURE 8.9: Definition of the THV by subtraction of the LVV with (top, images from Subject 3) and without temporal horns (bottom, circles); the latter is denominated LVV*. To the right, a coronal plane (straight line) defined by the posterior tip of the inferior colliculus (arrow) is shown on an axial section.

ular volume) and LVV*, respectively. The coronal plane defined by the posterior tip of the inferior colliculus and perpendicular to the AC-PC line was used as posterior boundary of the temporal horn (cf. Fig. 8.9 right).

A fully automated analysis of over-inclusive regional histograms, as described in Section 7.3 [HAHN *et al.*, 2004b], was applied to both segmentations (each containing two objects, left and right lateral ventricle) in order to robustly estimate LVV(L/R) and LVV*(L/R). The analysis is based on a trimodal Gaussian model (cerebrospinal fluid, white, and gray matter) that explicitly includes partial volume terms, so-called Mixed Gaussians. Assuming uniform partial volume effects, Mixed Gaussians take the form of plateau curves. The model is fitted to the histogram data by minimizing squared errors using a modified LEVENBERG-MARQUARDT method, such as described in Section 7.3.

Results

Individual left and right LVV and THV were successfully obtained in all data sets using the landmark-driven IWT within less than four minutes. No preprocessing of the image data was required, such as denoising or intensity homogenization. From the histograms, object volumes could be quantified highly reproducible for all twenty image acquisitions (cf. Table 8.2). The temporal horn index is defined by THI = THV / LVV, while temporal horn volumes are calculated by subtraction THV = LVV– LVV*.

Grand means were recorded for LVV: 12.90 ml (range min–max: 9.15–17.76 ml), THV: 0.32 ml (0.09–0.55 ml), and THI: 2.63 % (0.67–4.46 %). In order to assess inter-examination reproducibility for these three measures, mean standard deviations were evaluated

TABLE 8.2: Lateral ventricular volumes for five independent acquisitions, four volunteers, two sides (L/R), with (LVV) and without temporal horns (LVV*), resulting in $5\times8\times2=80$ measurements, given in ml. To assess inter-examination reproducibility, mean values are given besides standard deviations (SD in ml) and coefficients of variation (CV = SD/Mean, in %).

		V_1	V_2	V_3	V_4	V_5	Mean ± SD	CV
Subject 1	LVV(L)	13.66	13.44	13.79	14.07	13.87	13.77 ± 0.23	(1.7 %)
	LVV*(L)	13.17	12.89	13.38	13.36	13.30	13.22 ± 0.20	(1.5 %)
	LVV(R)	9.34	9.33	8.88	9.45	8.75	9.15 ± 0.31	(3.4 %)
	LVV*(R)	8.94	8.92	8.52	8.95	8.44	8.75 ± 0.25	(2.9 %)
Subject 2	LVV(L)	14.39	14.32	14.01	14.84	14.30	14.37 ± 0.29	(2.1 %)
	LVV*(L)	14.22	14.12	13.83	14.59	14.07	14.16 ± 0.28	(2.0 %)
	LVV(R)	17.93	17.70	17.70	17.82	17.67	17.76 ± 0.11	(0.6 %)
	LVV*(R)	17.55	17.42	17.41	17.46	17.28	17.42 ± 0.10	(0.6 %)
Subject 3	LVV(L)	13.93	14.02	14.16	14.06	14.42	14.11 ± 0.19	(1.3 %)
	LVV*(L)	13.86	13.85	14.05	13.97	14.39	14.02 ± 0.22	(1.6 %)
	LVV(R)	11.82	11.86	11.88	11.96	11.93	11.89 ± 0.06	(0.5 %)
	LVV*(R)	11.59	11.62	11.75	11.71	11.67	11.67 ± 0.07	(0.6 %)
Subject 4	LVV(L)	11.36	11.84	11.82	11.39	11.47	11.58 ± 0.24	(2.0 %)
	LVV*(L)	11.18	11.45	11.45	11.10	11.28	11.29 ± 0.16	(1.4 %)
	LVV(R)	10.25	10.45	10.97	10.69	10.35	10.54 ± 0.29	(2.8 %)
	LVV*(R)	9.88	10.06	10.37	10.09	9.95	10.07 ± 0.19	(1.9 %)

for LVV: 0.22 ml (range min–max: 0.06–0.31 ml), THV: 0.07 ml (0.03–0.12 ml), and THI: 0.56 % (0.21–0.99 %); n=5×8.

8.5 Discussion II

The inter-examination coefficients of variation for the lateral ventricular volumes are around two percent throughout (cf. Table 8.2), which is remarkable taking into account the small object volumes, as well as their elongated and complex shapes. Measuring THV by subtraction rather than directly has a twofold motivation. First, note that reproducibility is better for THV than for LVV, reflecting the fact that variations of LVV and LVV* are positively correlated (cf. p. 104). Second, regional histograms of solely the tiny temporal horns are extremely sparse. In comparison, the histograms for the two larger objects yield a higher robustness and reproducibility. From the user perspective, however, it is equivalent whether to directly segment the temporal horns or to exclude them from a previous lateral ventricle segmentation (cf. Fig. 8.9).

As shown in a previous study [HAHN *et al.*, 2004b], cerebral ventricular segmentation and volumetry based on the IWT and histogram analysis are largely independent of the landmark positions, such that both intra- and inter-observer variations are very small. Therefore, we concentrated on inter-examination characteristics, which are vital

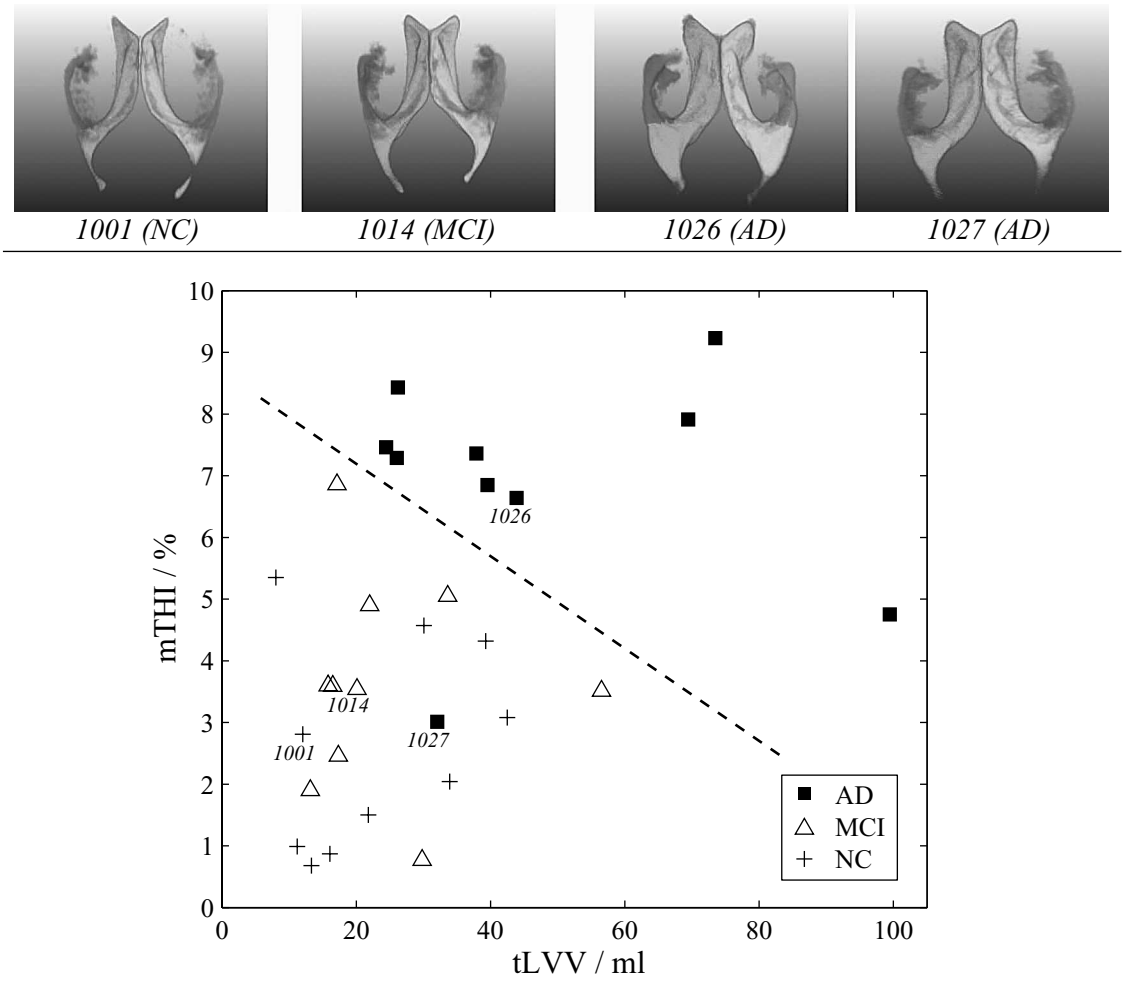


FIGURE 8.10: The combination of total lateral ventricular volume (tLVV) and mean Temporal Horn Index (mTHI) was able to significantly discriminate between Alzheimer's disease (AD) and both mild cognitive impairment (MCI) and normal control groups (NC), within a study comprising thirty subjects (ten of each group, cf. dashed line in scatter plot). *top*: Direct volume rendering of THI segmentation for four examples (cf. codes in scatter plot). (MR data courtesy of F. L. GIESEL, Heidelberg)

for longitudinal and follow-up studies. A mean THV standard deviation of only 0.07 ml, corresponding to a mean THI standard deviation of 0.56 %, provides a solid basis for reliably quantifying volumetric changes in the parahippocampal region. The mean THV in our four volunteers was small with only 0.32 ml. Yet, increased THV is observed in patients with mild or severe hippocampal atrophy. Given that the absolute precision of our method remains constant, which we expect for the subtraction method, the relative precision for patients will be superior in comparison to healthy volunteers. An ongoing study examines the significance of LVV, THV, and THI, besides other quantitative mark-

ers, to discriminate between AD, mild cognitive impairment (MCI), and normal controls (cf. Fig. 8.10).

Conclusion II

In a study comprising 192 subjects with probable AD, who underwent two MRI examinations with an interval of one year, correlations between image-based volumetric change and change in behavioral and cognitive measures were found to be even greater for the temporal horn than for the hippocampus [Jack *et al.*, 2003]. This further supports the suitability of THV/THI as a reliable biomarker in AD (cf. Fig. 8.10). Jack *et al.* demonstrated the technical feasibility of using "structural MRI measures as a surrogate endpoint of disease progression in therapeutic trials", resulting in "markedly lower estimated sample size requirements for clinical trials" [Jack *et al.*, 2003]. Our method is appropriate to be used in trials, which benefit from quantification of hippocampal or parahippocampal atrophy, and has the potential to replace direct hippocampus volumetry in many cases.

8.6 Precise Dynamic Brain Volumetry on MRI with Limited Image Quality

In cooperation with the Departments of Diagnostic Radiology (Michael Moche) and of Anesthesiology (Wolfgang Heinke) at the Leipzig University Hospital, the MBV method for semiautomatic quantification of brain volume as described in Section 8.2 was further extended to be applied to serial brain images from an open MRI system at 0.5 T (cf. Color Plate C.14). The reasons for this development have been:

- ▷ Volume changes over the course of a few minutes during various manipulations (e. g. hypo- and hyperventilation) have been of interest to our project.
- ▷ These involve fluctuations of about 1 % of the total volume, i. e. only slightly larger than the inter-examination reproducibility of the MBV method described above for 1.5 T.
- ▷ The available image quality at 0.5 T is much lower than on a closed 1.5 T system and was initially considered as too poor for accurate brain volumetry (cf. Fig. 8.11).

Hyperventilation is one of the commonly used methods in intensive care medicine to reduce potential intracranial pressure. No existing method for the direct quantification of brain volume reduction induced by hyperventilation is known to the author. Ventilation induced alterations of intracranial compliance are commonly measured indirectly through mechanical intracranial pressure probes.

Using an interventional MR scanner, Michael Moche hoped to be able to measure brain volumes accurately enough to show the hyperventilation effect. However, in order to monitor the progressive brain volume change, a single scan should not take longer than five minutes. Therefore, a compromise between image quality—mainly noise and resolution—and scanning time had to be accepted.

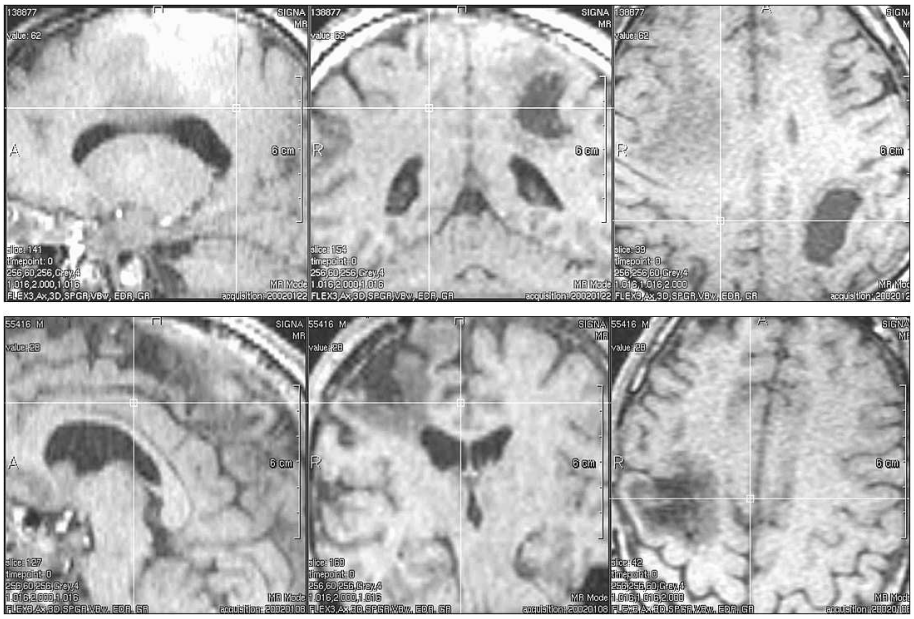


FIGURE 8.11: Original image data from interventional 0.5 T MR scanner (GE HEALTHCARE) used for dynamic brain volumetry. Image noise and nonuniformity associated with flexible surface coil (FLEX-3) is much higher than common for rigid head cage coils used in closed MR systems.

For patient measurements, four MR Series were acquired within twenty minutes total scanning time during normal ventilation and progressive hyperventilation at an end-tidal CO_2 concentration $p_{ET}(CO_2)$ of 42 (normal), 37, 32, and 27 mmHg. The original image data for the first time point of two patients are shown in Figure 8.11. With respect to quantitative image analysis the major problems were:

- ▷ Considerable pixel noise and image nonuniformity
- ▷ Further image artifacts, e. g. due to the Mayfield clamp
- ▷ Presence of brain pathology in all cases
- ▷ Patient head motion to a varying degree
- ▷ Partial brain coverage (only the supratentorial region was completely imaged)
- ▷ Subtle volume changes (cf. above).

Existing approaches to serial volumetry rely on either (i) boundary tracking, (ii) image registration, or (iii) tissue classification. All three approaches have critical problems on the given open MR images, mostly for two reasons:

(a) The image quality is too low for a reliable voxel-wise tissue classification.

(b) The subtle volume change is assumed to mainly have an effect on minimally widening all cortical sulci, which cannot be precisely located.

Morocz *et al.* [2001] managed to quantify brain swelling after hypobaric hypoxia exposure on high resolution 1.5 Tesla data and using the EM tissue classification algorithm presented by Wells *et al.* [1996]. Measured relative brain volume changes were relatively large with 2.8 %, plus the image quality was excellent compared to what we faced in our project. Moreover, the volumetry method was tested for reproducibility on one volunteer only, who underwent MRI three times. Nabavi *et al.* [2001] used the same kind of image data for serial intraoperative imaging of brain shift. As a side product they also calculated changes in brain volume due to swelling. The relative volume changes were in the range of several percent, measured via a balloon-like pial brain surface template.

The purpose of this project was to develop a precise and accurate method for directly measuring subtle brain volume changes that occur within several minutes on low quality MRI data, and to evaluate the robustness to patient movement and image noise. Such a method would be particularly interesting, but not limited to be used with an open MR system, e. g., to measure the volume reduction during various ventilation conditions.

Image Acquisition

We examined four volunteers, each with five subsequent acquisitions ($n_1 = 20$) and without head fixation. For simulation of patient movement and pixel noise, we shifted one original data set in six directions by 1 mm and 2 mm ($n_2 = 12$), and added uniform noise with an amplitude of up to 10 % of the WM image intensity ($n_3 = 12$). On six neurosurgical patients, we acquired four images each at the above listed ventilation conditions ($n_4 = 24$, total = 68), after patient positioning but before craniotomy. Patient heads were fixed in the surgical position using a Mayfield clamp, generating additional image artifacts, and resulting in non-standard oblique head positions.

Image acquisition was performed on a Genesis Signa SP/i (GE Healthcare) at 0.5 T (SPGR 3D, TE = 2.7 ms, TR = 13.3 ms). We used a FLEX-3 coil for acquiring 60 axial sections covering the supratentorial brain volume without gap at a slice thickness of 2.0 mm and an image matrix of 256×192. Acquisition time per data set was 5:20 min. The additional imaging effort has been justified ethically, since patients were anaesthetized, ventilated, and positioned in the open MR system anyhow for image guided neurosurgical interventions.

Image Analysis

Even though the image quality was discouraging for a quantitative analysis, a new two-step image analysis method is introduced to solve this particular image analysis problem, which is a modification of the MBV technique (cf. Color Plate C.15). On an arbitrary time point of a series of four (patients) to six (volunteers) images, a 3D IWT provides a rapid, robust, and reproducible segmentation of the intracranial cavity as described in Chapters 2 and 4. To quantify and compare brain volumes on all time points, a histogram analysis based on an adaptive mixed Gaussian model (pure tissues and partial volume effects) is automatically applied to the skull stripped individual images as described in Section 7.3.

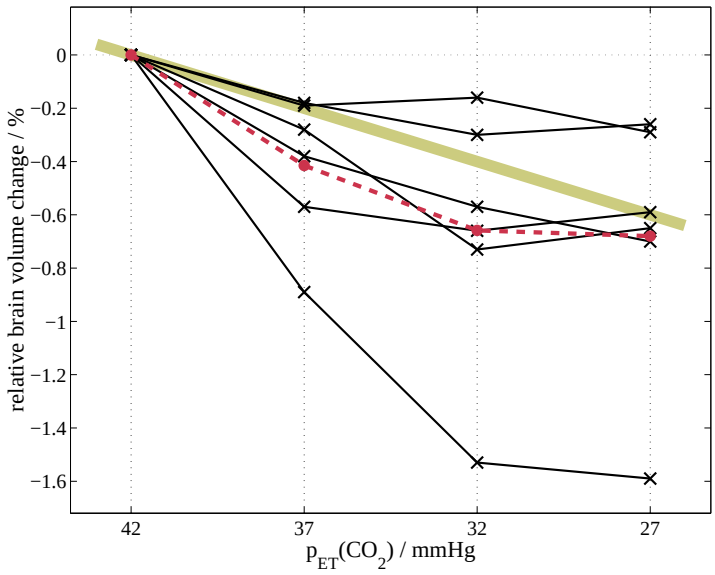


FIGURE 8.12: Relative dynamic brain volume change measured based on MRI on six patients (solid thin lines) and mean values (dashed line) compared to theoretically expected result (bold solid line), plotted as a function of end-tidal carbon dioxide concentrations after progressive hyperventilation

The innovative aspect is that the skull stripping is only performed on a single time point, whereas the resulting brain mask is equally applied to all time points. Restricted by this constant mask, histogram analysis is performed for all time points independently of each other. The uncertainty of skull stripping is thus largely eliminated from the relative volume measurement over time. Therefore, the success of our method is critically depending on both

(a) The robustness of the histogram analysis method to image artifacts and

(b) Its sensitivity to true brain volume reductions.

The skull stripping step only defines the region for which volume changes are to be computed, but does not directly influence the relative measurement. Also note that it is the *interactive* skull stripping that is less critical for the result, while full *automation* was chosen for the histogram analysis, not to be influenced by any user input (cf. Sec. 3.3 for a more general note on the interaction-automation problem).

Results

None of the volunteers who underwent normal ventilation showed a volume decrease greater than 0.05 % through all acquisitions. However, an increase between acquisitions has been measured to be less than 0.2 %. In all images, head motion was observed. Simulated movement up to 2 mm in any direction induced a measured volume increase by 0.0–0.2 %. Thus, movement seems to be the cause for the measured effects. Added noise at an amplitude of 10 % of the mean white matter gray value did not alter the result by

more than 0.2 %. Mean analysis time including interaction was 4:50 min (range: 2:40–7:10 min) for the comparison of two images.

Before discussing the results obtained on patients, it is important to estimate the expected brain volume reduction after successive hyperventilation. As ventilation induced volume changes only affect the vascular compartment, the expected relative brain volume change can be calculated directly from the cerebral blood volume (CBV) change. A reduction of arterial CO_2 partial pressure of 1 mmHg is assumed to induce a cerebral blood flow (CBF) reduction of 2 ml / (100 g · min), such that for normal CBF and CBV, we expect a corresponding CBV reduction of 0.04 ml per 100 g brain parenchyma [Schubert, 1997, pp. 291 ff.]. This means that the expected relative brain volume reduction per 1 mmHg $p(CO_2)$ is estimated by 0.04 ml / 100 ml = 0.04 %.

In our experiment, end-tidal CO_2 concentration was reduced by 5, 10, and 15 mmHg, leading to an expected relative brain volume reduction of 0.2, 0.4, and 0.6 %, respectively. Figure 8.12 compares these expected values to all six patient measurements plus mean values across patients.

8.7 Discussion III

At first sight on Figure 8.11, the low image quality of the open MR images (noise, nonuniformity, artifacts, etc.) seems to jeopardize any sensible volume measurement. Moreover, the volumetric differences between successive time points cannot be discerned visually (without Figure).

The measurement precision that was reported for the MBV method in Section 8.2 for independently repeated acquisitions and segmentations was significantly improved by synchronously segmenting all of the data sets involved. The accompanying histogram analyses are still performed independently. With this simple trick, changes of less than 0.2 % of the brain volume could be reproducibly measured. These initial results are based on volunteers and only a small number of patients. Further validation of the new dynamic volumetry method is now required to determine the lower error bounds for varying image quality and various acquisition techniques authoritatively.

Since one time point of each measurement series has to be chosen as a reference image for brain masking, we compared results for at least two different reference time points in all patients and volunteers, but did not find a significant influence of the reference time point in any case. This suggests that systematic errors of the computed relative volumes can be neglected.

Conclusion III

For the first time, brain volume changes of less than 1 % can be measured directly based on interventional MR images. Gadolinium or other T1 enhancing contrast agent must not be applied before imaging. The proposed method has a high sensitivity to volume change, and an very high specifity to volume reduction. Its relative precision has been determined to be 0.2 % for volume expansion and 0.05 % for volume reduction based on volunteer experiments. Image registration is only necessary if head movement exceeds 2 mm.

The method is suited for quantifying changes that are not perceptible by visual inspection, and has shown an extreme robustness to image artifacts. It was successfully applied to patients who underwent controlled ventilation changes from normal to uncritical hyperventilation immediately before brain surgery. Volume decrease during hyperventilation was measured reproducibly and despite inter-subject variability, it meets surprisingly well the expected decrease of cerebral blood volume due to CO_2 reduction.

Since the histogram analysis employed for this project is identical to the MBV histogram analysis, and since the quality of clinical brain scans is seldom worse than of the interventional scans, the high volumetric precision on the open MR data provides a further strong argument for the validity of the MBV method in general (cf. Sec. 8.3).

If it is possible to implement the ultra-fast intracranial CSF volumetry method described in Section 7.4 on an interventional MR system, the temporal resolution of dynamic brain volumetry could be dramatically improved, even if then as an indirect measure in accordance to Teasdale *et al.* [1988] (cf. comment on p. 119). Note that a relative variability of 1 % total intracranial CSF (cf. Table 7.2) roughly corresponds to only 0.1 % brain volume, such that indirect brain volume measurements with a very high precision were possible by such a method.

Publications. *The first part of this chapter was presented at MICCAI 2004, Saint Malo* [Hahn *et al.*, 2004a]. Benoît Jolly *carried out the methodological evaluation. A clinical evaluation of the whole brain volumetry method was published in* Neuroradiology *with* Carsten Lukas*, Dept. of Neurology, St. Josef Hospital Bochum, and co-workers* [Lukas *et al.*, 2004]. *The second part was presented at BVM 2004, Berlin* [Hahn *et al.*, 2004c]. *Figure 8.10 shows previously unpublished results from a larger study, which has been initiated by* Frederik L. Giesel *et al., German Cancer Research Center, Heidelberg. The third part has been presented at the RSNA 2002, Chicago* [Hahn *et al.*, 2002]. Michael Moche *was responsible for the radiologic part and was the initiator of the project, while* Wolfgang Heinke*, both Leipzig University Hospital, has been responsible for the anesthesiological part of the project.*

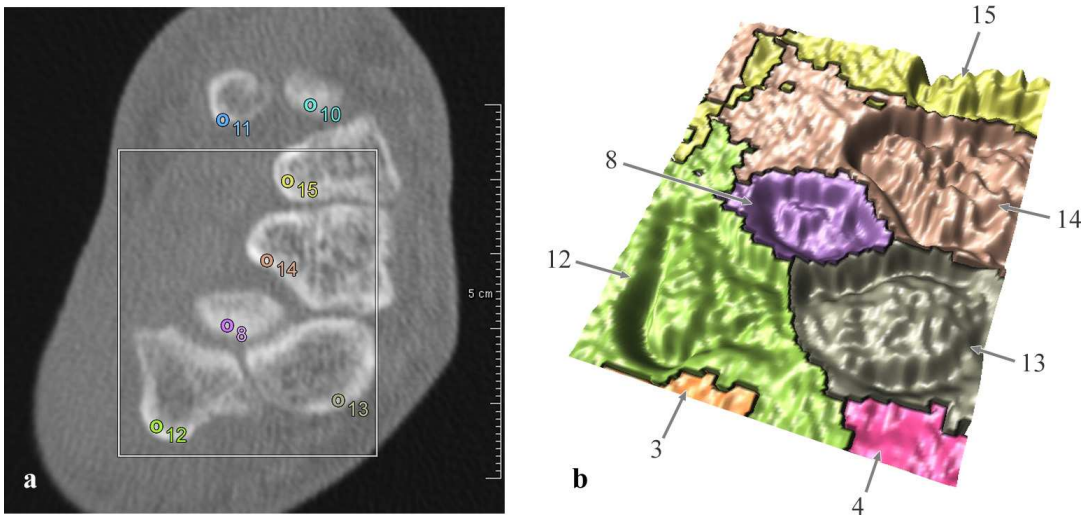
Es gibt weder große Entwicklungen
Noch wahre Fortschritte auf dieser Erde,
Solange noch ein unglückliches Kind auf ihr lebt.

—Albert Einstein

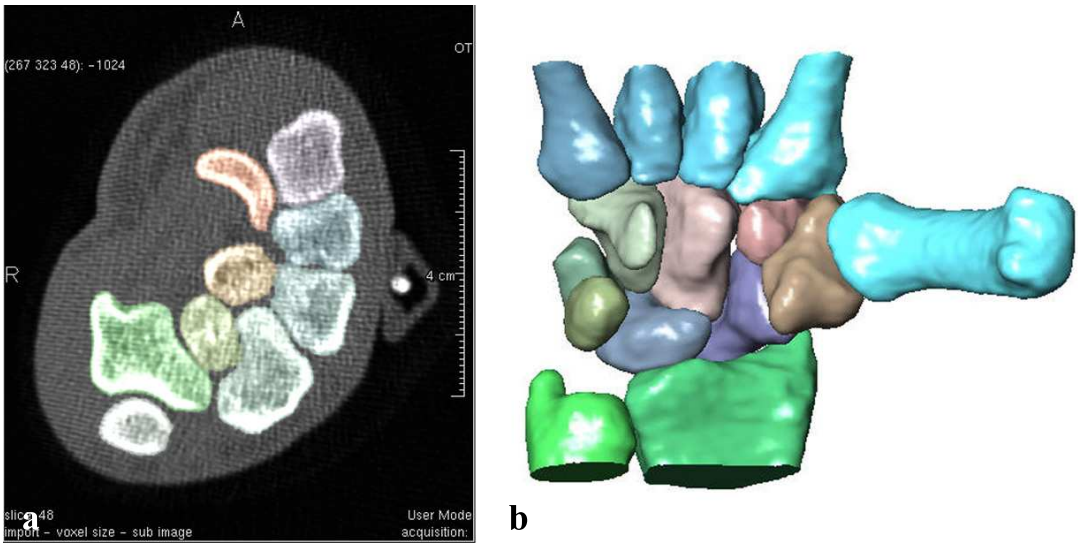
A poverty-free world would not be perfect,
But it would be the best approximation of the ideal.

—Muhammad Yunus

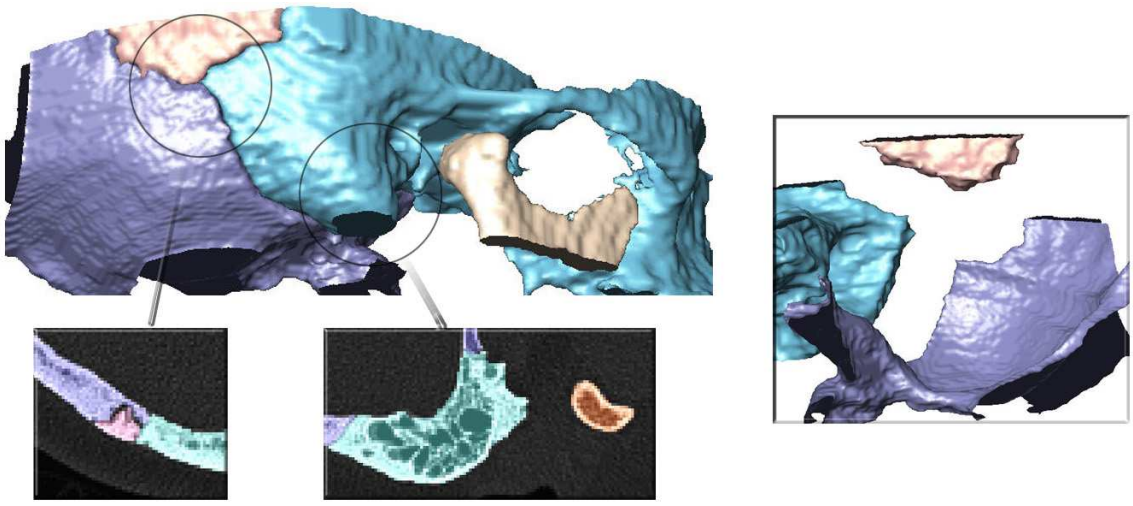
Color Plates



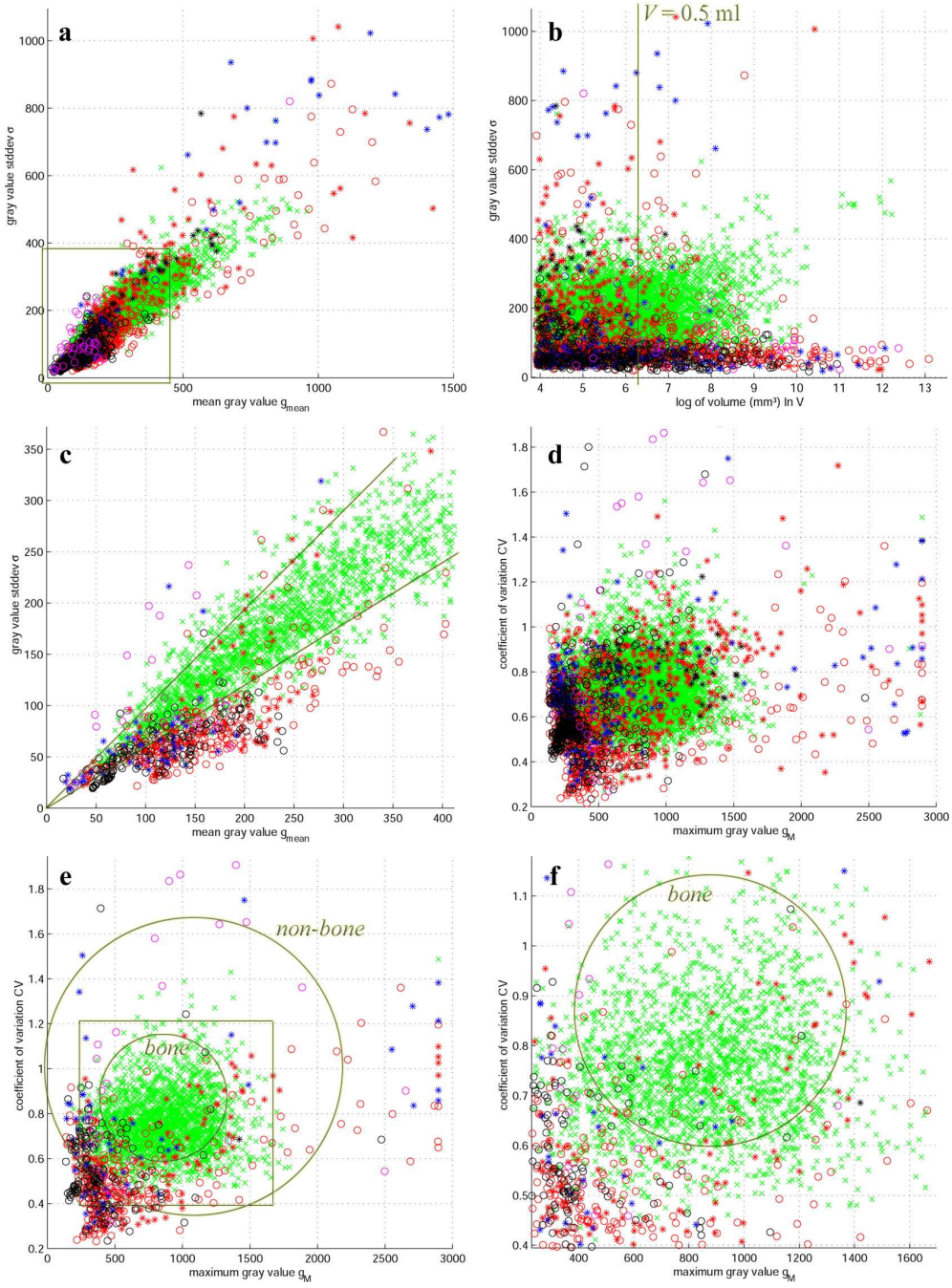
Color Plate C.1: *a*: Two-dimensional section through typical CT wrist image with ROI and possible marker positions. (Note that the markers can also be distributed over multiple slices.) *b*: Representation of the ROI as three-dimensional landscape and separation by IWT according to given markers. (This figure is a didactic projection of the actual four-dimensional landscape representing the complete data; labels cf. Fig. 5.1.)



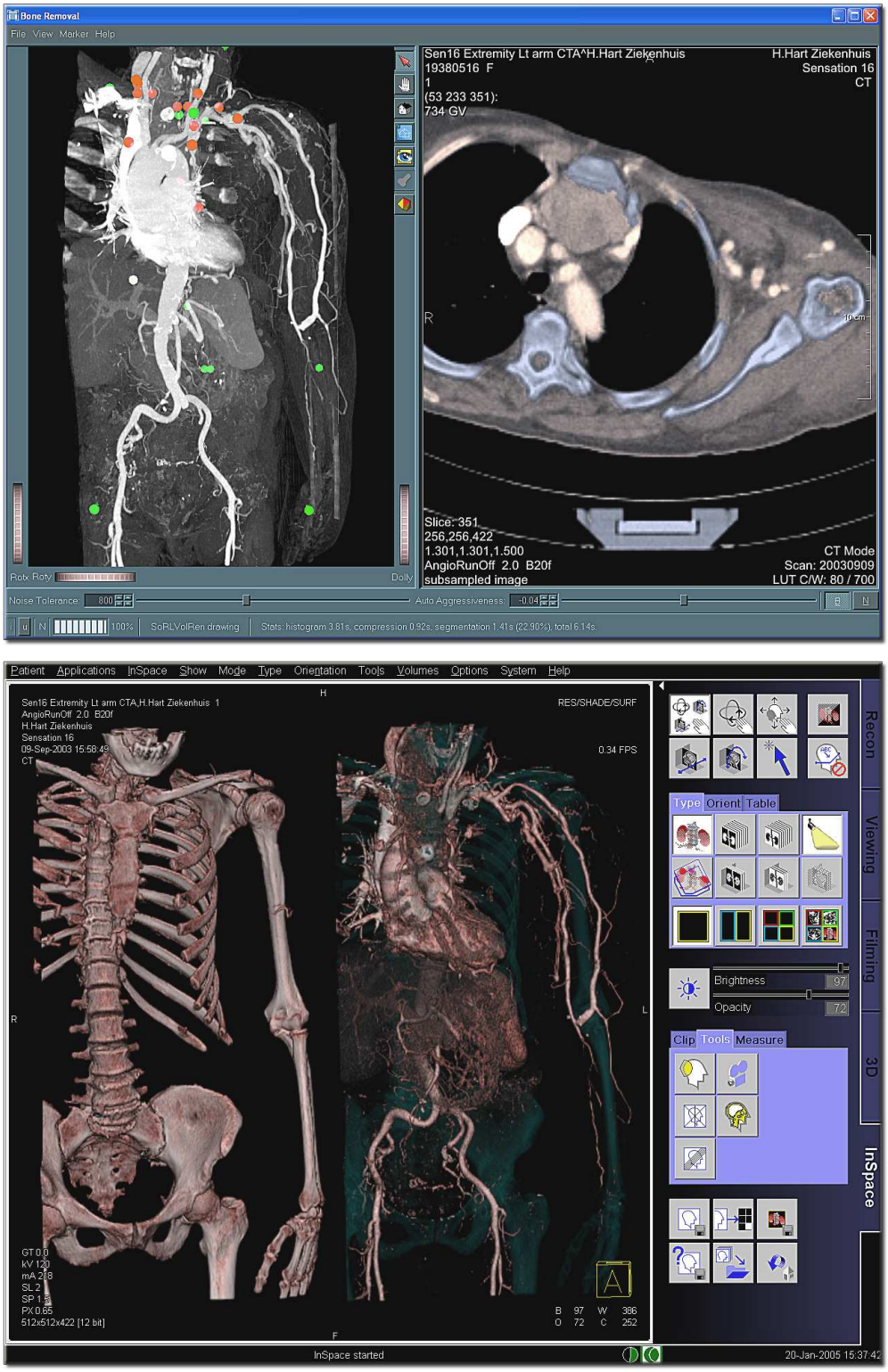
Color Plate C.2: Same data as Figure 5.3; solid bit masks for each individual bone were computed by *bone separation* using the IWT (cf. Color Plate C.1) and *bit morphological closing*.



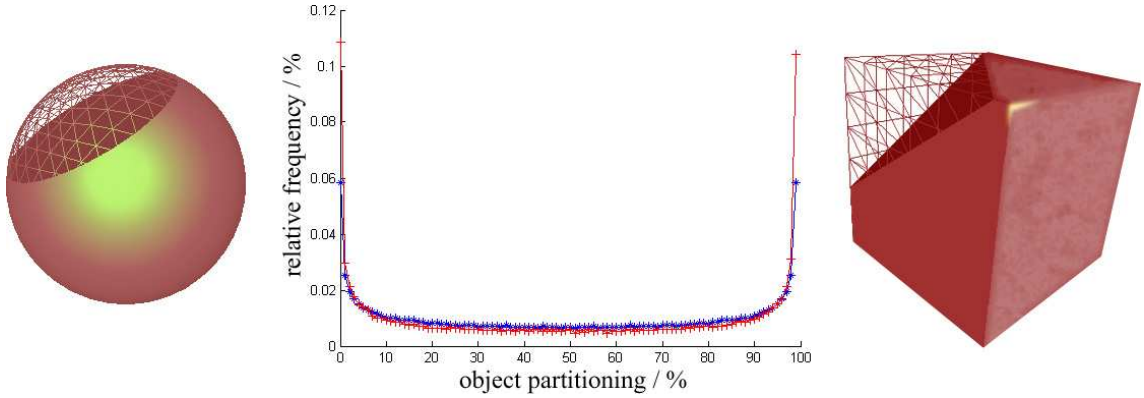
Color Plate C.3: *top left*: Surface rendering based on skull CT data (patient, F, 33 y), view from right inferior posterior. Four bones were segmented: occipital bone (violet), parietal bone (red), mandible (brown/orange), and temporal bone with maxilla and zygomatic bone (light blue). *bottom*: Close-up of 2D color overlay for two regions. *right*: Same objects viewed from intracranial with artificially translated bones. Note the good delineation of the squamosal and lambdoidal sutures. (CT data courtesy of T. Kahn, Leipzig)



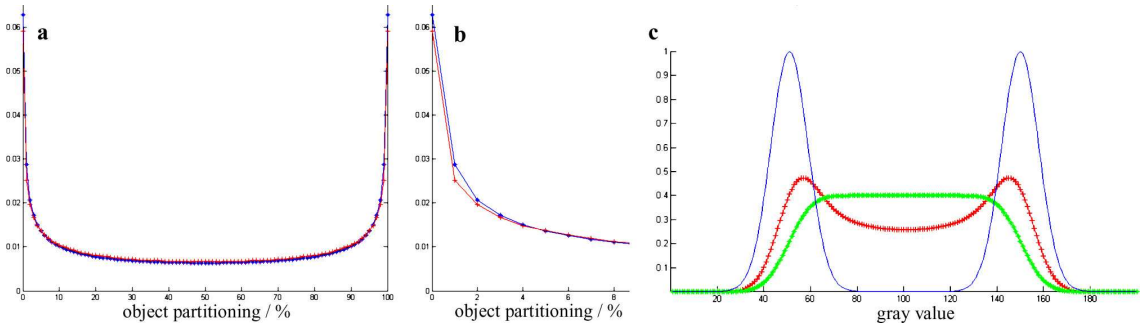
Color Plate C.4: Analysis of regional feature space from 14 representative multi-detector CT data sets, for which a reference segmentation was performed using the MIBS technique. Various symbols are used for bone (green cross) and non-bone structures (all other symbols, e. g. small, intermediate, and large vessels, cardiac ventricles, calcifications, high concentrations of contrast agent, etc.). *a*: Standard deviation (σ) vs. mean gray value (g_{mean}) for all basins. *b*: σ vs. basin volume (V) showing a clearer separation for larger volumes. *c*: Close-up, cf. rectangle in *a*, but only for basins above 0.5 ml volume (cf. straight line in *b*), showing a much clearer image; straight iso-lines indicate constant coefficient of variation (CV). *d*: CV vs. maximum gray value (g_{M}) for all basins. *e*: Same as *d*, but only for basins above 0.5 ml volume; the two circles indicate the empirical rules used for automatic bone and non-bone labeling, respectively. *f*: Close-up, cf. rectangle in *e*, showing 47 false-positives from all 14 data sets for bone labeling rule.



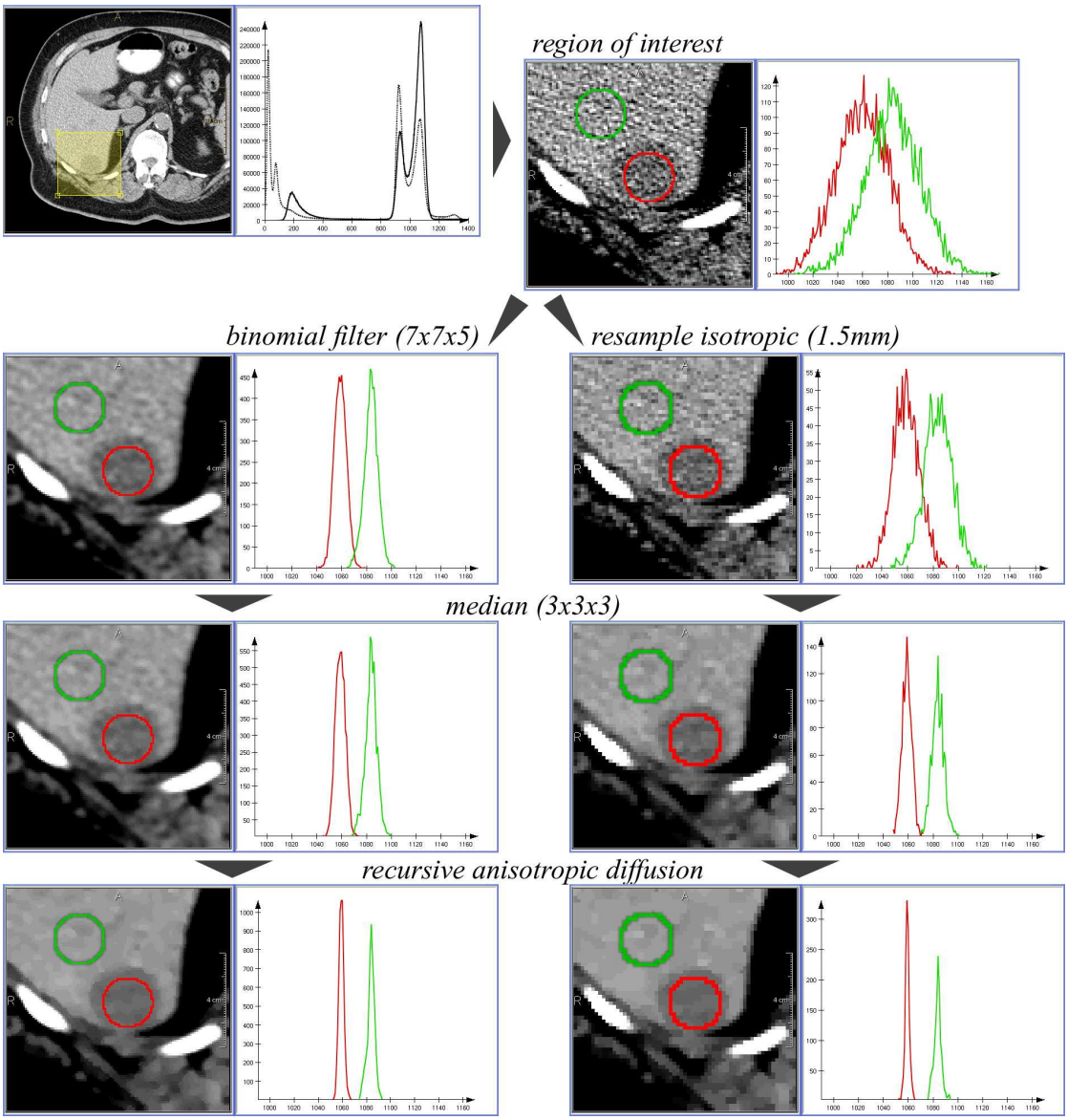
Color Plate C.5: Two screenshots of the proposed bone removal technique in action, with the same data set used in two different systems (CT data courtesy of Clara Soulié, Siemens Medical Solutions, Forchheim). *top*: Application prototype based on MeVisLab (May 2004). *bottom*: Integration into the Syngo InSpace 3D workstation (work-in-progress version, Nov 2004).



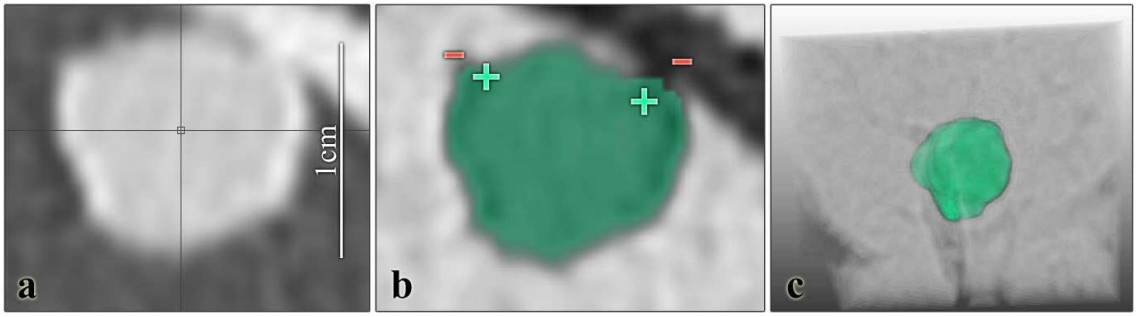
COLOR PLATE C.6: Monte Carlo simulation based on 100.000 uniform cuts (cf. body text, p. 100) of a sphere (blue line with '∗' markers) and a cube (red line with '+' markers) computed with the method by J. M. BECKER [JOLLY, 2004]. The value returned (abscissa) is the volume of the cut part divided by the whole initial volume. (Figures courtesy of BENOÎT JOLLY, Vienne, France)



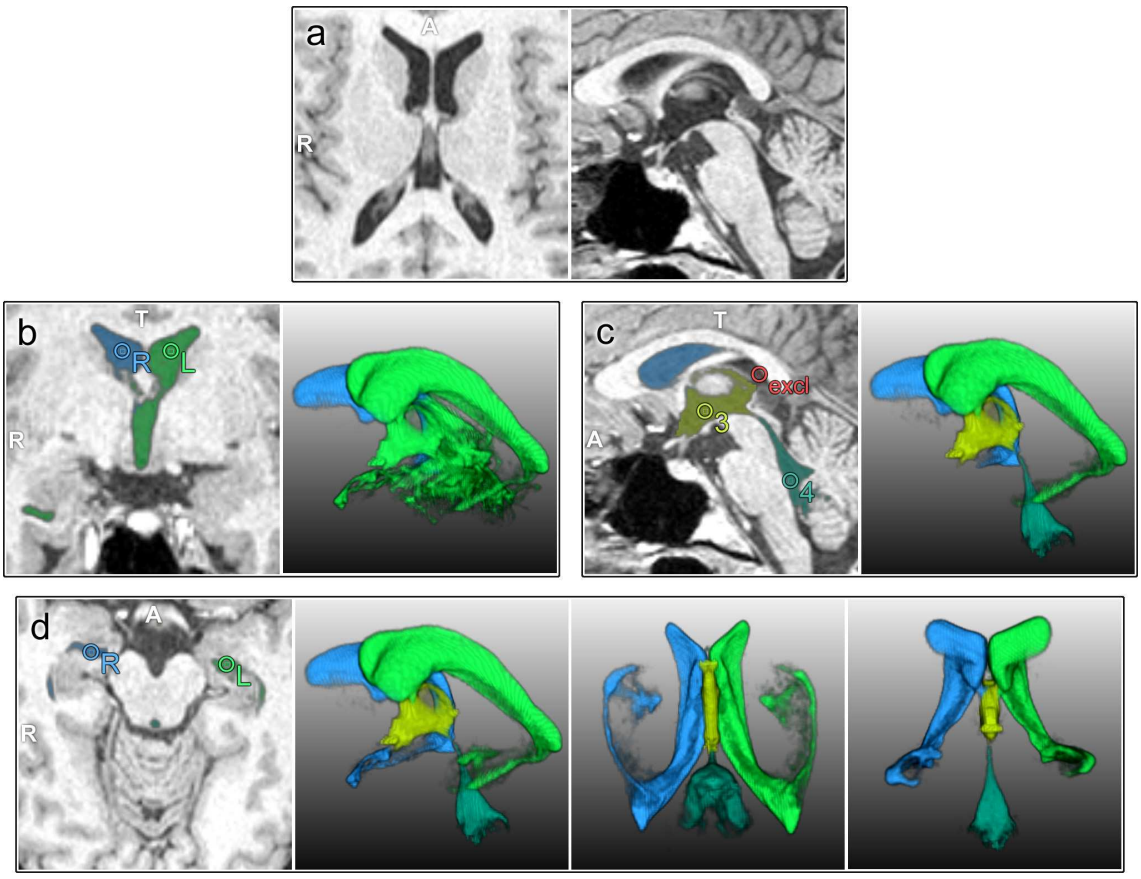
COLOR PLATE C.7: *a*: Theoretical partial volume distribution for the sphere computed analytically (blue line with '∗' markers, cf. body text) and numerically (red line with '+' markers, cf. Color Plate C.6). *b*: Detail of *a* for small partitionings. *c*: Theoretical partial volume distributions for two Gaussian pure tissue distributions (blue line) using either *uniform distribution* of voxel partitioning (green line with '∗' markers; 'flattened plateau') or *uniform cut* of a sphere (red line with '+' markers, 'U-shape'). (Figures courtesy of BENOÎT JOLLY, Vienne, France [JOLLY, 2004])



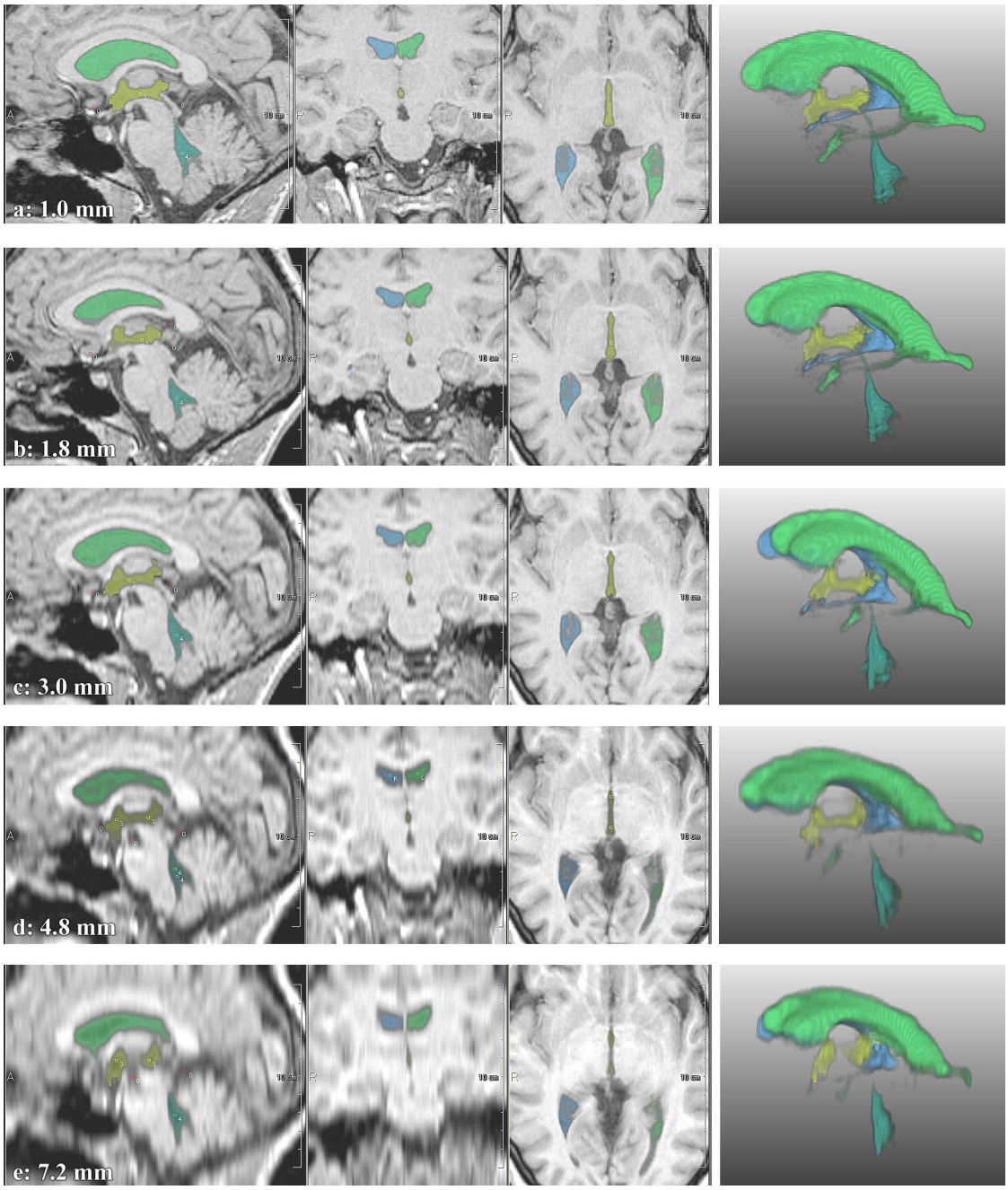
Color Plate C.8: Standard non-contrast thin-slice CT scan of a round liver metastasis at the mid-posterior tip of the right hepatic lobe. For two spherical regions, the regional image histograms are provided: inside (red) and outside (green) of the lesion. On the original data, heavily overlapping distributions corresponding to a very low SNR can be discerned (top right). In order to improve SNR, one can decrease the spatial resolution, either by binomial / Gaussian filtering (2^{nd} row, left) or image downsampling (right). Further SNR improvements are observed after median filtering (3^{rd} row) and successive anisotropic diffusion filtering (bottom). (CT image courtesy of Jon Wiener, Boca Raton)



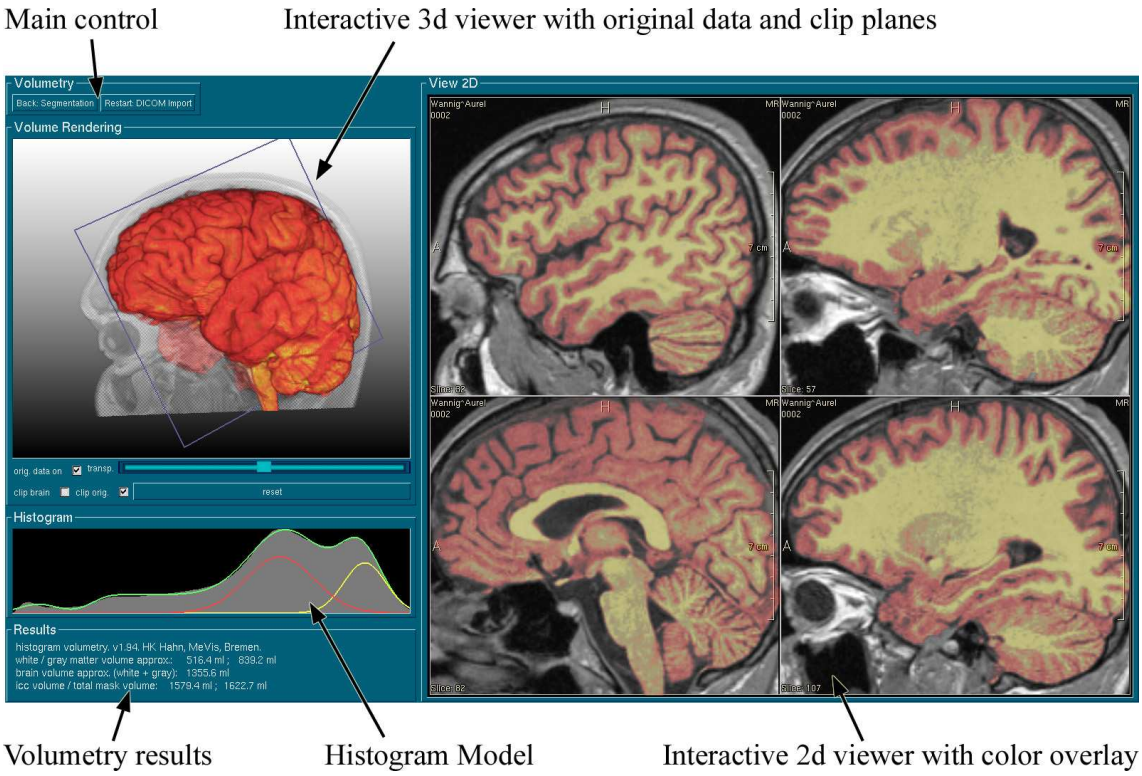
Color Plate C.9: CT image of a lung nodule (a). The nodule was separated from attached vessels using the marker-driven IWT applied to the original data interpreted as depth image (b). The volume of the nodule was computed using a model-based histogram analysis that automatically and robustly estimates partial volume effects (cf. Ch. 7). In (b), the difference between histogram-based border handling (smooth transition according to PVE) and marker-driven watershed enforcement (hard border) is discernible. In (c), a DVR of the segmented object (green) plus original data is shown. (CT image courtesy of Jon Wiener, Boca Raton)



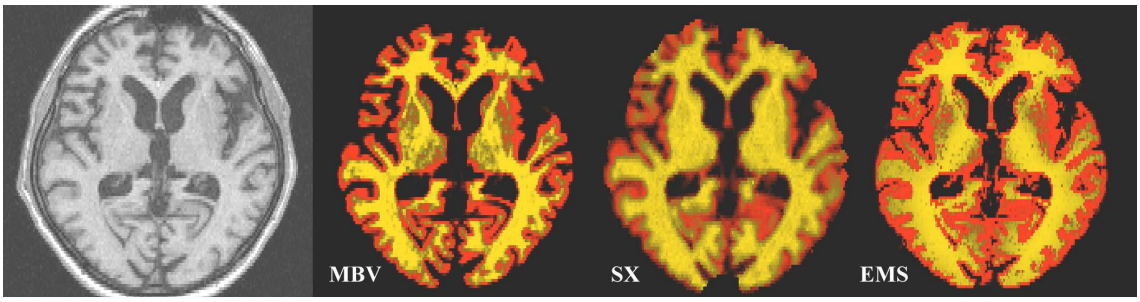
Color Plate C.10: Typical example of interactive segmentation procedure: *a*: original data, *b*: first segmentation step to define frontal horns with two markers, *c*: definition of third and fourth ventricles, additionally excluding non-ventricular CSF, *d*: final step to include lateral horns with two additional markers; in this case, seven markers suffice to accurately define the ventricular anatomy.



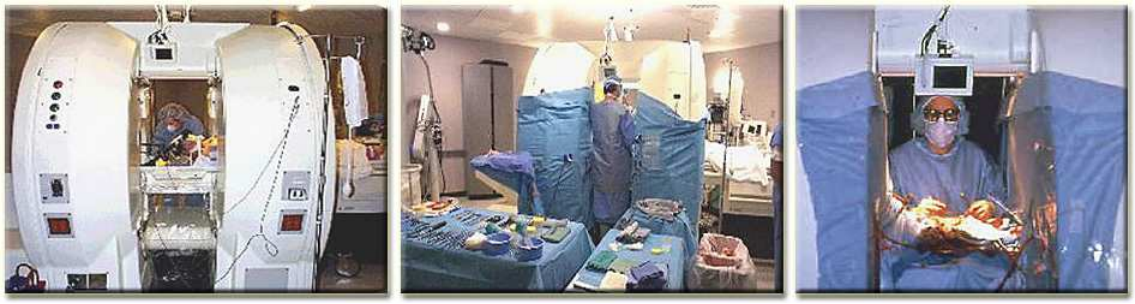
Color Plate C.11: Cerebral ventricular segmentation on a volunteer (F, 29 y) for five different resolutions (axial slicing): 1.0, 1.8, 3.0, 4.8, and 7.2 mm. *left*: Segmentation color overlay on original T1 weighted MR data (courtesy of B. Terwey, Bremen). *right*: Direct volume rendering.



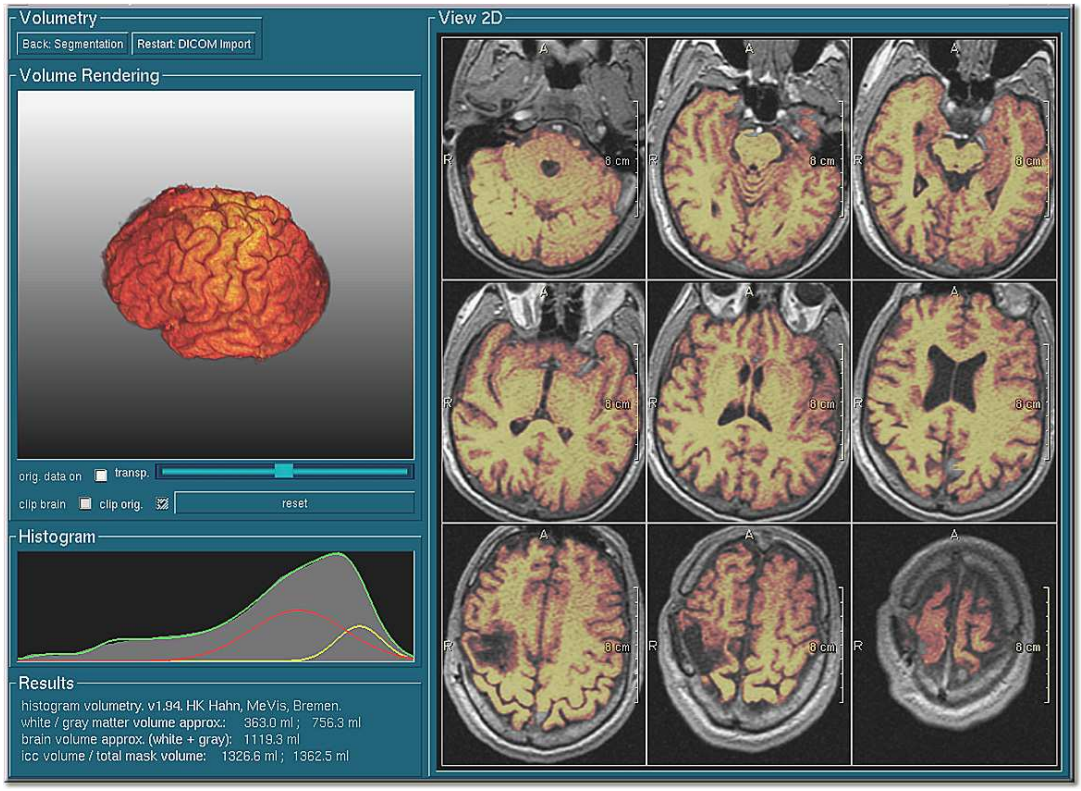
Color Plate C.12: Graphical user interface for exploration of brain volumetry results. Results of model-based histogram analysis as color-coding combined with original data in synchronized 2D and 3D viewers. Quantitative results and major functions of histogram model shown in the lower left corner.



Color Plate C.13: Segmentation results of three methods for AD patient (M, 60 y). Color coding of classified GM (red) and WM (yellow). Note that for MBV, partial volume voxels are not shown. The results were (TBV/GM/WM/BPF) 1004.2 / 560.4 / 443.8 / 68.81 % for MBV, 972.08 / 475.33 / 496.75 / 77.96 % for SIENAX (SX), and 1096.5 / 671.7 / 424.8 / 79.40 % for EMS.



Color Plate C.14: Photographs of the open magnetic resonance tomograph (Signa SP/i, GE Healthcare, 0.5 T "double doughnut", source: www.gehealthcare.com), operated until 2003 at the Department of Diagnostic Radiology, University Hospital Leipzig (head: T. Kahn).



Color Plate C.15: Results panel of a software assistant for analysis of serial brain scan from open MR scanner at 0.5 T. *bottom left*: quantitative results. *top left*: direct volume rendering of skull stripped brain. *left*: histogram of brain region. *right*: original image data with segmentation result as color overlay.

Acknowledgements

All that is needed for evil to triumph
Is for good men to do nothing.

– Edmund Burke

This thesis is the result of a long journey that started in a small office in the main building of M. Yunus' Grameen Bank in Dhaka, Bangladesh. On a pre-monsoon day in May 1999, with an invisible ocean filling the sweet air I received the invitation of Carl J. G. Evertsz and Heinz-Otto Peitgen to come and work with them in Bremen. Looking back, I am incredibly thankful to this being the start of further fortunate incidents and remarkable people crossing my life. Thinking of the ten thousand hours spent with MeVis since then, I received advice and feedback, support and all kinds of help from many colleagues and friends. At MeVis, the three first to name are Ilona Mehrtens, Guido Prause, and Markus Lang. Working without them is hardly imaginable—uncounted things would have been impossible.

Turning to this thesis, first of all my deep thanks go to Heinz-Otto Peitgen for teaching his way of science, for offering insight into his way of thinking, and for giving essential inspiration and much more. Second, I wish to thank Hans Reiber whom I take as a bright example in many ways and who kindly agreed upon reviewing my work. Third, my gratitude goes to all people within the MeVis group who provided help and an excellent basis to realize this thesis. There was virtually no question that remained unanswered and no technical problem that was not solved by one of them. These sometimes fundamentally chaotic strings of cooperation and communication are really fun.

Of the many good hands making up success, I want to name the following. Florian Link for being multitalented and pragmatic in its best sense.Bernhard Preim for showing me how a single person in eight hours can do what five in thirteen can't and for being a great teacher. Sven Kohle and Mathias Schlüter for standing and supporting me every day. Jan Rexilius for listening and never giving up. Olaf Klinghammer for being truly helpful. Olaf Konrad-Verse for his good spirit. Tobias Boskamp for many little things. Holger Bettag for detailed perfection. Sebastian Meyer for laughing, shouting, and singing. Thomas Schindewolf for beginning a short walk, which I prolonged. Andrea Schenk and Volker Dicken for inspiring me to think about the sphere in digital space. Dirk Selle for his patience and frankness. Wolf for teaching me many things, among which are how to make beautiful images and the practical world of image processing. Manfred Georg and Benoît Jolly for asking all these questions, without which the world would be darker. Sven

Hansen, Alexander Schubert, and Miriam Lee for being my students and for trying hard to understand me. Thomas Klink-Telieps and Billy Rustemeyer for being visual, professional, and yet effective. Tobias Preusser, Felix Ritter, Jan-Martin Kuhnigk, Bernd Kümmerlen, Hans Drexl, Stephan Kuhagen, Holger Bourquain, Ulrike Weers, Andreas Bohne, Anja Hennemuth, Gisela Gründl, Milo Hindennach, Richard Rascher-Frisenhausen, Jens Breitenborn, Dominik Böhm, and Jessica Brandt for their kind and indispensable support.

Outside of MeVis, I am indebted to Burckhard Terwey who is more than a radiologist and without whom some of these projects would not even have begun, to Klaus J. Klose who introduced me to his world of radiology, to Uwe Frese and J. Meissner who built the cerebral ventricular paraffin phantom, as well as to Clara Soulié and Axel Barth for discussing and testing the bone removal technique, and for making it work. I am thankful for the personal communication with Ralf Mekle and Tobias Block that was the basis for discussing the sources of error in magnetic resonance imaging. Furthermore, all image analysis methods of this thesis were integrated as modules, networks, scripts, and applications into MeVisLab, a research and development platform dedicated to rapid software prototyping in medical imaging—thanks to all who contributed. Special thanks go to Johannes Ranke, a good friend who provided me with the technical basis to write this manuscript.

I greatly acknowledge the cooperation with and contributions from numerous clinical and academic partners and institutions. Among these are Volker Diehl, Markus G. Lentschig, M. G. Henschel, and Burckhard Terwey, Bremen; Carsten Lukas, Barbara Bellenberg, Gebhard Schmid, Sebastian K. Schimrigk, Horst Przuntek, and Odo Köster, Bochum; Frederik L. Giesel, Bram Stieltjes, Philipp Thomann, M. Essig, and U. Kauczor, Heidelberg; Jon Wiener, Boca Raton; William S. Millar, P. K. Tulipano, C. Chiriboga, and V. J. Hinton, New York; Maureen S. Durkin, Madison; Michael Moche, Wolfgang Heinke, and T. Kahn, Leipzig; Klaus J. Klose, R. Leppek, H. Alfke, and Johannes H. Heverhagen, Marburg; David H. Laidlaw, J. Crisco, Mohammad A. Upal, and Georgeta E. Marai, Providence; Joachim Böttcher and W. A. Kaiser, Jena; R. J. van der Geest, Leiden; J. Debatin, Essen; Alexander Schubert, Leipzig; D. Krastel, Bremen; R. v. Kummer, Dresden; R. Robertson, Boston; P. Stoeter, Mainz; as well as Florent Segonne, Anders Dale, and Bruce Fischl, Charlestown.

It remains to thank Chrystalla Thoma and Carlos Jimenez for being a continuous source of help and inspiration, but most of all for being friends. Schließlich möchte ich denjenigen danken, die mich auf wunderbare Weise so weit gebracht haben, die mir erlaubten, ein Teil von ihnen zu sein, und die ich liebe—meiner Familie, meinen Eltern, und vor allem Dir, Christiane, und Dir, unserem Sohn Jonas.

Horst Karl Hahn
Bremen, 21. Jan. 2005

Bibliography

Hahn, H. K. (1999). Makroskopische Morphometrie und Dynamik der Lebergefäße. Diplomarbeit, Fakultät für Physik, Universität Heidelberg, MeVis internal report.

Hahn, H. K. and Peitgen, H.-O. (2000). The skull stripping problem in MRI solved by a single 3d watershed transform. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 1935, pages 134–143. Springer, Berlin.

Hahn, H. K., Selle, D., Evertsz, C. J. G., and Peitgen, H.-O. (2000a). Interaktive Visualisierung von Gefäßsystemen auf der Basis von Oberflächenprimitiven. In *Proc. SimVis—Simulation & Visualisierung*, pages 105–118. SCS, Ghent.

Hahn, H. K., Selle, D., Evertsz, C. J. G., and Peitgen, H.-O. (2000b). Strukturanalyse und Morphometrie interagierender Gefäßsysteme am Beispiel der menschlichen Leber. In *Proc. BVM*, pages 324–328. Springer, Berlin.

Hahn, H. K. (2001). *A Computer System and Method for Segmentation of a Digital Image.* US SN 09/972,030. patent pending.

Hahn, H. K., Deimling, M., Lentschig, M. G., Terwey, B., and Peitgen, H.-O. (2001a). Ultrafast MRI-based volumetry of intracranial liquor spaces. *NeuroImage*, 13(6, part 2):S140. Proc. HBM—Organization for Human Brain Mapping.

Hahn, H. K., Lentschig, M. G., Deimling, M., Terwey, B., and Peitgen, H.-O. (2001b). MRI-based volumetry of intra- and extracerebral CSF spaces. In *CARS—Computer Aided Radiology and Surgery*, ICS 1230, pages 384–389. Elsevier, Amsterdam.

Hahn, H. K., Lentschig, M. G., Terwey, B., and Peitgen, H.-O. (2001c). Clinical MRI-based volumetry: The cerebral ventricles. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 2208, pages 1291–1292. Springer, Berlin.

Hahn, H. K., Preim, B., Selle, D., and Peitgen, H.-O. (2001d). Visualization and interaction techniques for the exploration of vascular structures. In *Proc. IEEE Visualization*, pages 395–402.

Hahn, H. K., Millar, W. S., Durkin, M. S., Klinghammer, O., and Peitgen, H.-O. (2002a). Cerebral ventricular volumetry in pediatric neuroimaging. In *Proc. BVM*, pages 59–62. Springer, Berlin.

Hahn, H. K., Millar, W. S., Durkin, M. S., Klinghammer, O., Tulipano, P. K., and Peitgen, H.-O. (2002b). A new method for semi-automatic volumetry of cerebral ventricles in pediatric neuroimaging. In *Proc Intl Soc Mag Reson Med*, volume 10, page 358.

Hahn, H. K., Moche, M., Heinke, W., and Peitgen, H.-O. (2002c). New algorithm for precise dynamic brain volumetry on MRI with limited quality. *Radiology*, 255(P):451. Proc. RSNA.

Hahn, H. K., Evertsz, C. J. G., Fasel, J. H. D., and Peitgen, H.-O. (2003a). Fractal properties, segment anatomy and interdependence of the human portal vein and the hepatic vein in 3d. *Fractals*, 11(1):53–62.

Hahn, H. K., Link, F., and Peitgen, H.-O. (2003b). Concepts for rapid application prototyping in medical image analysis and visualization. In *Proc. SimVis—Simulation & Visualisierung*, pages 283–298. SCS, Erlangen.

Hahn, H. K. and Peitgen, H.-O. (2003). IWT – Interactive Watershed Transform: A hierarchical method for efficient interactive and automated segmentation of multidimensional grayscale images. In *Proc. SPIE Med Imaging, Image Processing*, volume 5032, pages 643–653. San Diego.

Hahn, H. K., Jolly, B., Lee, M., Krastel, D., Rexilius, J., Drexl, J., Schlüter, M., Terwey, B., and Peitgen, H.-O. (2004a). How accurate is brain volumetry? A methodological evaluation. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 3216, pages 335–342. Springer, Berlin.

Hahn, H. K., Millar, W. S., Klinghammer, O., Durkin, M. S., Tulipano, P. K., and Peitgen, H.-O. (2004b). A reliable and efficient method for cerebral ventricular volumetry in pediatric neuroimaging. *Methods Inf Med*, 43(4):376–382.

Hahn, H. K., Rexilius, J., Schlüter, M., Terwey, B., Stieltjes, B., Giesel, F. L., and Peitgen, H.-O. (2004c). Fast and robust quantification of parahippocampal atrophy via temporal horn index. In *Proc. BVM*, pages 371–375. Springer, Berlin.

Segonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., and Fischl, B. (2001). A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, 13(6, part 2):S241. Proc. HBM—Organization for Human Brain Mapping.

Schubert, A., Hahn, H. K., and Peitgen, H.-O. (2002). Robust fully automated brain segmentation based on a 3d watershed transform. In *Proc. BVM*, pages 193–196. Springer, Berlin. in German.

Kuhnigk, J.-M., Hahn, H. K., Hindennach, M., Dicken, V., Krass, S., and Peitgen, H.-O. (2003a). 3D-Lungenlappen-Segmentierung durch Kombination von Region Growing,

Distanz- und Wasserscheiden-Transformation. In *Proc. BVM*, pages 146–150. Springer, Berlin.

Kuhnigk, J.-M., Hahn, H. K., Hindennach, M., Dicken, V., Krass, S., and Peitgen, H.-O. (2003b). Lung lobe segmentation by anatomy-guided 3d watershed transform. In *Proc. SPIE Med Imaging, Image Processing*, volume 5032, pages 1482–1490. San Diego.

Rexilius, J., Hahn, H. K., Bourquain, H., and Peitgen, H.-O. (2003). Ground truth in MS lesion volumetry—A phantom study. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 2879, pages 546–553. Springer, Berlin.

Stieltjes, B., Schluter, M., Hahn, H. K., Wilhelm, T., and Essig, M. (2003). Diffusion tensor imaging. Theory, sequence optimization and application in Alzheimer's disease. *Radiologe*, 43(7):562–565. in German.

Apelt, D., Preim, B., Hahn, H. K., and Strauß, G. (2004). Bildanalyse und Visualisierung für die Planung von Nasennebenhöhlen-Operationen. In *Proc. BVM*, pages 194–198. Springer, Berlin.

Drexl, J., Knappe, V., Hahn, H. K., Lehmann, K. S., Frericks, B., Shin, H., and Peitgen, H.-O. (2004). Accuracy analysis of vessel segmentation for a litt dosimetry planning system. In *Proc. MRNV—Medical Robotics, Navigation and Visualization*, pages 204–212. World Scientific.

Heverhagen, J. T., Hahn, H. K., Wegmann, M., Herz, U., Shaffer Whitaker, C. D., Matschl, V., and Alfke, H. (2004). Volumetric analysis of mice lungs in a clinical magnetic resonance imaging scanner. *MAGMA*, 17(2):80–85.

Lukas, C., Hahn, H. K., Bellenberg, B., Rexilius, J., Schmid, G., Schimrigk, S. K., Przuntek, H., Köster, O., and Peitgen, H.-O. (2004). Sensitivity and reproducibility of a new fast 3d segmentation technique for clinical MR-based brain volumetry in multiple sclerosis. *Neuroradiology*, 46:906–915.

Rexilius, J., Hahn, H. K., Schlüter, M., Kohle, S., Bourquain, H., Böttcher, J., and Peitgen, H.-O. (2004). A framework for the generation of realistic brain tumor phantoms and applications. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 3217, pages 243–250. Springer, Berlin.

Schlüter, M., Stieltjes, B., Rexilius, J., Hahn, H. K., and Peitgen, H.-O. (2004). Unique planar color coding of fiber bundles and its application to fiber integrity quantification. In *Proc. ISBI—International Symposium on Biomedical Imaging*. IEEE.

Segonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., and Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, 22(3):1060–1075.

Giesel, F. L., Wüstenberg, T., Bongers, A., Weber, M. A., Zechmann, C., Baudendistel, K. T., von Tengg-Kobligk, H., Hahn, H. K., Essig, M., and Kauczor, H.-U. (2005). MR-based methods of the functional imaging of the CNS. *Fortschr Röntgenstr*, 177:714–730. in German.

Rexilius, J., Hahn, H. K., Schlüter, M., Bourquain, H., and Peitgen, H.-O. (2005). Evaluation of accuracy in MS lesion volumetry using realistic lesion phantoms. *Acad Radiol*, 12(1):17–24.

Schlüter, M., Konrad-Verse, O., Hahn, H. K., Stieltjes, B., Rexilius, J., and Peitgen, H.-O. (2005). White matter lesion phantom for diffusion tensor data and its application to the assessment of fiber tracking. In *Proc. SPIE Med Imaging, Physiology, Function, and Structure from Medical Images*, volume 5746, pages 835–844. San Diego.

Abbreviations

AD	Alzheimer's disease
AAM	Active Appearance Model
ASM	Active Shape Model
BOLD	Blood Oxygen Level Dependent
BPF	Brain Parenchymal Fraction
CA	Contrast Agent
CBF	Cerebral Blood Flow
CBV	Cerebral Blood Volume
CSF	Cerebrospinal Fluid
CT	Computed Tomography
CTA	CT Angiography
CV	Coefficient of Variation
DICOM	Digital Imaging and Communications in Medicine
DTI	Diffusion Tensor Imaging
DVR	Direct Volume Rendering
DWI	Diffusion Weighted Imaging
EPI	Echo Planar Imaging
EM	Expectation Maximization
FOV	Field of View

fMRI Functional MRI

FWT Fast Watershed Transform

GM Gray Matter

HU Hounsfield Units

IFT Image Foresting Transform

IR Inversion Recovery

IVH Intraventricular Hemorrhage

IWT Interactive Watershed Transform

LVV Lateral Ventricular Volume

MBV MeVisLab Brain Volumetry

MCI Mild Cognitive Impairment

MIBS Minimally Interactive Bone Segmentation

MIP Maximum Intensity Projection

MR Magnetic Resonance

MRF Markov Random Field

MRI Magnetic Resonance Imaging

MS Multiple Sclerosis

NPH Normal Pressure Hydrocephalus

NU Nonuniformity

PDF Probability Density Function

PET Positron Emission Tomography

PSF Point Spread Function

PVE Partial Volume Effects

PVL Periventricular Leukomalacia

QR Quantitative Radiology

QCA Quantitative Coronary Arteriography

QCT Quantitative CT

QCU Quantitative Coronary Ultrasound

QIA Quantitative Image Analysis

QLV Quantitative Left Ventricular Analysis

QVA Quantitative Vascular Arteriography

RECIST Response Evaluation Criteria in Solid Tumors

RF Radio Frequency

ROI Region of Interest

SBD Simulated Brain Database

SCDM Skeletally Coupled Deformable Models

SKIZ Skeleton of Influence Zones

SNR Signal-to-Noise Ratio

SPECT Single Photon Emission Computed Tomography

SSD Shaded Surface Display

TBV Total Brain Volume

TE Echo Time

THI Temporal Horn Index

THV Temporal Horn Volume

TI Inversion Time

TR Repetition Time

TSE Turbo Spin Echo

US Ultrasound

VOI Volume of Interest

WASS Watershed Based Skull Stripping

WHO World Health Organization

WM White Matter

WMD White Matter Damage

WT Watershed Transform

Bibliography

G B Aboutanos and B M Dawant. Automatic brain segmentation and validation: Image-based versus atlas-based deformable models. In *Proc. SPIE Med Imaging, Image Processing*, volume 3034, pp. 299–310, Feb 1997.

M Abramowitz and I A Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York. National Bureau of Standards: Applied Mathematics Series 55, Dec 1972.

A M Alyassin and G B Avinash. Semiautomatic bone removal technique from CT angiography data. In *Proc. SPIE Med Imaging*, volume 4322, pp. 1273–1283, 2001.

N C Andreasen, R Rajarethinam, T Cizadlo, S Arndt V W Swayze, 2nd, L A Flashman, D S O'Leary, J C Ehrhardt, and W T Yuh. Automatic atlas-based volume estimation of human brain regions from MR images. *J Comput Assist Tomogr*, 20(1):98–106, Jan 1996.

J B Arnold, J-S Liow, K A Schaper, J J Stern, J G Sled, D W Shattuck, A J Worth, M S Cohen, R M Leahy, J C Mazziotta, and D A Rottenberg. Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *NeuroImage*, 13:931–943, 2001.

M S Atkins and B T Mackiewich. Fully automatic segmentation of the brain in MRI. *IEEE Trans Med Imaging*, 17(1):98–107, Feb 1998.

M A G Ballester, A Zisserman, and M Brady. Segmentation and mesaurement of brain structures in MRI including confidence bounds. *Medical Image Analysis*, 4:189–200, 2000.

W A Barrett and E N Mortensen. Interactive live-wire boundary extraction. *Medical Image Analysis*, 1(4):331—341, 1997.

F Bello, A C F Colchester, and S A Röll. A geometry- and intensity-based partial volume correction for MRI volumetric studies. *Comput Med Imaging Graph*, 22(2):123–132, Mar 1998.

M E Blake, J A Soto, R A Hayes, and J T Ferrucci. Automated volumetry at CT colonography, A phantom study. 12(5):608–613, 2005.

J F Bonneville. Pituitary adenomas: Value of MR imaging. *J Radiol*, 81:939–942, 2000.

T Boskamp, D Rinck, F Link, B Kummerlen, G Stamm, and P Mildenberger. New vessel analysis tool for morphometric quantification and visualization of vessels in CT and MR imaging data sets. *Radiographics*, 24(1):287–297, Jan 2004.

F Brassow and K Baumann. Volume of brain ventricles in man determined by computer tomography. *Neuroradiology*, 16:187–189, 1978.

T R Brown, B M Kincaid, and K Ugurbil. NMR chemical shift imaging in three dimensions. 79:3523–3526, June 1982.

A Brunetti, A Postiglione, E Tedeschi, A Ciarmiello, M Quarantelli, E M Covelli, G Milan, M Larobina, A Soricelli, A Sodano, and B Alfano. Measurement of global brain atrophy in Alzheimer's disease with unsupervised segmentation of spin-echo MRI studies. *J Magn Reson Imaging*, 11(3):260–266, Mar 2000.

J W D Bull. The robert wartenberg memorial lecture: The volume of the cerebral ventricles. *Neurology*, 11:1–9, 1961.

J P Changeux. *Neuronal Man—The Biology of Mind*. Princeton Univ Press, 1985.

C A Cocosco, V Kollokian, R K S Kwan, and A C Evans. Online interface to a 3d MRI simulated brain database. *NeuroImage*, 5(4):part 2/4, S425, May 1997. Proc. HBM—3rd Int Conf on Functional Mapping of the Human Brain.

L D Cohen and I Cohen. Finite-element methods for active contour models and balloons for 2d and 3d images. *IEEE Trans Pattern Analysis Machine Intel*, 15(11):1131–1147, Nov 1993.

D L Collins, A P Zijdenbos, V Kollokian, J G Sled, N J Kabani, C J Holmes, and A C Evans. Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imaging*, 17(3):463–468, June 1998. BrainWeb: `http://www.bic.mni.mcgill.ca/brainweb/` .

B Condon, J Patterson, D Wyper, D Hadley, R Grant, G Teasdale, and J Rowan. Use of magnetic resonance imaging to measure intracranial cerebrospinal fluid volume. *Lancet*, 1(8494):1355–1357, 1986.

T F Cootes, G J Edwards, and C J Taylor. Active appearance models. *IEEE Trans Pattern Analysis Machine Intel*, 23(6):681–685, 2001.

T F Cootes, C J Taylor, D H Cooper, and J Graham. Active shape models—their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan 1995.

N Cristianini and J Shaw-Taylor. *An Introduction to Support Vector Machines and other Kernel-Based Learning Methods*. University Press, Cambridge, 2000.

A M Dale, B Fischl, and M I Sereno. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, 9:179–194, 1999.

R Damadian, R Goldsmith, and L Minkhoff. NMR in cancer. XVI. FONAR image of the live human body. *Physiol Chem Phys*, 9:97–108, 1977.

I de Scheerder, F de Man, M C Herregods, H de Geest, and J Piessens. The use of intracoronary ultrasound for quantitative assessment of coronary artery lumen diameter and for on-line evaluation of angioplasty results. *Acta Cardiol*, 48(2):171–181, 1993.

N De Stefano, P M Matthews, M Filippi, F Agosta, M De Luca, M L Bartolozzi, L Guidi, A Ghezzi, E Montanari, A Cifelli, A Federico, and S M Smith. Evidence of early cortical atrophy in MS: Relevance to white matter changes and disability. *Neurology*, 60(7):1157–1162, Apr 2003.

C DeCarli, J Maisog, D G Murphy, D Teichberg, S I Rapoport, and B Horwitz. Method for quantification of brain, ventricular, and subarachnoid CSF volumes from MR images. *J Comput Assist Tomogr*, 16(2):274–284, Mar 1992.

J Dequeker. Quantitative radiology: Radiogrammetry of cortical bone. *Br J Radiol*, 49 (587):912–920, Nov 1976.

H Digabel and C Lantuéjoul. Iterative algorithms. In J L Chermant, editor, *Proc. 2nd European Symp Quantitative Analysis of Micro-structures in Material Science, Biology and Medicine*, pp. 85–99. Riederer, Stuttgart, 1978.

J Ehrhardt, H Handels, B Strathmann, T Malina, W Plötz, and S J Pöppl. Atlas-based recognition of anatomical structures and landmarks to support the virtual three-dimensional planning of hip operations. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 2878, pp. 17–24. Springer, Berlin, Oct 2003.

P C Everett, E B Seldin, M Troulis, L B Kaban, and R Kikinis. A 3d system for planning and simulating minimally-invasive distraction osteogenesis of the facial skeleton. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 1935, pp. 1029–1039. Springer, Berlin, Oct 2000.

A X Falcao, J Stolfi, and R A Lotufo. The image foresting transform: Theory, algorithms, and applications. *IEEE Trans Pattern Analysis Machine Intel*, 26(1):19–29, Jan 2004.

M Fiebich, C M Straus, V Sehgal, B C Renger, K Doi, and K R Hoffmann. Automatic bone segmentation technique for CT angiographic studies. *J Comput Assist Tomogr*, 23 (1):155–161, Jan 1999.

J D Foley, A van Dam, S K Feiner, and J F Hughes. *Computer Graphics: Principles and Practice*. Addison-Wesley, Reading, MA, 2nd edition, 1990.

N C Fox, S Cousens, R Scahill, R J Harvey, and M N Rossor. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: Power

calculations and estimates of sample size to detect treatment effects. *Arch Neurol*, 57 (3):339–344, Mar 2000.

N C Fox, P A Freeborough, and M N Rossor. Visualisation and quantification of rates of atrophy in Alzheimer's disease. *Lancet*, 348(9020):94–97, July 1996.

R S J Frackowiak, K J Friston, C D Frith, R J Dolan, and J C Mazziotta. *Human Brain Function*. Academic Press, San Diego, 1997. Statistical Parametric Mapping SPM, `http://www.fil.ion.ucl.ac.uk/` .

P A Freeborough, N C Fox, and R I Kitney. Interactive algorithms for the segmentation and quantification of 3d MRI brain scans. *Comput Meth Progr Biomed*, 53(1):15–25, May 1997.

M Friedlinger, J Schroder, and L R Schad. Ultra-fast automated brain volumetry based on bispectral MR imaging data. *Comput Med Imaging Graph*, 23(6):331–337, Nov 1999.

J M Gauch. Image segmentation and analysis via multiscale gradient watershed hierarchies. *IEEE Trans Image Proc*, 8(1):69–79, Jan 1999.

A F Goldszal, C Davatzikos, D L Pham, M X H Yan, R N Bryan, and S M Resnick. An image processing system for qualitative and quantitative volumetric analysis of brain images. *J Comput Assist Tomogr*, 22(5):827–837, 1998.

E H Gombrich. *The Sense of Order—A Study in the Psychology of Decorative Art*. Phaidon Press, London, 2nd edition, 1984.

R Grant, B Condon, A Lawrence, D M Hadley, J Patterson, I Bone, and G M Teasdale. Human cranial CSF volumes measured by MRI: Sex and age influences. *Magn Reson Imaging*, 5:465–468, 1987.

V Grau, A U J Mewes, M Alcañiz, R Kikinis, and S K Warfield. Improved watershed transform for medical image segmentation using prior information. *IEEE Trans Med Imaging*, 23(4):447–458, Apr 2004.

H Gudbjartsson and S Patz. The Rician distribution of noisy MRI data. *Magn Reson Med*, 34(6):910–914, 1995. (erratum appears in Magn Reson Med 36(2): 332, Aug 1996).

E M Haacke, R W Brown, M R Thompson, and R Venkatesan. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley & Sons, New York, 1999.

H Hacker and H Artmann. The calculation of CSF spaces in CT. *Neuroradiology*, 16: 190–192, 1978.

H K Hahn and H-O Peitgen. The skull stripping problem in MRI solved by a single 3d watershed transform. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 1935, pp. 134–143. Springer, Berlin, 2000.

H K Hahn. *A Computer System and Method for Segmentation of a Digital Image*. US SN 09/972,030, 5 Oct 2001. patent pending.

H K Hahn, M G Lentschig, M Deimling, B Terwey, and H-O Peitgen. MRI-based volumetry of intra- and extracerebral CSF spaces. In *CARS—Computer Aided Radiology and Surgery*, ICS 1230, pp. 384–389. Elsevier, Amsterdam, 2001a.

H K Hahn, M G Lentschig, B Terwey, and H-O Peitgen. Clinical MRI-based volumetry: The cerebral ventricles. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 2208, pp. 1291–1292. Springer, Berlin, 2001b.

H K Hahn, M Moche, W Heinke, and H-O Peitgen. New algorithm for precise dynamic brain volumetry on MRI with limited quality. *Radiology*, 255(P):451, Nov 2002. Proc. RSNA.

H K Hahn, F Link, and H-O Peitgen. Concepts for rapid application prototyping in medical image analysis and visualization. In *Proc. SimVis—Simulation & Visualisierung*, pp. 283–298. SCS, Erlangen, Mar 2003.

H K Hahn and H-O Peitgen. IWT – Interactive Watershed Transform: A hierarchical method for efficient interactive and automated segmentation of multidimensional grayscale images. In *Proc. SPIE Med Imaging, Image Processing*, volume 5032, pp. 643–653. San Diego, Feb 2003.

H K Hahn, B Jolly, M Lee, D Krastel, J Rexilius, J Drexl, M Schlüter, B Terwey, and H-O Peitgen. How accurate is brain volumetry? A methodological evaluation. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 3216, pp. 335–342. Springer, Berlin, 2004a.

H K Hahn, W S Millar, O Klinghammer, M S Durkin, P K Tulipano, and H-O Peitgen. A reliable and efficient method for cerebral ventricular volumetry in pediatric neuroimaging. *Methods Inf Med*, 43(4):376–382, 2004b.

H K Hahn, J Rexilius, M Schlüter, B Terwey, B Stieltjes, F L Giesel, and H-O Peitgen. Fast and robust quantification of parahippocampal atrophy via temporal horn index. In *Proc. BVM*, pp. 371–375. Springer, Berlin, Mar 2004c.

K Haris, S N Efstratiadis, N Maglaveras, and A K Katsaggelos. Hybrid image segmentation using watersheds and fast region merging. *IEEE Trans Image Proc*, 7(12):1684–1699, Dec 1998.

J T Heverhagen, A Battmann, M Kirsch, D Boehm, R Eissele, K J Klose, and H J Wagner. Magnetic resonance hydrometry: Non-invasive quantification of the exocrine pancreatic function. *Fortschr Röntgenstr*, 174:291–296, 2002.

J T Heverhagen, D Müller, A Battmann, N Ishaque, D Boehm, M Katschinski, H J Wagner, and K J Klose. MR hydrometry to assess exocrine function of the pancreas: Initial results of noninvasive quantification of secretion. *Radiology*, 218:61–67, 2001.

K H Hoehne and W A Hanson. Interactive 3d segmentation of MRI and CT volumes using morphological operations. *J Comput Assist Tomogr*, 16(2):185–294, 1992.

R E Hogan, K E Mark, L Wang, S Joshi, M I Miller, and R D Bucholz. Mesial temporal sclerosis and temporal lobe epilepsy: MR imaging deformation-based segmentation of the hippocampus in five patients. *Radiology*, 216(1):291–297, July 2000.

S A Hojjatoleslami, F Kruggel, and D Y von Cramon. Segmentation of white matter lesions from volumetric MR images. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 1679, pp. 52–61. Springer, Berlin, 1999.

M Holden. *Registration of 3D serial MR brain images.* PhD thesis, King's College London, University of London, Mar 2001.

G N Holland, W S Moore, and R C Hawkes. Nuclear magnetic resonance tomography of the brain. *J Comput Assist Tomogr*, 4:1–3, 1980.

G N Hounsfield. Computerized transverse axial scanning (tomography): Part I. Description of a system. *Br J Radiol*, 46:1016–1022, 1973.

P S Huppi, S Warfield, R Kikinis, P D Barnes, G P Zientara, F A Jolesz, M K Tsuji, and J J Volpe. Quantitative magnetic resonance imaging of brain development in premature and mature newborns. *Ann Neurol*, 43(2):224–235, 1998.

T E Inder, P S Huppi, S Warfield, R Kikinis, G P Zientara, P D Barnes, F Jolesz, and J J Volpe. Periventricular white matter injury in the premature infant is followed by reduced cerebral cortical gray matter volume at term. *Ann Neurol*, 46(5):755–760, 1999.

E J Isaac and R C Singleton. Sorting by address calculation. *J ACM*, 3:169–174, 1956.

C R Jack, M Slomkowski, S Geacon, T M Hoover abd J P Felmlee, J Stewart, Y Xu, M Shiung, P C O'Brien, R Cha, D Knopman, and R C Petersen. MRI as a biomarker of disease progression in a therapeutic trial of milameline for AD. *Neurology*, 60(2): 253–260, Jan 2003.

L A Johnson, J D Pearlman, C A Miller, T I Young, and K R Thulborn. MR quantification of cerebral ventricular volume using a semiautomated algorithm. *Am J Neuroradiol*, 14: 1373–1378, Nov 1993.

B Jolly. *Cerebral Volumetry Computation by Histogram Analysis on T1 Weighted MR Scans.* Master's thesis, ESCPE—Ecole Supérieure de Chimie Physique et Electronique, Lyon, June 2004. Work carried out at MeVis, Bremen.

W A Kalender, D Felsenberg, O Louis, P Lopez, E Klotz, M Osteaux, and J Fraga. Reference values for trabecular and cortical vertebral bone density in single and dual-energy quantitative computed tomography. *Eur J Radiol*, 9(2):75–80, May 1989.

J Kallrath and A P Linnell. A new method to optimize parameters in solutions of eclipsing binary light curves. *The Astrophys J*, 313:346—357, Feb 1987.

Y Kang, K Engelke, and W A Kalender. A new accurate and precise 3-D segmentation method for skeletal structures in volumetric CT data. *IEEE Trans Med Imaging*, 22(5): 586–598, May 2003.

T Kapur, W E L Grimson, W M Wells, and R Kikinis. Segmentation of brain tissue from magnetic resonance images. *Medical Image Analysis*, 1(2):109–127, 1996.

M Kass, A Witkin, and D Terzopoulos. Snakes: Active contour models. *Int J Computer Vision*, 1(4):321–331, Jan 1988.

T E Keats and C Sistrom. *Atlas of Radiologic Measurement*. Harcourt Health Sciences, Philadelphia, PA, 7th edition, 2001.

R Kikinis, M E Shenton, G Gerig, J Martin, M Anderson, D Metcalf, C R G Guttmann, R W McCarley, W Lorensen, H Cline, and F A Jolesz. Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging. *J Magn Reson Imaging*, 2: 619–629, 1992.

K J Klose. *Personal communication with the author*. Marburg–Bremen, Apr–May 2002.

M I Kohn, N K Tanna, G T Herman, S M Resnick, P D Mozley, R E Gur, A Alavi, R A Zimmerman, and R C Gur. Analysis of brain and cerebrospinal fluid volumes with MR imaging. Part I. Methods, reliability, and validation. *Radiology*, 178:115–122, 1991.

O Konrad-Verse. *Interactive Refinement of a Balloon Segmentation on Medical Data*. Master's thesis, University of Bremen, Jan 2004.

J M Kuhnigk, V Dicken, L Bornemann, D Wormanns, S Krass, and H-O Peitgen. Fast automated segmentation and reproducible volumetry of pulmonary metastases in CT-scans for therapy monitoring. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 3217, pp. 933–941. Springer, Berlin, 2004.

D H Laidlaw. *Geometric model extraction from magnetic resonance volume data*. PhD thesis, Calif. Inst. Technol., Pasadena, 1995.

D H Laidlaw, K W Flescher, and H A Barr. Partial-volume bayesian classification of material mixtures in MR volume data using voxel histograms. *IEEE Trans Med Imaging*, 17:74–86, Feb 1998.

P C Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242:190–191, 1973.

B P F Lelieveldt, S C Mitchell, J G Bosch, R J van der Geest, M Sonka, and J H C Reiber. Time-continuous segmentation of cardiac image sequences using acrive appearance motion models. In *Proc. IPMI—Information Processing in Medical Imaging*, LNCS 2082, pp. 446–452. Springer, Berlin, 2001.

L Lemieux, G Hagemann, K Krakow, and F G Woermann. Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. *Magn Reson Med*, 42:127–135, 1999.

M M J Letteboer, W J Niessen, P W A Willems, E B Dam, and M A Viergever. Interactive multi-scale watershed segmentation of tumors in MR brain images. In *Proc. Workshop on Interactive Medical Image Visualization and Analysis*, pp. 11–16. Utrecht, Oct 2001.

B Likar, M A Viergever, and F Pernus. Retrospective correction of MR intensity inhomogeneity by information minimization. *IEEE Trans Med Imaging*, 20(12):1398–1410, Dec 2001.

C Lukas, H K Hahn, B Bellenberg, J Rexilius, G Schmid, S K Schimrigk, H Przuntek, O Köster, and H-O Peitgen. Sensitivity and reproducibility of a new fast 3d segmentation technique for clinical MR-based brain volumetry in multiple sclerosis. *Neuroradiology*, 46(11):906–915, 2004.

A Macovski. Noise in MRI. *Magn Reson Med*, 36(3):494–497, 1996.

F Maes, K Van Leemput, L E DeLisi, D Vandermeulen, and P Suetens. Quantification of cerebral gray and white matter asymmetry from MRI. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 1679, pp. 348–357. Springer, Berlin, 1999.

J F Mangin. Entropy minimization for automatic correction of intensity nonuniformity. In *IEEE Work. MMBIA*, pp. 162–169. IEEE Press, 2000.

P Marais and J W Brady. Detecting the brain surface in sparse MRI using boundary models. *Medical Image Analysis*, 4:283–302, 2000.

A Marmarou, K Shulman, and R M Rosende. A nonlinear analysis of the cerebrospinal fluid system and intracranial pressure dynamics. *J Neurosurg*, 48:332–344, 1978.

D Matthaei, J Frahm, A Haase, K D Merboldt, and W Hanicke. Multipurpose NMR imaging using stimulated echoes. *Magn Reson Med*, 3(4):554–561, Aug 1986.

E R McVeigh, R M Henkelman, and M J Bronskill. Noise and filtration in magnetic resonance imaging. *Medical Physics*, 12(5):586–591, 1985.

E R Melhem, A H Hoon, Jr., J T Ferrucci, Jr., C B Quinn, E M Reinhardt, S W Demetrides, and B M Freeman M V Johnston. Periventricular leukomalacia: Relationship between lateral ventricular volume on brain MR images and severity of cognitive and motor impairment. *Radiology*, 214(1):199–204, 2000.

MeVis. Bremen, 2004. MeVisLab development environment available at: `http://www.mevislab.de/` .

F Meyer. Topographic distance and watershed lines. *Signal Processing*, 38:113–125, 1994.

F Meyer and C Vachier. Image segmentation based on viscous flooding simulation. In H. Talbot and R. Bear, editors, *Proc. ISMM*, pp. 69–77. CSIRO, Sydney, Apr 2002.

G Mittelhaeusser and F Kruggel. Fast segmentation of brain magnetic resonance tomograms. In *Proc. CVR—1st Conf on Computer Vision, Virtual Reality and Robotics in Medicine*, LNCS 905. Springer, Berlin, 1995.

A N Moga. *Parallel Watershed Algorithms for Image Segmentation*. PhD thesis, Tampere Univ of Technology, Finland, Feb 1997.

A N Moga. Parallel multiresolution image segmentation with watershed transformation. In *Proc. ACPC*, LNCS 1557, pp. 226–235. Springer, Berlin, 1999.

E A Moore, J P Grieve, and H R Jäger. Robust processing of intracranial CT angiograms for 3D volume rendering. *Eur J Radiol*, 11(1):137–141, Jan 2001.

I A Morocz, G P Zientara, H Gudbjartsson, S Muza, T Lyons, P B Rock, R Kikinis, and F A Jolesz. Volumetric quantification of brain swelling after hypobaric hypoxia exposure. *Exp Neurol*, 168:96–104, 2001.

R Mullick, R Avila, J Knoplioch, Y Mallya, J Platt, and R Senzig. Automatic bone removal for abdomen CTA: A clinical review. In *Proc. RSNA*, Dec 2002.

A Nabavi, P M L Black, D T Gering, C F Westin, V Metha, R S Pergolizzi Jr, M Ferrant, S K Warfield, N Hata, R B Schwarts, W M Wells, R Kikinis, and F A Jolesz. Serial intraoperative mr imaging of brain shift. *Neurosurgery*, 48:787–798, 2001.

F H Netter. *Atlas of Human Anatomy*. Ciba-Geigy Corp., Summit, NJ, 1989.

C P Neu, R D McGovern, and J J Crisco. Kinematic accuracy of three surface registration methods in a 3d wrist bone study. *J Biomech Eng*, 122(5):528–533, Oct 2000.

H T Nguyen, M Worring, and R van den Boomgaard. Watersnakes: Energy-driven watershed segmentation. *IEEE Trans Pattern Analysis Machine Intel*, 25(3):330–342, Mar 2003.

D A Norman. *The Psychology of Everyday Things*. Basic Books, New York, 1988.

L G Nyul and J K Udupa. On standardizing the MR image intensity scale. *Magn Reson Med*, 42(6):1072–1081, Dec 1999.

S Ogawa, T M Lee, A R Kay, and D W Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci USA*, 87(24):9868–9872, Dec 1990.

S D Olabarriaga. *Human-Computer Interaction for the Segmentation of Medical Images*. PhD thesis, Univ of Amsterdam, Dec 1999. ISBN 90-5776-034-7.

S D Olabarriaga and A W M Smeulders. Setting the mind for intelligent interactive segmentation. In *Proc. IPMI—Information Processing in Medical Imaging*, LNCS 1230, pp. 417–422. Springer, Berlin, 1997.

S D Olabarriaga and A W M Smeulders. Interaction in the segmentation of medical images: A survey. *Medical Image Analysis*, 5:127–142, 2001.

W W Orrison, J D Lewine, J A Sanders, and M F Hartshorne. *Functional Brain Imaging*. Mosby–Year Book, St. Louis, 1995.

L Page and S Brin. *Google*. `http://www.google.com/` , since 1998.

K S Pentlow, D A Rottenberg, and M D F Deck. Partial volume summation: A simple approach to ventricular volume determination from CT. *Neuroradiology*, 16:130–132, 1978.

R S Pergolizzi Jr, A Nabavi, B B Schwartz, L Hsu, T Z Wong, C Martin, P M Black, and F A Jolesz. Intra-operative MR guidance during trans-sphenoidal pituitary resection: Preliminary results. *J Magn Reson Imaging*, 13:136–141, 2001.

B S Peterson, B Vohr, L H Staib, C J Cannistraci, A Dolberg, K C Schneider, K H Katz, M Westerveld, S Sparrow, A W Anderson, C C Duncan, R W Makuch, J C Gore, and L R Ment. Regional brain volume abnormalities and long-term cognitive outcome in preterm infants. *JAMA*, 284(15):1939–1947, 2000.

D L Pham and J L Prince. Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Trans Med Imaging*, 18(9):737–752, 1999.

J G Pipe. Motion correction with PROPELLER MRI: Application to head motion and free-breathing cardiac imaging. *Magn Reson Med*, 42(5):963–969, Nov 1999.

S M Pizer, C Sarang, P Joshi, T Fletcher, M Styner, G Tracton, and J Z Chen. Segmentation of single-figure objects by deformable M-reps. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 2208, pp. 862–871, 2001.

R Pohle. *Computerunterstützte Bildanalyse zur Auswertung medizinischer Bilddaten*. Habilitationsschrift, Otto-von-Guericke University, Magdeburg, Mar 2004.

M Pokric, N A Thacker, M L Scott, and A Jackson. The importance of partial voluming in multi-dimensional medical image segmentation. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 2208, pp. 1293–1294, 2001.

S R Prasad, K S Jhaveri, S Saini, P F Hahn, E F Halpern, and J E Sumner. CT tumor measurement for therapeutic response assessment: Comparison of unidimensional, bidimensional, and volumetric techniques—initial observations. *Radiology*, 225:416–419, 2002.

W H Press, S A Teukolsky, W T Vetterling, and B P Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988–1995.

R Raman, B Raman, W Hundt, D Stucker, S Napel, and G D Rubin. Improved speed of bone removal in CT angiography (CTA) using automated targeted morphological separation: Method and evaluation in CTA of lower extremity occlusive disease (LEOD). *Radiology*, 225(P):647, 2002.

K Rehm, D Shattuck, R Leahy, K Schaper, and D Rottenberg. Semi-automated stripping of T1 MRI volumes: I. Consensus of intensity- and edge-based methods. *NeuroImage*, 1999. Proc. HBM—5th Int Conf on Functional Mapping of the Human Brain, poster no. 86.

J H Reiber, B Goedhart, G J Brand, L Schiemanck, and P M van der Zwet. Quantitative coronary arteriography: Current status and future. *Heart Vessels*, suppl 12:209–211, 1997.

J H Reiber, P van Eldik-Helleman, C J Kooijman, J G Tijssen, and P W Serruys. How critical is frame selection in quantitative coronary angiographic studies? *Eur Heart J*, 10(Suppl F):54–59, Nov 1989.

J Rexilius, H K Hahn, H Bourquain, and H-O Peitgen. Ground truth in MS lesion volumetry—A phantom study. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 2879, pp. 546–553. Springer, Berlin, 2003.

J Rexilius, H K Hahn, M Schlüter, H Bourquain, and H-O Peitgen. Evaluation of accuracy in MS lesion volumetry using realistic lesion phantoms. *Acad Radiol*, 12(1):17–24, Jan 2005.

S O Rice. Statistical properties of a sine-wave plus random noise. *Bell System Technical Journal*, 27(1):109–157, 1948.

N Roberts, M J Puddephat, and V McNulty. The benefit of stereology for quantitative radiology. *Br J Radiol*, 73(871):679–697, July 2000. Review.

J B T M Roerdink and A Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*, 41:187–228, 2000.

R Rosen. *Essays on Life Itself*. Columbia University Press, New York, 2000.

R A Rudick, E Fisher, J C Lee, J Simon, and L Jacobs. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS. Multiple sclerosis collaborative research group. *Neurology*, 53(8):1698–1704, Nov 1999.

N Saeed, B K Puri, A Oatridge, J V Hajnal, and I R Young. Two methods for semi-automated quantification of changes in ventricular volume and their use in schizophrenia. *Magn Reson Imaging*, 16(10):1237–1247, 1998.

W D Sager, G Gell, G Ladurner, and P W Ascher. Calculation of cerebral tissue and cerebrospinal fluid space volumes from computer tomograms. *Neuroradiology*, 16:176–178, 1978.

S Sandor and R Leahy. Surface-based labeling of cortical anatomy using a deformable atlas. *IEEE Trans Med Imaging*, 16(1):41–54, Feb, 1997.

P Santago and H D Gage. Quantification of MR brain images by mixture density and partial volume modeling. *IEEE Trans Med Imaging*, 12(3):566–574, Sept 1993.

R Saxena, S G Zachariah, and J E Sanders. Processing computer tomography bone data for prosthetic finite element modeling: A technical note. *J Rehabilitation Research and Development*, 39(5):609–614, Sept 2002.

A Schenk, J Breitenborn, D Selle, T Schindewolf, D Böhm, W Spindler, H Jürgens, and H-O Peitgen. ILabMed-Workstation—A development platform for radiologic applications. In *Proc. BVM*, pp. 238–242. Springer, Berlin, Mar 1999. in German.

A Schenk, G Prause, and H-O Peitgen. Efficient semiautomatic segmentation of 3d objects in medical images. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 1935, pp. 186–195, 2000.

P Scheunders and J Sijbers. Multiscale watershed segmentation of multivalued images. In *Proc. 16th Int Conf on Pattern Recognition, Quebec, Canada*, Aug 2002.

L Schierlitz, P S Hüppi, M Jakab, R Kikinis, M C Frates, C McTempany, and F A Jolesz. Three-dimensional modeling and volume assessment of the fetal and neonatal intracranial ventricles. *Proc Intl Soc Mag Reson Med*, 9:402, 2001.

T Schindewolf, U Frese, and J Meissner. Segmentierung und Volumetrie der Hirnventrikel mit MRT-Datensätzen. In *Proc. BVM*, pp. 92–96. Springer, Berlin, Mar 1999.

T Schindewolf and H-O Peitgen. Interactive image segmentation of CT and MR images based on a modified hierarchical watershed transformation. In *Proc. BVM*, pp. 96–100. Springer, Berlin, Mar 2000.

M Schlüter, O Konrad-Verse, H K Hahn, B Stieltjes, J Rexilius, and H-O Peitgen. White matter lesion phantom for diffusion tensor data and its application to the assessment of fiber tracking. In *Proc. SPIE Med Imaging, Physiology, Function, and Structure from Medical Images*, volume 5746, pp. 835–844. San Diego, Feb 2005.

A Schubert. *Clinical Neuroanesthesia*. Butterworth-Heinemann, Washington, 1997.

A Schubert, H K Hahn, and H-O Peitgen. Robust fully automated brain segmentation based on a 3d watershed transform. In *Proc. BVM*, pp. 193–196. Springer, Berlin, Mar 2002. in German.

L H Schwartz, M S Ginsberg, D DeCorato, L N Rothenberg, S Einstein, P Kijewski, and D M Paniecek. Evaluation of tumor measurements in oncology: Use of film-based and electronic techniques. *J Clin Oncology*, 18(10):2179–2184, May 2000.

T B Sebastian, H Tek, J J Crisco, S W Wolfe, and B B Kimia. Segmentation of carpal bones from 3d CT images using skeletally coupled deformable models. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 1496, pp. 1184–1194. Springer, Berlin, Oct 1998.

F Segonne, A M Dale, E Busa, M Glessner, D Salat, H K Hahn, and B Fischl. A hybrid approach to the skull stripping problem in MRI. *NeuroImage*, 22(3):1060–1075, July 2004.

J Serra. *Image Analysis and Mathematical Morphology*. Academic, London, 1982.

D Shen, S Moffat, S Resnick, and C Davatzikos. Measuring size and shape of the hippocampus in MR images using a deformable shape model. *NeuroImage*, 12(2):422–434, Feb 2002.

B Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer-Interaction*. Addison-Wesley, Reading, MA, 3rd edition, 1997.

J Sijbers. *Signal and noise estimation from Magnetic Resonance Images*. PhD thesis, Vision Lab, University of Antwerp, May 1998.

L C Silbert, J F Quinn, M M Moore, E Corbridge, M J Ball, G Murdoch, G Sexton, and J A Kaye. Changes in premorbid brain volume predict Alzheimer's disease pathology. *Neurology*, 61(4):487–492, Aug 2003.

J G Sled. *A non-parametric method for automatic correction of intensity non-uniformity in MRI data*. Master's thesis, McGill University, Montreal, QC, May 1997.

J G Sled, A P Zijdenbos, and A C Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging*, 17(1):87–97, Feb 1998.

S Smith, Y Zhang, M Jenkinson, J Chen, P Matthews, A Federico, and N De Stefano. Accurate, robust and automated longitudinal and cross-sectional brain change analysis. *NeuroImage*, 17(1):479–489, 2002. `http://www.fmrib.ox.ac.uk/fsl/` .

J G Snel, H W Venema, and C A Grimbergen. Deformable triangular surfaces using fast 1D radial lagrangian dynamics—Segmentation of 3d MR and CT images of the wrist. *IEEE Trans Med Imaging*, 21(8):888–903, Aug 2002.

J G Snel, H W Venema, T M Moojen, , J P Ritt, C A Grimbergen, and G J den Heeten. Quantitative in vivo analysis of the kinematics of carpal bones from three-dimensional CT images using a deformable surface model and a three-dimensional matching technique. *Med Phys*, 29(9):2037–2047, Sept 2000.

P Soille. *Morphological Image Analysis*. Springer, Berlin, 2nd edition, 2003.

G Stalidis, N Maglaveras, S N Efstratiadis, A S Dimitriadis, and C Pappas. Model-based processing scheme for quantitative 4-d cardiac MRI analysis. *IEEE Trans Inf Technol Biomed*, 6(1):59–72, Mar 2002.

S Suryanarayanan, R Mullick, Y Mallya, C Wood, C McCullough, and K Thielen. Automatic bone removal for head CTA: A preliminary review. In *Proc. RSNA*, Dec 2003.

G M Teasdale, R Grant, B Condon, J Patterson, A Lawrence, D M Hadley, and D Wyper. Intracranial CSF volumes: Natural variations and physiological changes measured by MRI. *Acta Neurochir*, Suppl 42:230–235, 1988.

H T Thaler, P W Ferber, and D A Rottenberg. A statistical method for determining the proportions of gray matter, white matter, and CSF using computed tomography. *Neuroradiology*, 16:133–135, 1978.

P Therasse, S G Arbuk, E A Eisenhauer, J Wanders, R S Kaplan, L Rubinstein, J Verweij, M Van Glabbeke, A T van Oosterom, M C Christian, and S G Gwyther. New guidelines to evaluate response to treatment in solid tumors. *J Natl Cancer Inst*, 92:205–216, 2000.

A W Toga. *Brain Warping*. Academic Press, San Diego, 1999.

A Tsunoda, H Mitsuoka, K Sato, and S Kanayama. A quantitative index of intracranial cerebrospinal fluid distribution in normal pressure hydrocephalus using an MRI-based processing technique. *Neuroradiology*, 42:424–429, 2000.

R J van der Geest, V G M Buller, E Jansen, H J Lamb, J H B Baur, E E van der Wall, A de Roos, and J H C Reiber. Comparison between manual and semiautomated analysis of left ventricular volume parameters from short-axis MR images. *J Comput Assist Tomogr*, 21(5):756–765, Sept 1997a.

R J van der Geest, A de Roos, E E van der Wall, and J H C Reiber. Quantitative analysis of cardiovascular MR images. *Int J Card Imaging*, 13(3):247–258, June 1997b.

M S van der Knaap, C J Bakker, J A Faber, J Valk, W P Mali, J Willemse, and R H Gooskens. Comparison of skull circumference and linear measurements with CSF volume MR measurements in hydrocephalus. *J Comput Assist Tomogr*, 16(5):737–743, Sept 1992.

K Van Leemput, F Maes, D Vandermeulen, and P Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Trans Med Imaging*, 18(10):897–908, Oct 1999. `http://bilbo.esat.kuleuven.ac.be/web-pages/downloads/ems/` .

K Van Leemput, F Maes, D Vandermeulen, and P Suetens. A unifying framework for partial volume segmentation of brain MR images. *IEEE Trans Med Imaging*, 22(1): 105–119, Jan 2003.

L Vincent and P Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans Pattern Analysis Machine Intel*, 13(6):583–598, June 1991.

D Wang and D M Doddrell. A segmentation-based and partial-volume compensated method for an accurate measurement of lateral ventricular volumes on T1 weighted magnetic resonance images. *Magn Reson Imaging*, 19:267–272, 2001.

W M Wells, W E L Grimson, R Kikinis, and F A Jolesz. Adaptive segmentation of MRI data. *IEEE Trans Med Imaging*, 15:429–443, 1996.

C F Westin, S Warfield, A Bhalerao, L Mui, J Richolt, and R Kikinis. Tensor controlled local structure enhancement of CT images for bone segmentation. In *MICCAI—Medical Image Computing and Computer-Assisted Intervention*, LNCS 1496, pp. 1205–1212. Springer, Berlin, Oct 1998.

World Health Organization. *WHO handbook for reporting results of cancer treatment*, volume WHO offset publication. World Health Organization, Geneva, 1979.

A J Worth, N Makris, V S Caviness, jr., and D N Kennedy. Neuroanatomical segmentation in MRI: Technological objectives. *Int J Pattern Rec Art Int*, 11(8):1161–1187, 1998.

X Zeng, L H Staib, R T Schultz, and J S Duncan. Segmentation and measurement of the cortex from 3D MR images using coupled surfaces propagation. *IEEE Trans Med Imaging*, 18(10):100–111, Oct 1999.