ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 23 (2004) 625–637

A meta-algorithm for brain extraction in MRI

David E. Rex,[a] David W. Shattuck,[a] Roger P. Woods,[b] Katherine L. Narr,[a] Eileen Luders,[a] Kelly Rehm,[c] Sarah E. Stolzner,[b] David A. Rottenberg,[c] and Arthur W. Toga[a,*]

[a] *Laboratory of Neuro Imaging, Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095-1769, USA*
[b] *Department of Neurology, Neuropsychiatric Institute, Ahmanson-Lovelace Brain Mapping Center, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095-1769, USA*
[c] *Departments of Radiology and Neurology, Minneapolis VA Medical Center, University of Minnesota, Minneapolis, MN 55417, USA*

Received 20 February 2004; revised 8 June 2004; accepted 9 June 2004
Available online 12 September 2004

Accurate identification of brain tissue and cerebrospinal fluid (CSF) in a whole-head MRI is a critical first step in many neuroimaging studies. Automating this procedure can eliminate intra- and interrater variance and greatly increase throughput for a labor-intensive step. Many available procedures perform differently across anatomy and under different acquisition protocols. We developed the Brain Extraction Meta-Algorithm (BEMA) to address these concerns. It executes many extraction algorithms and a registration procedure in parallel to combine the results in an intelligent fashion and obtain improved results over any of the individual algorithms. Using an atlas space, BEMA performs a voxelwise analysis of training data to determine the optimal Boolean combination of extraction algorithms to produce the most accurate result for a given voxel. This allows the provided extractors to be used differentially across anatomy, increasing both the accuracy and robustness of the procedure. We tested BEMA using modified forms of BrainSuite's Brain Surface Extractor (BSE), FSL's Brain Extraction Tool (BET), AFNI's 3dIntracranial, and FreeSurfer's MRI Watershed as well as FSL's FLIRT for the registration procedure. Training was performed on T1-weighted scans of 136 subjects from five separate data sets with different acquisition parameters on separate scanners. Testing was performed on 135 separate subjects from the same data sets. BEMA outperformed the individual algorithms, as well as interrater results from a subset of the scans, when compared for the mean Dice coefficient, a rating of the similarity of output masks to the manually defined gold standards.

Keywords: Algorithm; Meta-algorithm; Brain extraction; Segmentation; Automated processing; MRI

Introduction

The accurate identification of the brain in a head MRI is crucial to many studies in neuroimaging. Low-level classification of the brain allows for the analysis of cortical structure (Fischl et al., 1999; Thompson et al., 2001) provides a measure of brain volume (Lawson et al., 2000; Smith et al., 2002), improves the localization of signal in magnetoencephalography and electroencephalography data (Baillet et al., 1999; Dale and Sereno, 1993), can initialize a more detailed segmentation of tissues (Shattuck et al., 2001; Zhang et al., 2001), and can be used to prepare data for accurate image registration (Woods et al., 1993). Automating the brain identification step in neuroimaging studies eliminates human rater variances and can allow for larger sample sizes by doing away with a labor-intensive step.

Brain extraction algorithms

Numerous algorithms have been written to perform brain extraction. Most are devised to work on T1-weighted MRI data, with several exceptions into other modalities (Alfano et al., 1997; Bedell and Narayana, 1998; Held et al., 1997). Various methodologies are used to achieve a semiautomated (Bomans et al., 1990; Hohne and Hanson, 1992) or fully automated (Dale et al., 1999; Smith, 2002) separation of brain from nonbrain tissue.

Atlas registration techniques for segmentation transfer brain labels to an individual subject (Bajcsy et al., 1983; Christensen et al., 1996; Collins et al., 1995; Davatzikos, 1997; Miller et al., 1993). Most perform well when defining deep structures of the brain but may fail at the cortical surface due to the large degree of intersubject variability in sulcal and gyral morphology. Improved atlas techniques joining low-level tissue classifications (gray matter, white matter, and cerebrospinal fluid) with image registration have had more success in demarcating anatomy (Collins et al., 1999; Kapur et al., 1996).

An early semiautomated technique for brain extraction uses edge detection to demarcate connected tissues within a slice (Bomans et al., 1990). Components that represent brain are manually selected and a morphological closing operation is performed on the selections to complete the process. Sandor and Leahy (1997) developed an automated edge-detection technique using anisotropic diffusion filtering, Marr–Hildreth edge detection, and a sequence of morphological processing steps to extract the brain in three dimensions. Shattuck et al. (2001) subsequently improved upon this technique.

Slice by slice brain identification based on gray matter and white matter intensity estimation, connected component determi-

* Corresponding author. Laboratory of Neuro Imaging, Department of Neurology, Room 4-238, 710 Westwood Plaza, Box 951769, Los Angeles, CA 90095-1769. Fax: +1-310-206-5518.
E-mail address: toga@loni.ucla.edu (A.W. Toga).
Available online on ScienceDirect (www.sciencedirect.com.)

1053-8119/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2004.06.019

nation, and morphology operations also have produced good results (Brummer et al., 1993; Lemieux et al., 1999; Ward, 1999). Lemieux et al. (2003) further extended these techniques to include CSF estimation for the inclusion of all intracranial CSF. Deformable templates guided by image intensity information, usually the search for the gray matter or CSF border, and smoothness constraints that mimic general properties of the brain also have been used (Dale et al., 1999; MacDonald et al., 1994; MacDonald et al., 2000; Smith, 2002).

Meta-algorithms

A meta-algorithm uses the results of individual algorithms for similar tasks, or subtasks, to perform the chosen task. Many meta-algorithms have been designed to achieve higher reliability or to attain greater accuracy using a trained system. Schroder et al. (1999) implemented a meta-algorithm for the deconvolution of disturbed data, called Munchhausen, to calculate blood volume using the intravascular concentration time course of an injected substance. They note that many deconvolution techniques vary in their performance depending on the type of data and the nature of the disturbance in the recorded values. Munchhausen uses a data-driven decision rule to select from its many deconvolution techniques to achieve more robust results than any individual algorithm.

Shaaban and Schalkoff (1995) and Schalkoff and Shaaban (1999) use a meta-algorithm to solve general image processing and feature extraction problems for two-dimensional images. A training set showing initial images and outlining the desired features to extract is used to solve a classification problem in an algorithm graph. Multiple algorithm paths exist through the graph and the training guide selection of the best processing path. The results can be applied to new data for identification of features defined by the training set.

Meta-algorithms also have been used previously for studies on MRI scans. Rehm et al. (1999) implemented and validated (Boesen et al., 2003) a meta-algorithm for brain extraction from an MRI volume called McStrip. It uses a polynomial registration (Woods et al., 1998) to provide a brain mask from an atlas and builds a threshold mask from estimates of tissue class boundaries. It also generates a BSE mask from the Brain Surface Extractor (Shattuck et al., 2001) using many parameter sets and choosing the mask in highest agreement with the threshold mask. The union of the threshold mask and the BSE mask provides the final output. McStrip outperformed three other algorithms, BSE, BET (Smith, 2002), and SPM (Ashburner and Friston, 2000) in both boundary similarity and misclassified tissue metrics for 15 test scans.

Collins et al. (1999) implemented a meta-algorithm for gross cerebral structure segmentation. A nonlinear registration is used to obtain tissue labels from an atlas, and a low-level tissue classification identifies regions of gray matter, white matter, and cerebrospinal fluid. The reconciliation of the two segmentations produces a more accurate identification of cerebral structures than either method produces on its own.

Brain extraction meta-algorithm

A single algorithm often will not adequately perform the neuroimaging task in every subject across an entire data set. Often, many different procedures must be attempted or manual intervention utilized to achieve acceptable results. An environment that presents many similar algorithms is a simple way to access and test various methods. A meta-algorithm that allows the specification of a general procedure and obtains a valid result, regardless of input data, would allow the task of deciphering the results from many algorithms and selecting the best procedures to be fully documented and automated.

Each of the aforementioned algorithms for brain identification possesses strengths and weaknesses that vary with scanning protocol, image characteristics such as contrast, signal-to-noise ratio, and resolution, and subject-specific characteristics like age and atrophy (Fennema-Notestine et al., 2003). Algorithms may also vary in their accuracy in different anatomic regions. The development of a meta-algorithm that intelligently utilizes the strengths of the contributing subalgorithms should obtain results that are, on average, superior to any individual algorithm. We developed and tested such a meta-algorithm using multiple extraction procedures in concert with a registration procedure. It achieves improved results using a variety of anatomically specified Boolean functions to combine the results of the extractors.

Methods

Brain extraction meta-algorithm

The Brain Extraction Meta-Algorithm (BEMA) uses four freely available brain extraction algorithms and a linear volume registration procedure, which does not require skull stripping, in concert to achieve its results (Fig. 1). In general, the registration procedure is used to bring a brain atlas into alignment with an individual subject scan being processed for brain extraction. The atlas contains information regarding which brain extraction algorithm, or combination of extractors, works best identifying brain in each anatomic region. The overall best combination of brain extractors for each region, based on a training set of scans and manual demarcations of brain, is then applied on a voxel by voxel basis.

The extractors used in BEMA include the Brain Surface Extractor (BSE) (Shattuck et al., 2001) from BrainSuite (Shattuck and Leahy, 2002), the Brain Extraction Tool (BET) (Smith, 2002) from FSL (Smith et al., 2001), 3dIntracranial (Ward, 1999) from AFNI (Cox, 1996; Cox and Hyde, 1997), and MRI Watershed from FreeSurfer (Dale et al., 1999). The volume registration procedure utilized is FLIRT (Jenkinson and Smith, 2001), also from FSL. The T1-weighted ICBM152 average MRI (Evans et al., 1994) in approximate Talairach space (Talairach and Tournoux, 1988) is utilized for the whole-head atlas space. All algorithms utilized are freely available on the World Wide Web and have been encapsulated in the LONI Pipeline Processing Environment (Rex et al., 2003), along with utility functions from the Laboratory of Neuro Imaging, AIR (Woods et al., 1998), and FSL.

Data preprocessing

Preprocessing of the data sets for each individual extraction algorithm was performed to provide the best possible results from each brain extractor. BEMA begins with a FLIRT registration of the ICBM152 average to the individual subject scan. A brain mask is resampled to the subject to identify a region that must contain the whole brain. This brain mask consists of voxels in ICBM152 space where any of 200 previously aligned subjects contained any brain tissue. The subject scan is masked and cropped to limit the search space for the subject's brain. The resulting volume is passed, in

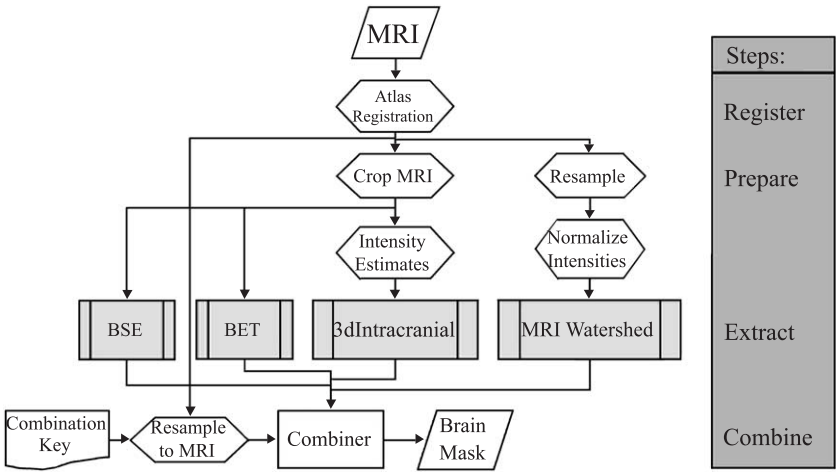

Fig. 1. A simplified diagram of the Brain Extraction Meta-Algorithm conceptually showing the data flow from a raw MRI to a completed brain mask. The steps involved in this extendable algorithm are registration, the preparation of data for each given extraction algorithm, the extraction of the brain from the MRI by each algorithm, and the combining of the results in an improved brain mask. BEMA is implemented in the LONI Pipeline Processing Environment and is available as a single fully encapsulated module for use within the environment.

parallel, to BSE and BET for extraction. The parameters utilized for BSE throughout this study are a sigma of 0.62 for the edge detection and three iterations of the anisotropic filter with a diffusion constant of 25. BET was run with its default parameters. The results of BSE and BET are placed back in the original subject space. AIR tools are used for all resizing and masking, ensuring data is not normalized or blurred through the process.

The 3dIntracranial branch of BEMA uses an initial registration mask to roughly estimate the subject's brain location. This more conservative mask identifies the brain using voxels in ICBM152 space where the brain was located 50% or more of the time for the aforementioned 200 aligned subjects. The purpose of this mask is to estimate gray and white matter intensity limits for 3dIntracranial. The Partial Volume Classifier (PVC) (Shattuck et al., 2001) is used to classify the estimated brain into gray matter, white matter, and CSF tissue classes. A robust maximum white matter intensity and a robust minimum gray matter intensity (Smith, 2002) are computed and input to 3dIntracranial. 3dIntracranial is then executed on the liberally masked volume from the BSE or BET preprocessing path described above.

The ICBM152 registration to the subject's native space is inverted and modified to resample the subject to the required FreeSurfer space for processing. Intensity normalization is performed on the volume using MRI Normalize (Dale et al., 1999). The normalized volume is processed with MRI Watershed and the resulting mask is resampled back to the native subject space with nearest neighbor interpolation.

Combining results

The results of the individual extraction paths through BEMA are combined to form the final brain mask. A Boolean function is stored at each voxel in atlas space that will be used to combine the binary results of the four input extractors. This combination key is resampled, using the FLIRT-derived ICBM152 transformation to subject native space, to the subject scan with a nearest neighbor interpolation and used with the extractor results to derive the BEMA brain mask. Varying the combination function across the voxels in atlas space allows different extractors and combinations of extractors to be utilized for various regions of anatomy.

Experience shows that individual extraction algorithms do not perform better across a data set when their results are reversed (labeled background is considered brain and labeled brain is considered background). Therefore, no Boolean functions with inversions are allowed and a voxel's identity is never determined by assuming an extractor, or group of extractors, may be wrong far more often than they are correct. With four inputs, this limits the number of Boolean functions to 168 possible combinations. They represent such combinations as (BSE or 3dIntracranial) identifying the specified voxel as brain for it to be included in the brain mask [(BET and MRI Watershed) or (BET and BSE)] identifying the voxel as brain, any three extractors needing to agree the identity of the voxel is brain for it to be labeled as brain, or all four extractors needing to agree the identity of the voxel is brain. They may also be as simple as using the results of a single extractor or always marking a voxel as brain or not brain based on the registration being more accurate than any other method. The choice of Boolean logic was made for this meta-extractor implementation because of its simplicity and the power to search the entire positive space for the four input extractors. In addition, it is equivalent to more complex optimization techniques, such as neural nets or statistical methods, when applied to this limited binary problem. For a problem with more inputs that utilizes the negative space or that works on continuous variables, an optimization technique would be better suited.

Training

To determine what combination of extractors works best at each anatomic location, as represented by a voxel in ICBM152 space, a training step was implemented. Training is performed on a representative set of scans that possess expertly determined masks demarcating the brain in the volumes. For optimal results, BEMA should be retrained for data sets with novel contrast and signal-to-noise characteristics. During training, the extractor paths in the pipeline are processed for each individual scan in the training set

Table 1
Scan acquisition information

Data set	Scanner	Voxel size (mm)	Acquisition plane	Acquisition sequence
ICBM	3 T General Electric	0.9375 × 0.9375 × 1.2	Sagittal	3D-SPGR, TR = 24 ms, TE = 4 ms, FA = 35°
IPDH	1.5 T General Electric	0.78125 × 0.78125 × 1.5	Coronal	3D-SPGR, TR = 35 ms, TE = 5 ms, FA = 35°
LIJMC	1.5 T General Electric	0.86 × 0.86 × 1.5	Coronal	3D-SPGR with inversion recovery, TR = 14.7 ms, TE = 5.5 ms
ZENIT	1.5 T General Electric	0.97 × 0.97 × 1.5	Sagittal	3D-SPGR, TR = 24 ms, TE = 8 ms, FA = 30°
NEUROVIA	1.5 T Siemens	0.86 × 0.86 × 1.0	Transverse	3D-FLASH, TR = 35 ms, TE = 6 ms, FA = 45°

ICBM—International Consortium for Brain Mapping, David Geffen School of Medicine at UCLA; IPDH—Institute of Psychiatry, Denmark Hill, London, UK; LIJMC—Long Island Jewish Medical Center, New York, New York; ZENIT—Center for Neuroscientific Innovation and Technology, Magdeburg, Germany; NEUROVIA—Neuroimaging, Visualization, and Data Analysis group at the Minneapolis VA Medical Center, University of Minnesota.
Scans were acquired from five different institutions on three different scanner types from two manufacturers and two field strengths with a variety of listed protocols.

and all brain masks for all extractors are resampled to the atlas space using the derived FLIRT transformations. The expertly demarcated masks are, respectively, resampled to the atlas space as well. The trainer program analyzes each voxel in the atlas space with each available Boolean function applied to the individual extractor results for all training scans. The Boolean function that most often determines the correct answer according to the expert masks is stored in the combination key. It represents the function to be used for that anatomic locale whenever this combination key is used in BEMA. A user modifiable window around each voxel is provided so that the voxel's Boolean function may be determined by the results of all voxels within the neighborhood of the voxel of interest. This provides a blurring of the anatomy and removes noise induced by registration.

Subjects

Two hundred and seventy-five subjects were amassed from five separate studies, three different scanner types, and five different scan protocols (Table 1) to provide five data sets consisting of 275 T1-weighted whole-head MRI volumes. Subjects from the International Consortium for Brain Mapping (ICBM) acquired at the University of California, Los Angeles, from the Center for Neuroscientific Innovation and Technology (ZENIT), Magdeburg, Germany, and from the Neuroimaging, Visualization, and Data Analysis group (NEUROVIA) at the University of Minnesota, Minneapolis VA Medical Center, were all healthy young adults. Subjects from the Institute of Psychiatry, Denmark Hill (IPDH), London, UK, were schizophrenia patients and normal controls. Subjects from Long Island Jewish Medical Center (LIJMC) were first episode schizophrenia patients and normal controls (Table 2). The first episode schizophrenia patients typically received 1–2 mg of oral lorazepam before the scan. All subjects provided written informed consent based on the institutional guidelines of the acquisition site.

All scans were manually assessed for voxels that correspond to brain tissue or cerebrospinal fluid (CSF). The ZENIT, IPDH, and LIJMC scans were assessed under the supervision of KLN using a voxel labeling of the scans in the coronal plane and reconciled in the sagittal and transverse planes. The ICBM and NEUROVIA scans were assessed under the supervision of RPW using contour tracings in the sagittal plane of the scans. KR performed additional brain demarcations of the NEUROVIA scans using contours. The contour demarcations were converted to voxel-based labels of the brains for use in BEMA training and assessment.

Assessment

BEMA was trained on 25 scans from the ICBM data set, 27 from the IPDH, 48 from LIJMC, 30 from ZENIT, and 10 from NEUROVIA. All training scans were randomly selected from their respective group. The data sets were separated into three separate groups for training and testing. The ICBM, IPDH, and LIJMC training scans were combined into one set of 100 scans to produce a single combination key for the three sets of scans. The ZENIT and NEUROVIA scans were kept separate from the first group and from each other to produce two additional combination keys for their respective data sets. A single additional training set was derived using 10 scans from each of the five data sets to test the meta-algorithm generalized across all data sets. The combination keys were produced with a training window of 5 × 5 × 5 mm. The NEUROVIA gold standard segmentations were taken from the KR segmentations of the scans, being the more internally consistent of the two human raters. The RPW-supervised segmentations of the NEUROVIA data were considered equally accurate for brain vs.

Table 2
Subject demographics

Data set	Number of subjects	Gender	Diagnosis	Age (mean ± SD, years)
ICBM	50	23 male, 27 female	All normal	Normal male = 23.7 ± 5.8; normal female = 25.0 ± 5.8
IPDH	53	30 male, 23 female	28 normal, 25 SZ	SZ male = 32.4 ± 7.9; normal male = 33.0 ± 10.1; SZ female = 39.9 ± 10.2; normal female = 35.2 ± 9.0
LIJMC	96	62 male, 34 female	34 normal, 62 SZ	SZ male = 24.1 ± 4.2; normal male = 35.5 ± 8.6; SZ female = 28.7 ± 5.1; normal female = 20.2 ± 11.0
ZENIT	60	30 male, 30 female	All normal	Normal male = 25.4 ± 4.7; normal female = 24.3 ± 4.4
NEUROVIA	16	8 male, 8 female	All normal	Normal male = 30.4 ± 6.9; normal female = 23.8 ± 4.5

Data sets from the five institutions varied in age, gender, and diagnosis (normal vs. schizophrenic).

nonbrain tissues but were not as internally consistent in the border reconciliation within the external CSF. They were used to determine human interrater measures.

The remaining 25 scans from the ICBM data set, 26 from IPDH, 48 from LIJMC, 30 from ZENIT, and 6 from NEUROVIA were used to test BEMA and the individual extractors that comprise it. MRI Watershed, 3dIntracranial, BET, and BSE were each run on the 135 test scans in their modified forms used in the BEMA algorithm—the FLIRT registrations and PVC preliminary tissue classification were used to enhance the output of each algorithm, as detailed above. Additionally, 3dIntracranial, BET, and BSE were executed in their raw form on each of the 135 scans, without any external aid from other programs. The raw approach was not used for MRI Watershed, as it is not how the authors

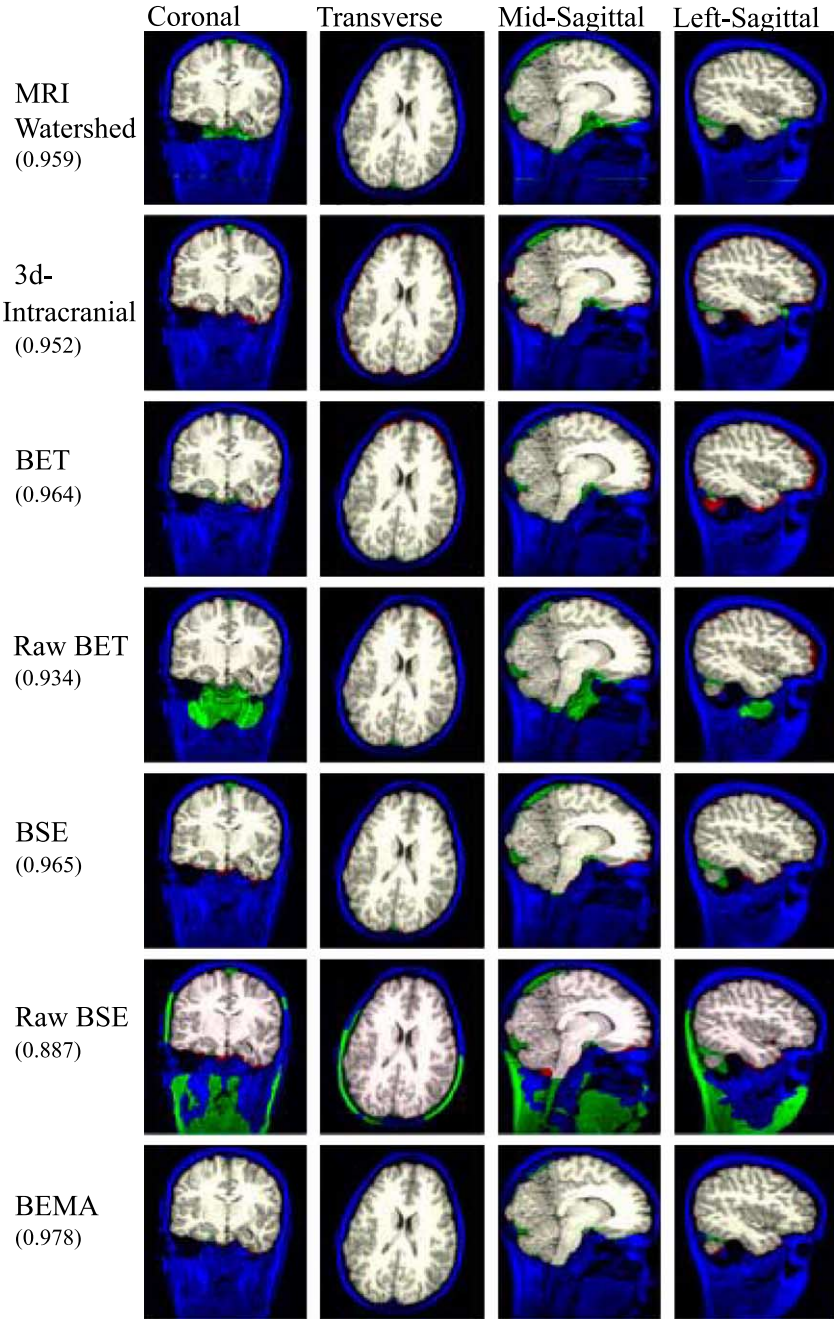

Fig. 2. The extraction results for subject number 91 (Fig. 3) from the LIJMC data set. Shown in blue is the subject's original T1-weighted MRI scan. The gray to white intensities show where the manual gold standard mask, and the automated extraction result agrees there is brain. Green represents where the automated extraction falsely classified voxels as brain. Red represents where the automated extraction falsely classified voxels as not brain. Dice coefficients comparing the automated methods to the gold standard are shown in parentheses. The BET and BSE methods are the registration-augmented versions used in the BEMA algorithm. The raw BET and raw BSE methods are the extractors run on their own with no preparation of the input data. The raw BSE result seen here is atypical but demonstrative of errors that sometimes occur and are fixed by the meta-algorithm. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

intended the algorithm to be used. It was designed for use after scan placement in a FreeSurfer volume and after scan inhomogeneity correction. BEMA was executed on the 99 test scans from ICBM, IPDH, and LIJMC using the combination key yielded from their combined training scans. BEMA was executed on the 30 ZENIT test scans using the ZENIT-derived combination key and on the 6 NEUROVIA scans using the NEUROVIA-derived combination key. Additionally, BEMA was executed using the pooled key from all five data sets on all 135 test scans to assess the ability of a generalized key.

The Dice coefficient was chosen as the set similarity metric to compare the extractor results with the manually derived gold standards. Its formula is $2V_C / (V_1 + V_2)$, where V_C is the number of voxels the two masks share in common, V_1 is the number of voxels in the first mask, and V_2 is the number of voxels in the second mask. The Dice coefficient is 1 if the masks are exactly the same and 0 if the masks share no common voxels. Each resultant extraction, from BEMA, the individual extractor subalgorithms, the raw individual extractor programs, and the NEUROVIA second human rater, was compared to the manual segmented gold standard to produce a Dice coefficient detailing the voxelwise similarity between the extracted mask and the gold standard. To test for differences between the results of the methods, *t* tests (paired by subject) were used.

Additional measures were made for voxelwise false-negative and false-positive rates of each brain extractor tested and for a gray- and white-matter-only Dice coefficient. False-negatives are removed voxels that were in the gold standard mask and false-positives are included voxels that were not in the gold standard mask. The average false-negative or false-positive percentage of each gold standard volume was calculated for each extractor methodology and volumetric maps were created in ICBM152 space to report where each extractor made false-negative or false-positive errors. The gray- and white-matter-only Dice coefficient excluded the effects of CSF in the accuracy of the brain extractors by using the Partial Volume Classifier (Shattuck et al., 2001) to determine which voxels corresponded to gray, white, or CSF types. The gray and white matter voxels were combined in a single mask of brain tissue for each gold standard and extractor mask produced, as well as for the second NEUROVIA human rater, and Dice coefficients comparing these masks were produced. Again, *t* tests paired by subject were used to test for differences between the error rates and the gray- and white-matter-only Dice coefficients.

Results

An example result, for a single subject from the LIJMC data set, is shown in Fig. 2. BEMA accounted for most of the inaccuracies of the individual brain extraction algorithms with information from the other extraction algorithms and the registration procedure. As shown, a small difference in Dice coefficients relates to a noticeable difference in brain masks. BEMA was able to fix nearly all of the false-negative voxels and most of the false-positive voxels. Notable exceptions include small fractions of the superior sagittal sinus and the transverse sinus.

BEMA's average Dice coefficient across all subjects tested was 0.975 with a standard deviation of 0.00529. It performed superior to, possessed significantly higher Dice coefficients than, each individual component extraction and than the extractors run on the raw input data when compared to the manually defined brain masks ($P \ll 0.001$) (Table 3). BEMA also performed superior to

Table 3
Average Dice coefficients

	MRI Watershed	3dIntracranial	Raw BET	BET	Raw BSE	BSE	BEMA (three keys)	BEMA (pooled key)	Human
ICBM	0.951 ** ± 0.0116	0.922 ** ± 0.117	0.959 ** ± 0.00632	0.961 ** ± 0.00410	0.953 ** ± 0.00640	0.952 ** ± 0.00668	0.970 ± 0.00381	0.968 ** ± 0.00592	
IPDH	0.958 ** ± 0.00721	0.954 ** ± 0.00537	0.967 ** ± 0.00526	0.968 ** ± 0.00393	0.944 ** ± 0.0491	0.965 ** ± 0.00347	0.977 ± 0.00259	0.976 ** ± 0.00272	
LIJMC	0.944 ** ± 0.0145	0.943 ** ± 0.0140	0.895 ** ± 0.0753	0.959 ** ± 0.00820	0.922 ** ± 0.0982	0.956 ** ± 0.0116	0.974 ± 0.00470	0.972 ** ± 0.00439	
ZENIT	0.932 ** ± 0.00796	0.975 ** ± 0.00363	0.895 ** ± 0.0153	0.948 ** ± 0.00512	0.973 ** ± 0.00492	0.974 ** ± 0.00507	0.980 ± 0.00374	0.962 ** ± 0.00558	
NEUROVIA	0.952 * ± 0.0216	0.947 * ± 0.0233	0.882 * ± 0.0755	0.970 * ± 0.00508	0.810 * ± 0.117	0.860 * ± 0.133	0.978 ± 0.00221	0.967 ± 0.00962	0.968 * ± 0.00523
All data	0.946 ** ± 0.0147	0.949 ** ± 0.0536	0.920 ** ± 0.0583	0.959 ** ± 0.00933	0.938 ** ± 0.0743	0.957 ** ± 0.035	0.975 ± 0.00529	0.969 ** ± 0.00696	0.970 * ± 0.00521

The average Dice coefficient results for the extraction procedures across all data sets and separated by data set. The errors are given as a single standard deviation. A perfect score, where the automated procedure and gold standard agree on every voxel, would be 1. A score of 0 signifies a situation when the automated procedure and manual gold standard never agree. The Human result is the comparison of the first human rater, the gold standard for the extractor comparisons, to the second human rater. The human rater comparison was only available for the NEUROVIA data set. It is the average of the six test subjects for the NEUROVIA result and the average of all 16 NEUROVIA subjects for the all data result. BSE and BET are the registration prepared results from BEMA and the raw BSE and raw BET results use the algorithms with no preprocessing of the input data. BEMA was run with the three data-set-specific keys to produce its optimal results and with the single pooled key to produce a generalized result.
* $P < 0.01$ for the optimal BEMA results (three keys) having a higher mean Dice coefficient than the extraction method signified.
** $P << 0.001$ for the optimal BEMA results (three keys) having a higher mean Dice coefficient than the extraction method signified.

the second human rater in a pairwise comparison across the six overlapping subjects from the NEUROVIA data ($P < 0.01$) or when compared across all subjects tested ($P \ll 0.001$, not paired). These results were also valid when tested under each individual data set (Table 3). BEMA had a higher Dice coefficient than any other method tested on 133 of the 135 subjects studied (Fig. 3). The two subjects where BEMA did not result in the closest match to the manually derived masks were subjects 77 and 124, from the LIJMC and ZENIT data sets, respectively. On subject 77, BET had a Dice coefficient of 0.957, besting BEMA's result of 0.951. This was BEMA's worst result, still superior to the worst results from all other automated methods. On subject 124, 3dIntracranial had a Dice coefficient of 0.975[1], just beating BEMA's result of 0.974[7]. Raw 3dIntracranial failed to extract any of the presented volumes correctly due to intensity histograms that were not accounted for in its development. Its results have been omitted.

The standard deviation in BEMA's Dice coefficients was smaller than any other automated method when compared across all data sets. The standard deviation of the human interrater Dice coefficients across the NEUROVIA data was slightly smaller than the BEMA result, 0.00521 vs. 0.00529. When compared across the six overlapping subjects, however, BEMA possessed a smaller standard deviation than the interrater results (Table 3). Within data sets, BEMA possessed the smallest standard deviation across Dice coefficients in all cases but one. 3dIntracranial had a slightly smaller standard deviation in the ZENIT data set, although BEMA possessed the higher average Dice coefficient (Table 3).

The pooled key version of BEMA was significantly less accurate than BEMA trained for each individual grouping of the data (Table 3). The pooled key BEMA did perform better than any contributing extractor for the ICBM, IPDH, and LIJMC data sets and was better overall than the contributing algorithms. It suffered on the ZENIT and NEUROVIA data, performing worse than BSE and 3dIntracranial on the ZENIT data set and indistinguishably worse than BET on the NEUROVIA data set.

BEMA performed well regarding both the false-negative and false-positive rates with a false-negative rate of 2.68% and a false-positive rate of 2.21%. For the false-negative rate, only MRI Watershed, 2.17%, outperformed BEMA ($P \ll 0.001$) (Fig. 4). BEMA had a better false-negative rate than 3dIntracranial, BET, BSE, and raw BSE ($P << 0.001$) and was statistically indistinguishable from raw BET. BEMA had a better false-positive rate than all other automated extraction methods ($P \ll 0.001$; BSE, $P < 0.05$) (Fig. 4). BEMA's standard deviations for the false-negative and false-positive rates were 1.26% and 0.989%, respectively, much less than any other method.

The derived maps in ICBM152 space show quantitatively where each algorithm made errors across the 135 test subjects (Fig. 5). BEMA possessed smaller error rates across its map than the other algorithms. BEMA's only consistent error was leaving in tissue along regions consistent with venous sinuses, an error possessed by all the automated methods. BEMA's rate of false-positives along the venous sinuses was noticeably less than the other methods. Other common errors included MRI Watershed leaving in extra tissue along the ventral aspect of the brain, 3dIntracranial leaving out tissue all along the border of the brain, BET leaving out tissue along the anterior borders of the frontal and temporal lobes, raw BET leaving in ventral tissue anterior to the brainstem as well as leaving out tissue along the anterior frontal lobe (though less than BET running in BEMA), BSE leaving out tissue all along the border of the brain, and raw BSE failing to remove tissue from various regions of several brains and possessing the errors of BSE running in BEMA.

The gray- and white-matter-only Dice coefficients were higher on average for BEMA across all data sets, 0.990, than for any other automated method ($P \ll 0.001$) (Table 4). BEMA's results were also statistically indistinguishable from the human interrater results at 0.989 for the entire NEUROVIA data set. Additionally, BEMA possessed the lowest standard deviation across the data sets, 0.00525, for any automated method. The interrater result had a

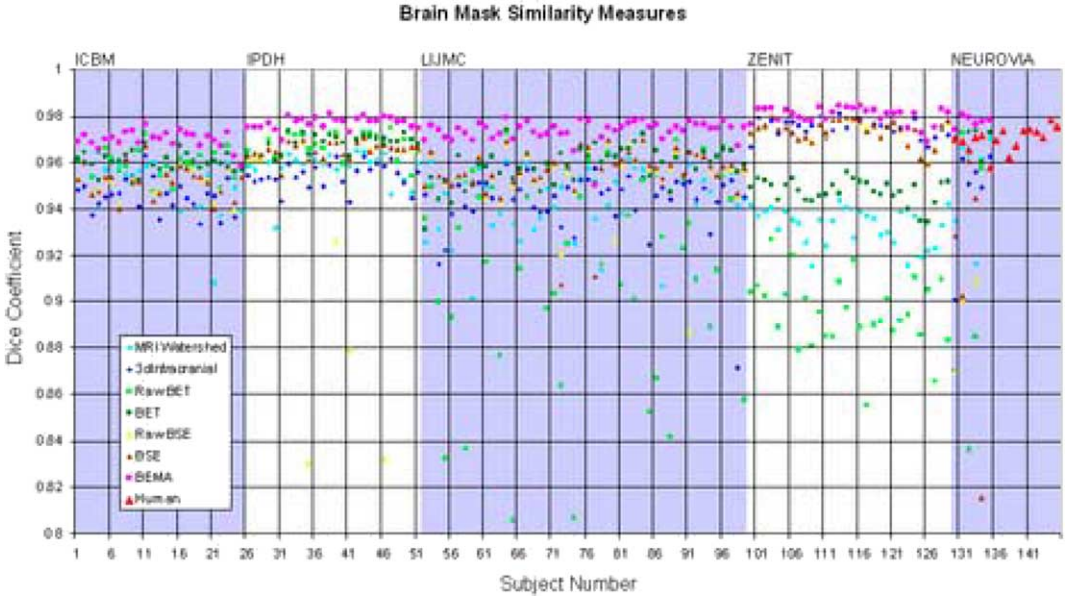

Fig. 3. Dice coefficient results for every test subject studied. BEMA performed superior to every other method in every case except for subjects 77 and 124. The BEMA results were noticeably more consistent across subjects and data sets than any other method. The BSE and BET procedures used in the BEMA algorithm, aided by a preregistration procedure, faired better than the raw BSE and raw BET procedures in most situations.

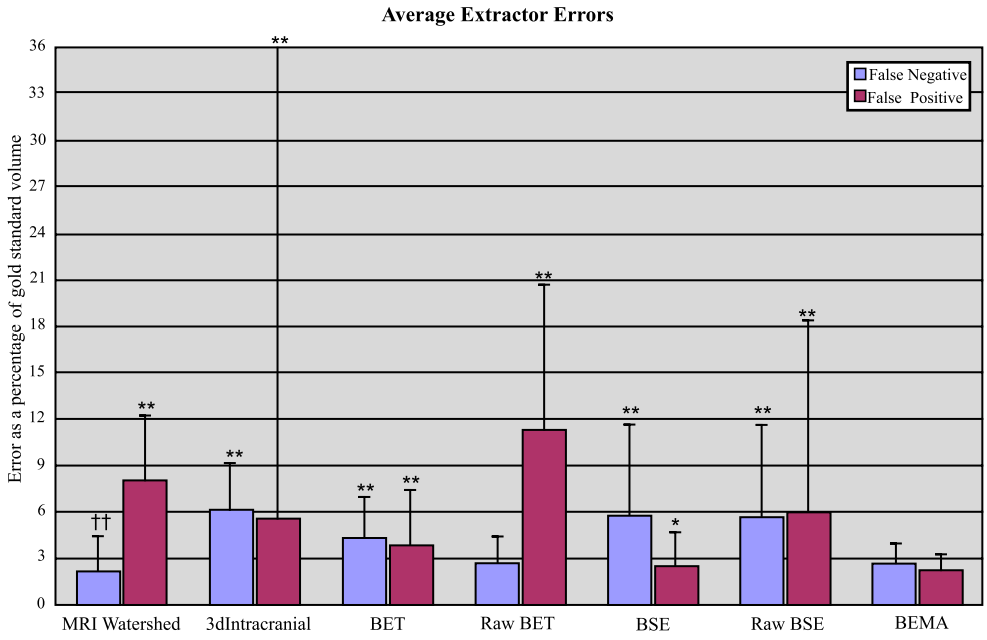

Fig. 4. The false-negative and false-positive results for each extractor averaged across all test subjects studied. Error bars are a single standard deviation. BEMA's results are the only method below the 3% rate for both categories. $^{**}P < 0.001$ and $^{*}P < 0.05$ for BEMA having a lower rate than the extractor signified. $^{\dagger\dagger}P < 0.001$ for the extractor signified having a lower rate than BEMA. Note that MRI Watershed possesses a lower false-negative rate than BEMA though BEMA has a lower standard deviation of its false-negative results.

better standard deviation than BEMA at 0.00367. BEMA did consistently better than the other automated methods within data sets as well (Table 4). BEMA had a higher average gray- and white-only Dice coefficient than the human interrater result across the six NEUROVIA test subjects, but it was not a statistically significant result. The standard deviations of BEMA and the interrater results within the NEUROVIA data were only slightly different, 0.00375 and 0.00348, respectively. Raw BET and raw BSE were left out of this portion of the study due to large errors in the brain extractions for multiple subjects leading to improper tissue classifications.

The training of BEMA produced three very different combination keys for the data sets they represent (Fig. 6). These combination keys were specific to each of the data sets and would cause BEMA to produce less than optimal results when used on the other data sets. The registration for the ZENIT and NEUROVIA keys was robust, leaving large areas where there is never brain and areas where brain is always found. Other areas, near the surface of the brain, used a single algorithm or a Boolean combination of many algorithms to accurately identify the region. These Boolean functions extract regions that the linear registration does not clearly separate.

The ICBM/IPDH/LIJMC key is not as homogeneous as the other keys because it had three failures of registrations while processing the 100 training scans. These registration errors were left in the training because they represent real pitfalls in the data. BEMA successfully dealt with these errors by using the brain extraction algorithms instead of the registrations to find the deep areas of the brain. Higher order Boolean combinations of extractors are also found in the deep regions. They represent the surface regions of the misregistered subjects.

Running times for the training algorithm, using a single MIPS R12000 processor, were approximately 70 h for the ICBM/IPDH/LIJMC key (100 subjects), 19 h for the ZENIT key (30 subjects), and 5 h for the NEUROVIA key (10 subjects). The running times for BEMA and its associated programs and subalgorithms, or pipelets, including necessary pre- and postprocessing of individual algorithm results, are shown in Table 5 for a Silicon Graphics Inc. Origin 3000 providing 50 400-MHz MIPS R12000 processors through a LONI Pipeline Server. BEMA took much more time, approximately 30 min, to execute than any of the individual extractor programs used within it. However, with multiple processors available, BEMA ran only 5 s longer than the MRI Watershed subalgorithm that requires a FLIRT-based registration, an MRI Normalize step, and a couple of format conversions. FreeSurfer's MRI Normalize takes up most of the running time at approximately 20 min. A meta-algorithm can only run as fast as its slowest path. Utilizing the parallel nature of the LONI Pipeline Processing Environment, BEMA used 3.5% more time to simultaneously extract three subjects, approximately 31 min, than it used to extract one subject.

Discussion

Results

The results suggest that BEMA is consistently superior at matching results with the gold standard examples of brain masks than any of the individual extractors that are combined to form it. BEMA even edges out the interrater results from the NEUROVIA data when looking at the Dice coefficients of the raw masks. Furthermore, BEMA possesses the lowest standard deviation in its results of any automated method tested and is on par with the human results, even besting the interrater results when compared within the NEUROVIA data set. This suggests that BEMA is more robust, producing more reliable results more often than other methods. Additionally, BEMA's lowest Dice coefficient was 0.951. This is better than the poorest result from MRI Watershed

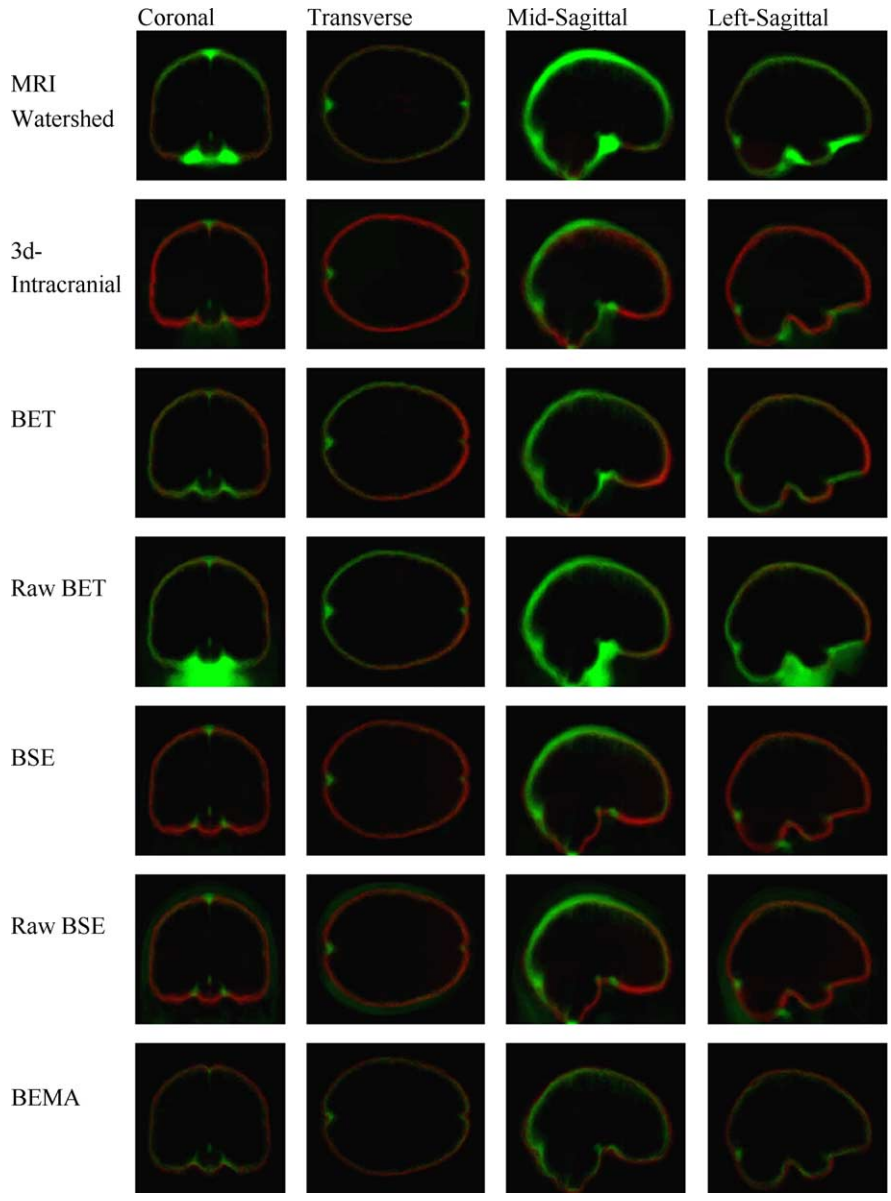

Fig. 5. False-negative (red) and false-positive (green) maps of errors for each of the algorithms across all subjects tested. The maps are scaled from 0% to 100% errors for a given voxel location in ICBM152 space. Note: Raw BSE possesses a few failed extractions showing up as false-positives that fall below detectable levels at this scaling. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(0.902), 3dIntracranial (0.359), raw BET (0.612), BET (0.931), raw BSE (0.531), and BSE (0.609). Only the human interrater result fared better in its lowest coefficient (0.958). It is also important to note that for our purposes BSE was run with a single set of parameters optimized across all data sets simultaneously. The BSE algorithm's results, specifically for the NEUROVIA data, may improve when a different parameter set is chosen.

The BEMA combination keys produced were optimal for the particular data sets that they were trained on. Three (ICBM, IPDH, and LIJMC) of the five data sets were combined to produce one combination key among them. The results of BEMA using this combination key for all three data sets were consistently better than the contributing extractors. This first combination key was attempted on the ZENIT and NEUROVIA data sets. It produced results that were on par with the contributing algorithms, but no better. The pooled combination key was fashioned for the data sets that used 10 training scans from all five data sets. This key fared better on the ZENIT and NEUROVIA data but still performed suboptimally and was marginally worse on the ICBM, IPDH, and LIJMC data sets. New, separate combination keys were formed for the ZENIT and NEUROVIA data sets. These combination keys performed best on their respective test data. These results suggest that the scanner and acquisition protocol contribute greatly to the results of the brain extraction algorithms on the data volumes. Some data sets were combined without diminishing the results, the ICBM, IPDH, and LIJMC data sets, but others possessed properties that kept them from being grouped with the previous data sets, the ZENIT and NEUROVIA data sets. Noticeable differences in the data sets included the level of noise in the scans and the contrast between the tissue types.

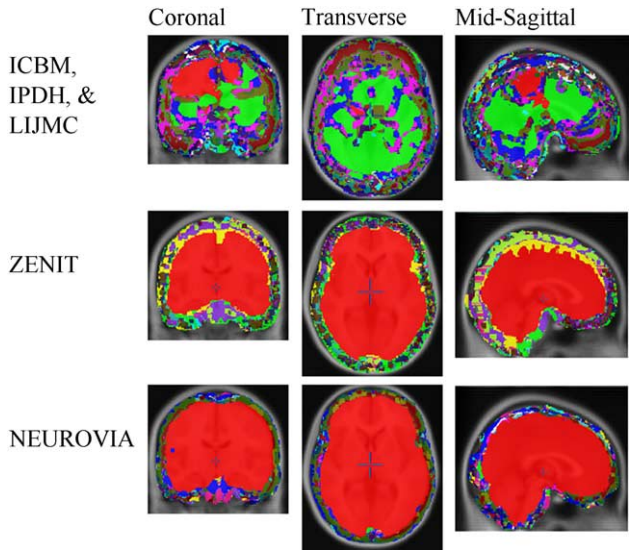

Fig. 6. The combination keys generated from each of the training sets overlaid on the ICBM152 average. Based on the registration results, red denotes regions that are always brain and clear voxels are regions that are never brain. Other colors represent regions of a single extractor being used or multiple extractors being used in one of the 162 other Boolean functions available. The noticeable lack of a large red region in the ICBM/IPDH/LIJMC combination key is due to a few bad registrations forcing the trainer to use the brain extractors to elucidate regions that are usually found by the registration procedure. The deep regions using multiple methods are around the surfaces of the misregistered brains. These misregistrations did not have a detectable impact on the test cases. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

The preprocessing of the data sets helped the individual extractors that contribute to BEMA immensely. Both BET and BSE were made much more robust with more accurate overall results by using registration to crop out obvious nonbrain tissues. BET's results, however, were seemingly hampered in the anterior frontal and temporal regions by our methodology. This can be addressed in future BEMA versions by providing both the raw BET and modified BET results to BEMA to enable a regional choice. Finally, in its raw mode, 3dIntracranial could not process the data sets we used. It was enhanced, and in fact given the ability to function on these data sets, by providing it with needed intensity parameters about the data through further automated methods.

The running time of the BEMA approach is slightly greater than the longest subalgorithm in the pipeline. In this case, the FreeSurfer approach requires registration and intensity normalizations that take up most of the 30-min execution time. Given the parallel nature of the LONI Pipeline Processing Environment, however, the number of subjects to extract can be increased greatly to approximately 45 subjects on our 50 processor Pipeline Server before a noticeable increase in execution time occurs. Substantial running time is required for the BEMA Trainer when using a large number of subjects. The results have shown, however, that accurate results were also achieved for a 30- and 10-subject training session, greatly reducing the required time to train BEMA. Additional time savings can be obtained by reducing the per voxel window size in the training session. A 3-mm isotropic window ran 5.5 times faster than the 5-mm window, and no window ran approximately 140 times faster than the 5-mm window training session. The average Dice coefficient decreased by only 0.0003 for the 3-mm window and 0.002 for no window for the ICBM/IPDH/LIJMC data sets.

BEMA's current implementation is within the LONI Pipeline Processing Environment (http://www.loni.ucla.edu/Software/) and is available through the LONI Pipeline Server to the neuroimaging community (http://www.loni.ucla.edu/NCRR/Application/Collaborator_Application.jsp). The pipeline implementation was made possible by the previous inclusion of all needed algorithms and processing modules on the LONI Pipeline Server. Recreation of BEMA in other environments or a scripting language is also possible but requires the acquisition and compilation of all the required processing packages and accessories.

Improvements

As individual automated brain extraction algorithms improve, so will BEMA. Given a representative training set and BEMA's use of a generalized strategy, the overall best algorithm, or group of algorithms, for extracting a region will be employed. If an algorithm that is used in a BEMA combination key is improved, the results of BEMA also will improve. Retraining for a new

Table 4
Average gray- or white-matter-only Dice coefficients

	MRI Watershed	3dIntracranial	BET	BSE	BEMA	Human
ICBM	0.977 ** ± 0.0126	0.930 ** ± 0.194	0.980 ** ± 0.00461	0.968 ** ± 0.00999	0.983 ± 0.00549	
IPDH	0.945 ** ± 0.193	0.986 ** ± 0.00546	0.989 ** ± 0.00292	0.987 ** ± 0.00390	0.993 ± 0.00206	
LIJMC	0.968 ** ± 0.0138	0.974 ** ± 0.0130	0.980 ** ± 0.00674	0.975 ** ± 0.0107	0.989 ± 0.00428	
ZENIT	0.976 ** ± 0.00549	0.989 ** ± 0.00393	0.982 ** ± 0.00484	0.991 ** ± 0.00321	0.993 ± 0.00221	
NEUROVIA	0.967 * ± 0.0230	0.968 * ± 0.0172	0.987 ± 0.00593	0.878 * ± 0.138	0.990 ± 0.00375	0.986 ± 0.00348
All data	0.967 ** ± 0.0848	0.971 ** ± 0.0852	0.983 ** ± 0.00638	0.975 ** ± 0.0360	0.990 ± 0.00525	0.989 ± 0.00367

Average Dice coefficient results, across all subjects studied and separated by data set, for masks including only gray matter and white matter voxels. Errors are a single standard deviation. Extracted brains from each procedure, and the gold standard masks, were classified for gray, white, and CSF voxels using the Partial Volume Classifier. Only voxels of gray or white matter were kept. Dice coefficients were computed on these submasks. Results for each extraction method were compared to the BEMA results. Comparison of the first human rater to the second human rater was only available for the NEUROVIA data set. The Human result for the NEUROVIA data is only for the six test subjects and the result for all data is from all 16 NEUROVIA subjects. The human result was statistically indistinguishable from the BEMA approach, as was the BET approach for only the NEUROVIA data set. Raw BET and Raw BSE were omitted due to multiple severe extraction errors confounding PVC's ability to correctly classify voxels.

* $P < 0.01$ for BEMA having a higher mean Dice coefficient than the extractor signified.

** $P << 0.001$ for BEMA having a higher mean Dice coefficient than the extractor signified.

Table 5
Execution times

	One subject, one processor	One subject, six processors	Three subjects, one processor	Three subjects, 18 processors
Raw BET	29 s	29 s	1 min, 7 s	29 s
Raw BSE	21 s	21 s	1 min, 7 s	23 s
Program in BEMA				
MRI Watershed	1 min, 15 s	1 min, 15 s	3 min, 49 s	1 min, 24 s
3dIntracranial	16 s	16 s	47 s	16 s
BET	16 s	16 s	51 s	18 s
BSE	15 s	15 s	40 s	17 s
MRI Normalize	19 min, 52 s	19 min, 52 s	51 min, 4 s	20 min, 8 s
FLIRT	3 min, 32 s	3 min, 32 s	9 min, 39 s	3 min, 32 s
Pipelet in BEMA				
MRI Watershed	30 min, 5 s	30 min, 5 s	82 min, 29 s	31 min, 3 s
3dIntracranial	6 min, 5 s	5 min, 8 s	17 min, 52 s	5 min, 22 s
BET	4 min, 17 s	4 min, 17 s	12 min, 6 s	4 min, 21 s
BSE	4 min, 16 s	4 min, 16 s	12 min, 3 s	4 min, 20 s
BEMA Total	34 min, 7 s	30 min, 10 s	95 min, 33 s	31 min, 14 s

The running times for BEMA, its associated critical programs, and its pipelets or subalgorithms with the enhancing preprocessing steps including registration, tissue classification, volume resampling, format conversions, as well as other ancillary steps. All results are via the LONI Pipeline Processing Environment and a LONI Pipeline Server running on a Silicon Graphics Inc. Origin 3000 providing 50 400-MHz MIPS R12000 processors. The number of processors shown in the table is the maximum number of processors the environment was allowed to use in an execution—associated speed increases reflect the parallel nature of the environment. The single subject is the LIJMC subject used for the example results in Fig. 2. The three subject numbers use two more random subjects from the LIJMC data set.

combination key may garner even more improvements. Additional algorithms that are specifically good at a particular region of anatomy can also help immensely. If one algorithm identifies the superior sagittal and/or transverse sinuses well, then it can be utilized by BEMA solely in those regions to correct that common error.

Improvements to the registration technique will also improve BEMA. The errors seen in registration for the ICBM/IPDH/LIJMC data sets and their combination key did not dramatically affect the outcomes of the test cases because they were evenly distributed in the training and test sets, three and two failures of registration, respectively. This allowed the meta-algorithm to overcome the registration failures and correctly characterize regions based on extractor results. If the failures of registration did not occur during training, the two misregistered test cases would possess inaccurate extractions with BEMA. Utilizing a more robust and accurate, possibly nonlinear, registration algorithm will improve BEMA. Registration improvements may also include using different atlas spaces that conform better to a specific subject's anatomy. That is, an Alzheimer's disease atlas space (Thompson et al., 2001) could better identify regions in an Alzheimer's patient than an atlas derived from healthy young adults.

To efficiently expand BEMA past four input brain extractors, a different approach must be taken. The total number of Boolean functions available with n input extractors is 2^{2^n}. With an increasing number of input extraction algorithms, this quickly becomes an intractable number of functions to search, even when not allowing inverses to occur. This is especially true when a separate search is done for each voxel in the volume. Instead of trying to search this space for large numbers of extraction algorithms, a pruning step is implemented to keep the search space small. At each voxel location, the four most accurate extraction algorithms for the defined neighborhood are used as the inputs to find the optimal four-input Boolean function. Which four algorithms are chosen are stored in a selection key volume that accompanies the combination key. Together, these two keys and their atlas space determine which algorithms to use and how to combine their results to get the best possible overall outcome.

The BEMA approach can be extended to work on data sets from varying scanners, modalities, and protocols with different dimensions, contrast, and noise levels by being provided with additional brain extraction algorithms and new training sets to produce unique combination keys. The keys will utilize various combinations of the input algorithms and may even use completely different algorithms in some cases. This approach will not only provide for the scans of varying contrast and noise seen here, but it can also group extractors that work on vastly different modalities under one generalized extraction protocol. The appropriate key for T1-weighted, T2-weighted, PD, DTI, or multimodality data sets would be provided to allow BEMA to use the correct procedures. Separate keys should also exist for various subject groups, such as children, young adults, elderly, and atrophic subjects, as they have also been shown to be a factor in the accuracy of various extraction algorithms (Fennema-Notestine et al., 2003). With the aid of the LONI Pipeline Processing Environment (Rex et al., 2003), a single module will be presented for brain extraction, and the correct methodology will be selected by the provided combination key.

Conclusions

In our tests, BEMA was able to produce results that were of superior accuracy and increased robustness when compared to any of the brain extraction algorithms used in its processing. BEMA was also on par with, and in some measures better than, the human interrater results for the NEUROVIA data set. One drawback of BEMA is to gain optimal performance, there exists a need to train the meta-algorithm for new data when a data set is sufficiently different, in contrast, noise, resolution, or possibly tissue atrophy, from the previously trained data sets. However, new scanners and protocols can be utilized to train a BEMA algorithm when data acquisition begins and then the algorithm may be used with new acquired data. Keys generated for similar scanners and protocols at other institutions may also be useful for newly acquired data. Additionally, as the number of informative algorithms that are available to BEMA increases, and the quality of their results increases, BEMA will become even more robust and capture the best possible results for more data sets from a larger variety of acquisitions and subject populations.

Acknowledgments

This work was generously supported by grants from the National Institute of Mental Health (1 P20 MH65166 and 5 P01 MH52176) and the National Center for Research Resources (2 P41 RR13642 and 2 M01 RR00865), with a supplement for the Biomedical Informatics Research Network (2 P41 RR13642) (http://www.nbirn.net/). DER is supported, in part, by an ARCS Foundation scholarship and a National Institutes of General Medical Sciences Medical Scientist Training Program grant (GM08042). The authors wish to thank Drs. Robert Bilder, John Mazziotta, and Tonmoy Sharma for providing the LIJMC, ICBM, and IPDH data sets, respectively. The authors also wish to thank the members of the Laboratory of Neuro Imaging for their help and support.

References

Alfano, B., Brunetti, A., et al., 1997. Unsupervised, automated segmentation of the normal brain using a multispectral relaxometric magnetic resonance approach. Magn. Reson. Med. 37, 84–93.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry—The methods. NeuroImage 11 (6 Pt 1), 805–821.

Baillet, S., Mosher, J., et al., 1999. Brainstorm: a Matlab toolbox for the processing of MEG and EEG signals. NeuroImage 9, S246.

Bajcsy, R., Lieberson, R., et al., 1983. A computerized system for elastic matching of deformed radiographic images to idealized atlas images. J. Comput. Assist. Tomogr. 7, 618–625.

Bedell, B.J., Narayana, P.A., 1998. Volumetric analysis of white matter, grey matter, and CSF using fractional volume analysis. Magn. Reson. Med. 39, 961–969.

Boesen, K., Rehm, K., et al., 2003. Quantitative comparison of four brain extraction algorithms. NeuroImage Abs. 19 (2) (CD-ROM).

Bomans, M., Hohne, K., et al., 1990. 3-D segmentation of MR images of the head for 3-D display. IEEE Trans. Med. Imag. 9 (2), 177–183.

Brummer, M.E., Merseau, R.M., et al., 1993. Automatic detection of brain contours in MRI data sets. IEEE Trans. Med. Imag. 12, 153–166.

Christensen, G.E., Rabbitt, R.D., et al., 1996. Deformable templates using large deformation kinematics. IEEE Trans. Image Process. 5, 402–417.

Collins, D.L., Holmes, C.J., et al., 1995. Automatic 3D model-based neuroanatomical segmentation. Hum. Brain Mapp. 3 (3), 190–208.

Collins, D.L., Zijdenbos, A.P., et al., 1999. ANIMAL+INSECT: improved cortical structure segmentation. Proc. Annu. Symp. Inf. Process. Med. Imag. 1613, 210–223.

Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance Neuroimages. Comput. Biomed. Res. 29 (3), 162–173.

Cox, R.W., Hyde, J.S., 1997. Software tools for analysis and visualization of fMRI data. NMR Biomed. 10 (4–5), 171–178 (pii).

Dale, A.M., Sereno, M.I., 1993. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: a linear approach. J. Cogn. Neurosci. 5, 162–176.

Dale, A.M., Fischl, B., et al., 1999. Cortical surface-based analysis: I. Segmentation and surface reconstruction. NeuroImage 9 (2), 179–194.

Davatzikos, C., 1997. Spatial transformation and registration of brain images using elastically deformable models. Comput. Vis. Image Underst. 66, 207–222.

Evans, A.C., Collins, D.L., et al., 1994. Three-dimensional correlative imaging: applications in Human brain mapping. In: Huerta, M. (Ed.), Functional Neuroimaging: Technical Foundations. Academic Press, San Diego, pp. 145–162.

Fennema-Notestine, C., Ozyurt, I.B., et al., 2003. Bias correction, pulse sequence, and neurodegeneration influence performance of automated skull-stripping methods. Society for Neuroscience, New Orleans.

Fischl, B., Sereno, M.I., et al., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8 (4), 272–284.

Held, K., Kops, E.R., et al., 1997. Markov random field segmentation of brain MR images. IEEE Med. Imag. 16, 878–886.

Hohne, K.H., Hanson, W.A., 1992. Interactive 3D segmentation of MRI and CT volumes using morphological operations. J. Comput. Assist. Tomogr. 16 (2), 285–294.

Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5 (2), 143–156.

Kapur, T., Grimson, W.E.L., et al., 1996. Segmentation of brain tissue from magnetic resonance images. Med. Image Anal. 1, 109–127.

Lawson, J.A., Vogrin, S., et al., 2000. Cerebral and cerebellar volume reduction in children with intractable epilepsy. Epilepsia 41 (11), 1456–1462.

Lemieux, L., Hagemann, G., et al., 1999. Fast, accurate and reproducible automatic segmentation of the brain in T1-weighted volume magnetic resonance image data. Magn. Reson. Med. 42, 127–135.

Lemieux, L., Hammers, A., et al., 2003. Automatic segmentation of the brain and intracranial cerebrospinal fluid in T1-weighted volume MRI scans of the head, and its application to serial cerebral and intracranial volumetry. Magn. Reson. Med. 49 (5), 872–884.

MacDonald, D., Avis, D., et al., 1994. Multiple surface identification and matching in magnetic resonance images. Proc. Vis. Biomed. Comput. 2359, 160–169.

MacDonald, D., Kabani, N., et al., 2000. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. NeuroImage 12 (3), 340–356.

Miller, M.I., Christensen, G.E., et al., 1993. Mathematical textbook of deformable neuroanatomies. Proc. Natl. Acad. Sci. 90 (24), 11944–11948.

Rehm, K., Shattuck, D., et al., 1999. Semi-automated stripping of T1 MRI volumes: I. Consensus of intensity- and edge-based methods. NeuroImage Abs. 9 (6), S86.

Rex, D.E., Ma, J.Q., et al., 2003. The LONI pipeline processing environment. NeuroImage 19 (3), 1033–1048.

Sandor, S., Leahy, R., 1997. Surface-based labeling of cortical anatomy using a deformable database. IEEE Trans. Med. Imag. 16, 41–54.

Schalkoff, R.J., Shaaban, K.M., 1999. Image processing meta-algorithm development via genetic manipulation of existing algorithm graphs. SPIE Proc. Vis. Inf. Process. VIII 3716, 61–70.

Schroder, T., Rosler, U., et al., 1999. Optimizing deconvolution techniques by the application of the Munchhausen meta algorithm. Biomed. Tech. (Berl) 44 (11), 308–313.

Shaaban, K.M., Schalkoff, R.J., 1995. Image processing and computer vision algorithm selection and refinement using an operator-assisted meta-algorithm. SPIE Proc. Vis. Inf. Process. IV 2488, 77–87.

Shattuck, D.W., Sandor-Leahy, S.R., et al., 2001. Magnetic resonance image tissue classification using a partial volume model. NeuroImage 13 (5), 856–876.

Shattuck, D.W., Leahy, R.M., 2002. BrainSuite: an automated cortical surface identification tool. Med. Image Anal. 6 (2), 129–142.

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17 (3), 143–155.

Smith, S., Bannister, P., et al., 2001. FSL: new tools for functional and structural brain image analysis. Seventh International Conference on Functional Mapping of the Human Brain, Brighton, UK, NeuroImage Abstracts.

Smith, S.M., Zhang, Y., et al., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. NeuroImage 17 (1), 479–489.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotaxic Atlas of the Human Brain. Thieme, New York.

Thompson, P.M., Mega, M.S., et al., 2001. Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. Cereb. Cortex 11 (1), 1–16.

Ward, B.D., 1999. Intracranial Segmentation. Technical report. Medical College of Wisconsin: http://afni.nimh.nih.gov/pub/dist/doc/3dIntracranial.pdf.

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

Woods, R.P., Grafton, S.T., et al., 1998. Automated image registration: II. Intersubject validation of linear and nonlinear models. J. Comput. Assist. Tomogr. 22 (1), 153–165.

Zhang, Y., Brady, M., et al., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation–maximization algorithm. IEEE Trans. Med. Imag. 20 (1), 45–57.