

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

# Medical Image Analysis

journal homepage: [www.elsevier.com/locate/media](http://www.elsevier.com/locate/media)

## Detection and measurement of coverage loss in interleaved multi-acquisition brain MRIs due to motion-induced inter-slice misalignment

Roger C. Tam<sup>a,b,\*</sup>, Andrew Riddehough<sup>b</sup>, David K.B. Li<sup>a,b</sup><sup>a</sup> Department of Radiology, University of British Columbia, 211-2386 East Mall, Vancouver, British Columbia, Canada V6T 1Z3<sup>b</sup> MS/MRI Research Group, Division of Neurology, Department of Medicine, University of British Columbia, 211-2386 East Mall, Vancouver, British Columbia, Canada V6T 1Z3

### ARTICLE INFO

#### Article history:

Received 17 September 2007  
 Received in revised form 30 July 2008  
 Accepted 12 December 2008  
 Available online 8 January 2009

#### Keywords:

MRI  
 Interleaving  
 Motion artifact  
 Inter-slice misalignment

### ABSTRACT

In MRI scans that are acquired in a slice-by-slice manner, patient motion during scanning can cause adjacent slices to overlap, resulting in duplicate coverage in some areas and missing coverage in others. Scans in which multiple slices are acquired simultaneously and interleaved with other sets of slices are particularly vulnerable because a single movement can result in the misalignment and overlap of many slices. Despite the fact that considerable data losses can occur even with few visible artifacts, this problem has received very little attention from MRI researchers. The primary goals of this paper are: (1) to raise awareness of the problem in the MRI community and (2) to present an efficient multiscale algorithm that accurately quantifies the amount of data loss. Validation of the algorithm's accuracy is performed on 200 scans with simulated patient motion so that the true amount of data loss is known for each scan. The motion parameters are chosen to simulate scans that have significant data loss (mean missing coverage = 14.39% of head volume, SD = 6.61%, range = 2.76–32.98%) but with few visual indications of the problem. The algorithm is shown to be very accurate, yielding estimates that differ from the true values by a mean of only 1.1% point (SD = 0.98 pt, range = 0.00–6.54 pt). The algorithm is also shown to be consistent and robust when tested on a large set of scans from a recent multiple sclerosis clinical trial.

© 2009 Elsevier B.V. All rights reserved.

### 1. Introduction

Most magnetic resonance image (MRI) scans are composed of 2D slices that are stacked together to form a volume. In general, multiple slices are acquired at the same time to increase efficiency. However, the number of slices that can be imaged simultaneously is limited by a number of factors, such as the pulse sequence parameters that are used to control the relative intensities of various tissues in the scans. Because of this limitation, multiple *acquisitions* (other synonymous terms used by scanner manufacturers include *concatenations* and *packages*) are often required in order to obtain the number of slices necessary for full anatomical coverage.

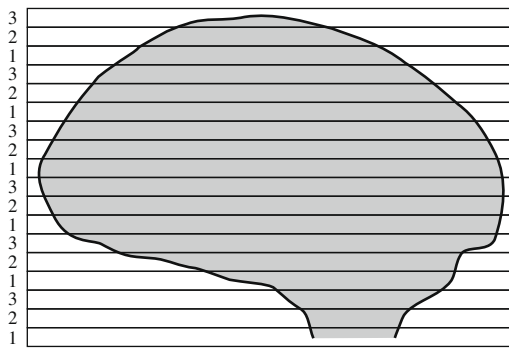
The way in which the slices are grouped into packages can have a strong effect on image quality, because the excitation pulse targeted at a particular slice usually causes some cross excitation and residual energy transfer to the adjacent slice locations. To avoid the artifacts that can result from this “stray energy”, a process called *interleaving* is frequently used (Westbrook et al., 2005). In an interleaved scan, each acquisition only targets slices

that are not adjacent to one another. For example, a two-acquisition interleaved scan can capture the odd slices in the first pass, and the even slices in the second. For scans with three or more acquisitions, the most common way of packaging the slices is  $[s_{1,1}, \dots, s_{q,1}][s_{1,2}, \dots, s_{q,2}] \dots [s_{1,n}, \dots, s_{q,n}]$  where  $s_{ij}$  is the  $j$ th slice of the  $i$ th acquisition,  $n$  is the number of slices per acquisition, and  $q$  is the total number of acquisitions. Fig. 1 illustrates an example with three acquisitions. This is the method of slice ordering that we will assume for the remainder of this article.

One of the problems arising from the use of interleaving is that patient motion during the scan can cause the slices from one acquisition to be misaligned with another. Any component of the movement that is perpendicular to the slice plane can cause two or more acquisitions to overlap, resulting in duplicate coverage in parts of the volume but incomplete coverage in others. For the sake of brevity, in this article we shall refer to this type of error simply as *interleave error*. Sudden non-repeating movements potentially cause more data loss than slow periodic movements because the former can result in a complete lack of signal from some areas, whereas the latter usually yields at least a noisy averaged signal. Fig. 2a illustrates the misalignment caused by a sudden movement between the odd and even slices in a two-acquisition scan. In the most extreme case, a complete overlap between the odd and even slices in a two-acquisition scan would result in the loss of 50% of the data. Large overlaps in coverage are typically visible in the

\* Corresponding author. Address: Department of Radiology, University of British Columbia, 211-2386 East Mall, Vancouver, British Columbia, Canada V6T 1Z3. Tel.: +1 604 827 5381; fax: +1 604 822 7877.

E-mail address: [roger@msmri.medicine.ubc.ca](mailto:roger@msmri.medicine.ubc.ca) (R.C. Tam).



**Fig. 1.** The most common slice ordering for multi-acquisition MRIs. The slices with the same labels are acquired simultaneously. This example shows three interleaved acquisitions.

image data as repeated anatomy in adjacent slices (Fig. 2b), or distorted anatomy when the volume is reformatted to a view parallel to the slice plane (Fig. 2c). However, in more subtle cases, such problems can be very difficult to detect visually. It should be noted that while this article focuses on scans with axial slices, which are the most common type of brain MRI, the same problem applies to sagittal and coronal scans that are acquired in a slice-by-slice manner.

Even though other MRI motion artifacts such as blurring and ghosting have been extensively studied, with a number of correction methods having been proposed (e.g., Atkinson et al., 1999; Lin et al., 2005), we are not aware of any studies focusing on the loss of coverage due to interleave error. A movement causing slice

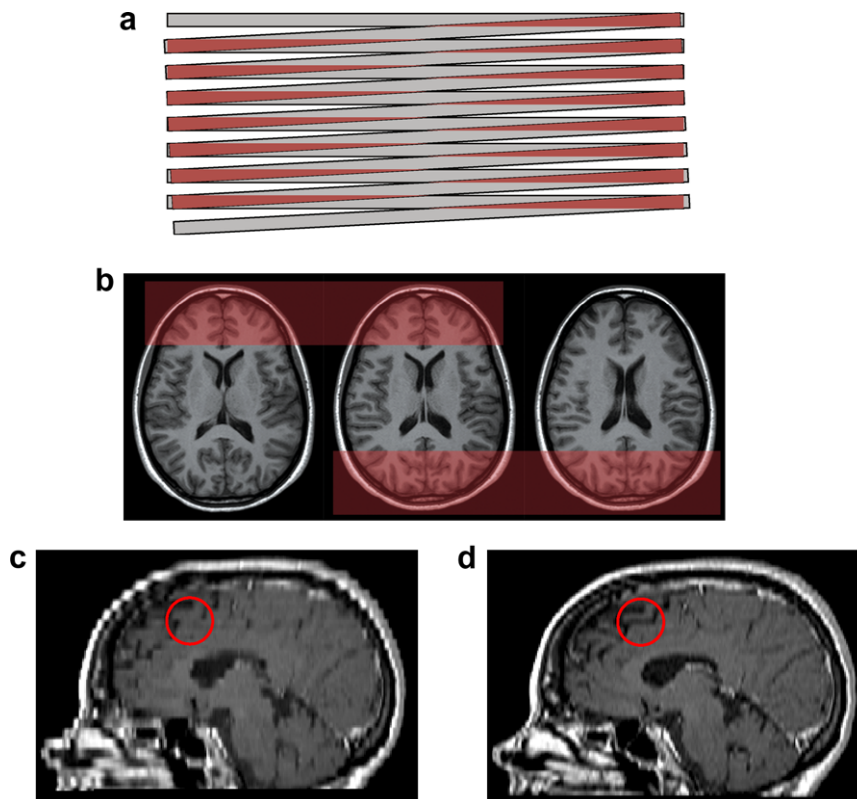
overlap is potentially more detrimental than blurring or ghosting because the loss of data precludes corrective action other than re-acquiring the scan. In addition, blurring and ghosting artifacts are typically visually apparent when their magnitude is sufficient to be problematic. In contrast, a small movement that does not cause strong visible aberrations can still cause significant slice overlap and loss of data.

The impact of interleave errors on the quantitative analysis of MRIs is potentially very significant. For example, the volumetric measurement of white matter lesions in the MRIs of multiple sclerosis (MS) patients is widely used in clinical trials as an indicator of treatment efficacy (Traboulsee et al., 2005), but requires analysis methods that are highly accurate and reproducible (Filippi et al., 1998a). A number of studies have shown that even visually subtle factors such as crosstalk between slices (Filippi et al., 1998b), patient repositioning (Gawne-Cain et al., 1996), and small changes in slice thickness (Firbank et al., 1999) can significantly alter lesion load measurements. With these results in mind, it seems prudent to investigate interleave errors as a potentially important confounding factor in quantitative MRI assessments.

The primary goals of this article are: (1) to increase awareness of the interleave error problem in the MRI community and (2) to present an efficient multiscale algorithm that accurately quantifies the amount of data loss. The algorithm is validated on brain MRIs with synthetically generated inter-slice misalignments, as well as real clinical scans from a large recent MS clinical trial.

## 2. Interleave error quantification

We have developed an algorithm to accurately compute the coverage lost due to inter-slice misalignment. In earlier experi-



**Fig. 2.** Examples of inter-slice misalignment (a) Schematic of a slice stack exhibiting a misalignment between the odd and even slices of a two-acquisition scan. The red areas indicate duplicate coverage. (b) MRI scan showing duplicate coverage in the overlapping areas. The red areas show repeated anatomy in adjacent slices. By comparison, the centres of the slices show relatively normal anatomical progression. (c) Sagittal view of another axially-acquired scan with a large misalignment between the odd and even slices. The boundaries appear ragged, and there are even missing horizontal edges, such as that in the red circle, when compared to a normal scan of the same patient, shown in (d). (a), (b), (c) Misaligned scan, and (d) Normal scan.

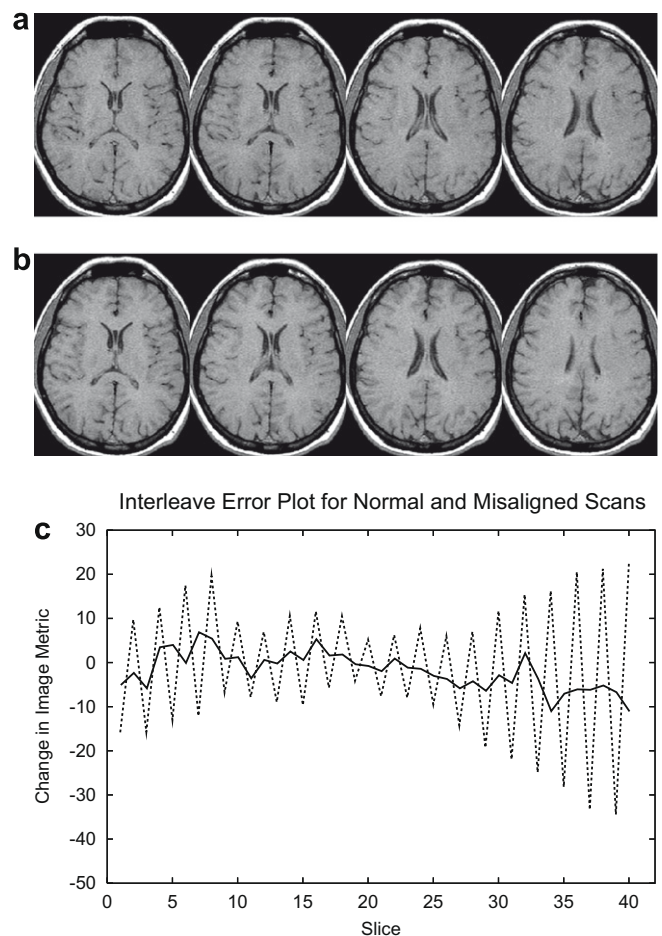
ments, we attempted to determine the missing coverage using rigid registration to a 3D reference scan. The main idea is that positional differences between the acquisitions of the scan in question can be determined by registering each acquisition independently to a high resolution 3D scan, which tends to average motion errors instead of losing coverage because it is acquired volumetrically instead of a slice-by-slice manner. While the method works reasonably well for scans with a low number of acquisitions (2–3), the accuracy of the method decreases considerably with more acquisitions, especially for histogram-based registration methods such as mutual information (Pluim et al., 2003), because there is less data available per acquisition for registration. In addition, if the number of acquisitions is not known beforehand, the computational cost of the method is high because many registrations are required to check the different sets of slices that may be out of place. These limitations, along with the dependence on a reference scan, provided strong motivation to develop a novel technique. The algorithm that we propose in this article works well on scans with a wide range of acquisitions without a priori knowledge of the number of acquisitions present, does not require a reference scan and is very computationally efficient.

The main theoretical basis of our algorithm is that a mis-positioned slice is likely to be unusually similar, at least regionally, to one of its neighbouring slices. In many cases there is also a corresponding amount of increased dissimilarity to the other neighbouring slice. It is the asymmetry in the similarity between a given slice and its neighbours that we use to characterize the degree of interleave error. Our algorithm applies an image similarity metric to all consecutive pairs of slices to measure how much each slice differs from its neighbours. The result is that each slice, except for the end slices, has two associated similarity values. The algorithm then takes the difference between the two values to quantify the asymmetry for each slice; this difference can also be thought of as the *change* in similarity associated with each slice as the slice stack is traversed. As detailed in Section 2.2, our algorithm estimates a baseline that represents normal symmetry, which can be used to detect unusually strong extrema. The algorithm computes the symmetry value for each slice, and identifies bad slices by searching for extrema in the slice stack that are regularly spaced, which is suggestive of multiple slices moving together. The set of extrema with the strongest combination of magnitude and regularity is used to compute an interleave error value that is representative of the overall error in the scan.

Fig. 3 shows a simple example from two real T1-weighted clinical scans. In this case, a repeat scan, shown in Fig. 3b, was performed immediately after an unspecified problem was noticed with the first scan, shown in Fig. 3a. In Fig. 3c, the black line is a plot of the change in the image similarity metric, as defined in Section 2.2, through the slice stack for the second scan, which as evidenced by the small extrema (relative to the overall trend) and the lack of a distinct pattern, is very likely to have little to no interleave error. The dotted line represents the first scan, and shows a strong pattern of alternating peaks and valleys, which is indicative of an overlap in coverage between the odd and even slices in a two-acquisition scan.

### 2.1. Algorithm assumptions

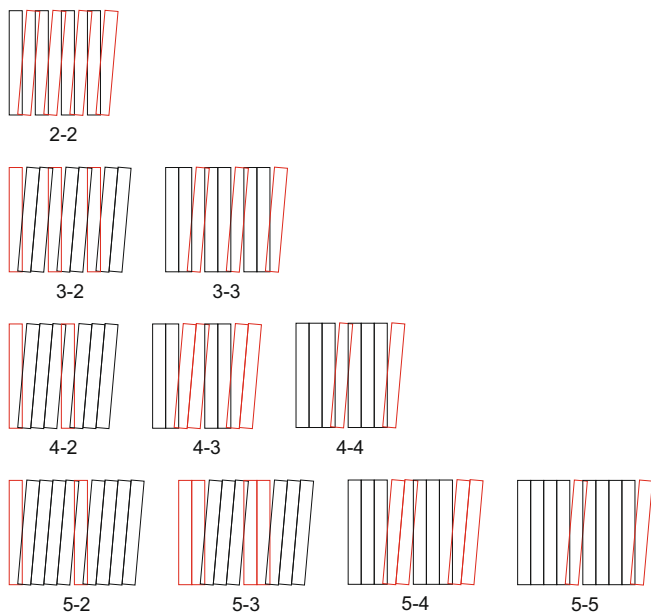
It is easy to imagine a case where multiple drastic movements during a scan can make determining the missing coverage an intractable problem. Without a reference scan for comparison, the scan must have a sufficient number of slices in common alignment in order to act as a *self-reference*. In general, we assume that the quality of the scan is reflective of a clinical environment in which proper procedures are in place to minimize movement, including asking the patient to hold still and, where necessary,



**Fig. 3.** Comparison of two clinical scans. (a) MRI with an inter-slice alignment error. (b) MRI of the same patient, but with no detectable interleave error. The differences between the two scans are visually subtle (the normal scan exhibits more “steady” anatomical progression from slice to slice, most noticeably in the ventricles). (c) Interleave error plots of the two scans, as computed by our algorithm. The black line represents the normal scan, and the dotted line represents the scan with interleave error. A strong pattern of alternating maxima and minima indicates a misalignment between the odd and even slices.

restraining the part of the body being scanned in as comfortable a manner as possible. In technical terms, our assumptions on the movement present can be summed up as follows:

- (1) The movement consists primarily of a single dominant non-repeating motion. We assume this simple case because we need a basic starting point from which to develop our algorithm, and because of anecdotal evidence, gathered by visual inspection of a large set of images, that suggests most clinical brain MRIs with interleave errors can be modelled in this manner. Our experiments with a large MS clinical trial data set, the results of which are summarized in Section 4.2, provide further evidence that an algorithm designed with this assumption can be applied effectively to a large majority of scans in a real clinical environment, even with physically compromised patients. We also assume that the movement occurs over a relatively short period of time, so that any resultant blurring or ghosting is minor enough to be corrected or ignored. Given the above assumptions, Fig. 4 illustrates how the timing of the movement can affect different slices in a scan.
- (2) The majority of the voxels in each misaligned slice do not have a positional error perpendicular to the correct slice



**Fig. 4.** Misalignment patterns caused by a single motion in scans with two to five acquisitions. Each label  $q-p$  indicates the scan has  $q$  total acquisitions, and a movement occurring before or during the  $p$ th acquisition has occurred, thereby affecting all subsequent ( $p + 1$  to  $q$ ) acquisitions. The angled slices represent those acquired after the movement. The red slices indicate those that would be used by our algorithm to detect and measure the amount of data loss (by symmetry, red and black can be swapped in 2–2 and 4–3).

plane greater than the thickness of one slice. The reasoning is that an out-of-plane error of one slice is already a large error, and that any slice with a significant percentage of voxels that are mis-positioned beyond the assumed range should have large areas that are visually apparent as being repeated or out of place relative to an adjacent slice. On the other hand, more subtle errors would be more difficult to see, in which case an automatic method should prove advantageous. In addition, this assumption allows a theoretical maximum error to be computed for each slice, which is used to normalize the actual interleave error found (details in Step 9 of Section 2.2). Although we have used the limit of one slice thickness to develop the algorithm, the method is designed to perform well even when a significant percentage of voxels are beyond the expected error range, as shown by our results in Section 4.1.

## 2.2. Algorithm details

The algorithm is a two-phase approach. The first phase (Steps (1)–(8)) identifies a set of sample slices that are likely to be misaligned. The slices associated with the largest asymmetry extrema that exhibit a regular pattern are selected for further analysis. The second phase (Steps (9)–(10)) quantifies the error for each slice selected, and computes an overall score for the scan. If the number of acquisitions is known, then the algorithm is applied only once; otherwise, it is applied for a range of acquisitions (2–5 is the default), and the strongest interleave error is reported.

For clarity, we first describe the core algorithm, which assumes that any movement detected is a straight stack shift (*i.e.*, purely perpendicular to the slice plane). Section 2.2.1 describes how a recursive multiscale application of the core algorithm is used to detect and measure more complex motion resulting from head rotation.

Assuming the number of acquisitions is  $q$ , the following steps are performed:

- (1) For each slice, a bounding box is extracted that defines a rectangular region occupied primarily by the foreground. Otsu thresholding (Otsu, 1979) and connected component analysis are used to find the largest component of the darkest region, which is typically the background. Only the area inside the bounding box is used for the image similarity computations.
- (2) To avoid working with slices with little content, a start slice  $s_s$  and end slice  $s_e$  are determined depending on the ratio of the slice's bounding box size to that of the largest bounding box in the slice stack. The default threshold is 0.25.
- (3) (Optional) To correct for in-plane motion that does not contribute to slice overlap but may impact the accuracy of the algorithm, inter-slice registration is applied to straighten the slice stack. From experiments with our large data sets, in-plane movement is usually not large enough to cause dramatic changes in the reported interleave error value, even if left uncorrected. However, it is conceivable that scans in which the patient has made large side-to-side movements would benefit from inter-slice registration.
- (4) Using an image similarity metric  $d$ , the differences between all neighbouring slices are computed. Letting  $d_{ij} = d(s_i, s_j)$  denote the measured value between slices  $s_i$  and  $s_j$ , the resulting sequence can be written as

$$d_{i,i+1}, i = s_s, \dots, s_e - 1.$$

The similarity measure that we use is the *mean absolute error* (MAE), which is one of the most commonly used similarity measures in image processing. For two  $m \times n$  images  $I$  and  $J$ , the MAE is defined as

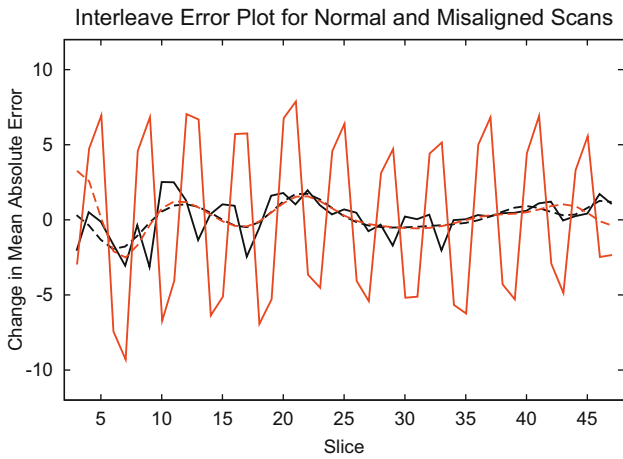
$$\text{MAE} = \frac{\sum_{x=1}^m \sum_{y=1}^n |I(x, y) - J(x, y)|}{m \times n} \quad (1)$$

where  $I(x, y)$  and  $J(x, y)$  are the intensity values of pixel  $(x, y)$  in images  $I$  and  $J$ , respectively.

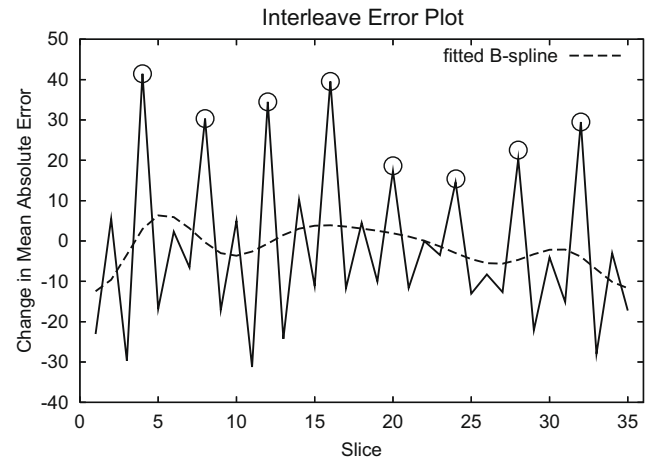
- (5) The symmetry value for each slice is calculated by computing the change in the image metric between the two neighbours:

$$\Delta d_i = d_{i-1,i} - d_{i,i+1}, i = s_s + 1, \dots, s_e - 1.$$

- (6) Treating the  $\Delta d_i$ 's as equally spaced samples, the data is fitted with a cubic B-spline curve using least squares approximation (Schneider and Eberly, 2003). The control points are sparsely placed to capture only the gross fluctuations in the symmetry values, and be robust to the sharp changes associated with interleave errors. Using the simulated scans described in Section 3.1, we have found that, for scans with 3 mm slice thickness, placing the control points three slices apart to model smoothly progressing anatomy gives a good tradeoff between approximation accuracy and spline smoothness. Although we have not yet experimented with other slice thicknesses, we expect that keeping the control points approximately 9–10 mm apart would work well in the general case. The results of an experiment demonstrating the stability of the fitted B-spline are presented in Section 4.1. Fig. 5 illustrates an example of this stability, which is vital for the B-spline's primary function as a baseline by which the significance of each extremum can be measured. The use of the spline also helps to mitigate the effects of measurement noise, and normalizes the interleave error estimates across slices so that they may be combined to yield an overall result. The B-spline curve is sampled at



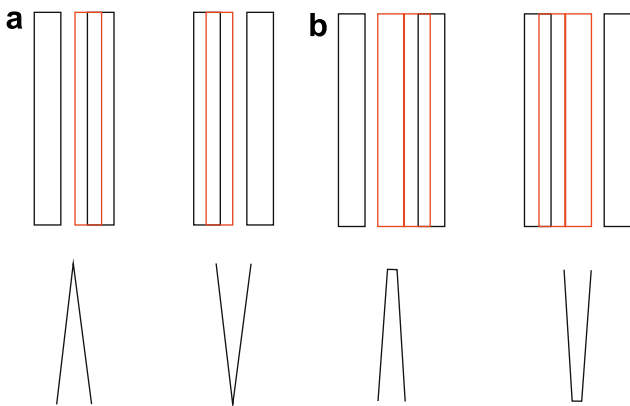
**Fig. 5.** Two interleave error plots with fitted B-splines. The data shown in black is from a normal 3 mm scan, and the data shown in red is from the same scan with a synthetically generated interleave error (a 1 mm slice stack shift in a 4–3 pattern). The closeness of the two splines demonstrates the stability of the spline representation.



**Fig. 7.** Interleave error plot and fitted B-spline of a clinical scan. The spline is used to provide a baseline for measuring the magnitude of the extrema. In this case, the strongest pattern of interleave misalignment, as marked by the circles, is a large peak at every fourth slice.

the same intervals as the  $\Delta d_i$ 's to give a set of values  $b_i, i = s_s + 1, \dots, s_e - 1$ .

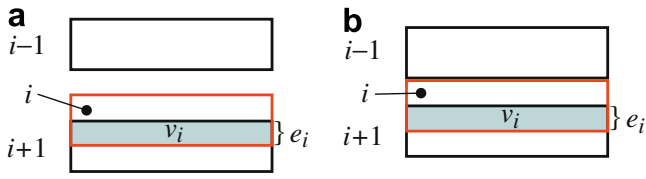
- (7) From the computed symmetry values, all of the local extrema of the two types illustrated in Fig. 6 are identified. The first type (Fig. 6a) consists of a single peak value induced by a slice that is displaced relative to both of its neighbours. This type of error is found in the 2–2, 3–2, 3–3, 4–2, 4–4, 5–2 and 5–5 misalignment patterns illustrated in Fig. 4. To identify maxima of this type, all of the values in the  $\Delta d_i$ 's that satisfy these three conditions are selected: (1)  $\Delta d_i > b_i$ , (2)  $\Delta d_{i-1} < b_{i-1}$  and (3)  $\Delta d_{i+1} < b_{i+1}$ . The minima are found similarly. The second type of local extremum (Fig. 6b) consists of a broader, two-point peak associated with two neighbouring slices that are aligned with each other, but are out of place relative to their other neighbours. This type of error is found in the 4–3, 5–3 and 5–4 misalignment patterns illustrated in Fig. 4. To identify maxima of this type, all of the values in the  $\Delta d_i$ 's that satisfy these three conditions are selected: (1)  $\Delta d_i > b_i$ , (2)  $\Delta d_{i-1} > b_{i-1}$  and (3)  $\Delta d_{i+1} < b_{i+1}$ . These conditions, which are sufficient but not completely specific, pick out the second value in any dual-value peak. The minima are found similarly, except that the first value in the pair is chosen instead. For both maxima and minima, the  $\Delta d_i$  cho-



**Fig. 6.** The two types of extrema used for interleave error quantification. The misalignment of the red slices result in the extrema shown. (a) Single-value extrema. (b) Dual-value extrema.

sen corresponds to the slice that is overlapping in coverage with an adjacent slice. The maxima and minima lists are then ordered from largest magnitude to smallest. The magnitude of each extremum is defined relative to the baseline defined by the B-spline (i.e.,  $|\Delta d_i - b_i|$ ). The greatest values in each list, up to a maximum number equivalent to the size of one acquisition  $((s_e - s_s - 1)/q)$ , are kept for further analysis. The number of slices selected in each list is limited to the size of one acquisition because this allows the sample to span the slice stack while avoiding the inclusion of slices that may obfuscate the strongest pattern. This sample size should be sufficient for all single-movement scenarios. In Section 4.2, we show by our experiments with clinical data that in the large majority of cases, the interleave error of a scan is well represented by the chosen set of sample slices.

- (8) The remaining maxima and minima lists are sorted by slice number. All of the values in each resulting list that are  $q$  slices apart are identified as potentially bad slices. Fig. 7 shows an example of an interleave error plot of a clinical scan, with the most significant peaks identified with  $q = 4$ .
- (9) After the bad slices are identified, the magnitude of the error is computed for each slice. Each slice  $s_i$  in consideration either overlaps in coverage with slice  $s_{i-1}$  or  $s_{i+1}$ ; which one it is can be easily determined by checking whether the associated extremum is a maximum ( $s_{i+1}$ ) or minimum ( $s_{i-1}$ ). The volume  $v_i$  of the overlap, which is equivalent to the amount of missing coverage, is proportional to the positional error  $e_i$  (Fig. 8). According to the most commonly used partial volume model for MRI, each voxel is composed of a linear combination of tissue signals (Choi et al., 1991); thus a change in position should cause a proportional shift in tissue signal in the direction of movement. The result is that the offset  $e_i$  causes a proportional change in  $\Delta d_i$ . We aim to compute an estimate of  $e_i$ , denoted  $\tilde{e}_i$ , that is normalized with respect to the slice thickness to possess a value between 0 and 1. We refer to this value as the *interleave severity*. Having computed  $\Delta d_i$  for each bad slice,  $\tilde{e}_i$  is calculated by using derived baseline and maximum values to normalize  $\Delta d_i$  to the range [0,1]. The baseline is given by the B-spline, as explained above. The maximum  $\Delta d_i^{\max}$  is obtained by computing  $\Delta d$  for a stack shift of one complete slice, which as described earlier is the maximum error assumed. In this case, slice  $s_i$  would overlap completely with either



**Fig. 8.** The goal is to estimate the volume  $v_i$  of the overlap, which corresponds to the amount of missing data, and is proportional to  $e_i$ . (a) and (b) correspond to the two types of extrema shown in Fig. 6.

$s_{i-1}$  or  $s_{i+1}$ . Assuming the overlap is with  $s_{i+1}$  (i.e.,  $d_{i,i+1} = d_{i+1,i+1} = 0$ ), the maximum value for a peak of the first type (Fig. 8a), for which the distance between  $s_{i-1}$  and  $s_{i+1}$  can provide a fixed reference, is computed as  $\Delta d_i^{\max} = d_{i-1,i} - d_{i,i+1} = d_{i-1,i+1} - d_{i+1,i+1} = d_{i-1,i+1}$ . For a peak of the second type (Fig. 8b),  $s_i$  and  $s_{i-1}$  provide the reference distance, and  $\Delta d_i^{\max} = d_{i-1,i} - d_{i,i+1} = d_{i-1,i} - d_{i+1,i+1} = d_{i-1,i}$ . For both types of peaks, an overlap with slice  $s_{i-1}$  entails a similar computation, with the resulting value given a negative sign. Using the maximum value  $\Delta d_i^{\max}$ ,  $\Delta d_i$  is normalized to give  $\tilde{e}_i = (\Delta d_i - b_i) / (\Delta d_i^{\max} - b_i)$ . It is clear that  $\tilde{e}_i$  approaches 1 as  $\Delta d_i$  approaches  $\Delta d_i^{\max}$ . In addition, for the first type of extremum (single peak), because the MAE is a true metric, the triangle inequality ensures that the value never exceeds 1:  $d_{i-1,i} + d_{i,i+1} \geq d_{i-1,i+1} \Rightarrow |d_{i-1,i} - d_{i,i+1}| \leq d_{i-1,i+1} \Rightarrow |\Delta d_i| \leq |\Delta d_i^{\max}|$ . For the second type of extremum (dual-value peak),  $|\Delta d_i^{\max}|$  is always the larger of  $d_{i-1,i}$  and  $d_{i,i+1}$ , so  $|d_{i-1,i} - d_{i,i+1}| \leq |\Delta d_i^{\max}|$ . Since  $\Delta d_i$  and  $\Delta d_i^{\max}$  have the same sign and  $\Delta d_i / \Delta d_i^{\max} \leq 1$ ,  $\tilde{e}_i$  is in the range  $[0, 1]$ . Although we have made the general assumption that the positional error does not exceed one slice thickness, we implement an explicit check for violations to this condition to make the algorithm more robust, because in such cases  $\tilde{e}_i$  as computed above may not give a representative value. Letting slice  $s_i^{\text{overlap}}$  be the neighbour with which  $s_i$  overlaps, and slice  $s_i^{\text{non}}$  be the non-overlapping neighbour, the check is performed by determining if the MAE distance between  $s_i$  and  $s_i^{\text{non}}$  is greater than that between the two neighbours  $s_i^{\text{overlap}}$  and  $s_i^{\text{non}}$ . As evident from Fig. 8, for both types of extrema, if  $d(s_i, s_i^{\text{non}}) > d(s_i^{\text{overlap}}, s_i^{\text{non}})$ ,  $s_i$  has likely moved beyond expected parameters. In this case, the information in the slice is considered to be completely lost, and the error is assigned a value of 1. Thus, the final interleave severity for each slice is computed as

$$\tilde{e}_i = \begin{cases} \frac{\Delta d_i - b_i}{\Delta d_i^{\max} - b_i} & : d(s_i, s_i^{\text{non}}) < d(s_i^{\text{overlap}}, s_i^{\text{non}}) \\ 1 & : \text{otherwise} \end{cases} \quad (2)$$

- (10) The result of the above steps are two sets of  $\tilde{e}_i$ 's, one computed from the  $\Delta d_i$  maxima, and the other from the  $\Delta d_i$  minima. An overall value for each list is then computed by taking the sum of the  $\tilde{e}_i$ 's and dividing by the maximum possible number of sample slices as determined in Step (7):

$$\tilde{e}_{\text{overall}} = \frac{\sum \tilde{e}_i}{(s_e - s_s - 1)/q} \quad (3)$$

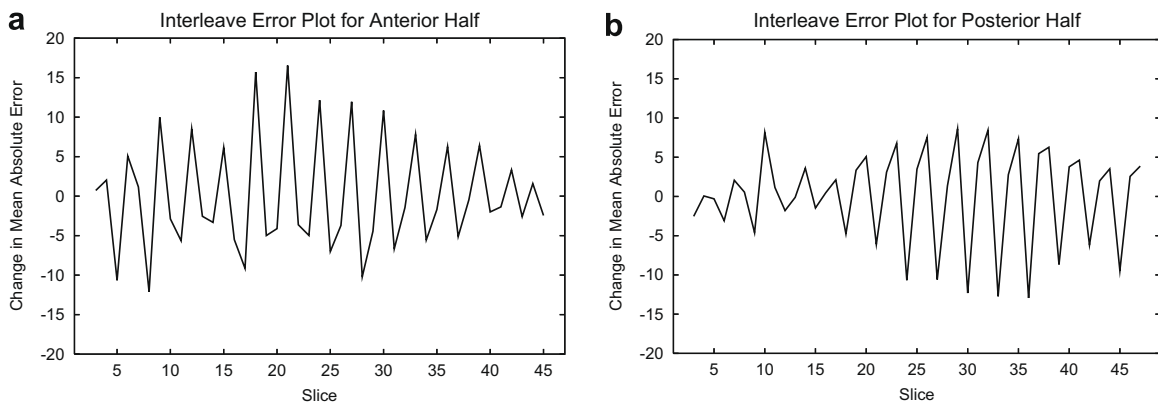
For  $q = 2$ , the two overall values are averaged to give the final interleave severity for the scan, because when there are only two acquisitions present a peak in  $\Delta d_i$  in one acquisition typically corresponds to a valley of similar magnitude in the other. For  $q > 2$ , the greater of the two overall values is used. The final overall score can be interpreted as the amount of data that has been lost, computed from a maximal sampling of slices and normalized to the amount of data in one acquisition. An overall interleave error value of one would indicate that an amount of data equivalent to an entire acquisition is missing from the scan. We can thus think of interleave severity as being in units of acquisitions, or acq in short form. The percentage of data missing from the scan can be computed from the final severity by dividing the value by  $q$ .

If the number of acquisitions present in the scan is not known, Steps (7)–(10) can be performed for each potential  $q$ , and the largest severity value found for all  $q$  tested is reported as the final error. The efficiency of the algorithm is such that a wide range, such as  $q = 2, \dots, 5$ , can be used to do a thorough assessment.

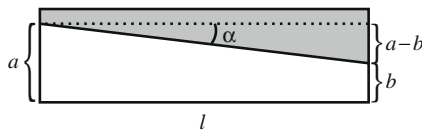
### 2.2.1. Adaptive subdivision

While the core algorithm is effective for measuring errors due to a straight stack shift, more complex motions such as rotation need a more advanced strategy. For example, if a patient drops his chin during an axial scan, the method described above may underestimate interleave error because the anterior and posterior halves of the image move in opposite directions (into versus out of the image plane), and the overall asymmetry score for each slice may not change very much. Fig. 9 illustrates how the interleave error plots for the anterior and posterior halves of such a scan have distinctly different patterns.

To estimate the more complex motions, we apply an adaptive subdivision scheme to the image so that, at the scale of each result-



**Fig. 9.** Interleave error plots for the (a) anterior and (b) posterior halves of a scan in axial orientation. A single chin drop motion during the third acquisition causes the anterior and posterior halves of every third slice to move in opposite directions, causing the single peaks to be largely positive in the anterior plot and negative in the posterior plot.



**Fig. 10.** Angled overlap between slices (the shaded area is the overlap). For any given angle  $\alpha$ ,  $a-b$  is proportional to  $l$ , with  $\tan \alpha$  being the constant of proportionality.

ing region, the misalignment is a close approximation to a straight stack shift. Fig. 10 is a 2D illustration of an angled overlap between slices. The goal is to minimize the difference  $a-b$  by reducing  $l$ . It is clear that the higher the angle  $\alpha$ , the smaller  $l$  has to be for a good approximation.

The subdivision process recursively partitions the image and applies the core algorithm to each subregion. A region is no longer divided when its computed interleave severity is not significantly different from that of its parent region. Starting with the largest foreground bounding box and a given tolerance  $\tau$ , the subdivision proceeds as follows:

- (1) The interleave error  $\tilde{e}$  of the region is computed. Let  $r_x$  and  $r_y$  denote the dimensions of the region.
- (2) The region is divided into four equal subregions of size  $r_x/2 \times r_y/2$ .
- (3) The interleave severity of each subregion is computed independently to give  $\tilde{e}_r, r = 1, \dots, 4$ .
- (4) For each subregion  $r$ , if  $|\tilde{e}_r - \tilde{e}| > \tau$ , the subdivision process is repeated (Steps 2–4). If  $|\tilde{e}_r - \tilde{e}| \leq \tau$  or if the subregion has reached a lower size limit (measured in physical units, taking into account the pixel size),  $\tilde{e}_r$  is assigned to subregion  $r$  as the final interleave error for that area.
- (5) The result of the process is a set of regions, each with its own computed interleave error. The overall severity is taken to be the mean severity from all of the regions, with each value weighted by the relative size of the region.

The number of levels of subdivision is limited by region size because if the regions are allowed to get too small, the measurements performed on them would be confounded by even small anatomical changes between slices, and the accuracy of the overall estimate may degrade. Using simulated misalignment errors applied to a variety of axial head scans with 3 mm slice thickness, we have determined that enforcing a lower limit of approximately 20 mm  $\times$  20 mm generally provides accurate estimates for this type of image.

### 3. Experimental validation

To validate the algorithm, we performed a number of experiments on a variety of axial brain scans. To gauge the accuracy of the method, we applied the algorithm to 200 scans with simulated interleave errors. In addition, we used 400 scans from a recent MS clinical trial to check the consistency and robustness of our method and to determine the validity of our single-movement assumption. The study protocols for both sets of data were approved by the institutional ethics review board.

#### 3.1. Interleave error simulation

The accuracy of our approach can be determined by modifying error-free scans to simulate interleave alignment errors. This is the only way to acquire a data set for which the interleave error of each scan is known accurately, because such errors are extremely diffi-

cult to quantify visually, and expert input cannot be relied upon to provide a gold standard. We started with 20 3D T1-weighted spoiled gradient echo (SPGR) scans acquired on a Philips Achieva 3.0 Tesla scanner (Philips Healthcare, Andover, Massachusetts, United States) with parameters TE/TR = 3.6/7.7 ms, a resolution of 256  $\times$  256  $\times$  170 and isotropic (1 mm<sup>3</sup>) voxel size. As mentioned in Section 2, this type of scan is acquired volumetrically, instead of slice-by-slice, and is therefore free of interleaving artifacts. In addition, this data set was acquired with healthy volunteers, who tend to have fewer problems holding still than patients with serious diseases.

To simulate scans with interleave errors, each volume was resampled 10 times with different parameters to give 10 new volumes. The resampling consisted of reformatting to 3 mm slices with 1 mm<sup>2</sup> pixels, and transforming a subset of the slices with randomly generated parameter values. The new voxel values were computed with cubic B-spline interpolation, which has been experimentally determined to be one of the most accurate methods for interpolating medical images (Meijering et al., 2001).

To simulate a  $q-p$  misalignment pattern, that is, a displacement starting with the  $p$ th acquisition of a scan with  $q$  total acquisitions, every slice  $i$  for which  $i \pmod{q} \geq p$  was transformed with the same parameters. We used parameters that largely conform to our assumptions on the amount of movement present, as described in Section 2.1, but exceed those assumptions enough to test the limits of the method. As specified in Table 1, the transformation has three degrees of freedom (two rotations and one translation). We did not simulate in-plane motion explicitly for this experiment, because such motions do not cause overlapping coverage and can be corrected by inter-slice registration. The coordinate system is such that in an axial scan, the  $x$ -axis is horizontal, the  $y$ -axis is vertical, and the  $z$ -axis is perpendicular to the image plane. The parameter ranges have been chosen to cause a significant interleave error that is not strongly visible in the average case, but nearer the limits would overlap with the ranges that would result in visually salient artifacts. For example, Fig. 11 shows six slices in a scan taken from our simulated data set. The  $x$ -axis rotation,  $y$ -axis rotation and  $z$ -axis translation parameters are 95%, 50% and 76% to their respective limits, resulting in an interleave severity of 0.57 acq. A check for duplicate coverage in adjacent slices makes the problem plainly visible, even without the benefit of an error-free reference scan.

Each voxel  $(x, y, z)$  that is transformed has a new value sampled from its new coordinates  $(x_t, y_t, z_t)$ . The voxel now overlaps in coverage with another slice, and because the pixels are small, the magnitude of this error can be estimated by taking the orthogonal distance between the transformed coordinates and the original slice plane (i.e.,  $d_{\perp}((x, y, z), (x_t, y_t, z_t)) = |z - z_t|$ ). If  $|z - z_t| > 1$ , the value is clamped to 1. The overall severity for a volume with a slice resolution of  $m \times n$  is defined as

$$\frac{\sum_{x=1}^m \sum_{y=1}^n \sum_{z \in T} d_{\perp}((x, y, z), (x_t, y_t, z_t))}{m \times n \times \text{card}(T)} \quad (4)$$

where  $T$  is the set of resampled slices, and  $\text{card}(T)$  is its cardinality. This value represents the overlap between the resampled slices and the other slices, normalized to a maximum of 1 by dividing by the number of voxels transformed. The value 1 represents a positional

**Table 1**  
Parameters for interleave error simulation.

Parameter	Range
Number of acquisitions	2–5
$x$ -axis rotation	–2 to +2°
$y$ -axis rotation	–2 to +2°
$z$ -axis translation	–2 to +2 mm



**Fig. 11.** An example of a simulated two-acquisition scan with visible misalignment error. The odd slices have been transformed with randomly generated parameters within the ranges specified in Table 1. In this case, the parameters are  $x$ -axis rotation =  $-1.89^\circ$ ,  $y$ -axis rotation =  $0.99^\circ$  and  $z$ -axis translation =  $-1.52$  mm. Even an untrained observer can likely see the improper slice-to-slice anatomical progression in the anterior-most 25% of the volume; the first–second, third–fourth, and fifth–sixth slice pairs are all unusually similar in this region.

error of one slice thickness or more, or loss of all of the transformed voxels. The percentage of data loss is calculated by dividing the numerator by the total number of voxels in the scan.

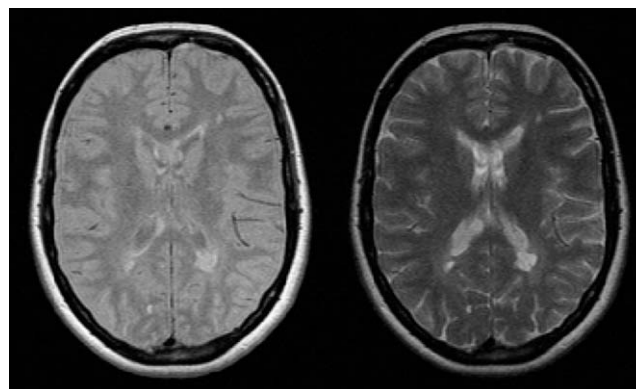
Table 2 shows the statistics for the severity and data loss for the simulated scans. As a result of our parameter choices, the majority of the scans (141/200) have at least some voxels with a positional error perpendicular to the correct image plane of one slice thickness or more. For these scans, a mean of 10.32% (SD = 8.17%, range = 0.01–31.58%) of the transformed voxels have an out-of-plane error of 3 mm or greater.

This data was first used to examine the stability of the fitted B-spline which, as described in Section 2.2, is used to provide a baseline by which the magnitude of interleave errors can be computed. The B-spline values computed from the simulated scans at each scale were compared to those of the error-free scans, using the same subdivided regions. After the stability property was established, the data was used to determine the accuracy of the method by comparing the estimated severity and data loss of each scan to the true values.

### 3.2. MS clinical trial data

A group of 200 patients from 46 sites were randomly selected from a recent clinical trial evaluating the efficacy of a treatment for the secondary progressive form of MS. For each patient, a proton density-weighted (PDw) and a T2-weighted (T2w) scan were used for our experiments. The TE and TR values varied depending on the site: TE1 = 8.4–23.3 ms, TE2 = 60.5–98.0 ms, and TR = 2000–3400 ms. The majority of the volumes have dimensions of  $256 \times 256 \times 50$  voxels, with a voxel size of  $0.937 \text{ mm} \times 0.937 \text{ mm} \times 3.000 \text{ mm}$ . About 10% of the scans have a higher in-plane resolution ( $512 \times 512$ ), with a corresponding decrease in voxel size ( $0.468 \text{ mm} \times 0.468 \text{ mm} \times 3.000 \text{ mm}$ ). No a priori knowledge of the number of acquisitions used for each scan was available.

The PDw and T2w pair for each patient are inherently co-registered, because they were acquired together in a dual-echo sequence. Although the registration is often not perfect due to chemical shift artifacts (Westbrook et al., 2005), the degree of interleave error is generally expected to be very similar between a PDw and its corresponding T2w scan. This property is very useful



**Fig. 12.** PDw and T2w pair acquired simultaneously in a dual-echo sequence. While the scans contain similar information, the tissue contrast is very different, which is a property that is useful for validating the consistency and robustness of our algorithm.

for checking the consistency of our algorithm when applied to scans with significant differences in tissue contrast (Fig. 12). While we do not expect the severity measurements to be exactly the same for PDw and T2w, because the detectability of interleave errors is unavoidably affected by tissue contrast,<sup>1</sup> in general the scores should be close.

We also used the clinical trial data to validate our assumption of a single dominant movement and the key design decision described in Step (7) of Section 2.2, which is to use a sample set of slices, up to the size of one acquisition, to compute the interleave severity at each scale. After running the algorithm on each scan, we repeated the procedure but tagged the sample slices used at each scale in the first round so they would not be identified as candidate slices. The *residual* interleave severity computed in the second round is indicative of how much of the error is not represented by the first measurement. The computation of the residual severity is only done as part of our validation process, and is not required in routine use of the algorithm.

## 4. Results

To interpret the results, it is helpful to have an idea of the minimum error that can be measured by our method. To this end, we applied the algorithm to the SPGR scans that had been resampled to 3 mm thickness, without simulating misalignment, to estimate the amount of baseline noise. The results show a mean severity score of 0.11 acq with a standard deviation of 0.01 acq. We con-

<sup>1</sup> In the extreme case, an interleave error in a scan of a completely homogeneous object would be virtually impossible to detect.

**Table 2**  
Statistics for simulated interleave data.

Statistic	Value
Mean severity	0.44 acq
Standard deviation of severity	0.12 acq
Severity range	0.12–0.68 acq
Mean missing data	14.39%
Standard deviation of missing data	6.61%
Missing data range	2.76–32.98%

clude that the interleave error of a scan would have to equal or exceed approximately 0.15 acq in order to be measured by our algorithm.

#### 4.1. Simulated scans

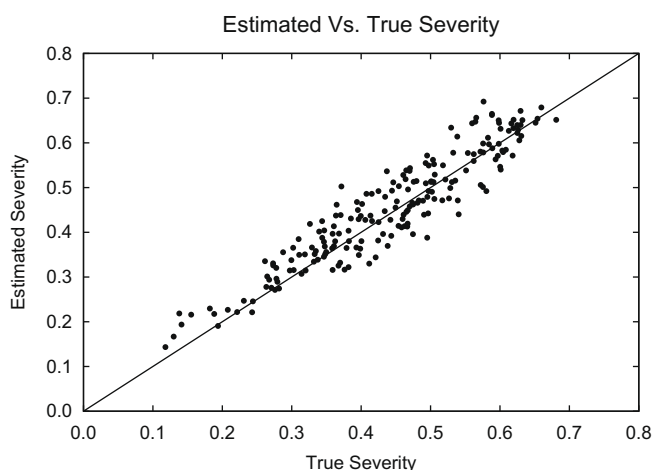
The simulated data was first used to confirm the stability of the fitted B-spline. The differences in the spline values between each simulated scan and its corresponding error-free scan, relative to the differences in the symmetry values, were computed. The results yield a mean of 0.07 (SD = 0.09) for the 200 scans tested, with most of the larger differences being at the ends of the splines, where there is less data available for computing a smooth fit. Overall, the differences are very small, which is a positive indication that the B-spline can be used to provide a confident baseline.

Table 3 presents the summary statistics for the comparison between the measured severity values and the true values for the 200 scans with simulated interleave errors. In general, the estimates and the real values differ only by small amounts; about 98% of the estimates differ from the true values by less than 0.10 acq, with the rest differing by 0.13 acq or less. Fig. 13 shows the scatter plot of the estimated vs. true severity; a strongly linear relationship is the dominant feature. The randomly generated parameters used for the simulation have produced a set of interleave errors that are evenly distributed over a wide range of values and are thus useful for determining whether the algorithm's accuracy is affected by the magnitude of the error being measured. To reveal such potential biases, a Bland–Altman plot of the severity values (differences vs. means) was created and is shown in Fig. 14. The mean of the differences is only 0.01 acq (SD = 0.04 acq). No appreciable bias is evident.

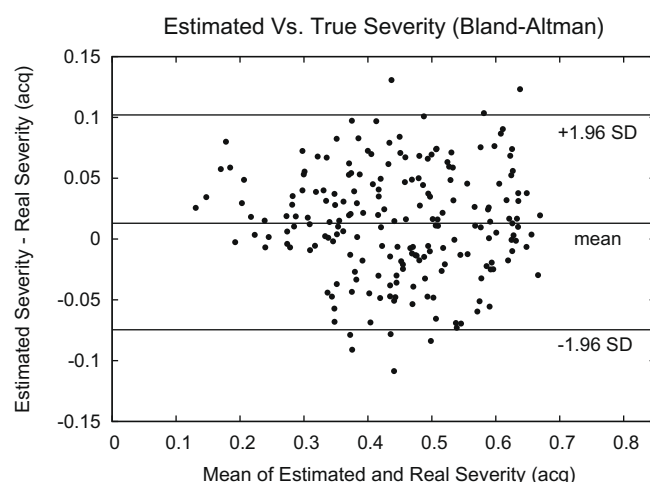
Table 4 gives the summary statistics for the comparison between the measured data loss and the true values for the scans with simulated errors. As mentioned in Section 2.2, the data loss

**Table 3**  
Results of comparing the estimated severity to the true values.

Statistic	Value
Mean absolute difference	0.04 acq
Standard deviation of absolute difference	0.03 acq
Difference range	0.00–0.13 acq
Linear regression slope (95% confidence interval)	0.92 (0.87, 0.97)
Linear regression intercept (95% confidence interval)	0.05 (0.02, 0.07)
Pearson correlation	0.93



**Fig. 13.** Scatter plot of the estimated vs. true severity values.



**Fig. 14.** Bland–Altman Plot comparing the estimated and true severity values. No dependence of the algorithm's accuracy on the magnitude of the error is evident.

**Table 4**  
Results of comparing the estimated percentage data loss values to the true values (pt = percentage point).

Statistic	Value
Mean difference	1.12 pt
Standard deviation of difference	0.98 pt
Difference range	0.00–6.54 pt
Linear regression slope (95% confidence interval)	1.02 (0.99, 1.05)
Linear regression intercept (95% confidence interval)	0.22 (-0.27, 0.71)
Pearson correlation	0.98

for each scan is computed by dividing the severity value by the number of acquisitions found. As with the severity values, the differences between the estimates and real values are generally small. The values also show a strongly linear relationship (Fig. 15 shows the scatter plot), which indicates that the algorithm is able to accurately determine the number of acquisitions comprising each scan. Like the severity values, there is no apparent dependence of accuracy on the magnitude of the real error.

#### 4.2. MS clinical trial data

Tables 5 and 6 contain the summary statistics for the interleave severity and data loss values from the MS clinical trial data set. Despite procedures in place to minimize patient motion, including the use of head straps, a very significant percentage ( $\approx 30\%$ ) of the scans have a measurable amount of interleave error.<sup>2</sup>

Table 7 gives the summary statistics for the comparison between the PDw and T2w severity values. As evident from the numbers, there is excellent agreement between the PDw and T2w measurements. There is only one pair with a difference greater than or equal to 0.10. This provides solid evidence that the method is consistent and not overly sensitive to variations in tissue contrast.

Table 8 shows the summary statistics for the comparison between the PDw and T2w data loss values. These numbers indicate that, in addition to the strong agreement between the interleave severity values, the algorithm is consistent in determining the number of acquisitions present in the scans.

<sup>2</sup> A more thorough analysis of interleave errors in a clinical setting, including a hypothesized link to patient fatigue, will be presented in a more clinically oriented article.



Fig. 15. Scatter plot of the estimated vs. true percentage data loss values.

Table 5  
Interleave severity statistics for MS clinical trial data (units = acq).

Scan	Mean	SD	q1	q2	q3	Range	%>0.15
PDw	0.15	0.07	0.11	0.13	0.16	0.05–0.54	31%
T2w	0.14	0.06	0.11	0.12	0.16	0.05–0.50	29%

Table 6  
Percentage data loss statistics for MS clinical trial data.

Scan	Mean (%)	SD (%)	q1 (%)	q2 (%)	q3 (%)	Range (%)
PDw	4.40	3.53	2.54	3.23	4.88	1.12–27.18
T2w	4.17	3.21	2.52	3.18	4.52	1.21–24.47

Table 7  
Comparison of PDw/T2w severity results.

Statistic	Value
Mean difference	0.02 acq
Standard deviation of difference	0.02 acq
Difference range	0.00–0.11 acq
Linear regression slope (95% confidence interval)	1.04 (0.98, 1.10)
Linear regression intercept (95% confidence interval)	0.00 (–0.01, 0.01)
Pearson correlation	0.93

Table 8  
Comparison of PDw/T2w percentage data loss results (pt = percentage point).

Statistic	Value
Mean difference	0.70 pt
Standard deviation of difference	0.80 pt
Difference range	0.01–5.02 pt
Linear regression slope (95% confidence interval)	1.05 (1.01, 1.10)
Linear regression intercept (95% confidence interval)	0.01 (–0.22, 0.24)
Pearson correlation	0.96

Table 9 contains the summary statistics for the residual severity values from the MS clinical trial data set. The residual severity values were computed by applying the algorithm to the images a second time, but purposely not using the same sample slices that were detected as significant at each scale in the first round. Because the most salient information is being purposely ignored, the residual values are likely to be more affected by measurement noise and

Table 9  
Residual severity statistics for MS clinical trial data (units = acq).

Scan	Mean	SD	q1	q2	q3	Range	%>0.15
PDw (residual)	0.10	0.03	0.08	0.09	0.11	0.03–0.29	6%
T2w (residual)	0.09	0.03	0.07	0.08	0.10	0.02–0.23	5%

not expected to be as accurate as the first- round values. However, they should still be useful indicators of the remaining error. In this case, a large majority (95%) of the scans have a residual severity less than the significance threshold of 0.15 acq. The small magnitude of the residual values strongly suggests that in the large majority of cases, the sample slices selected by our algorithm can be used to compute an interleave error that is representative of the entire scan.

### 5. Discussion and future work

We envision an algorithm for measuring interleave errors to become an essential tool for assessing the quality of images destined for quantitative analysis. For example, in a large clinical trial, where hundreds or even thousands of scans need to be reviewed for quality control, the task of checking every scan manually for visually subtle interleave errors and judging them in an unbiased manner is a daunting one. In this case, our algorithm can be used with a preset threshold to automatically identify scans with large errors. Another environment in which the method can be potentially deployed is the scanner site. The computational overhead of our method is low, and its speed allows immediate feedback to the MR technologist. Our current implementation only takes approximately 4 s to process a  $256 \times 256 \times 60$  scan on an Intel Xeon CPU running at 1.86 GHz.

We have demonstrated our algorithm to be sensitive, accurate and stable. However, as revealed by some of the larger deviations from the expected values, the method is not perfect, and these errors can be attributed to a number of factors, the most conspicuous of which are: (1) violations to our two main assumptions in designing the method; (2) the existence of image artifacts such as noise, blurring and field inhomogeneity, and perhaps most fundamentally; (3) normal variations in the asymmetry values that can only be approximated, in our case by a smoothing spline. Despite these limitations, we assert that the method is sufficiently reliable for practical use.

To further improve the algorithm, there are a number of avenues for future work that we plan to pursue. For the sake of simplicity, so far we have only simulated the motion of one acquisition per scan in our synthetic experiments. However, a patient may move more than once during a scan, especially if they are suffering from spasms as a result of disease. Our experiments with the clinical trial data indicate that in the large majority of cases our method is able to capture most of the motion, even with patients who are suffering from an advanced form of MS. However, we have not determined conclusively whether the effectiveness of the algorithm is due primarily to the way the patients tend to move or properties of the algorithm. Although the slice sampling steps of the algorithm (Steps 7 and 8 in Section 2.2) clearly should work best when there is only one movement, the method does not assume the slices selected all come from the same acquisition, and therefore has some ability to pick up multiple movements. We hypothesize that the strong results obtained are likely attributable to a combination of motion tendencies and algorithmic robustness. Further work with simulated multiple movements should prove informative, especially for determining the effects of significant violations to the assumption that there are sufficient slices in common alignment for a scan to act as a self-reference.

To gauge the impact of other image artifacts that may confound interleave error measurements, various correction methods can be applied for comparison testing. For example, there are existing techniques, many well-established, for noise removal, intensity normalization between acquisitions, bias field correction, and minimization of blurring and ghosting effects due to motion. In addition, further investigation into other image similarity measures that can potentially be more accurate and stable than the MAE when applied to degraded data, such as some of those surveyed by Avcibaş et al. (2002), may prove fruitful.

So far, we have only applied the algorithm to scans with no slice gap, because the focus is on data acquired for quantitative analysis, such as determining lesion volume, in which case gaps can negatively impact the accuracy of the measurements. However, many clinical MR protocols include a slice gap. In theory, our algorithm can be adapted without much difficulty to handle such scans, but we have not yet tried this modification.

## 6. Conclusion

We have designed, implemented and validated an algorithm for measuring the severity of motion-induced inter-slice alignment errors on brain MRIs. We have shown that using an image similarity metric to measure the symmetry of a slice relative to its two neighbours is a valid approach. We performed a number of experiments on scans with synthetically generated interleave errors, as well as a large number of scans from a recent multiple sclerosis clinical trial. The simulated data was used to evaluate the accuracy of the method by comparison to a known ground truth. The clinical data was used to show that the method is robust, consistent and practical for use on real data. We hope that this work will provide the foundation for future investigation into this type of motion artifact.

## Acknowledgements

The authors gratefully acknowledge Yan Cheng, Trudy Harris, Erick Wong, Guojun Zhao and the anonymous reviewers for their

valuable input. Our implementation of the algorithm uses the Wild Magic 3.11 software library (<http://www.geometrictools.com>) to perform the B-spline fitting.

## References

- Atkinson, D., Hill, D.L.G., Stoyke, P.N.R., Summers, P.E., Clare, S., Botwell, R., Keevil, S.F., 1999. Automatic compensation of motion artifacts in MRI. *Magnetic Resonance in Medicine* 41 (1), 163–170.
- Avcibaş, I., Sankur, B., Sayood, K., 2002. Statistical evaluation of image quality measures. *Journal of Electronic Imaging* 11 (2), 206–223.
- Choi, H.S., Haynor, D.R., Kim, Y., 1991. Partial volume tissue classification of multichannel magnetic resonance images – A mixel model. *IEEE Transactions on Medical Imaging* 10 (3), 395–407.
- Filippi, M., Horsfield, M.A., Ader, H.J., Barkhof, F., Bruzzi, P., Evans, A., Frank, J.A., Grossman, R.I., McFarland, H.F., Molyneux, P., Paty, D.W., Simon, J., Tofts, P.S., Wolinsky, J.S., Miller, D.H., 1998a. Guidelines for using quantitative measures of brain magnetic resonance imaging abnormalities in monitoring the treatment of multiple sclerosis. *Annals of Neurology* 43 (4), 499–506.
- Filippi, M., Youstry, T.A., Rocca, M.A., Pereira, C., Alkadhi, H., Comi, G., 1998b. The effect of cross-talk on MRI lesion numbers and volumes in multiple sclerosis using conventional and turbo spin-echo. *Multiple Sclerosis* 4 (6), 471–474.
- Firbank, M.J., Coulthard, A., Harrison, R.M., Williams, E.D., 1999. Partial volume effects in MRI studies of multiple sclerosis. *Magnetic Resonance Imaging* 17 (4), 593–601.
- Gawne-Cain, M.L., Webb, S., Tofts, P., Miller, D.H., 1996. Lesion volume measurement in multiple sclerosis: How important is accurate repositioning? *Journal of Magnetic Resonance Imaging* 6 (5), 705–713.
- Lin, W., Wehrli, F.W., Song, H.K., 2005. Correcting bulk in-plane motion artifacts in MRI using the point spread function. *IEEE Transactions on Medical Imaging* 24 (9), 1170–1176.
- Meijering, E.H.W., Niessen, W.J., Viergever, M.A., 2001. Quantitative evaluation of convolution-based methods for medical image interpolation. *Medical Image Analysis* 5, 111–126.
- Otsu, N., 1979. A threshold selection method from gray level histograms. *IEEE Transactions on Systems, Man and Cybernetics* 9, 63–66.
- Pluim, J., Maintz, J., Viergever, M., 2003. Mutual information based registration of medical images: A survey. *IEEE Transactions on Medical Imaging* 22 (8), 986–1004.
- Schneider, P.J., Eberly, D.H., 2003. *Geometric Tools for Computer Graphics*. Morgan Kaufmann Publishers.
- Traboulsee, A., Zhao, G., Li, D.K., 2005. Neuroimaging in multiple sclerosis. *Neurologic Clinics* 23 (1), 131–148.
- Westbrook, C., Kaut Roth, C., Talbot, J., 2005. *MRI in Practice*, third ed. Blackwell Publishing.