# Statistical Morphological Skull Stripping of Adult and Infant MRI Data

John Chiverton<sup>a,\*</sup>, Kevin Wells<sup>a</sup>, Emma Lewis<sup>a</sup>, Chao Chen<sup>a</sup>, Barbara Podda<sup>b</sup> and Declan Johnson<sup>c</sup>

<sup>a</sup>Centre for Vision, Speech and Signal Processing, School of Electronics and Physical Sciences, University of Surrey, GU2 7XH, U.K.

<sup>b</sup>Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Italy.

<sup>c</sup>St. George's Hospital Medical School, Cranmer Terrace, London, SW17 0RE, UK.

#### Abstract

This paper describes a novel automatic Statistical Morphology Skull-Stripper (SMSS). The result of applying SMSS to 12 MRI test data set volumes, including scans of both adult and infant subjects is also described. Quantitative performance assessment was undertaken with the use of brain masks provided by a brain segmentation expert. The performance is compared with an alternative technique known as Brain Extraction Tool (BET). The results suggest that SMSS is capable of skull-stripping neurological data with small amounts of over and under-segmentation. The results also suggest that both techniques provide similar performance for the adult and infant test data set volumes.

*Key words:* mathematical morphology, image processing, skull-stripping, MRI, neurological, infant

## 1 Introduction

The statistical analysis of neurological Magnetic Resonance Imaging (MRI) data [1,2] is often made easier when voxels corresponding to non-Central Nervous System (CNS) tissue compartments are removed from the MRI data set, as illustrated in figure 1. This figure (1) illustrates how the CNS tissue compartments of White Matter (WM), Gray Matter (GM) and Cerebro-Spinal Fluid (CSF) can be identified in the intensity histogram of the skull-stripped data volume. It also illustrates how the non-CNS tissue compartment voxels contribute towards similar ranges of intensities to the CNS tissue compartment voxels, thereby confounding the identification of individual CNS tissue compartment voxels when using intensity based analysis.

Skull stripping techniques have previously included the use of surface deformation techniques [3–5], the watershed transform [6] or techniques mostly reliant on morphological operations [7–11]. More sophisticated techniques have been developed that combine surface deformation techniques with morphological operations [12–14]; watershed transforms with surface deformation techniques [15]; and techniques based on a consensus of publicly available skull-stripping techniques such as [16] that uses the publicly available code associated with [3,17–19]. References [17,20] take a slightly different approach, both of which use a combination of edge detection and morphological processing, although [17] introduced a denoising step in the form of anisotropic diffusion filtering.

Artificial Intelligence techniques, such as the use of multiple software agents have also been used in [21], where the agents specialize in differing image features including region, edge and intensity features. Lee *et al* in [22] automatically isolate a midsagittal slice from sagittally acquired data sets utilizing landmarks such as the

<sup>\*</sup> Corresponding Author.

Email address: j.chiverton@surrey.ac.uk (John Chiverton).

tip of the nose (a possible limitation in some data sets) and is extended to skull-strip the entire CNS in 3-D in [23]. Reference [24] is unusual as it uses a multiresolution contour tracking algorithm to extract the contours of the human brain in both X-ray Computed Tomography (CT) and MRI neurological data sets.

This work is primarily concerned with presenting a fully automatic methodology utilizing statistical techniques including fitting of probabilistic functions and thresholding, and further image processing intensive operations including region-growing and mathematical morphological operations. A fully automatic approach has been developed because of the need in some studies for unsupervised processing where large numbers of data sets are utilized, for example, in longitudinal studies of brain atrophy [25]. The technique is deemed to be flexible in contrast to other skullstripping techniques dominated by mathematical morphological operations due to the use of a 2-D brain mask and non-dependency on landmarks or the plane of acquisition. The use of a 2-D brain mask for template matching the entire volume is an approach based on a manually assisted segmentation method as described in [26]. The 2-D brain mask enables a variable number of erosions and dilations to be performed (reducing dependence on the relative sizes of the anatomical structures of interest), in contrast to techniques such as [10] that use a fixed number of morphological operations. Similar work in [7] utilized a variable number of erosions and dilations, but these were applied to a thresholded 3-D data volume, where the iterations ceased when a major disconnection event occurred. Our approach, by contrast differs significantly from the approach taken in [7]. SMSS has also been designed so as to include CSF voxels. This enables better subsequent analysis to be undertaken. for example, probabilistic models of the three main CNS component classes to be built for the purposes of Partial Volume (PV) modeling, in techniques such as [1,2].

This newly fully automated technique (SMSS) is then applied to a number of neurological MRI scans that include infant and adult data sets (T1 and T2 weighted).

These results are then compared with the results of an alternative popular skull stripping technique known as BET, as discussed in [3]. BET is a simple but effective and popular deformable surface skull-stripping technique.

As far as the authors are aware, no other work concerning skull-stripping has previously been published that attempts to skull-strip infant neurological data. Infant neurological MRI data is considered to be a more difficult skull-stripping problem in comparison to adult data sets due to the relative difference in size of the corresponding anatomy (acquired at the same resolution to adult data, or worse) and the different statistical properties of the signal in the CNS voxels. As a point of reference, we also include a number of adult neurological MRI data sets and observe the differences in the results between the two methodologies.

[Fig. 1 about here.]

# 2 Methodology

Skull stripping techniques based solely on mathematical morphological operations usually require some user interaction. In contrast, the methodology presented here is primarily based on mathematical morphological operations with additional statistical techniques that fully automate the skull stripping process.

The main contribution of this paper lies in the methodology described below and its demonstrated application to adult as well as infant brain data. In summary the method proceeds as follows: initially, background voxels are removed via automated region growing using statistical termination criteria; parameter estimation is then undertaken to fit an intensity Gaussian mixture model to the histogram of a subset of the data defined by a further fitting operation; from this a target slice is selected which is segmented by automatic thresholding and region growing to create a 2-D target mask. The entire 3-D volume is then subjected to a series of 3-D morphological operations. The process halts when all the brain voxels have been disconnected from the unwanted surrounding non-brain voxels in the target slice using the 2-D target mask (initially described in [26]). A set of further 3-D morphological operations are then undertaken to compensate for the over-erosion of the brain volume. The methodology is described in more detail in the sections below (also summarized in figure 2).

## 2.1 Background Removal

An initial first step in this skull stripping algorithm is the removal of voxels that contain zero or very little Nuclear Magnetic Resonance (NMR) signal. These voxels usually arise either due to the air surrounding a patient or from additional voxels that have been added to the data volume to obtain a convenient set of dimensions for processing with Fourier methods (zero-padding).

Initially, individual slices are considered independently to remove zero-padded voxels, via a simple region-growing operation. This is automatically initiated by selecting seed points corresponding to the four corners of each slice, terminated when no connected zero value voxels remain (using 2-D eight way connectivity).

The second stage is a more complex region-growing operation to remove voxels with little NMR signal present that correspond to air surrounding the patient. These voxel intensities are usually considered to possess a Rayleigh Probability Distribution Function (PDF) and voxels that correspond to non-air regions with a larger NMR signal can be approximated with a Gaussian PDF [27]. Therefore, during the background region growing process, the boundary of non-air voxels can be detected when the statistics of the border region changes. From this, it was decided that a region growing operation that iteratively grew the background region using a process that increased the intensity threshold but halted when the standard deviation of the border region goes through an inflection. This would signal the point at which the surrounding background region "floods" into the foreground object corresponding to the point at which the statistics of the background Rayleigh density are affected by the relative increase in brightness of the Gaussian distributed foreground. Once the termination criterion has been met, then the values of the voxels defined as background, through being included in the region grown, are set to a zero intensity value. This enables subsequent operations to exclude those voxels based solely on their zero intensity value *label*.

## 2.2 Parameter Estimation

This stage attempts to determine approximate values for the parameters of the PDFs of the GM and WM CNS tissue component voxels. Approximate values for these parameters will aid subsequent stages in the skull-stripping algorithm.

Initially, slices in the data volume need to be identified that contain a substantial number of CNS tissue component voxels. These slices are identified by comparing the Root Mean Square (RMS) error between the normalized histogram for the entire data volume following the background removal stage (described previously), and the normalized histogram for individual image slices. The schema relies on the assumption that the CNS tissue component voxels are the most populous tissue component voxels within the data volume. This implies that the RMS error will decrease for image slices with a substantial population of CNS tissue component voxels.

It was initially proposed that an appropriate populous CNS slice could be found to be defined by the region where the first RMS histogram error minima occurs (below the mean error for the entire data volume). In practice, it transpires that further appropriate slices could be found from this minima and up to where the RMS error became greater than the mean error as illustrated in figure (3).

Once these slices have been identified, the number,  $\kappa$ , of major Gaussian components are estimated by analyzing the number of peaks in the aggregate histogram,  $f_{data}(g)$ , of these slices. In this instance,  $\kappa$  was estimated by determining zero crossing points in the second derivative of the aggregate histogram by applying a suitable filter mask with a width corresponding to 20 intensity histogram bins for 8 bit data.

The Nelder-Mead downhill simplex algorithm [28] is then used to fit a  $\kappa$  class finite Gaussian mixture model to the histogram (excluding the background voxels) to estimate the parameters of each major Gaussian component that contributed to the histogram of the entire set of selected 2-D slices. The Nelder-Mead downhill simplex algorithm attempts to minimize the mean square error,  $\mathcal{E}(\mathcal{I})$  defined by:

$$\mathcal{E}(\mathcal{I}) = \sqrt{\sum_{g=1}^{g_{max}} \left( f_{data}(g) - f_{model}(g|\mathcal{I}) \right)^2},\tag{1}$$

where g = 1 to  $g_{max}$  represents the voxel intensity range.  $\mathcal{I}$  is the set of parameters of the  $\kappa$  class Gaussian mixture model:

$$\mathcal{I} = \{\mu_j, \sigma_j, P(i_j) | 0 \le j \le \kappa\},\tag{2}$$

where  $\mu_j$ ,  $\sigma_j$  and  $P(i_j)$  are the mean, standard deviation and *a priori* probability for the  $j^{th}$  class component of the mixture model respectively. The mixture model,  $f_{model}(g|\mathcal{I})$ , is defined by:

$$f_{model}(g|\mathcal{I}) = \sum_{j=1}^{\kappa} \left( \frac{P(i_j)}{\sqrt{2.\pi . \sigma_j^2}} \cdot \exp\left(-\frac{(g-\mu_j)^2}{2.\sigma_j^2}\right) \right).$$
(3)

The initial values for the set of parameters in  $\mathcal{I}$  were calculated from  $f_{data}(g)$ . For

each class component, j, the corresponding peak found in the histogram (e.g. see figure 1) was used to initialize the mean value,  $\mu_j$ . To initialize the standard deviation values,  $\sigma_j$ , the histogram was divided into  $\kappa$  regions of equal intensity width and the values taken by each  $\sigma_j$  was a third of the size of each of these regions. Similarly, the *a priori* probability values,  $P(i_j)$  were initialized by determining the total population of non-background voxels, dividing this value by  $\kappa$  and normalizing those values with respect to the total voxel population.

The algorithm was terminated when a minima was found by the optimization algorithm. A subset of the resulting parameter estimates, corresponding to the GM and WM CNS tissue components, were used in the subsequent slice selection process, as described shortly.

# 2.3 Automated Slice Selection

Following removal of background voxels where no NMR signal is present, a suitable transverse slice (the target slice) has to be selected utilizing results from the aforementioned parameter estimation step. In previous work, [26], this has required manual intervention in the segmentation process, to select the slice, and undertake the required thresholding. However, following the background isolation and parameter estimation steps described above, this process can now be achieved automatically. There are some unique properties associated with a suitable target slice that can be utilized in the automation of this process: (i) the CNS tissue component voxels represent the largest group of all the tissue components in the ideal target slice; (ii) the frequency of occurrence of all tissue component voxels (CNS and non-CNS) are at a local maxima in the region of the target slice.

Using these observations the formulation will now be discussed in detail.

A data set that corresponds to an MRI scan of the human head,  $\Theta$ , composed of a number of transverse slices,  $\theta_a$ , ordered consecutively so that the first slice, a = 1, corresponds to the top most slice (furthest from the feet in the direction of the head) is represented by the following expression:

$$\Theta = \{\theta_a | 1 \le a \le M\},\tag{4}$$

where M is the number of transverse slices in the data set. An image slice,  $\theta_a$ , is a mapping from a 2-D point,  $\boldsymbol{\omega} = (x \ y)^T$ , to a scalar intensity value, i.e.  $\theta_a(\boldsymbol{\omega}) =$ g. A suitable target slice,  $\theta_b$ , (to be used in subsequent morphological processing steps), is then defined as the slice that maximizes the accumulated tissue class,  $i_j$ , probabilities,  $\psi_b$ :

$$\psi_b \ge \psi_a \; \forall a \neq b,\tag{5}$$

where

$$\psi_a = \sum_{\forall \boldsymbol{\omega}} \sum_{\forall j} P(i_j) . p(g = \theta_a(\boldsymbol{\omega}) | i_j).$$
(6)

 $p(g|i_j)$  is the tacitly assumed Gaussian PDF associated with tissue class j, the parameters of which are taken from the aforementioned parameter estimation step,  $\mathcal{I}$  defined by equation (2). Note, that only a limited number of classes are taken from the available selection. For adult T1 data, it was found that the two components with the largest *a priori* probabilities found by the previously defined fitting operation correspond to the GM and WM CNS tissue component classes. This finding proved highly convenient as these are the tissue component classes that must be maximized in the desired target slice, as will be discussed.

## 2.4 Seed Point and Threshold Value Detection

After the 2-D target slice has been identified, a binary 2-D target mask is created representing the CNS tissue classes using an automated process. This target mask is used in subsequent steps in the algorithm to determine whether the CNS tissue component voxels have been successfully disconnected from the extraneous non-CNS tissue component voxels, e.g. at the optic nerve.

To create the mask, the CNS tissue voxels in the target slice were isolated using region growing operations. Intensity based region growing operations usually require user-specified initial seed points, and intensity values that act as thresholds to terminate the growth of the region. This section describes how the seed points and threshold values used in this work can be determined automatically. First, the method used to determine threshold values needed for region growing are described below.

The CNS tissue component voxels within the target slice represent the largest single class of voxels. This observation leads to the following formalized approach for identifying suitable values for automatically initializing threshold and seed point intensities:

The target slice,  $\theta_b$ , containing intensities, g, can be formally described by a normalized histogram:

$$h(g|\theta_b) = \frac{|f(g|\theta_b)| .\nu}{X.Y},\tag{7}$$

where

$$f(g|\theta_b) = \{ \boldsymbol{\omega} | \theta_b(\boldsymbol{\omega}) = g, \forall \boldsymbol{\omega} \},$$
(8)

X and Y are the width and height, in voxels, of the image slice,  $\theta_b$  and  $\nu$  is the bin width (usually  $\nu = 1$ ).

## [Fig. 4 about here.]

The histogram is then used to identify the largest cluster of voxel intensities representing the majority of points in  $\theta_b$ ; these possess a limited range of intensity values, which will correspond to the vast majority of CNS tissue component voxel values. A center-of-cluster function,  $X(g'|\theta_b)$ , is calculated using a moving window over  $h(g|\theta_b)$ :

$$X(g'|\theta_b) = \frac{1}{(\tau_2 - \tau_1)} \sum_{g=g'-\tau_1}^{g'+\tau_2} h(g|\theta_b),$$
(9)

where  $(\tau_2 - \tau_1)$  is the width of the moving window and is calculated from the distance between the two means of the GM and WM CNS tissue component classes selected from  $\mathcal{I}$ . A suitable value for the seed point voxel in  $\theta_b$ , is given when the maximum for (9) is found at  $g' = g_{peak}$ .

The region growing operation that follows, also requires upper,  $\tau_U$  and lower,  $\tau_L$ threshold intensity values to limit region growth.  $\tau_U$  can be determined from  $h(g|\theta_b)$ : a suitable value for  $\tau_U$  occurs at the highest intensity *knee* of the right most CNS tissue component class in the target slice histogram,  $h(g|\theta_b)$ . This is illustrated in figure (4) and can be defined when the following becomes true:

$$h(\tau_U = g|\theta_b) \le \alpha X(g' = g_{peak}|\theta_b) \text{ for } g \ge g_{peak}, \tag{10}$$

where  $\alpha$  defines the location of the *knee* and whose value can be obtained empirically. In practice  $\alpha = 0.05$  was found to give reliable results.

The lower threshold,  $\tau_L$  can then be assumed to be symmetric about  $g_{peak}$ , therefore,  $\tau_L = g_{peak} - \tau_U$ , which will enable the region growing algorithm to include the vast majority of the population of CNS tissue component voxels in the target slice,  $\theta_b$ .

Having determined appropriate threshold values, a region growing process is automatically initiated using the following formalism: a 2-D target mask,  $I_{TM}$ , (a set of 3-D points that correspond to a single image slice, b), is defined by:

$$I_{TM} = \text{Connected}_{2D}(C, T), \tag{11}$$

where  $\text{Connected}_{2D}(\Lambda_1, \Lambda_2)$  is a function that takes a set of 3-D seed points in set  $\Lambda_1$  and using 2-D eight-neighborhood connectivity, determines the corresponding connected points in set  $\Lambda_2$ . T is a set of 3-D points and can be considered as a

binary representation of the target slice:

$$T = \{ (x \ y \ b)^T | \tau_L \le \theta_b(\boldsymbol{\omega}) \le \tau_U, \forall \boldsymbol{\omega} = (x \ y)^T \},$$
(12)

and C is a set of seed points, defined by points taking a gray level value equal to  $g_{peak}$ , (as illustrated in figure 4 and described by equation 10), and located surrounding the centroid of the points in T, (assuming the CNS is centrally located in the image data).

### 2.5 Morphological Operations

After the target mask has been defined, a set of morphological operations are executed to disconnect the CNS tissue component voxels from non-CNS tissue component voxels. These connections arise due to the various anatomical parts of the head such as the optic nerve and the meninges. This configuration of morphological operations is similar to the configurations initially suggested in [26], but with some modifications, primarily to include the majority of the CSF tissue component voxels.

The reader is reminded to refer to figure 2 to obtain a detailed overview of the steps in this latter stage. Symbols used in the various equations that follow can be cross-referenced with the symbols in figure 2. Initially, the entire 3-D data set,  $\Theta$ , is transformed to a binary representation using the previously determined intensity threshold values,  $\tau_L$  and  $\tau_U$ , to produce a binary mask volume  $L_0$ :

$$L_0 = \{ (x \ y \ a)^T | \tau_L \le \theta_a(\boldsymbol{\omega}) \le \tau_U, \forall \ a, \forall \boldsymbol{\omega} = (x \ y)^T \}.$$
(13)

A number of iterations are then performed to disconnect the set of voxels in  $L_0$ , that correspond to CNS tissue compartment voxels from non-CNS voxels. This process starts with a series of erosion operations to remove small links between major compartments. First, a 3-D mathematical morphology erosion operation is applied to the binary mask volume, commencing at iteration r = 1:

$$L_1^{[r]} = L_2^{[r-1]} \ominus B, (14)$$

where initially,

$$L_2^{[r=0]} = L_0, (15)$$

and B is a cubic morphological structuring element of fixed size,  $3 \times 3 \times 3$  voxels. The next stage in the current iteration utilizes a 26-way connectivity analysis to determine the set of points that are still connected to the points in the target mask. Essentially, all the points in  $L_1^{[r]}$ , following the erosion, that are connected to any of the points that are jointly in the target mask are kept, that is:

$$L_2^{[r]} = \text{Connected}_{3D}(I_{TM}, L_1^{[r]}), \qquad (16)$$

where  $\text{Connected}_{3D}(-,-)$  is a similar connectivity function to  $\text{Connected}_{2D}(-,-)$  as used in equation 11, but using 3-D 26-way connectivity instead of 2-D 8-way connectivity.

A test is now performed to determine whether the current iteration should terminate. If the membership of the current  $L_2^{[r]}$  results in a set of points for the target slice,  $I_{TS}$ , (corresponding to image slice b) that are a subset of the set of points in the target mask,  $I_{TM}$ , then it is deemed that the algorithm has disconnected the entire set of CNS tissue component voxels from the non-CNS tissue component voxels. This termination criterion can be formalized as when:

$$(I_{TS} \cap L_2^{[r=w]}) \subseteq I_{TM},\tag{17}$$

becomes true, where the set of points in the target slice is given by:

$$I_{TS} = I_b = \{ (x \ y \ b)^T | \forall x, \forall y \}.$$

$$(18)$$

Otherwise r is incremented and (14) is re-applied.

Once this process has completed, a number of dilations (2.w times) of the resultant set of points  $L_2^{[r=w]}$  are performed next, to recover voxels corresponding to CNS tissue component voxels inadvertently lost in the disconnection process including additional CSF component voxels:

$$L_2^{[r+1]} = L_2^{[r]} \oplus B, \ w < r \le 3.w.$$
<sup>(19)</sup>

After the dilations have been performed at r = 3.w, the set of voxels  $L_2^{[r=3.w]}$ , will still contain a significant number of residual voids. E.g. for T1 MRI data, the CSF component voxels (such as the ventricles) will usually possess intensity values outside the threshold range  $(\tau_L, \tau_U)$  and therefore will not have been included in the target mask. Some skull-stripping techniques are not concerned with CSF regions, but the statistical modeling of the entire CNS tissue component volume can benefit from the inclusion of the CSF tissue component voxels, especially when modeling the PV effect [2]. Therefore, these voids are filled using the following steps. A set of 3-D points,  $\Phi_1$ , within the image space and not in  $L_2^{[r=3.w]}$  is defined by the following expression:

$$\Phi_1 = \{ (x \ y \ a)^T | (x \ y \ a)^T \in I_a, \notin L_2^{[r=3.w]}, \forall a \}.$$
(20)

This effectively results in an inversion of the segmented binary data volume. The previously defined background voxels can now be used in a region growing operation to grow around the outside of the CNS tissue component region. This process avoids "holes" such as the ventricles.

Therefore a set of points,  $\Phi_2$ , that are connected to the set of background seed points, S (as discussed in the first section of this paper), within  $\Phi_1$ , have to be determined:

$$\Phi_2 = \text{Connected}(\Phi_1, S). \tag{21}$$

The segmented CNS tissue component voxels,  $\Phi_3$ , can then be defined as the inverse

of  $\Phi_2$ :

$$\Phi_3 = \{ \boldsymbol{\omega} = (x \ y \ a)^T | \boldsymbol{\omega} \in I_a, \notin \Phi_2, \forall a \}.$$
(22)

Finally, a set of erosion operations are then performed. These erosions remove any non-CNS tissue component voxels that might have been included within the segmented data volume during the application of equation (19):

$$\Phi_4^{[r+1]} = \Phi_4^{[r]} \ominus B, \ 3.w < r \le 4.w;$$
(23)

where  $\Phi_4^{[r=3.w]} = \Phi_3 \ominus B$ . The resulting volume,  $\Phi_4^{[r=3.w]}$ , then represents the derived skull-stripped CNS voxels component.

#### **3** Performance Assessment

[Fig. 5 about here.]

The methodology presented here was assessed using a number of MRI test data set volumes with different dimensions and of variable quality. These data sets and their properties are summarized in table 1. Adult data set V1 is from a volunteer at the Centre for Vision, Speech and Signal Processing, University of Surrey, U.K. The infant data set volumes (V4 to V9) are from a single subject and are illustrated in figure 5. Adult data set volume V3 is publicly available with a copyright notice<sup>\*</sup>. Adult data set volumes (V10 to V12) are also publicly available<sup>†</sup>. Adult data set

<sup>\*</sup> V3 copyright notice: "Courtesy of, and ©by, Mark Bentum, bentum@wsrt00.nfra.nl, Netherlands Foundation for Research in Astronomy".

<sup>&</sup>lt;sup>†</sup> Data sets V10 to V12 come from a set of 20 normal MR brain data sets and their manual segmentations were provided by the Center for Morphometric Analysis at Massachusetts General Hospital and are available at http://www.cma.mgh.harvard.edu/ibsr/.

V2 is part of the Chapel Hill Volume Rendering Test Data Sets.

The performance of the automatic skull stripper was compared with the performance of BET [3]. BET can be summarized as a skull-stripping technique primarily based upon a surface deformation algorithm. Initially intensity thresholds for the CNS image data are estimated from the intensity frequency histogram. Secondly, the center-of-gravity for voxels with these intensity threshold values is calculated which is then used to initialize the deformable surface consisting of a spherical triangular tessellation. The tessellated sphere starts from a point that should be within the CNS spatial limits and is then allowed to deform with smoothness constraints until the CNS limits have been determined. This process is sometimes repeated if pre-defined smoothness constraint constants of the surface have not been met.

Quantitative assessment of the skull-stripping performance was undertaken with the use of ground truth masks generated by an expert segmenter using software described in [29]. The performance metrics included the Dice coefficient [30,15,16] as this gives a scalar measure [0, 1] for the performance of a segmentation overall (where 1 is representative of a perfect match). False Positives (FP) and False Negatives (FN) were also calculated as these give an indication of the relative amount of under and over segmentation respectively. These metrics were determined for the BET segmentations and the SMSS segmentations in relation to the ground truth. Default parameter settings were used for BET.

# 4 Results and Discussion

The quantitative results of applying the skull stripping methodology to the test data sets summarized in table 1 can be seen in table 2 and summarized in table 3 using the arithmetic mean of each of the performance metrics. Figures 7 and 9 show the corresponding image slices from each of the resulting skull stripped data volumes utilizing SMSS, while figures 8 and 10 show the results obtained for BET [3].

## [Table 2 about here.]

Through visual comparison of the segmentation results with the quantitative results, one could state that the quantitative results reflect well the variable ability of the skull-stripping techniques to segment the CNS from the non-CNS tissue voxels.

### [Table 3 about here.]

If a segmentation result can be classed as good quality through an arbitrarily high threshold on the Dice coefficient, such as Dice> 0.90, then SMSS has good performance for 75% of the test data sets. This can be compared with BET that has good performance for only 25% of the data volumes. The numerical results obtained for BET are in general agreement with previous publications, e.g. [31].

BET consistently under-segments the CNS from the non-CNS voxels (mean FP=0.24, mean FN=0.00), i.e. non-CNS voxels are not removed from the segmented data volume. SMSS provides segmentation results that tend to also under-segment (mean FP=0.11), whilst also over-segmenting a small amount (mean FN=0.04). BET provides a parameter that can be adjusted to control the quality of the segmentation, but it was found that this parameter increased over-segmentation whilst only reducing the under-segmentation by a small amount.

As can be seen from figures 7 and 9, SMSS has produced some results that appear to be of very good quality, in particular with reference to test data sets: V1, V3, V4, (see figure 7); V10 and V12, (see figure 9); especially when compared with the results obtained for BET (see figures 8 and 10). Moderately good results with some over segmentation on a limited number of image slices can be seen for data set volumes: V5, V6, V7, V9 and V11; in contrast to BET producing under segmented results for these particular data set volumes.

Skull stripping of infant data presents its own unique set of challenges because of the relatively poor voxel dimensions of the structure under study. Another important consideration with infant data is the dramatic changes in the WM and GM CNS tissue composition that occur in the first 2 years of life. For this reason, the window  $(\tau_2 - \tau_1)$  utilized by equation (9) defined by the means of the GM and WM CNS tissue components could not be determined. This range of intensities, was then modeled as a single component Gaussian during the fitting process and as a result of this, the window width was derived from the standard deviation of this fitted Gaussian. The effect of this on the level of automation of the skull stripping algorithm is minimal, as programmatically, the age of the patient is available as part of the medical imaging file format data structure. The mean performance quantities for adults and infants in table 3, are presented independently, illustrating that the mean Dice coefficient for SMSS is approximately equal for both adults and infants, similarly for BET, suggesting that both BET and SMSS are as capable of overcoming the challenges of skull-stripping infant neurological data in relation to adult neurological data.

As can be seen from the results presented in figures 7 and 9, SMSS has successfully removed the skull, subcutaneous fat and other extraneous non-CNS tissue component voxels from all of the test data sets. However, the algorithm has over segmented the CNS tissue component voxels on a number of the test data sets, particularly test data sets V8 and V11, while a number of the other data sets have had CNS tissue component voxels removed in a limited number of image slices. The under and over segmentation of the CNS tissue component voxels can be attributed to several factors. The segmentation results of test data sets V2, V8 and V11 are now discussed.

For test data set V2, it was found that the upper threshold value,  $\tau_U$  was initially overestimated, (see results in brackets in tables 2 and 3). This was mainly due to the subcutaneous fat component voxels possessing a higher than usually found range of intensities and higher relative frequency. This prevented the *knee* of the WM tissue component from being determined accurately and therefore caused the  $\tau_U$ upper threshold value to be over-estimated. To address this, a considerably improved result was achieved by using a modified value of  $\alpha = 0.2$ , (from  $\alpha = 0.05$ ) in equation (10). The original segmentation result can be seen in figure 6. Although this affects the level of automation of the algorithm, this could possibly be overcome programmatically by determining the imaging sequence parameters of the MRI scan being skull-stripped and adjusting  $\alpha$  accordingly.

The over segmentation of test data set V8 on the sagittal image slice is due to a large population of CSF component voxels that have not been included in the segmented result. For many applications it is not necessary to include all of the CSF voxels within the segmented result.

One of the assumptions used in these experiments is that the effect of Radio Frequency (RF) field inhomogeneity on the image data is negligible. However, it is apparent that test data set V11 has been affected by the RF field inhomogeneity, affecting the stationarity of the GM and WM intensity PDFs. As a result of this non-stationarity, the skull-stripping algorithm has over segmented test data set V11. Therefore an RF field inhomogeneity correction (e.g. [32]) should be applied to the image data prior to the application of the skull stripping process.

# 5 Conclusion

The work in this paper has demonstrated a newly developed fully automatic skull stripping methodology, SMSS. The results obtained using SMSS have shown that the technique could be used as part of a wider neurological tissue analysis framework, possibly as part of a consensus of skull-strippers, particularly as BET appears to consistently under-segment whereas SMSS under-segments less and over-segments more. This is confirmed by the quantitative analysis in relation to the expert segmented ground truth obtained for each of the test data sets. The application of the two techniques to the problem of skull-stripping of infant data volumes has suggested that both the new methodology and BET are adaptable to various relative sizes in anatomy found in CNS of the infant in contrast to the usual problem of skull-stripping of adult MRI T1 weighted data sets.

#### Acknowledgment

JPC would like to thank the UK EPSRC for the doctoral training grant number GR/P00956/01. All MRI image slices through out this work were produced using MRIcro, as described in [33].

# References

- J. Chiverton, K. Wells, Estimation of partial volume mixtures with a confidence measure combining intensity and gradient magnitude information, in: MIUA Conf. Rec., British Machine Vision Association, Bristol, UK, 2005.
- [2] J. Chiverton, K. Wells, Volumetric partial volume quantification via a statistical model of 3-D voxel gradient magnitude, in: IEEE Nuc. Sci. Sym. Med. Imag. Conf. Rec., IEEE, Rome, Italy, 2004.
- [3] S. Smith, Fast robust automated brain extraction, Hum. Brain Mapp. 17 (2002) 143– 155.

- [4] D. MacDonald, N. Kabani, D. Avis, A. Evans, Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI, Neuroimage 12 (2000) 340–356.
- [5] M. Atkins, B. Mackiewich, Fully automatic segmentation of the brain in MRI, IEEE Trans. Med. Imaging 17 (1998) 98–107.
- [6] H. Hahn, H.-O. Peitgen, The skull stripping problem in MRI solved by a single 3D watershed transform, in: MICCAI 2000, Springer, 2000, pp. 134–143.
- [7] R. Stokking, K. Vincken, M. Viergever, Automatic morphology-based brain segmentation (MBRASE) from MRI-T1 data, Neuroimage 12 (2000) 726–738.
- [8] L. Lemieux, G. Hagemann, K. Krakow, F. Woermann, Fast, accurate, and reproducible automatic segmentation of the brain in T1-weighted volume MRI data, Magn. Reson. Med. 42 (1999) 127–135.
- [9] Z. Shan, G. Yue, J. Liu, Auomated histogram-based brain segmentation in T1weighted three-dimensional magnetic resonance head images, Neuroimage 17 (2002) 1587–1598.
- [10] N. Miura, A. Taneda, K. Shida, R. Kawashima, Y. Kawazoe, H. Fukuda, T. Shimizu, Automatic brain tissue extraction method using erosion-dilation treatment (BREED) from three-dimensional magnetic resonance imaging T1-weighted data, J. Comput. Assist. Tomogr. 26 (2002) 927–932.
- [11] M. Brummer, R. Mersereau, R. Eisner, R. Lewine, Automatic detection of brain contours in MRI data sets, IEEE Trans. Med. Imaging 12 (1993) 153–166.
- [12] T. Kapur, W. Grimson, W. W. III, R. Kikinis, Segmentation of brain tissue from magnetic resonance images, Med. Image Anal. 1 (1996) 109–127.
- [13] K. Rehm, K. Schaper, J. Anderson, R. Woods, S. Stoltzner, D. Rottenberg, Putting our heads together: a consensus approach to brain/non-brain segmentation in T1weighted mr volumes, Neuroimage 22 (2004) 1264–1270.
- [14] S. Sandor, R. Leahy, Surface-based labeling of cortical anatomy using a deformable atlas, IEEE Trans. Med. Imaging 16 (1997) 41–54.

- [15] F. Segonne, A. Dale, E. Busa, M. Glessner, D. Salat, H. Hahn, B. Fischl, A hybrid approach to the skull stripping problem in MRI, Neuroimage 22 (2004) 1060–1075.
- [16] D. Rex, D. Shattuck, R. Woods, K. Narr, E. Luders, K. Rehm, S. Stolzner, D. Rottenberg, A. Toga, A meta-algorithm for brain extraction in MRI, Neuroimage 23 (2004) 625–637.
- [17] D. Shattuck, S. Sandor-Leahy, K. Schaper, D. Rottenberg, R. Leahy, Magnetic resonance image tissue classification using a partial volume model, Neuroimage 13 (2004) 856–876.
- [18] A. Dale, B. Fischl, M. Sereno, Cortical surface-based analysis, Neuroimage 9 (1999) 179–194.
- [19] R. Cox, Afni: Software for analysis and visualization of functional magnetic resonance neuroimages, Comput. Biomed. Res. 29 (1996) 162–173.
- [20] M. Bomans, K.-H. Hohne, U. Tiede, M. Riemer, 3-D segmentation of mr images of the head for 3-D display, IEEE Trans. Med. Imaging 9 (1990) 177–183.
- [21] L. Germond, M. Dojat, C. Taylor, C. Garbay, A coorperative framework for segmentation of MRI brain scans, Artif. Intell. Med. 20 (2000) 77–93.
- [22] C. Lee, S. Huh, T. Ketter, M. Unser, Unsupervised connectivity-based thresholding segmentation of midsagittal brain mr images, Comput. Biol. Med. 28 (1998) 309–338.
- [23] S. Huh, T. Ketter, K. H. Sohn, C. Lee, Automated cerebrum segmentation from threedimensional sagittal brain mr images, Comput. Biol. Med. 32 (2002) 311–328.
- [24] H. Soltanian-Zadeh, J. Windham, A multiresolution approach for contour extraction from brain images, Med. Phys. 24 (1997) 1844–1853.
- [25] J. Barnes, R. Scahill, R. Boyes, C. Frost, E. Lewis, C. Rossor, M. Rossor, N. Fox, Differentiating AD from aging using semiautomated measurement of hippocampal atrophy rates, Neuroimage 23 (2004) 574–581.
- [26] R. Robb, Biomedical Imaging, Visualization and Analysis, Wiley-Liss, USA, 2000.

- [27] H. Gudbjartsson, S. Patz, The rician distribution of noisy MRI data, Magn. Reson. Med. 34 (1995) 910–914.
- [28] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes in C, Cambridge University Press, Cambridge, UK, 1992.
- [29] P. Freeborough, N. Fox, R. Kitney, Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans., Comput. Methods Programs Biomed. 53 (1997) 15–25.
- [30] L. Dice, Measures of the amount of ecologic association between species, Ecology 26 (1945) 297–302.
- [31] J.-M. Lee, U. Yoon, S. H. Ham, J. Kim, I.-Y. Kim, S. Kim, Evaluation of automated and semi-automated skull-stripping algorithms using similarity index and segmentation error, Comput. Biol. Med. 33 (2003) 495–507.
- [32] E. Lewis, N. Fox, Correction of differential intensity inhomogeneity in longitudinal MR images, Neuroimage 23 (2004) 75–83.
- [33] C. Rorden, M. Brett, Stereotaxic display of brain lesions, Behav. Neurol. 12 (2000) 191–200.

[Fig. 7 about here.][Fig. 8 about here.][Fig. 9 about here.]

[Fig. 10 about here.]



Fig. 1. Illustration of the benefits of preprocessing neurological MRI data. The histograms correspond to an un-processed T1 MRI head scan  $(\Box)$ , and a skull stripped T1 MRI head scan  $(\times)$ . The histogram corresponding to the skull-stripped data illustrates the benefit of removing the irrelevant information from the data set, where the peaks corresponding to GM, WM and CSF can be seen more easily. This is due to the removal of unwanted non-CNS tissue component voxels in an MRI scan of a human head that share a similar intensity range as CNS tissue compartment voxels.



Fig. 2. Diagram of the steps involved in the skull-stripping algorithm. The symbols in the diagram can be cross-referenced with the symbols used in the text.



Fig. 3. Illustration of the process of determining image slices that contain a substantial population of CNS tissue component voxels. The mean RMS histogram error is calculated for the entire data volume after the background removal stage, while the solid black line corresponds to the histogram errors for individual image slices for the same volume.



Fig. 4. Illustration of the determination of the upper threshold intensity value,  $\tau_U$ . The histogram,  $h(g|\theta_b)$  (plane line) and the center-of-cluster function,  $X(g'|\theta_b)$ , (dotted line) are illustrated.  $\tau_U$  can be determined close to the *knee* of the right most peak of  $h(g|\theta_b)$ , occurring at approximately  $\alpha$  of the peak value of  $X(g'|\theta_b)$ .



Fig. 5. Transverse, Sagittal and Coronal exemplar un-processed image slices taken from approximate mid-data set points in the infant data set volumes (V4-V9). This illustrates the quality of the infant data sets and the limited available anatomical image information due to the poor resolution in the various different non-acquisition planes.



Fig. 6. Segmentation results obtained for test data set volume V2, initially for  $\alpha = 0.05$ , then corrected by adjusting  $\alpha = 0.20$  in equation (10) to take account of the offset caused by voxels containing subcutaneous fat, where the NMR signal from the fat had not been suppressed through the image acquisition parameters.



Fig. 7. Middle transverse, middle sagittal and middle coronal image slices taken from the results of the automatic skull stripper presented in this paper for data sets V1 to V6. Each row represents a different data set, starting from V1 for row one and finishing with V6 for the final row (see table 1 for further details).



Fig. 8. Middle transverse, middle sagittal and middle coronal image slices taken from the results of the BET skull stripper [3], for data sets V1 to V6. Each row represents a different data set, starting from V1 for row one and finishing with V6 for the final row (see table 1 for further details).



Fig. 9. Middle transverse, middle sagittal and middle coronal image slices taken from the results of the automatic skull stripper presented in this paper for data sets V7 to V12. Each row represents a different data set, starting from V7 for row one and finishing with V12 for the final row (see table 1 for further details).



Fig. 10. Middle transverse, middle sagittal and middle coronal image slices taken from the results of the BET skull stripper [3], for data sets V7 to V12. Each row represents a different data set, starting from V7 for row one and finishing with V12 for the final row (see table 1 for further details).

Table 1 Summary of MRI test data sets used to assess the performance of the skull-stripping algorithm.

ID	Description	No of Voxels	Voxel Dimensions $\rm mm^3$
V1	Adult,T1	$256{\times}107{\times}256$	$1.00 \times 1.50 \times 1.00$
V2	Adult,T1	$256{\times}109{\times}256$	$1.00 \times 1.50 \times 1.00$
V3	Adult,T1	$256{\times}256{\times}256$	$1.00 \times 1.00 \times 1.00$
V4	Infant,T2	$256{\times}20{\times}256$	$0.86 \times 6.00 \times 0.86$
V5	Infant,T1	$256{\times}20{\times}256$	$0.86 \times 6.00 \times 0.86$
V6	Infant,T2	$256{\times}20{\times}256$	$0.86 \times 6.80 \times 0.86$
V7	Infant,T1	$256{\times}20{\times}256$	$0.86 \times 6.80 \times 0.86$
V8	Infant,T2	$256{\times}256{\times}21$	$0.86 \times 0.86 \times 6.00$
V9	Infant,T1	$256{\times}256{\times}21$	$0.86 \times 0.86 \times 6.00$
V10	Adult,T1	$256{\times}63{\times}256$	$1.00 \times 3.00 \times 1.00$
V11	Adult,T1	$256{\times}63{\times}256$	$1.00 \times 3.00 \times 1.00$
V12	Adult,T1	$256{\times}63{\times}256$	$1.00 \times 3.00 \times 1.00$

Table 2

Quantitative skull stripping results. Bracket values indicate alternative segmentation result for an alternative value of  $\alpha$  (see text).

ID	SMSS Dice	BET Dice	SMSS FP	SMSS FN	BET FP	BET FN
V1	0.95	0.76	0.10	0.00	0.38	0.01
V2	0.89(0.64)	0.86	0.19(0.53)	0.00(0.00)	0.24	0.00
V3	0.90	0.86	0.18	0.00	0.25	0.00
V4	0.94	0.91	0.10	0.01	0.16	0.00
V5	0.92	0.90	0.09	0.07	0.18	0.00
V6	0.93	0.88	0.11	0.03	0.22	0.00
V7	0.90	0.86	0.11	0.06	0.25	0.00
V8	0.92	0.93	0.14	0.01	0.14	0.00
V9	0.85	0.88	0.07	0.18	0.21	0.00
V10	0.94	0.80	0.11	0.01	0.33	0.00
V11	0.90	0.87	0.03	0.15	0.24	0.00
V12	0.94	0.84	0.10	0.01	0.28	0.00

Table 3

Arithmetic mean skull stripping results. Bracket values indicate effect due to alternative segmentation result for alternative  $\alpha$  value (see text).

Key	SMSS Dice	BET Dice	SMSS FP	SMSS FN	BET FP	BET FN
Infants	0.91	0.87	0.10	0.07	0.23	0.00
Adults	0.92(0.88)	0.86	0.13(0.18)	0.02(0.02)	0.25	0.00
Overall	0.91(0.89)	0.86	0.11(0.14)	0.04(0.04)	0.24	0.00