

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

Horst K. Hahn and Heinz-Otto Peitgen

MeVis – Center for Medical Diagnostic Systems and Visualization
Universitätsallee 29, 28359 Bremen, Germany
e-mail: {hahn, peitgen}@mevis.de

Abstract. 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 that is perfectly suited to locate the brain, including the cerebellum and the spinal cord.

The main advantages of our method lie in its simplicity and robustness. It is simple since neither preprocessing of the MRI data nor contour refinement is required. Furthermore, the skull stripping solely relies 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 presence of severe intensity non-uniformity and noise. For that purpose, the watershed algorithm has been modified by the concept of pre-flooding, which helps to prevent over-segmentation, depending on a single parameter. The automatic selection of the optimal parameter as well as the applicability are discussed based on the results of phantom and clinical brain studies.

Keywords. Whole brain segmentation, skull stripping, 3D watershed transform, pre-flooding, neurological image processing.

1 Introduction

The skull stripping problem is well-known and has been studied in [1–10]. It is equivalent to the segmentation of the whole brain or the removal of non-cerebral

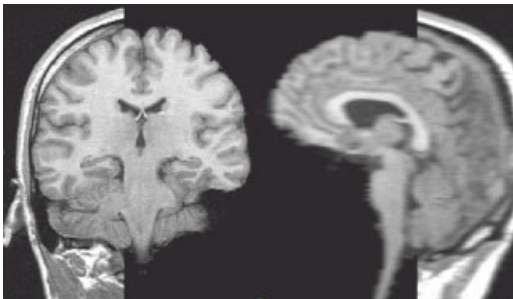


Fig. 1. Fusion of a successfully skull-stripped image with the original data in two projections (healthy volunteer, coronal slicing, thickness 3 mm).

tissue such as skull, scalp, veins or meninges (Fig. 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 protocols, that belong to the fastest MRI protocols available, are often preferred, since they offer a good contrast between gray (GM) and white cerebral matter (WM) as well as between GM and cerebrospinal fluid (CSF).

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 [5], 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), such as statistical brain tissue segmentation methods [2, 11, 12]. The latter offer a good basis for volumetric or morphometric examinations, e.g. cortex reconstructions [8], and yield promising results in accounting for intensity non-uniformity [11, 12], however, their convergence is distorted by the presence of non-brain tissue [7, 11, 13]. 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 [14].

The paper is organized as follows. The next sections will give a brief overview of existing methods and describe the motivation and goals of the novel approach. Then, the method will be derived in detail. An evaluating section assesses the 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 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 et al. [6]. 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 [1, 2, 3, 5, 7, 10].
- *Hybrid methods.* In order to account for the shortcomings of morphological multi-step approaches, they have been combined with edge-based methods [2, 4, 6, 9]. For example, two-dimensional contours such as snakes are applied to the morphological segmentation result in a final step [6].
- *Template-based.* More recently, some investigators succeeded in fitting a balloon-like surface to the intensity-normalized MR data in order to separate the brain from surrounding structures [8]. Another three-dimensional template based on volume data is described in [15].

Our goal in this paper 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 against considerable radio frequency (RF) non-uniformity as well as noise. This is not the case for most of the existing techniques. Thresholding and region growing based approaches exhibit the problem of *leaking out*, as discussed in [16], and are highly sensible to image non-uniformities. Deformable templates are susceptible to image noise that has to be removed in a preprocessing step.

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, so 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. We do not 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 an automatic segmentation procedure. Finally, the computational costs should be approximately of the order of the number of voxels, to be able to deal with future imaging resolutions.

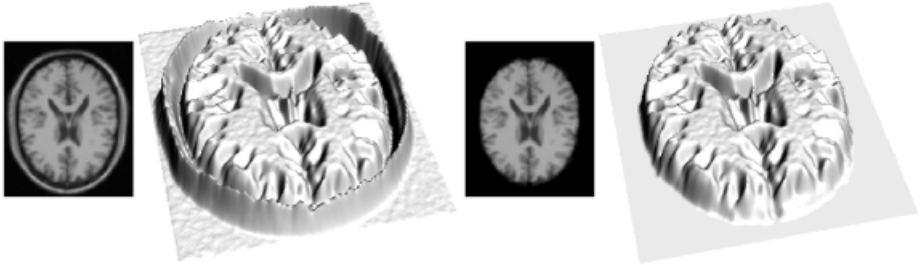


Fig. 2. Axial slice of the MNI phantom (simulated with slice thickness 3 mm, noise level 3%, RF non-uniformity 20%) and interpretation of gray value as height information. **left:** original image. **right:** skull stripped with $h_{pf} = 0.16 I_{max}$ and only one connected hill left, representing the brain.

3 Methods

3.1 Basic Ideas

Similar to other approaches [1, 2, 7], our basic assumption regarding brain anatomy is the *connectivity of white matter*. We define two points of a gray level image to be *connected*, if a path of adjacent voxels exists between them that is at least as bright as the darker one of the two points. This concept can be easily understood by interpreting the image intensity as height information. For each image slice, 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 left). According to the above definition of connectivity, two regions are disconnected, if they are separated by a valley, otherwise they are connected.

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. Therefore, we prefer a segmentation procedure that works fully in 3D, rather than 2D techniques that may not account for three-dimensional connectivity.

Assuming that the brain is surrounded by CSF and all non-brain tissue that needs to be removed shows brighter image intensities than CSF, skull stripping becomes equivalent to isolating a single hill in the four-dimensional landscape. (Figures 2 and 3 represent didactical three-dimensional projections of the four-dimensional problem.) 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 right).

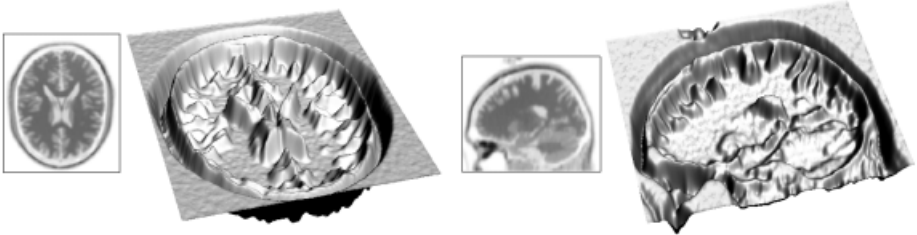


Fig. 3. Same data as Figure 2, but gray level inverted. Hills in the original image are now represented by basins, that can be separated by a watershed transform. **left:** axial, **right:** sagittal view.

3.2 Solution Based on a 3D Watershed Transform

We now consider the gray level inverted T1-weighted brain image data. Under this transformation, hills become basins and valleys become crest lines (Fig. 3). The morphological operation that partitions an image into regions, each of them corresponding to a connected basin, is *the watershed transform* [17]. Thus, all that needs to be done for a whole brain segmentation, is to perform *a 3D watershed transform on the inverted original data*¹ and to choose the catchment basin representing the brain.

Before introducing the fast watershed algorithm, we will have a more detailed look at the 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 non-uniformity 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 pre-flooding*: Prior to the transformation, each catchment basin is flooded up to a certain height above its *bottom* (i.e. the darkest voxel) called *pre-flooding 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.

¹ It has to be mentioned, that usually the watershed transform is applied to the gradient image.

3.3 Algorithm

The applied fast watershed transform starts by sorting all voxels (of the gray level inverted image, Fig. 3) according to their intensity in an ascending order. Each voxel with regard to its three-dimensional 6-neighborhood is processed exactly once, until the brightest voxels have been processed. If the voxel has some already processed neighbors (i.e. voxels of same or less 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 approach is similar to the one described by MITELHAEUSSER and KRUGGEL [18], but simpler in that the voxel-basin merging is non-conditional. We now introduce two criteria for the merging procedures (not published in [18]):

- *Voxel-basin merging*: Each voxel will be merged with the deepest neighboring basin, i.e. the basin with the darkest bottom voxel.
- *Basin-basin merging*: All neighboring basins whose depth relative to the current voxel intensity is less or equal to the pre-flooding height h_{pf} will be merged with the same basin as the voxel itself.

After the transform with an appropriate pre-flooding height one basin should exist that represents the whole brain with all parts that are connected via WM, i.e. cerebellum and spinal cord. This basin usually is the largest existing one, unless the field of view (FOV) has been selected too large. In that case, one could choose the basin containing the center of FOV or a manually selected basin.

The number of basins is monotonically decreasing whereas its sizes are increasing with increasing h_{pf} . Figure 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 continuously with varying h_{pf} but, depending on the image quality, a more or less broad *range of proper h_{pf} values* exists.

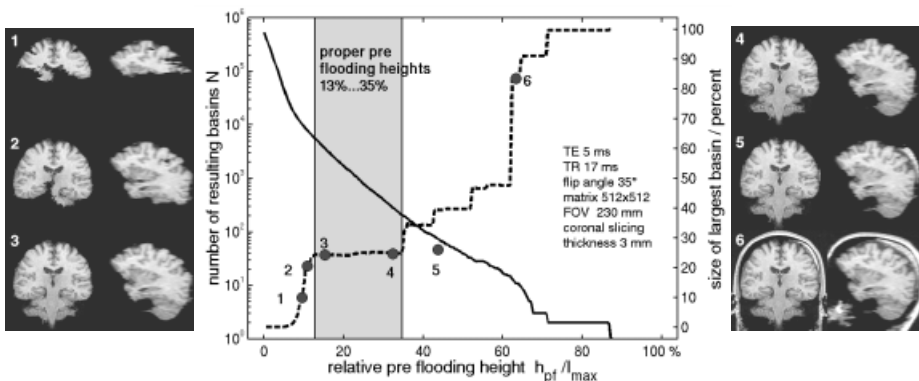


Fig. 4. Algorithmic behavior of the modified watershed transform with increasing pre-flooding height, number of resulting catchment basins (solid line) and size of largest basin (dashed line). For the six given h_{pf} values (filled circles) the result is shown in coronal and axial view. In case no. 5, where under-segmentation did already occur, the largest basin did not contain the brain.

4 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 different scanners as well as phantom data to answer the following questions:

- Is it possible to successfully segment the whole brain with our method?
- If yes, what is the range of proper pre-flooding 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 and pre-flooding heights h_{pf} will be given in units relative to the maximum image intensity I_{max} .

For clinical MRI segmentation studies no *gold standards* are available. Our “gold standard” was defined by an expert radiologist who classified the 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 7 manually segmented datasets the overlap has been more than 96%.

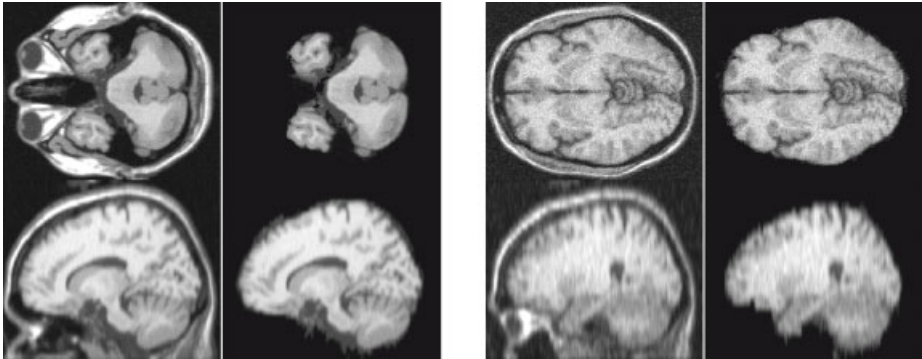


Fig. 5. Two images of the MNI simulated brain database and application of the skull stripping procedure, axial and sagittal view. Image simulation parameters (slice thickness / noise level / RF non-uniformity): **left.** 3 mm / 3% / 20%. **right.** 5 mm / 9% / 40%.

4.1 Phantom Studies

To offer a reproducible data source, we extensively tested our method on the Simulated Brain Database (SBD) [19, 20] from the Montréal Neurological Institute (MNI), McGill University. All available 90 T1-weighted datasets of the normal brain database have been processed. The results were satisfactory: All datasets were successfully segmented by our method, even for extreme slice thickness (9 mm), noise (9%) and RF non-uniformity (40%). Two examples are shown in Figure 5.

4.2 Clinical Brain Studies

For further evaluation, 43 clinical brain images of healthy volunteers and patients have been acquired with two MR scanners: A Siemens Magnetom Vision Plus 1.5 T and a Philips S15-ACS 1.5 T, using various 3D T1-weighted acquisition protocols (TR 8.1–17 ms, TE 4–5 ms, flip angle 12° – 35°) and different 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 6 for examples. However, in two cases the pre-flooding value h_{pf} for correct segmentation turned out to be unexpectedly high, see Section 5.

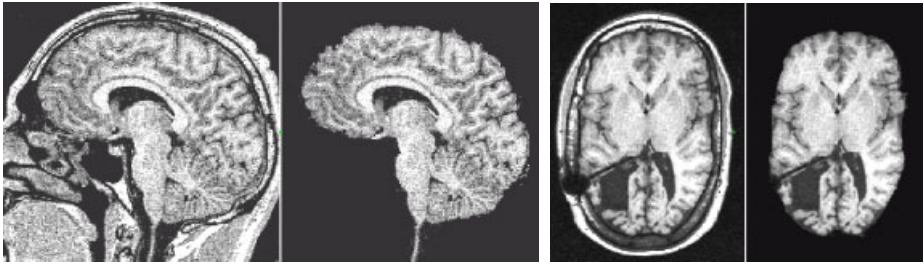


Fig. 6. Clinical brain datasets and application of skull stripping procedure. Two problems are addressed: **left.** extreme image noise level (acquisition parameters: slice thickness 1 mm, TR 8.1 ms, TE 4 ms). **right.** pathological abnormality (enlarged left ventricle, with drainage).

4.3 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 datasets (Fig. 7). The proper h_{pf} values roughly scale linearly with the noise level. The dotted line represents the equation $h_{pf} = 0.11 I_{max} + 3.5 n$ chosen such that all 45 datasets are correctly segmented. This behavior makes it possible to automatically select an appropriate pre-flooding height. The broader the h_{pf} range, the more robust is the segmentation to further noise and inhomogeneity (Table 1).

Table 1. Quantitative evaluation of robustness in terms of the h_{pf} range.

Range of proper pre-flooding heights ($h_{pf,max} - h_{pf,min}$) / I_{max}	# datasets	
	phantom	clinical
0 (failure)	0	0
< 5%	6	2
5–10%	25	9
> 10%	59	32

Table 2. Average sorting and transformation times for three different image sizes on two standard SGI machines (Onyx2: 2xR10000, 195 MHz, O2: R5000, 180 MHz).

# voxels	SGI Onyx2	SGI O2
1 Mio.	2s / 5s	5s / 9s
4 Mio.	11s / 21s	25s / 43s
10 Mio.	33s / 51s	69s / 103s

4.4 Numerical Issues

The complexity of the modified watershed transform is linear in the number of voxels N for the voxel processing, and in the worst case proportional to $N \log N$ for the sorting (about 30% of the total computing time, see Table 2). In the average case our algorithm shows *linear complexity in N* . All algorithms have been implemented in C++ and are available on the MeVis medical imaging platform ILabMed [21]. For sorting the voxels according to their intensity, we used a QuickSort algorithm [22].

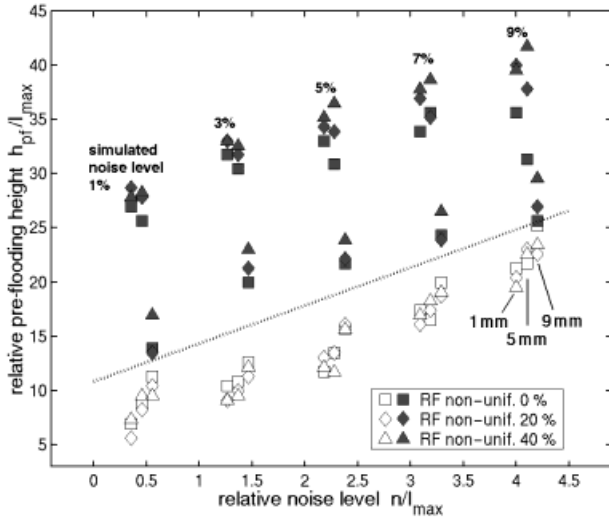


Fig. 7. Range of proper pre-flooding heights (minimum: white, maximum: black markers) plotted against noise level for 45 datasets of the MNI database. For each noise level, slice thickness is increasing from left to right (1 / 5 / 9 mm).

5 Discussion

The described algorithm is able to successfully segment the whole brain in all 133 datasets, without any preprocessing. For any given image the parameter h_{pf} can be varied over a certain range without changing the output, this being a measure for robustness of the method. The sensitivity in comparison to manual expert segmentation is estimated to 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. Only in two cases, parts of the superior sagittal sinus have been included.

The results were satisfactory even where the range of proper pre-flooding heights h_{pf} has been rather narrow. This has been 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 (compare Figure 5 right), and in two of 43 clinical datasets (see Table 1). When inspecting these two images, we found that some isolated bright voxels existed in different parts of the brain, thus increasing the required h_{pf} value.

The merging rules introduced in Section 3.3 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 (Table 2). 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.

6 Conclusion

The modified watershed transform presented here is a powerful tool for segmenting the whole brain from MRI datasets. The application to the original MRI data rather than the gradient image is a new approach which works 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 pre-flooding on the inverted original data bears several advantages compared to existing skull stripping techniques, in that it is extremely robust against RF non-uniformity 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 of the brain surface are made and no initialization of a template is required. Finally, interaction techniques such as region selection or watershed placement can easily be incorporated.

The presented algorithm provides the basis for a standardized brain segmentation procedure that increases reliability and reproducibility in the field of neuroimaging.

Acknowledgments

We would like to thank our clinical partners Prof. Dr. B. Terwey, Bremen, and Prof. Dr. P. Stoeter, Mainz, for kindly providing us with brain data and Wolf Spindler, Bremen, for his support on visualization issues.

References

1. R.Kikinis, M.E.Shenton, G.Gerig, J.Martin, M.Anderson, D.Metcalf, C.R.G.Guttman, R.W.McCarley, W.Lorensen, H.Cline, and F.A.Jolesz. "Routine Quantitative Analysis of Brain and Cerebrospinal Fluid Spaces with MR Imaging" *J. Magnetic Resonance Imaging* 2: 619-629, 1992.
2. 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.
3. 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.
4. G.B.Aboutanos and B.M.Dawant. "Automatic Brain Segmentation and Validation: Image-Based versus Atlas-Based Deformable Models" *Proc. SPIE Med. Imaging '97*, 299-310, Feb. 1997.

5. P.A.Freeborough, N.C.Fox, and R.I.Kitney. "Interactive Algorithms for the Segmentation and Quantification of 3D MRI Brain Scans" *Comput. Metho. Progr. Biomed.* 53(1): 15-25, May 1997.
6. 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.
7. A.F.Goldszal, C.Davatzikos, D.L.Pharm, 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.
8. A.M.Dale, B.Fischl, and M.I.Sereno. "Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction" *NeuroImage* 9: 179-194, 1999.
9. 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" *Proc. 5th Int. Conf. on Functional Mapping of the Human Brain HBM '99*, Düsseldorf, poster no. 86, 1999, abstract pub. in *NeuroImage*.
10. S.A.Hojjatoleslami, F.Kruggel, and D.Y.von Cramon. "Segmentation of White Matter Lesions from Volumetric MR Images" *Proc. Medical Image Computing and Computer-Assisted Intervention MICCAI '99*, Cambridge. Springer LNCS, vol. 1679: 52-61, 1999.
11. 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.
12. D.L.Pharm, J.L.Prince. "Adaptive Fuzzy Segmentation of Magnetic Resonance Images" *IEEE Trans. Med. Imaging* 18(9):737-752, Sept. 1999.
13. F.Maes, K.Van Leemput, L.E.DeLisi, D.Vandermeulen, and P.Suetens. "Quantification of Cerebral Gray and White Matter Asymmetry from MRI" *Proc. Medical Image Computing and Computer-Assisted Intervention MICCAI '99*, Cambridge. Springer LNCS, vol. 1679: 348-357, 1999.
14. A.W.Toga. *Brain Warping*, Academic Press, San Diego, 1999.
15. 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/>
16. 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.
17. J.Serra. *Image Analysis and Mathematical Morphology*. London Academic, 1982.
18. G.Mittelhaeuser and F.Kruggel. "Fast Segmentation of Brain Magnetic Resonance Tomograms" *Proc. 1st Conf. on Computer Vision, Virtual Reality and Robotics in Medicine CVR '95*. Springer LNCS, vol. 905, 1995.
19. C.A.Cocosco, V.Kollokian, R.K.-S.Kwan, and A.C.Evans. BrainWeb: "Online Interface to a 3D MRI Simulated Brain Database" *Proc. 3rd Int. Conf. on Functional Mapping of the Human Brain HBM '97*, Copenhagen. *NeuroImage* 5(4): part 2/4, S425, May 1997. <http://www.bic.mni.mcgill.ca/brainweb/>
20. 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.
21. A.Schenk, J.Breitenborn, D.Selle, T.Schindewolf, D.Böhm, W.Spindler, H.Jürgens, and H.-O. Peitgen. "ILabMed-Workstation – Eine Entwicklungsumgebung für radiologische Anwendungen" *Proc. Bildverarbeitung für die Medizin '99*, 238-242, Springer, 1999.
22. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988, 1995.