ELSEVIER

NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 22 (2004) 1255–1261

Quantitative comparison of four brain extraction algorithms

Kristi Boesen,[a,*] Kelly Rehm,[b] Kirt Schaper,[a] Sarah Stoltzner,[c] Roger Woods,[c]
Eileen Lüders,[c] and David Rottenberg[a,b]

[a] *Department of Neurology, University of Minnesota, Minneapolis, MN 55417, USA*
[b] *Department of Radiology, University of Minnesota, Minneapolis, MN 55417, USA*
[c] *Department of Neurology, University of California, Los Angeles, CA 90095-1361, USA*

Received 24 November 2003; revised 2 March 2004; accepted 3 March 2004

In a companion paper (Rehm et al., 2004), we introduced Minneapolis Consensus Strip (McStrip), a hybrid algorithm for brain/non-brain segmentation. In this paper, we compare the performance of McStrip and three brain extraction algorithms (BEAs) in widespread use within the neuroimaging community—Statistical Parametric Mapping v.2 (SPM2), Brain Extraction Tool (BET), and Brain Surface Extractor (BSE)—to the "gold standard" of manually stripped T1-weighted MRI brain volumes. Our comparison was based on quantitative boundary and volume metrics, reproducibility across repeat scans of a single subject, and assessments of performance consistency across datasets acquired on different scanners at different institutions. McStrip, a hybrid method incorporating warping to a template, intensity thresholding, and edge detection, consistently outperformed SPM2, BET, and BSE, all of which rely on a single algorithmic strategy.

Keywords: Extraction algorithms; Brain; McStrip

Introduction

Segmentation of brain/non-brain tissue is one of the most time-consuming preprocessing steps performed in neuroimaging laboratories, and numerous brain extraction algorithms (BEAs) have been developed to perform this step automatically. While BEAs speed up overall image processing, their output varies greatly and can affect the results of subsequent image analysis. In a companion paper (Rehm et al., 2004), we introduced a consensus approach to brain/non-brain segmentation, and in this paper, we provide a quantitative comparison of the performance of four BEAs against the "gold standard" of expert manual brain extraction using high-resolution T1-weighted MRI brain volumes.

Methods for brain/non-brain segmentation currently available to the neuroimaging community are based on a priori information, deformable models, edge detection, and intensity thresholding. To date, few BEA comparisons have been published in the scientific literature. When a new BEA is implemented, its performance is often compared to that of one or more existing BEAs (Lemeiux et al., 1999; Shan et al., 2002; Smith, 2002). Lee et al. (2003) reported the most comprehensive comparison published to date: using two synthetic datasets they compared two automated BEAs [Brain Extraction Tool, BET: Smith (2002) and Brain Surface Extractor, BSE: Shattuck et al. (2001)] and two semiautomated methods [Analyze 4.0: Richard (2000) and mRG: Yoon et al. (2001)] with respect to a similarity index and tissue classification errors. While this study provides an unbiased assessment of the relative performance of the algorithms tested, the relevance of the authors' conclusions to brain/non-brain segmentation is limited by the fact that BEA performance was not evaluated using real MRI brain volumes.

We compared our own [Minneapolis Consensus Strip, McStrip: Rehm et al. (1999, 2004)] and three popular BEAs [Statistical Parametric Mapping v.2 (SPM2): Ashburner and Friston (2000), BET, and BSE] with regard to (i) the reproducibility of their output when tested on repeat scans of the same subject and (ii) their ability to accurately capture both brain volume and brain boundary. The selected BEAs were developed in academic neuroimaging centers, have been published in extenso (with the exception of McStrip), and are freely available for download; moreover, each of the four represents a different algorithmic approach to the problem of brain/non-brain segmentation—hybrid, a priori information, deformable model, and edge detection, respectively. We did not include an algorithm that relies exclusively on intensity thresholding because all of the BEAs compared make use of intensity thresholding at some stage in their operation.

Materials and methods

Datasets

Three sets of T1-weighted MRI scans of normal volunteer subjects were used for our comparison. Datasets #1 and #2 correspond to Datasets #1 and #2 discussed in the companion paper. Dataset #3 was acquired at the Center for Neuroscientific

---

* Corresponding author. International Neuroimaging Consortium (11P), University of Minnesota, 1 Veterans Drive, Minneapolis, MN 55417. Fax: +1-612-725-2068.

E-mail address: kristi@neurovia.umn.edu (K. Boesen).

1053-8119/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2004.03.010

Innovation and Technology (ZENIT) at the University of Magdeburg (Magdeburg, Germany).

Dataset #1 (six repeat scans of a single subject) was used for a reproducibility study in which BEA performance was assessed by comparing the brain masks created by each algorithm, while Datasets #2 and #3 were used for a comparison of BEA performance against the gold standard of expert manual brain extraction. Each volume in Dataset #2 was stripped manually by two of the authors (KR and SS); KR manually corrected brain masks created by an early version of McStrip, and SS manually delineated the cortical envelope for each scan (Fig. 1). Dataset #3 consisted of 16 high-resolution scans acquired using a spoiled gradient echo pulse sequence on a 1.5T GE scanner (TR = 24 ms, TE = 8 ms, flip angle = 30°; voxel dimensions = 0.98 × 0.98 × 1.5 mm). Volumes from 16 subjects were randomly selected from a larger group of normal, healthy volunteers participating in another study (30 males, mean age 25; 30 females, mean age 24). EL manually corrected brain masks created by BSE (Fig. 1).

Algorithms

SPM2

SPM2 does not explicitly generate a strip mask; however, one can be created as the binarized sum of the grey matter (GM) and white matter (WM) compartments after segmentation (J. Ashburner, personal communication, 2003; Fig. 2). The *Realign* and *Normalize* routines were employed to transform the volumes into Talaraich space, and the *Segment* routine was employed to create GM, WM, and CSF compartments. Brain masks were then created using Interactive Data Language (IDL, Research Systems, Inc., Boulder, CO) as the binarized sum of the GM and WM compartments, and *Normalize* was employed to transform the masks back into native space for comparison against the manual masks. To maintain a constant volume size throughout this "stripping" process, bounding boxes employed by *Normalize* were customized for the two transforms, and voxel sizes were inputted.

BET

BET makes an intensity-based estimation of the brain/non-brain threshold, determines the center of gravity of the head, defines an initial sphere based on the center of gravity, and deforms the tessellated sphere outward toward the brain surface (Smith, 2002). Two parameters are user-adjustable: fractional intensity threshold (FIT, default = 0.50) and threshold gradient (TG, default = 0.0).

Fixed parameters. One volume within Dataset #1 and two volumes within Datasets #2 and #3 were selected at random to initialize the parameters. Those parameters resulting in the best strip (removal of scalp, skull, CSF, and dura with preservation of brain tissue as determined by visual inspection) were selected and applied to all volumes within their respective datasets. Using IDL the volumes were cropped to lie within bounding boxes consistent with those used for the BSE stage of McStrip, and the parameters listed in Table 1 were applied to the cropped volumes.

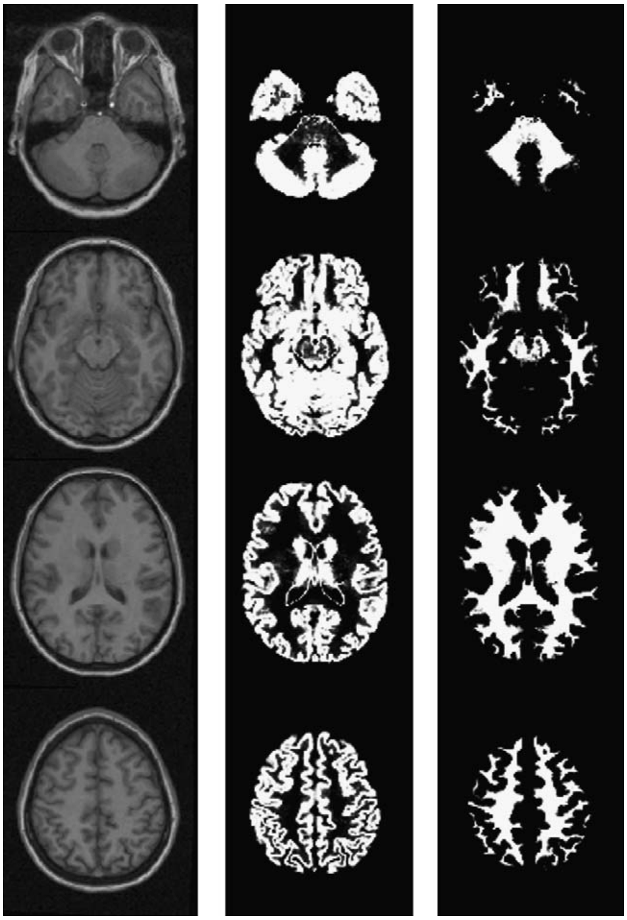

Fig. 2. Output from the *Segment* routine in SPM2. Columns left to right: T1 MRI volume, GM compartment, WM compartment.

Omniscient BET (oBET). The "omniscient user" serves as a proxy for an expert user of BET. In Datasets #2 and #3, parameters were individually selected for each volume after performing an exhaustive search against the manual masks for the parameter combination that produced the least Misclassified Tissue (see Performance metrics below).

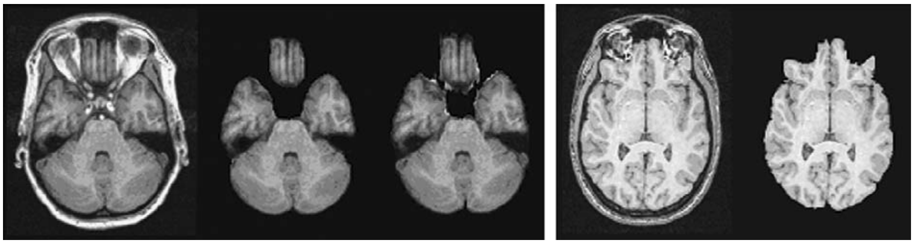

Fig. 1. Manual masks used as the gold standard for performance evaluation. Left panel: Dataset #2, Subject A. From left to right: raw T1 MRI volume, manual mask (KR), manual mask (SS). Right panel: Dataset #3, Subject A. From left to right: raw T1 MRI volume, manual mask (EL).

Table 1
BEA parameters: default values and parameters used for each dataset

BEA[a]	Default values	Dataset #1	Dataset #2	Dataset #3
BET (fixed parameters)				
Fractional intensity threshold	0.50	0.55	0.45	0.50
Threshold gradient	0.0	0.0	0.0	0.0
BSE (fixed parameters)				
Anisotropic smoothing kernels	5	15	15	15
Iterations	3	3	5	3
Edge detection σ's	0.75	0.90	0.75	0.060
McStrip				
Warp mask	Third-order polynomial	Third-order polynomial	Third-order polynomial	Third-order polynomial
Dilation kernel (voxels)	7 × 7 × 7	15 × 15 × 13	7 × 7 × 7	7 × 3 × 3
Grey threshold (%)	15–35	25–35	15–35	30–50
Smoothing kernel FWHM (mm)	2	3	3	2
BSE parameters (within McStrip)				
Anisotropic smoothing kernels	5, 10, 15	5, 10, 15	5, 10, 15	5, 10, 15
Iterations	3	3	3	3
Edge detection σ's	0.60, 0.64, 0.70, 0.80, 0.90	0.60, 0.64, 0.70, 0.80, 0.90	0.60, 0.64, 0.70, 0.80, 0.90	0.60, 0.64, 0.70, 0.80, 0.90
oBET				
Fractional intensity threshold	0.50	–	0.30–0.70	0.30–0.70
Threshold gradient	0.0	–	−0.20–0.20	−0.20–0.20
oBSE				
Anisotropic smoothing kernels	5	–	0, 5, 10, 15	0, 5, 10, 15
Iterations	3	–	1, 3, 5	1, 3, 5
Edge detection σ's	0.75	–	0.50–0.95	0.50–0.95

[a] SPM2 is excluded from this table as the masking process in SPM2 does not involve any initial parameter settings.

BSE

BSE is an edge-based method that employs anisotropic diffusion filtering. Edge detection is implemented using a 2D Marr–Hildreth operator employing low-pass filtering with a Gaussian kernel and localization of zero crossings in the Laplacian of the filtered image; the final step is morphological processing of the edge map (Shattuck et al., 2001). Three parameters are user-adjustable: anisotropic smoothing kernel (ASK, default = 5), number of iterations (ITER, default = 3), and edge detection σ (default = 0.75).

Fixed parameters. One volume within Dataset #1 and two volumes within Datasets #2 and #3 and were selected at random to initialize the parameters. Volumes were individually cropped to lie within bounding boxes consistent with those used for the BSE stage of McStrip. Those parameters resulting in the best strip were selected and applied to all volumes within their respective datasets.

Omniscient BSE (oBSE). In Datasets #2 and #3, parameters were individually selected for each volume as described above for oBET.

McStrip

McStrip is an automatic hybrid algorithm implemented in IDL that incorporates BSE and requires no user intervention; it relies on warping to a template, intensity thresholding, and edge detection (Rehm et al., 2004).

It should be noted that these four algorithms were not developed with the same design goals. BET, BSE, and McStrip algorithms are designed to remove muscle, fat, skin, bone, and dura to produce a cortical envelope for use in subsequent data analysis, while SPM2 masks are built from the results of a GM–WM–CSF tissue segmentation operation. A summary of the parameter settings used for each BEA is provided in Table 1.

Performance metrics

Volumes from Dataset #1 were used to assess reproducibility and were compared against each other within BEA groups. Manual masks created by KR and SS for Dataset #2 and by EL for Dataset #3 were used as the gold standard against which BEA masks were compared.

Volume metrics

Misclassified Tissue describes the percentage of high-intensity voxels (i.e., voxels above the GM–CSF threshold) incorrectly

Table 2
Dataset #1; BEA performance on repeat scans of a single subject

BEA	Misclassified Tissue (%)	Similarity Index	Correct Boundary (%)	Pertinent Boundary (%)	Processing Time
SPM2	2.5	0.986	77.9	81.5	35 min
BET (fixed parameter)	4.6	0.976	64.1	65.7	**40 s**
BSE (fixed parameter)	4.3	0.977	81.9	82.2	1 min
McStrip	**1.2**	**0.994**	**83.1**	**86.2**	13 min

The means for five scans are reported, and all percentages are based on comparisons of Masks #2–#6 to Mask #1 (see text).
The best performance for each metric is bolded.

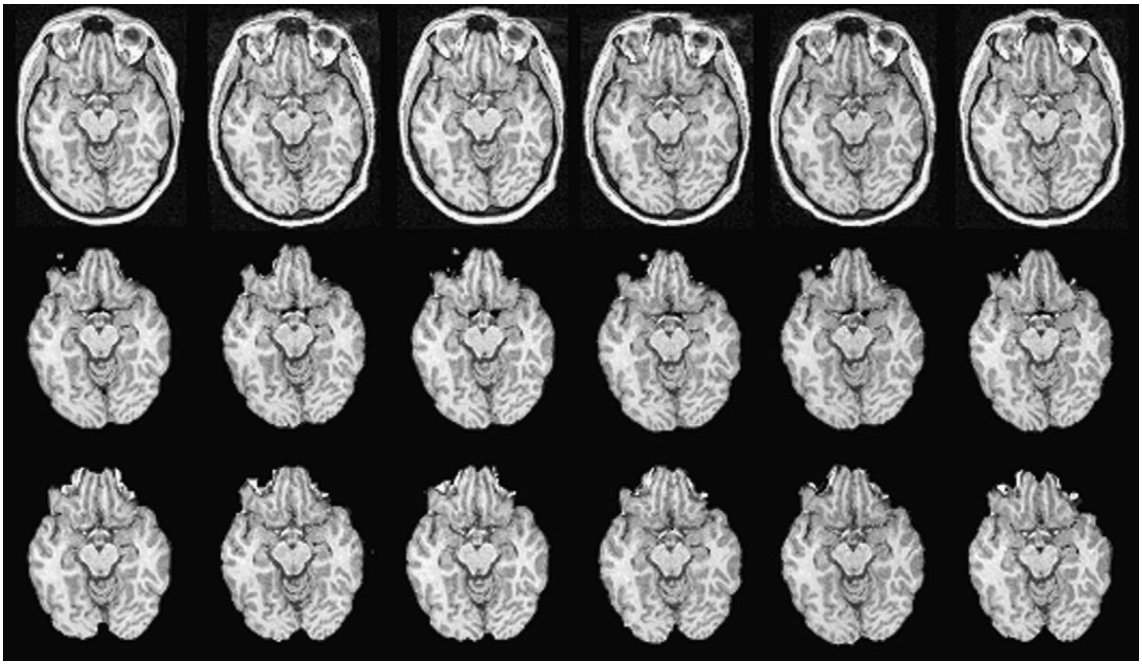

Fig. 3. Dataset #1, Subject A. Raw volumes and BEA masks of the same subject after intrasubject alignment. First row, raw T1 MRI volumes; second row, McStrip (best performing BEA on all volume and boundary metrics); third row, BET with fixed parameters (worst performing BEA on all volume and boundary metrics). Left to right in each row: Scan #1 (target mask), Scan #2, Scan #3, Scan #4, Scan #5, Scan #6.

included in or excluded from the BEA mask relative to the manual mask. Similarity Index describes the amount of overlap of high-intensity voxels between the BEA mask and the manual mask. This metric was calculated using the Dice Similarity metric defined as: $2\ |M1 \cap M2|\ /\ (|M1| + |M2|)$, where M1 is the BEA mask and M2 is the manual mask (Zijdenbos et al., 1994).

Boundary metrics

The boundary metrics used in this study were introduced in the companion paper. Because large volume errors can often be corrected with minor editing along the boundary, we calculated the overlap of the BEA mask boundary relative to the manual mask boundary. Correct Boundary describes the percentage of BEA mask boundary voxels that correspond to voxels in the

Table 3
Datasets #2 and #3; BEA performance vs. manual masks

Dataset #2

BEA	Misclassified Tissue (%)		Similarity Index		Correct Boundary (%)		Pertinent Boundary (%)		Processing Time
	KR	SS	KR	SS	KR	SS	KR	SS	
SPM2	3.6	3.8	0.980	0.978	61.1	37.0	65.3	41.8	35 min
BET (fixed parameter)	2.7	2.8	0.985	0.983	61.9	40.0	64.4	44.7	40 s
BSE (fixed parameter)	10.1	10.1	0.945	0.942	33.4	15.7	40.6	20.6	1 min
McStrip	**1.7**	2.1	**0.990**	0.988	**85.7**	**62.9**	**86.6**	**68.7**	11 min
oBET	2.3	**2.0**	0.988	**0.989**	68.9	55.2	71.6	62.3	–
oBSE	2.6	4.3	0.985	0.976	45.8	20.1	53.1	24.8	–

Dataset #3

BEA	Misclassified Tissue (%)	Similarity Index	Correct Boundary (%)	Pertinent Boundary (%)	Processing Time
SPM2	6.9	0.965	77.9	78.2	35 min
BET (fixed parameter)	5.9	0.971	54.3	58.5	**40 s**
BSE (fixed parameter)	3.3	**0.984**	**82.9**	**83.1**	1 min
McStrip	4.3	0.978	80.4	80.6	16 min
oBET	4.7	0.976	66.8	69.9	–
oBSE	**3.2**	**0.984**	82.8	83.0	–

The means for 16 scans are reported, and all percentages are based on comparisons of BEA masks to manual masks (see text).
The best performance for each metric is bolded.

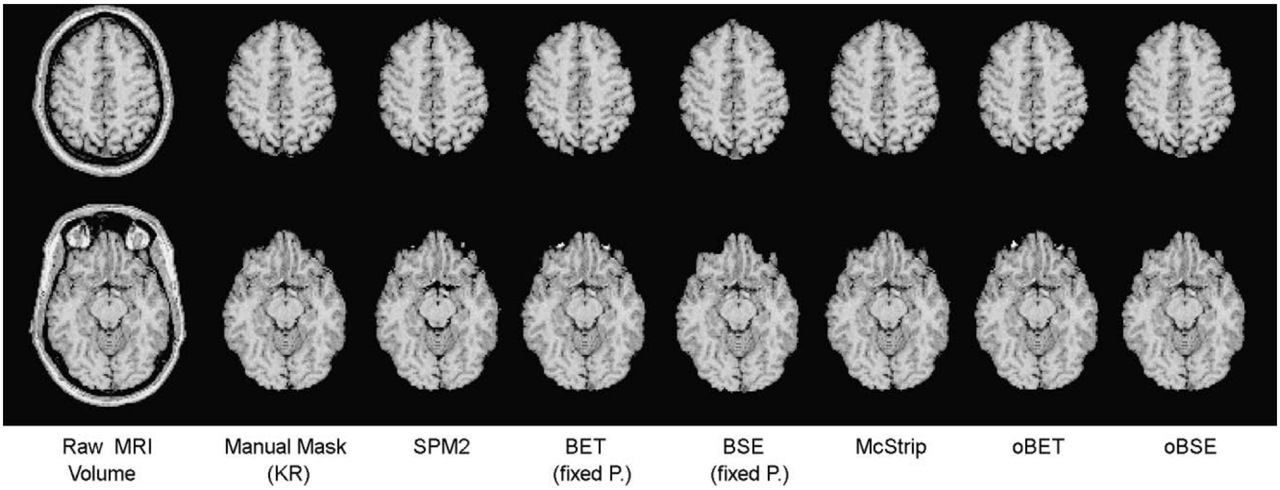

Fig. 4. Dataset #2, Subject B. Masks for axial brain slices above (top row) and below (bottom row) the tentorium created by hand and by four BEAs.

manual-mask boundary—regardless of intensity. As a measure of leniency, the Euclidean distance between voxels on the two boundaries was calculated, and those lying within 1.41 voxels ($\sqrt{2}$) of one another were considered acceptable. Pertinent Boundary is a less rigid metric than Correct Boundary, the rationale being that boundary errors involving suprathreshold voxels were more critical than errors involving low-intensity voxels. This metric disregards errors that occurred below the GM–CSF threshold; Pertinent Boundary captured will always be greater than or equal to Correct Boundary.

Consistency

Informally, we assessed consistency of BEA performance across all datasets by ranking best performances. The "best performance" on a metric is defined as the highest mean percentage for Correct Boundary, Pertinent Boundary, and Similarity Index and the lowest mean percentage for Misclassified Tissue. Rank scores were assigned to BEAs according to their performance on each metric (1 = best performance, 4 = worst performance).

Processing Time

The time necessary for generating the final mask from each group was recorded.

Results

Reproducibility

On volume metrics, McStrip produced the least Misclassified Tissue, closely followed by SPM2. On boundary metrics, McStrip performed best followed by BSE. Overall, the results reported for volume and boundary metrics indicate that all four algorithms provide reproducible results (Table 2 and Fig. 3).

Volume metrics

In Dataset #2, McStrip performed best when compared to KR's manual mask and oBET performed best when compared to SS's manual mask with regard to the Misclassified Tissue and Similarity

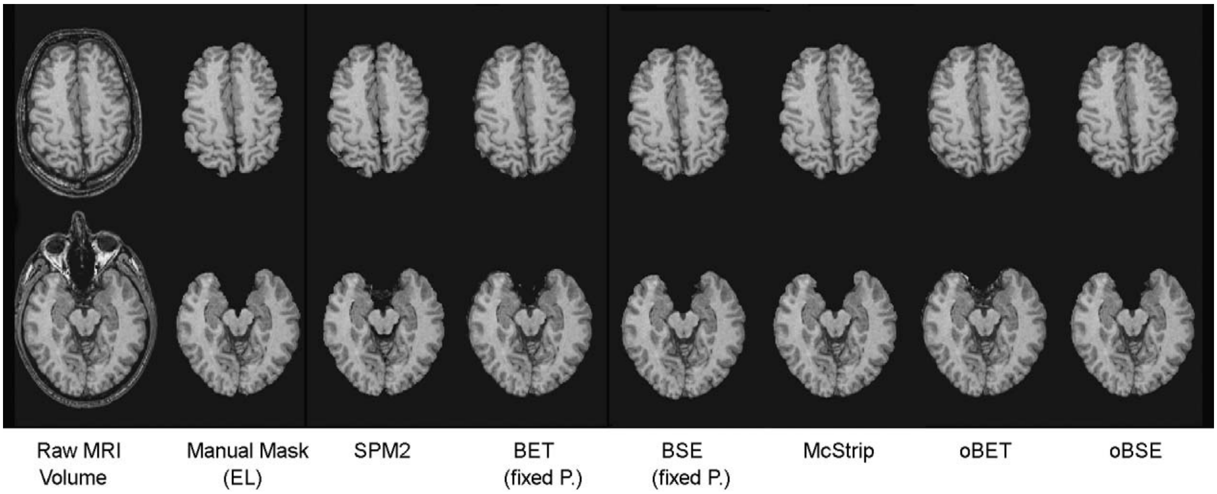

Fig. 5. Dataset #3, Subject B. Masks for axial brain slices above (top row) and below (bottom row) the tentorium created by hand and by four BEAs.

Index metrics (Table 3 and Fig. 4). In Dataset #3, oBSE performed best with regard to the Misclassified Tissue metric, while oBSE and BSE with fixed parameters performed best with regard to the Similarity Index metric (Table 3 and Fig. 5).

Boundary metrics

In Dataset #2, McStrip performed best with regard to the Correct Boundary and Pertinent Boundary metrics (Table 3 and Fig. 4). In Dataset #3, BSE with fixed parameters performed best with regard to the Correct Boundary and Pertinent Boundary metrics (Table 3 and Fig. 5); recall that oBET and oBSE were optimized with respect to Misclassified Tissue.

Consistency

Rank scores for BEA performance across datasets and metrics are summarized in Table 4. Omniscient BET and BSE were excluded from rank score calculation because they were not automated algorithms. McStrip had the best average rank score (5.0), outperforming the other three algorithms (SPM2, 11.8; BET with fixed parameters, 11.8; BSE with fixed parameters, 11.5).

Processing Time

All algorithms were run on a 2-GHz Linux workstation with sufficient memory available (1 GB) such that the use of swap space was not necessary. BET with fixed parameters performed best on this metric, creating masks in less than 1 min (Tables 2 and 3).

Table 4
Performance rankings of BEAs on four metrics

BEA	Dataset #1: Reproducibility	Dataset #2: KR	Dataset #2: SS	Dataset #3: EL	Total score
Metric: Misclassified Tissue					
SPM2	2	3	3	4	12
BET (fixed parameter)	4	2	2	3	11
BSE (fixed parameter)	3	4	4	1	12
McStrip	1	1	1	2	5
Metric: Similarity Index					
SPM2	2	3	3	4	12
BET (fixed parameter)	4	2	2	3	11
BSE (fixed parameter)	3	4	4	1	12
McStrip	1	1	1	2	5
Metric: Correct Boundary					
SPM2	3	3	3	3	12
BET (fixed parameter)	4	2	2	4	12
BSE (fixed parameter)	2	4	4	1	11
McStrip	1	1	1	2	5
Metric: Pertinent Boundary					
SPM2	3	2	3	3	11
BET (fixed parameter)	4	3	2	4	13
BSE (fixed parameter)	2	4	4	1	11
McStrip	1	1	1	2	5

Scores 1–4 refer to rankings of "good performance" as defined by low mean percentage of Misclassified Tissue, high Similarity Index, and high mean percentage of Correct and Pertinent Boundaries. A score of 1 signifies best performance, a score of 4 worst performance.

Performance is reported for BET and BSE with fixed parameters; omniscient users (oBET, oBSE) were excluded from rank score calculation.

Discussion

BEA performance is affected by a variety of machine- and subject-dependent variables such as contrast to noise ratio, geometric distortion, image intensity nonuniformity, susceptibility artifacts, etc. Thus, BEAs may perform differently on datasets acquired on different scanners using the same pulse sequence, on datasets acquired at different sites using the same scanner and pulse sequence, or even on datasets acquired using the same scanner and pulse sequence on different days of the week. In this study, BSE with fixed parameters performed poorly on Dataset #2 yet produced the most accurate brain masks for Dataset #3. One might expect that a hybrid BEA, such as McStrip, that does not rely on a single algorithmic strategy would better compensate for differences in signal structure than a single-strategy algorithm such as BSE, and this proved to be the case; McStrip performed well on all three datasets.

Accuracy

The gold standard datasets against which BEA performance was evaluated were created by anatomically sophisticated investigators (KR, EL, and SS) who manually corrected the masks generated by an early version of McStrip (KR, Dataset #2) or BSE with fixed parameters (EL, Dataset #3), or manually delineated the cortical surface without the aid of a BEA (SS, Dataset #2). As might be expected, McStrip produced the best results on all volume and boundary metrics for Dataset #2 when compared to KR's manual mask ($P < 0.01$); on Dataset #3, oBSE and BSE produced virtually indistinguishable results; BSE (with fixed parameters and with omniscient user) significantly outperformed the other BEAs on all metrics ($P < 0.01$). On Dataset #2 when compared to SS's manual mask, McStrip produced the best results on boundary metrics ($P < 0.01$) while oBET produced the best results on volume metrics ($P < 0.01$; Table 3). Although the accuracy required of a BEA ultimately depends on the end-product of an image-processing chain, for example, intersubject registration, brain tissue segmentation, surface or volume rendering, etc., and no one algorithm may best serve the needs of all users, McStrip permits the sophisticated user to vary the quality of its output to suit his or her needs.

Reproducibility

Although the hybrid method McStrip performed best with respect to the reproducibility metrics (Table 2), the range of means reported on two of the four metrics was quite small, indicating that all four algorithms produced reproducible brain masks; Misclassified Tissue ranged from 1.2% (McStrip) to 4.6% (BET with fixed parameters), and the Similarity Index ranged from 0.994 (McStrip) to 0.976 (BET with fixed parameters). McStrip and BET with fixed parameters were also the best and worst performers, respectively, on the Correct Boundary and Pertinent Boundary metrics, though BET and BSE outperformed McStrip on the Processing Time metric: 13 min for McStrip vs. 40 s for BET and 1 min for BSE.

Consistency

Consistency of performance was assessed by averaging BEA rankings across datasets and performance metrics to produce a Total Score (Table 4). McStrip far outperformed all the other automated BEAs on this aggregate metric, and, as noted above, this performance advantage was likely due to its hybrid algorithm, which offers some protection against a variety of common sources of error, for example, image intensity nonuniformity, low contrast to noise, and geometric distortion.

Automaticity

A major goal of automating the process of brain extraction has been to speed up image analysis by decreasing the amount of time an operator must spend interacting with his or her BEA of choice. As with other image processing applications, a reduction in user interaction time usually entails an increase in algorithmic complexity, which in turn may result in an increase in the number of user-adjustable parameters. Although BEAs with a few user-adjustable parameters, e.g., BET, may reduce user interaction time, they also limit the user's ability to influence the extraction process. Whereas BEAs with five or more parameters, for e.g., McStrip, require more user interaction, they provide for greater flexibility and more control over the extraction process. McStrip has successfully addressed the automaticity–complexity trade-off by incorporating a grid search for the optimal combination of adjustable parameters. See the companion paper by Rehm et al. for a discussion of parameter estimation in McStrip.

Processing speed

Processing speed is dependent on several factors, some of which, such as code optimization and CPU speed, bear little or no relation to the intrinsic properties of the BEA. Nevertheless, our data serve to confirm our expectation that single-strategy algorithms run faster than hybrids (Tables 2 and 3). However, for most image processing purposes, the run-time differences that we observed would seldom, if ever, determine one's choice of a BEA.

Web service

We have created a Brain Extraction Evaluation Web Service (BEE) available at http://www.neurovia.umn.edu. To use this service, an investigator downloads a set of 15 anonymized T1-weighted MRI volumes, strips them by hand or with an in-house BEA, and uploads the resultant strip masks to the service website. The uploaded masks are then compared to our manual masks using the metrics described in the Materials and methods section, and the results of the comparison are automatically e-mailed to the user together with comparable results for McStrip, BSE, BET, and SPM2.

Conclusions

Our comparison of one new (McStrip) and three popular BEAs against expertly hand-stripped T1-weighted MRI brain volumes revealed that McStrip, a hybrid algorithm incorporating intensity thresholding, nonlinear warping, and edge detection modules, consistently outperformed SPM2, BET, and BSE, all of which rely on a single algorithmic strategy. Indeed, the hybrid algorithm's performance was comparable to that achieved by "omniscient users" of the other BEAs. Based on their rankings on four quantitative metrics (Misclassified Tissue, Similarity Index, Correct Boundary, and Pertinent Boundary), those BEAs that employed a single algorithmic strategy performed less satisfactorily than McStrip. As with other image-processing applications, there were trade-offs between algorithmic complexity (i.e., the number of tunable parameters), flexibility, user control, and processing speed. Ultimately, the choice of BEA may depend on the demands of an image processing pipeline or user-defined criteria for accuracy, reproducibility, and consistency over time or across datasets acquired on different scanners.

Acknowledgment

Supported in part by Human Brain Project grant P20 EB02013.

References

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry: the methods. NeuroImage 11, 805–821.

Lee, J.-M., Yoon, U., Nam, S.H., Kim, J.-H., Kim, I.-Y., Kim, S.I., 2003. Evaluation of automated and semi-automated skull-stripping algorithms using similarity index and segmentation error. Comput. Biol. Med. 33, 409–507.

Lemeiux, L., Hagemann, G., Krakow, K., Woermann, F.G., 1999. Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data. Magn. Reson. Med. 42, 127–135.

Rehm, K., Shattuck, D., Leahy, R., Schaper, K., Rottenberg, D.A., 1999. Semi-automated stripping of T1 MRI volumes: I. Consensus of intensity- and edge-based methods. NeuroImage 9, S86.

Rehm, K., Schaper, K., Anderson, J., Woods, R., Stoltzner, S., Rottenberg, D., 2004. Putting our heads together: a consensus approach to brain/non-brain segmentation in T1-weighted MR volumes. NeuroImage (in press).

Richard, A., 2000. Biomedical Imaging, Visualization, and Analysis. Wiley, New York.

Shan, Z.Y., Yue, G.H., Liu, J.Z., 2002. Automated histogram-based brain segmentation in T1-weighted three-dimensional magnetic resonance head images. NeuroImage 17, 1587–1598.

Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K.A., Rottenberg, D.A., Leahy, R.M., 2001. Magnetic resonance image tissue classification using a partial volume model. NeuroImage 13, 856–876.

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

Yoon, U.C., Kim, J.S., Kim, J.S., Kim, I.Y., Kim, S.I., 2001. Adaptable fuzzy c-means for improved classification as a preprocessing procedure of brain parcellation. J. Digit. Imaging 14, 238–240.

Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C., 1994. Morphometric analysis of white matter lesions in MR images. IEEE Trans. Med. Imag. 13, 716–724.