HHS Public Access
Author manuscript
Comput Biol Med. Author manuscript; available in PMC 2016 September 01.

Published in final edited form as:
Comput Biol Med. 2015 September ; 64: 12–23. doi:10.1016/j.compbiomed.2015.06.003.

Towards measuring neuroimage misalignment

Revanth Reddy Garlapati[a], **Ahmed Mostayed**[a], **Grand Roman Joldes**[a], **Adam Wittek**[a], **Barry Doyle**[a,b], and **Karol Miller**[a,c]

[a]Intelligent Systems for Medicine Laboratory, The University of Western Australia, Perth, Australia

[b]Centre for Cardiovascular Science, The University of Edinburgh, Edinburgh, United Kingdom

[c]Institute of Mechanics and Advanced Materials, School of Engineering, Cardiff University, Cardiff, Wales, United Kingdom

Abstract

To enhance neuro-navigation, high quality pre-operative images must be registered onto intra-operative configuration of the brain. Therefore evaluation of the degree to which structures may remain misaligned after registration is critically important. We consider two Hausdorff Distance (HD)-based evaluation approaches: the edge-based HD (EBHD) metric and the Robust HD (RHD) metric as well as various commonly used intensity-based similarity metrics such as Mutual Information (MI), Normalized Mutual Information (NMI), Entropy Correlation Coefficient (ECC), Kullback-Leibler distance (KLD) and Correlation Ratio (CR). We conducted the evaluation by applying known deformations to simple sample images and real cases of brain shift. We conclude that the intensity-based similarity metrics such as MI, NMI, ECC, KLD and CR do not correlate well with actual alignment errors, and hence are not useful for assessing misalignment. On the contrary, the EBHD and the RHD metrics correlated well with actual alignment errors; however, they have been found to underestimate the actual misalignment. We also note that it is beneficial to present HD results as a percentile-HD curve rather than a single number such as the 95-percentile HD. Percentile-HD curves present full range of alignment errors and also facilitate the comparison of results obtained using different approaches. Furthermore, the qualities that should be possessed by an ideal evaluation metric were highlighted. Future studies could focus on developing such an evaluation metric.

Keywords

Non-rigid registration; intra-operative registration; brain deformation; Hausdorff distance; image similarity metrics

Corresponding author: Winthrop Prof. Karol Miller, Address: School of Mechanical and Chemical Engineering, The University of Western Australia (M050), 35 Stirling Highway, Crawley WA 6009, Australia, Phone: (61) 8 6488 8545, Fax: (61) 8 6488 1024, karol.miller@uwa.edu.au.

Conflict of interest statement. None declared.

1 Introduction

Our overall objective is to bring the well-proven benefits of image-guided neurosurgery for brain tumours to sufferers throughout the world (1). Rather than using very expensive and often cumbersome intra-operative Magnetic Resonance (MR) scanners, we plan to include realistic computation of brain deformations, based on a fully non-linear biomechanical model, in a system to improve intra-operative visualisation, navigation and monitoring. The system will create an augmented reality visualisation of the intra-operative configuration of the patient's brain merged with high resolution pre-operative imaging data, including diffusion tensor imaging (DTI) and functional MR imaging (fMRI), in order to better localise the tumour and critical healthy tissues. We accomplish this by registering high quality pre-operative neuroimages onto the current, intra-operative configuration of the patient's brain; however, we do not use an intra-operative image as a target (2, 3)(Fig. 1).

We compute the deformation fields within the entire brain volume and use them to warp high-quality pre-operative MR images so that they correspond to the current, intra-operative configuration, thus compensating for the brain shift (4). The ability to objectively evaluate the accuracy of such biomechanics-based registration against a gold standard – an intra-operative MRI – is of immense importance for the acceptance of biomechanics-based approaches by medical image analysis community and ultimately the clinicians. Moreover, as it is now recognised that precise localisation of the target is the first principle of modern neurosurgery — the measurement of the accuracy of intra-operative image registration (whether biomechanics-based or purely image analysis-based) is of primary significance (5).

There is no satisfactory gold standard for evaluating the accuracy of non-rigid registration, even though a number of widely used methods have been developed for a simpler task of rigid registration of neuroimages (6, 7). The registered pre-operative images are evaluated using intra-operative images as ground truth (8). Evaluating the accuracy of non-rigid registration is inherently difficult and has attracted a lot of attention in medical image analysis literature (7, 9, 10). In particular, inter-subject non-rigid registration accuracy evaluation of brain MR images has been investigated in great detail (10, 11). Unfortunately, the conclusions of these papers are rather sober: even though a number of registration approaches performed well as measured by a variety of image similarity indices, most of these evaluation metrics are unable to measure the misalignment of structures reliably (7). However the accuracy of structure alignment is the key requirement for intra-subject, intra-operative neuroimage registration aimed at producing reliable data for neuro-navigation.

Our objective is to generate accurate data for neuro-navigation and therefore the estimation of the extent to which the warped (registered) pre-operative images are misaligned relative to the current, intra-operative configuration of patient's brain is of great importance to us. We require measures that provide the misalignment (alignment error) estimates in millimetres (mm), for the evaluation results to be of any practical use. It appears, therefore, that we need to focus on approaches allowing comparison of two sets of feature points. One such measure, often used in image analysis, is the Hausdorff Distance (HD) (12).

The choice of the “features” whose locations in images are compared appears as probably the most important practical aspect of application of HD-based approaches. For example, Ferrant et al. used 400 manually selected landmarks, and identified the misalignment (in mm) of the corresponding landmarks before and after registration (13). A similar approach was used by Clatz et al. (14). These approaches are labour-intensive and probably unsuitable for large amounts of image data. The practical difficulty of having an expert carry this out reproducibly needs to be highlighted, and therefore replacing that expert landmark selection with an automated process would be beneficial (15).

Another suggested approach is to conduct segmentations of small regions of the brain (10). These regional segmentations are like landmarks, but instead of points, they are small regions. A lot of effort has been put into obtaining regional segmentations, and the best techniques are competitive with expert segmentations (16–18). However, the precision of small regions for assessing accuracy may be limited by the region size.

We also applied HD measures to contours of structures (ventricles and tumours) segmented in warped pre-operative and intra-operative images (19, 20). The published results look very convincing, but their validity crucially depends on the accuracy and reproducibility of manual segmentation, which unfortunately is always in doubt.

HD-based approaches that do not rely on manual segmentation or (expert) identification of landmarks require automatic, repeatable identification of corresponding features in compared images. It has been demonstrated that point-sets forming edges or contours of segmented regions, may be used together with HD-based methods to estimate the misalignment of neuroimages (8, 20, 21). In this paper we analyse results obtained using two such methods: Edge-based Hausdorff Distance (EBHD) (21) and Robust Hausdorff Distance (RHD) (8) metrics.

We conducted the evaluation by applying known deformations to simple artificial sample images, and MR images of real cases of brain shift. Additionally, we computed other commonly used evaluation metrics such as Mutual Information (MI), Normalized Mutual Information (NMI), Entropy Correlation Coefficient (ECC), Kullback-Leibler distance (KLD) and Correlation Ratio (CR) to demonstrate that their utility in assessing the quality of registration, and in particular the degree of structure misalignment and accuracy of surgical target localisation, is limited, as was also suggested by Rohlfing (7).

The paper is organised as follows: in Section 2, a brief description of the evaluation metrics used for assessing registration accuracy is provided; in Section 3 we present verification results for simple sample images, and five real craniotomy-induced brain shift cases; and finally, the discussion and conclusions are given in Section 4.

2 Materials and Methods

2.1 Medical imaging data

We chose the pre-operative MR images which we previously used to build biomechanical brain models (20). These five image datasets with brain tumours (cerebral gliomas) were randomly selected from a retrospective database of 859 intracranial tumour cases available at the Harvard Medical School's clinical affiliate Brigham and Women's Hospital in Boston (22). Images were acquired using a 0.5T open MR system in the neurosurgical suite. The resolution of the images is 0.85 mm × 0.85 mm × 2.5 mm. Consent for the use of the anonymised retrospective image database was obtained in accordance with the Institutional Review Board of the Children's Hospital (Harvard Medical School's clinical affiliate) in Boston (whose researchers accessed the database).

2.2 Image registration evaluation metrics

Edge-Based Hausdorff Distance evaluation procedure—The choice of the "features" whose locations in images are compared is critically important. It is preferable that this choice be objective and repeatable. For these reasons we use automatically detected Canny edges (23) (Fig. 2).

Round-trip distance and consistent pixels: The concepts of round-trip distance and consistent pixels are necessary for defining the EBHD metric and for pre-processing of edges. Let us consider two sets of Canny edges, **A** and **B**, extracted from two images being compared. Each pixel and the corresponding point in space are denoted using the same notation in this paper. For each non-zero pixel p in the binary image **A**, we find the nearest non-zero pixel q in the other image **B**. If the same procedure is followed from the point q in image B, we expect to arrive at the starting point p from image **A**, if the images are perfectly aligned. However, the images that are being compared are not perfectly aligned in general, and therefore we reach an end point r which is different from the initial point p. The distance between the initial point and the end point in image **A** is defined as the round-trip distance.

The pixels that have a round-trip distance higher than a prescribed threshold are omitted from the evaluation process, as they are less likely to have corresponding pixels in the other image. This threshold is termed as the round-trip consistency threshold criterion, while the pixels that are not omitted are called consistent pixels. Since pixels that have a round-trip distance greater than the threshold criterion get omitted from the evaluation process, it is important to choose an optimal threshold such that the consistency is maximized, while minimizing the number of features excluded. The round-trip threshold criterion for computing the EBHD metric was chosen as 1mm, that is, twice the in-plane resolution of the images being compared. Only the consistent pixels are involved in the computation of directed distance for each edge while computing the EBHD metric.

Pre-processing: The skull was excluded from the images prior to the evaluation process by applying a mask based on the outer boundary of the brain that was segmented from the images using in-built functionality of Slicer3D. The edges shorter than 5 mm were deleted from the analysed images as their dimension is inconsequential in comparison to the relevant brain dimension, which is approximately 150 mm. The remaining edges were made consistent with respect to each other by applying the round-trip threshold criterion. The round-trip threshold criterion for pre-processing is selected as 2 mm, that is, 4 times the in-plane resolution of the resampled images. The purpose of this round-trip threshold criterion is to remove features that have no correspondence in the other image. During pre-processing, the choice of a lower (stringent) threshold causes some of the features with correspondence in the other image to be omitted. The threshold during pre-processing is greater (less stringent) than during the EBHD metric computation, because pixels that are not consistent are only ignored during the computation of EBHD metric between the edges, whereas such pixels are removed from the images during pre-processing. Higher (less stringent) threshold during preprocessing enables us to remove inconsistent features while retaining most of those that actually have correspondence. When edges are made consistent with respect to each other using the round-trip threshold, more short edges are generated. The short edges are deleted from the images and the consistency procedure is applied one more time. The remaining edges were found to represent features that better correspond to each other in the images being compared (Fig. 3). For more detailed information on round-trip consistency procedure, the reader is referred to our previous work (21).

Definition of the EBHD metric: Let the images $\mathbf{A} = \{\mathbf{a_1}, \ldots, \mathbf{a_m}\}$ and $\mathbf{B} = \{\mathbf{b_1}, \ldots, \mathbf{b_n}\}$ indicate finite edge sets, with each element of the set as a Canny edge (23)(Fig. 2). The HD is the maximum directed distance computed using the edge sets as demonstrated in Eq. 1. The procedure for computing the directed distance between edge sets is different from the point-based directed Euclidean distance, as illustrated in Eqs. 2–4 (21).

$$\mathrm{H}(\mathbf{A},\mathbf{B})=\max\ (\mathrm{h}(\mathbf{A},\mathbf{B}),\mathrm{h}(\mathbf{B},\mathbf{A})) \quad (1)$$

$$\mathrm{h}(\mathbf{A},\mathbf{B})=\max_{\mathbf{a}\in\mathbf{A}}\min_{\mathbf{b}\in\mathbf{B}}\mathrm{h}^{'}(\mathbf{a},\mathbf{b}) \quad (2)$$

Let edges **a** and **b** belong to edge sets **A** and **B** respectively. The point sets $\mathbf{a}=\left\{\mathbf{a}_{\mathbf{1}}^{'},\ldots,\mathbf{a}_{\mathbf{m}}^{'}\right\}$ and $\mathbf{b}=\left\{\mathbf{b}_{\mathbf{1}}^{'},\ldots,\mathbf{b}_{\mathbf{n}}^{'}\right\}$ contain the points that form these edges. These point sets are made consistent with respect to each other preceding the computation of the EBHD metric using the round-trip consistency procedure explained above. Therefore, only the consistent subsets of edges **a** and **b** are used to compute the edge-based directed distance. Let the consistent point sub-sets be denoted by $\tilde{\mathbf{a}}=\left\{\mathbf{a}_{\mathbf{1}}^{'},\ldots,\mathbf{a}_{\tilde{\mathbf{m}}}^{'}\right\}$, $m \lesssim m$, and $\tilde{\mathbf{b}}=\left\{\mathbf{b}_{\mathbf{1}}^{'},\ldots,\mathbf{b}_{\tilde{\mathbf{n}}}^{'}\right\}$, $\tilde{n} \le n$. The edge-based directed distance is determined using the point-based HD metric between the consistent point subsets $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ (Eqs. 3–4):

$$\mathrm{h}^{'}(\mathbf{a},\mathbf{b})=\max\ \left(\mathrm{h}^{''}(\tilde{\mathbf{a}},\tilde{\mathbf{b}}),\mathrm{h}^{''}(\tilde{\mathbf{b}},\tilde{\mathbf{a}})\right) \quad (3)$$

$$\mathrm{h}^{''}(\tilde{\mathbf{a}},\tilde{\mathbf{b}})=\max_{a'\in\tilde{\mathbf{a}}}\min_{b'\in\tilde{\mathbf{b}}}\|a^{'}-b^{'}\| \quad (4)$$

Application to three dimensional (3D) MR images: In this study two dimensional (2D) image sequences were created in the axial and coronal planes (mutually perpendicular planes) from each of the 3D MR images using 3D Slicer (www.slicer.org)(24). Our EBHD metric is used to evaluate the registration accuracy by comparing each warped pre-operative 2D image with the corresponding intra-operative image.

The original resolution of the images is 0.85 mm x 0.85 mm x 2.5 mm. The pre-operative and intra-operative images were resampled to a resolution of 0.5 mm x 0.5 mm x 2.5 mm, in order to enhance the precision in locating the edges in each 2D (axial/coronal) image. We coded our evaluation algorithm in Matlab environment (www.mathworks.com). Subsequently, the built-in Canny edge detector in Matlab using default lower and upper bounds, as well as Gaussian smoothing parameter of 4, is used for 2D Canny edge detection (Fig. 2). The EBHD metric is used to compute the alignment error between each 2D image pair in the region of interest (craniotomy area). The results of all image pairs are combined together for the percentile-EBHD metric analysis in each plane. The *n* percentile-EBHD metric is defined as the value that is higher than *n* percent of the total number of directed distances belonging to edges of either image. The EBHD metric is computed for full percentile range (0–100). The round-trip consistency procedure ensures that most edges used for EBHD computation have a corresponding edge in the other image. Nevertheless, some pairs of edges do not represent the same anatomical feature in the two images, and thus lack correspondence (Fig. 3). The presence of such pairs of edges (outliers) leads to unusually large values of the EBHD, which are not representative of the actual registration accuracy. In order to eliminate such outliers, the registration accuracy analysis involves the computation of percentile-EBHD metric values in conjunction with visualisation of overlay of edges for 2D images containing potential outliers (21)(Fig. 4). The n-percentile EBHD metric is defined as the value that is greater than n percent of the total number of directed distances belonging to edges of either image.

Robust Hausdorff Distance evaluation procedure—This evaluation method is based on the RHD metric defined by Fedorov et al. and contains a number of robust modifications to the conventional point-based HD metric (8). Here, rather than Canny edges, we used the boundaries (contours) of the corresponding segmented anatomical regions as features whose alignment is to be compared. Automatic atlas-based segmentation was employed to determine these contours (25). The Maximum A Posteriori (MAP) formulation of the Simultaneous truth and performance level estimation (STAPLE) algorithm, referred to as MAP STAPLE (26) was used for segmentation. This is one of the leading segmentation algorithms and we are not aware of an algorithm that could detect gray matter boundaries more accurately. This approach is beneficial because the ordinary edge detection methods such as Canny edge detection may fail to extract the relevant boundaries (contours) for low contrast and low resolution intra-operative images (Fig. 5). Therefore, we extract the boundaries of different tissue types, i.e., grey/white matter, from the tissue segmentation of the images (25–27), and compare corresponding boundaries using the RHD measure (Fig. 5) (8). Some boundaries extracted using Canny edge detector are partially detected, which disrupts the edges, or are incorrectly connected to each other, which are falsely considered as one boundary (Fig 5.a). Furthermore, if there are skull or cerebrospinal fluid residues remaining after the brain extraction, the boundaries of the remaining residues are incorrectly detected as the brain tissue boundaries (Fig 5.a).

While the robust modifications to the standard HD metric reduce the number of outliers, they do not entirely prevent them (8). A typical example of outliers occurring in this evaluation process is presented in Fig. 6. This proposed approach also provides local estimates and visual assessment of the registration error (Fig. 7). In order to compare the results of contour-based RHD metric to those of our edge-based HD metric, we compute the alignment errors for full percentile range (0–100) from the RHD error estimate maps.

The RHD metric between images **A** and **B** depends on the definition of greyscale local HD (GH_{loc} (x)), defined as follows (8):

$$\forall x \in R^2 : \mathrm{GH}_{\mathrm{loc}}(x) = \left| 1_{\mathrm{A}(x)} - 1_{\mathrm{B}(x)} \right| \times \max\left(\mathrm{d}\left(x_g, \tilde{\mathrm{A}}\right), \mathrm{d}\left(x_g, \tilde{\mathrm{B}}\right) \right) \quad (5)$$

The RHD is calculated as the average of the ordered values of GH_{loc} (x) in the fixed size window centered at point/pixel x, after discarding top 20% percent of the distance values within this window (trimmed mean value). $1_{A(x)}$ is a function that has value of 1 if A(x) is non-zero, and 0 otherwise; d(**a**, B) = $\min_{b \in \mathbf{B}} \|a - b\|$ for pixels/points a∈A and b∈B.

Intensity-based similarity metrics—Measures of image similarity such as MI, NMI, ECC, KLD and CR that inspect the intensity probability distribution of the images are widely used as surrogates for evaluating image registration accuracy (7, 28–32). MI, NMI, ECC and KLD minimise various definitions of joint entropy between the registered images (28, 29, 31, 32); while CR measures the degree to which one image is a single-valued function of the other (30). None of these measures provide the alignment error in terms of physical distance (in mm) and serve only as a qualitative indicator of registration accuracy. The unreliability of such measures was discussed in a recent study (7).

2.3 Measurement of image misalignment using HD-based methods

Verification of our EBHD metric and the RHD metric using simple artificial images

2D verification: The first step in the verification of our edge-based evaluation algorithm was the creation of a simple artificial brain-like 2D image (Fig. 8). A new image was generated by rigidly translating this artificial image by 5 pixels along the x-axis. Subsequently, the percentile-EBHD and percentile-RHD metric values were computed by comparing the translated image with the original image, in order to test whether these methods can recover such alignment errors.

3D verification: Similar to that of 2D verification process, an artificial image of an ellipsoid with dimensions 100 mm x 150 mm x 100 mm (200 pixels x 300 pixels x 200 pixels) was generated. A new image was created by rigidly displacing this ellipsoid by 5 mm along X, Y and Z axes. Subsequently, the percentile-EBHD and percentile-RHD metric values were computed in axial and coronal planes by comparing the displaced image with the original image, in order to test whether these methods can recover such alignment errors.

Verification of the similarity metrics using real brain MR images and known deformation fields—In one of our recent studies, patient-specific biomechanical models were used to compute the deformation fields used to warp the pre-operative images onto their intra-operative configurations (20). Pre-operative image from each of these five real neurosurgery cases is warped with three distinct known (non-rigid) 3D deformation fields using algorithms developed by our group previously (4, 33). These deformation fields for each case are obtained by scaling known deformation fields by factors 0.25, 1.0 and 1.75. The EBHD metric, RHD metric, MI, NMI, ECC, KLD and CR are then used to assess the difference between the warped pre-operative images and the original pre-operative image, and subsequently these results are compared.

3 Results

3.1 2D Verification results for our EBHD metric and the RHD metric using a simple brain-like object image

The EBHD metric values at higher percentiles (65–100) were found to be equal to the actual alignment error, that is, 5 pixels (equivalent distance is 2.5 mm, as the image resolution is 0.5 mm) (Fig. 9). Similarly, the RHD metric values between 78–100 percentiles are equal to the actual alignment error. A repetition of this procedure using a 5 pixel translation along the y-axis, led to a similar results, demonstrating that these methods can recover alignment errors associated with rigid translation in the plane of the image.

3.2 3D Verification results for our EBHD metric and the RHD metric using an ellipsoid image

The EBHD metric values at higher percentiles (90–100) in axial and coronal planes were found to be close to the actual (planar) alignment error in either planes, that is, 7 mm (Fig. 10 and Fig. 11). Whereas, the RHD metric values between 95–100 percentiles are close to 8 mm for the axial and coronal planes respectively, implying that actual alignment errors were recovered, along with a few outliers (Fig. 10 and Fig. 11). The results demonstrate that these evaluation methods can accurately recover at least a large fraction of alignments errors associated with rigid translation in 3D.

3.3 Verification results for HD-based evaluation metrics using real images

The percentile-EBHD metric and percentile-RHD metric plots associated with various imposed deformation fields are presented for the five cases of real brain MR images in Fig. 12 and Fig. 13 for axial and coronal planes respectively. It can be observed that the EBHD and RHD metric values at higher percentiles (70–100), generally increase (with the exception of outliers) with an increase in the value of the maximum 3D deformation (alignment error). This shows that a good correlation exists between the percentile-EBHD/percentile-RHD metric and the maximum deformation (alignment error), and therefore the percentile-EBHD/percentile-RHD metric (at high percentiles) can be utilised as an indicator of the degree of misalignment. The coefficient of determination (R^2) values for the plot of 90-percentile EBHD metric in axial plane against 90-percentile deformation are 0.95, 0.91, 0.99, 0.89 and 0.98 for Cases 1–5 respectively. The coefficient of determination (R^2) values for the plot of 90-percentile EBHD metric in coronal plane against 90-percentile deformation are 0.97, 1.0, 0.89, 0.96 and 1.0 for Cases 1–5 respectively. The coefficient of determination (R^2) values for the plot of 90-percentile RHD metric in axial plane against 90-percentile deformation are 0.97, 0.96, 0.81, 0.92 and 0.81 for Cases 1–5 respectively. The coefficient of determination (R^2) values for the plot of 90-percentile RHD metric in coronal plane against 90-percentile deformation are 0.98, 0.96, 0.92, 0.92 and 0.92 for Cases 1–5 respectively.

However, in general, both the percentile-EBHD metric and the percentile-RHD metric values deviate from the applied maximum deformation values (alignment errors), as their definitions do not include point-to-point correspondence (8, 21). Furthermore, it can be observed that the percentile-EBHD and percentile-RHD curves in the either axial or coronal planes are more sensitive to increase in maximum alignment error in a particular plane in case a larger component of the alignment error is concentrated in that plane (Fig. 12 and Fig. 13). For example, it is apparent that Case 4 has larger alignment error in the coronal plane than axial plane (Fig. 12 and Fig. 13). Additionally, we presented overlaid Canny edges corresponding to EBHD/RHD metric plots for Case 1 (Fig. 14). The deformation field ranges for Cases 1–5 in the craniotomy region were presented in Fig. 15.

3.4 Verification results for intensity-based evaluation metrics using real brain images

From the intensity-based similarity metric results for real images, it can be observed that all the similarity metrics such as MI, NMI, ECC and CR increase with decrease in the alignment error (Tab. 1). The contrary trend is observed in KLD as it is measure of dissimilarity between images, different from other intensity-based similarity metrics (Tab. 1). In addition, none of the intensity-based metrics provide an estimate of alignment error in terms of physical distance (mm). Furthermore, the intensity-based similarity metrics do not correlate well with alignment errors, in comparison to HD-based measures. In fact, the similarity metric values such as MI, NMI, ECC and KLD only serve as a relative qualitative indicator of alignment error, as their absolute value does not give an indication of alignment error. However, the absolute value of CR serves as a qualitative indicator of alignment error, though it is not possible to get an approximate indication of alignment error from its value. The CR values close to 1 indicate lower alignment errors, and values closer to 0 (lower than 0.5) indicate large values of alignment error. The values of other similarity metrics such as MI, NMI, ECC and KLD can be very different for same alignment errors depending on the intensity ranges of the images being compared (Tab. 1).

4 Discussion

The aim of this paper is to highlight the need for an objective measure of misalignment in registered neuroimages and suggest ways in which it can be constructed. In Section 3, we demonstrated using simple brain-like object images that our EBHD evaluation metric and the RHD metric can at least recover large component of alignment errors associated with rigid translation. Subsequently, we confirmed that the HD-based evaluation metrics correlate well with (non-rigid) alignment errors. However, neither of these methods can completely avoid outliers. It is also important to note that all HD-based evaluation methods suffer from the limitation of underestimation of alignment error, as the registration error can only be measured in the direction perpendicular to the edge, which leads to uncertainty of the error estimate tangential to the edge (7). In order for HD-based evaluation metrics to recover all components of alignment error, we must be able to compute the point-to-point correspondence between the dense set of features being compared (8, 21). This is a challenging problem, and has not been addressed satisfactorily by the medical image processing community yet. Moreover, our methods evaluate edges within slices across the entire volume and in perpendicular cross-sections, rather than in three dimensions. Unfortunately it is currently not possible to reliably identify 3D edges in an MRI volume. Perhaps this imperfection of the methods considered here is another reason why they underestimate the alignment error – they never fully capture the displacements normal to the slices considered. This problem can be partly addressed by combining results obtained in perpendicular planes.

Additionally, it is important to note that a comparison of the EBHD and the RHD metrics was facilitated by the presentation of the alignment error results using percentile-HD curves. One of the key advantages of this form of reporting alignment error results is that it provides a full range of alignment errors, enabling more detailed comparison between results obtained using different approaches. Furthermore, the percentile-HD curves enable us to estimate the percentage of edges (or points) that have been successfully registered by depicting the percentile at which the alignment errors are lower than the acceptable alignment error, that is, 2 pixels (1. 7 mm in this study). For example, if the alignment errors below the 80th percentile are lower than the acceptable alignment error, then the estimate of the percentage of edges that have been successfully registered is 80 %.

It was also demonstrated in the results section that the intensity-based evaluation (surrogate) metrics such as MI, NMI, ECC, KLD and CR are not very useful for evaluating image registration accuracy when the primary concern is the magnitude of misalignment of anatomical structures. This is because they neither provide the registration error estimate in terms of physical distance (mm) nor do they correlate well with actual alignment errors. It is important to distinguish between image similarity measures, which are used by registration optimisation process, and the Euclidean/physical distance measures, which are necessary for registration evaluation in the context of intra-operative registration. It could be argued that these findings are well-known and obvious. However, there are several studies in the literature that continue to use image similarity measures as evaluation metrics even though it is now known that they are not reliable for such purpose. A recent study by Rohlfing illustrated the unreliability of several intensity–based similarity metrics using a different approach, and recommended the use of distance-based metrics instead (7). That study illustrated that the intensity-based similarity metrics indicate relatively good quality registration of images, despite of the images being registered by an obviously poor quality registration tool (7). Additionally, it was suggested that the evaluation metric should use a dense set of landmarks corresponding to anatomical structures in the region of interest (7).

While fully comprehending the enormity of the task at hand we suggest here the qualities of a yet to be developed ideal evaluation method for measuring neuroimage registration accuracy:

1. The method’s output should be in terms of distance, i.e., in mm.
2. It should rely on a dense set of features.
3. It should be automatic and fast, enabling evaluation of large amounts of image data.

4. It should be objective, i.e., not depend on manual segmentation or manual landmark definition and subjective selection of tuneable algorithmic parameters.
5. It should be able to compute the actual error, i.e., include point-to-point correspondence between the features being compared.

5 Conclusions

Our EBHD metric and the RHD metric possess the first four qualities from the aforementioned list. Although our EBHD and the RHD metrics perform similarly in terms of alignment error recovery (Fig. 12 and Fig. 13), the EBHD metric is easier to use as it is based on features extracted using the user-friendly and widely available Canny edge detector, and does not entail specialist knowledge required to segment images. There is no evaluation method that possesses the fifth quality yet. Perhaps, future studies could focus on developing evaluation methods that incorporate all the aforementioned five qualities.

Acknowledgments

The first author is a recipient of a SIRF (Scholarship for International Research Fees) scholarship and gratefully acknowledges the support of The University of Western Australia (Research Collaboration Award). The financial support of the National Health and Medical Research Council (Grant No. APP1006031), National Institute of Health (Grants R01 EB008015 and R01 LM010033), Australian Research Council (DP 1092893) and Children's Hospital Boston Translational Research Program is also gratefully acknowledged. Additionally, the authors also acknowledge the financial support of the Neuroimage Analysis Center (NIH P41 EB015902), National Center for Image Guided Therapy (NIH U41 RR019703) and the National Alliance for Medical Image Computing (NAMIC), supported by the National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149. Information on the National Centers for Biomedical Computing can be acquired from http://nihroadmap.nih.gov/bioinformatics. None of the aforementioned funding agencies had any direct involvement in this study. The authors would like to thank Dr. Vahid Taimouri and Prof. Simon K. Warfield of Computational Radiology Lab (CRL), Harvard Medical School for their assistance in the computation of the evaluation results associated with the RHD metric. Authors declare that the study sponsors were not involved in the study design, in the collection, analysis and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

Abbreviations

HD	Hausdorff Distance
EBHD	edge-based HD
RHD	Robust HD
MI	Mutual Information
NMI	Normalized Mutual Information
ECC	Entropy Correlation Coefficient
KLD	Kullback-Leibler distance
CR	Correlation Ratio

References

1. Miller, K. Biomechanics of the Brain. New York: Springer; 2011.

2. Garlapati RR, Roy A, Joldes GR, Wittek A, Mostayed A, Doyle B, et al. More accurate neuronavigation data provided by biomechanical modeling instead of rigid registration. J Neurosurg. 2014; 120(6):1477–83. [PubMed: 24460486]
3. Mostayed A, Garlapati RR, Joldes GR, Wittek A, Roy A, Kikinis R, et al. Biomechanical Model as a Registration Tool for Image-Guided Neurosurgery: Evaluation Against BSpline Registration. Ann Biomed Eng. 2013; 41(11):2409–25. [PubMed: 23771299]
4. Joldes, GR.; Wittek, A.; Warfield, SK.; Miller, K. Performing Brain Image Warping Using the Deformation Field Predicted by a Biomechanical Model. In: Nielsen, PMF.; Wittek, A.; Miller, K., editors. Computational Biomechanics for Medicine. New York: Springer; 2012. p. 89-96.
5. Nakaji P, Spetzler R. Innovations in surgical approach: the marriage of technique, technology, and judgment. Clin Neurosurg. 2004; 51:177–85. [PubMed: 15571143]
6. Murphy K, van Ginneken B, Klein S, Staring M, de Hoop B, Viergever MA, et al. Semi-automatic construction of reference standards for evaluation of image registration. Med Image Anal. 2011; 15(1):71–84. [PubMed: 20709592]
7. Rohlfing T. Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable. IEEE Trans Med Imaging. 2012; 31(2):153–63. [PubMed: 21827972]
8. Fedorov, A.; Billet, E.; Prastawa, M.; Gerig, G.; Radmanesh, A.; Warfield, S., et al. Evaluation of Brain MRI Alignment with the Robust Hausdorff Distance Measures. In: Bebis, G.; Boyle, R.; Parvin, B.; Koracin, D.; Remagnino, P.; Porikli, F., et al., editors. Advances in Visual Computing. Lecture Notes in Computer Science. Berlin: Springer; 2008. p. 594-603.
9. Wei, Y.; Christensen, GE.; Song, JH.; Rudrauf, D.; Bruss, J.; Kuhl, JG., et al. Evaluation of five non-rigid image registration algorithms using the NIREP framework. In: Dawant, BM.; Haynor, DR., editors. Medical Imaging 2010; San Diego, USA. Proceedings of the SPIE; 2010. p. 76232L-L-10
10. Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang M, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage. 2009; 46(3):786–802. [PubMed: 19195496]
11. Song, J.; Christensen, G.; Hawley, J.; Wei, Y.; Kuhl, J. Evaluating Image Registration Using NIREP. In: Fischer, B.; Dawant, B.; Lorenz, C., editors. Biomedical Image Registration. Lecture Notes in Computer Science. Vol. 6204. Berlin: Springer; 2010. p. 140-50.
12. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing Images Using the Hausdorff Distance. IEEE Trans Pattern Anal Mach Intell. 1993; 15(9):850–63.
13. Ferrant M, Nabavi A, Macq B, Black PM, Jolesz FA, Kikinis R, et al. Serial registration of intraoperative MR images of the brain. Medical Image Analysis. 2002; 6(4):337–59. [PubMed: 12426109]
14. Clatz O. Robust nonrigid registration to capture brain shift from intraoperative MRI. IEEE Trans Med Imaging. 2005; 24(11):1417–27. [PubMed: 16279079]
15. Rohr K, Stiehl HS, Sprengel R, Buzug TM, Weese J, Kuhn MH. Landmark-based elastic registration using approximating thin-plate splines. IEEE Trans Med Imaging. 2001; 20(6):526–34. [PubMed: 11437112]
16. Yang F, Kruggel F. Automatic segmentation of human brain sulci. Med Image Anal. 2008; 12(4):442–51. [PubMed: 18325826]
17. Tu Z. Brain anatomical structure segmentation by hybrid discriminative/generative models. IEEE Trans Med Imaging. 2008; 27(4):495–508. [PubMed: 18390346]
18. Barber D, Hose D. Automatic segmentation of medical images using image registration: diagnostic and simulation applications. J Med Eng Technol. 2005; 29(2):53–63. [PubMed: 15804853]
19. Wittek A, Miller K, Kikinis R, Warfield SK. Patient-specific model of brain deformation: Application to medical image registration. J Biomech. 2007; 40(4):919–29. [PubMed: 16678834]
20. Wittek A, Joldes G, Couton M, Warfield SK, Miller K. Patient-specific non-linear finite element modelling for predicting soft organ deformation in real-time; Application to non-rigid neuroimage registration. Prog Biophys Mol Biol. 2010; 103(2–3):292–303. [PubMed: 20868706]
21. Garlapati, RR.; Joldes, GR.; Wittek, A.; Lam, J.; Weisenfeld, N.; Hans, A., et al. Objective evaluation of accuracy of intra-operative neuroimage registration. In: Wittek, A.; Miller, K.;

Nielsen, PMF., editors. Computational Biomechanics for Medicine. New York: Springer; 2013. p. 87-99.

22. Warfield SK, Haker SJ, Talos IF, Kemper CA, Weisenfeld N, Mewes AUJ, et al. Capturing intraoperative deformations: research experience at Brigham and Womens's hospital. Med Image Anal. 2005; 9(2):145–62. [PubMed: 15721230]
23. Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Machine Intell. 1986; 8(6):679–98.
24. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. 2012; 30(9):1323–41. [PubMed: 22770690]
25. Akhondi-Asl A, Warfield S. Simultaneous truth and performance level estimation through fusion of probabilistic segmentations. IEEE Trans Med Imaging. 2013; 32(10):1840–52. [PubMed: 23744673]
26. Commowick O, Akhondi-Asl A, Warfield SK. Estimating a reference standard segmentation with spatially varying performance parameters: Local MAP STAPLE. IEEE Transactions on Medical Imaging. 2012; 31(8):1593–606. [PubMed: 22562727]
27. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004; 23(7):903–21. [PubMed: 15250643]
28. So, RW.; Chung, AC. Multi-modal non-rigid image registration based on similarity and dissimilarity with the prior joint intensity distributions. IEEE International Symposium on Biomedical Imaging: From Nano to Macro; Rotterdam: IEEE; 2010. p. 368-71.
29. Pluim JP, Maintz JA, Viergever MA. Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging. 2003; 22(8):986–1004. [PubMed: 12906253]
30. Roche, A.; Malandain, G.; Pennec, X.; Ayache, N. The correlation ratio as a new similarity measure for multimodal image registration. In: Wells, WM.; Colchester, A.; Delp, S., editors. Medical Image Computing and Computer-Assisted Intervention—MICCAI'98. Berlin: Springer; 1998. p. 1115-24.
31. Studholme C, Hill DL, Hawkes DJ. An overlap invariant entropy measure of 3D medical image alignment. Pattern Recogn. 1999; 32(1):71–86.
32. Wells W, Viola P, Atsumi H, Nakajima S, Kikinis R. Multi-modal volume registration by maximization of mutual information. Med Image Anal. 1996; 1(1):35–51. [PubMed: 9873920]
33. Li M, Wittek A, Miller K. Efficient Inverse Isoparametric Mapping Algorithm for Whole-Body Computed Tomography Registration Using Deformations Predicted by Nonlinear Finite Element Modeling. J Biomech Eng. 2014; 136(8) in press.

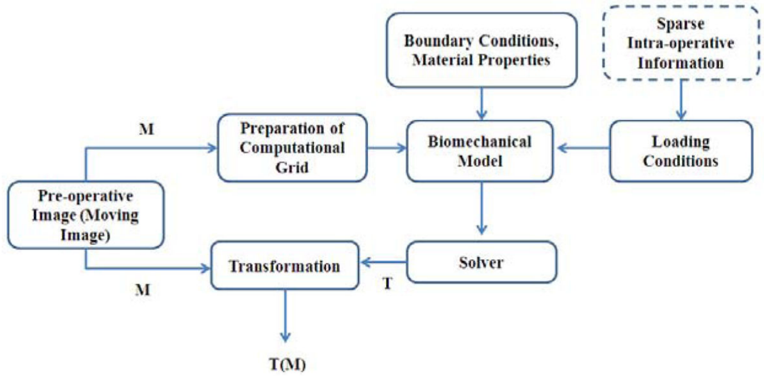


Fig. 1.

Registration process based on a biomechanical model. The flow chart depicts various steps used in registering the pre-operative images onto their intra-operative configuration using biomechanical models. M is the moving image (pre-operative image). T is the transform that registers the pre-operative image onto current intra-operative configuration of the brain. T(M) is the transformed moving image (warped pre-operative image).

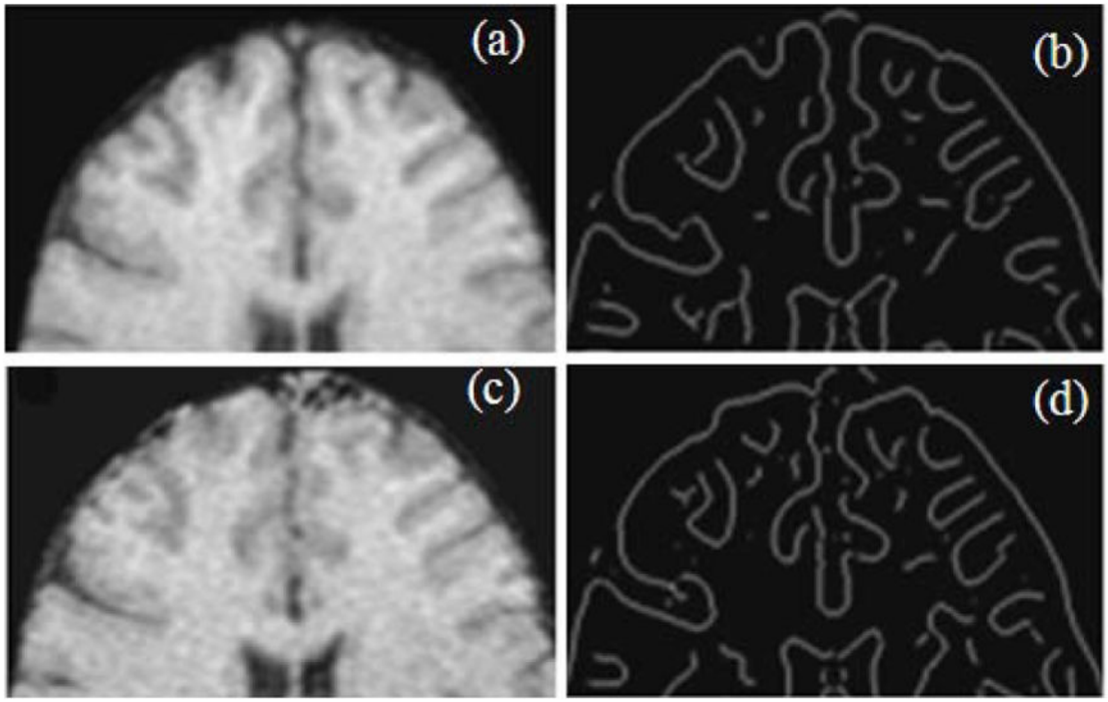


Fig. 2.
The Canny edges of a deformed (warped) pre-operative and the corresponding intra-operative images. (a) Warped pre-operative image. (b) Canny edges of the warped pre-operative image. (c) Intra-operative image. (d) Canny edges of the intra-operative image.

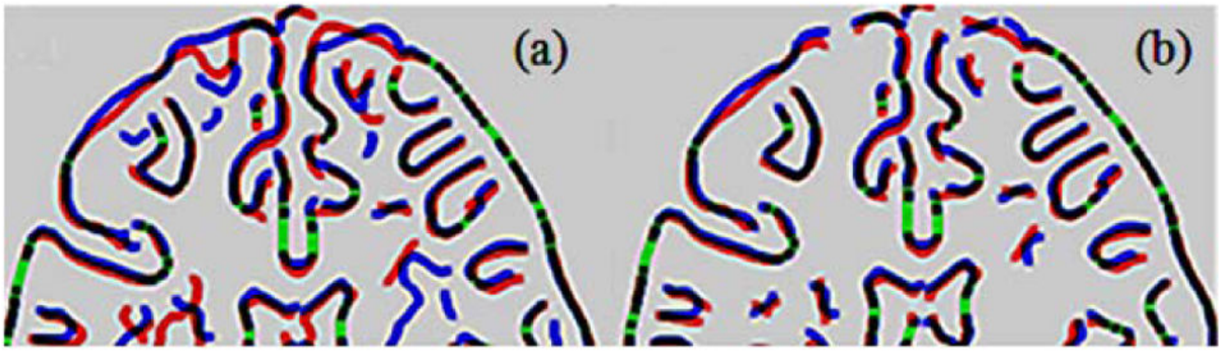


Fig. 3.

The overlaid Canny edges of the warped pre-operative image and the corresponding intra-operative image before (a) and after (b) pre-processing for an example case (Case 1). The green coloured portion represents overlapping edges, the blue coloured part identifies non-overlapping edges of the warped pre-operative image, and the red coloured portion identifies non-overlapping edges of the intra-operative image.

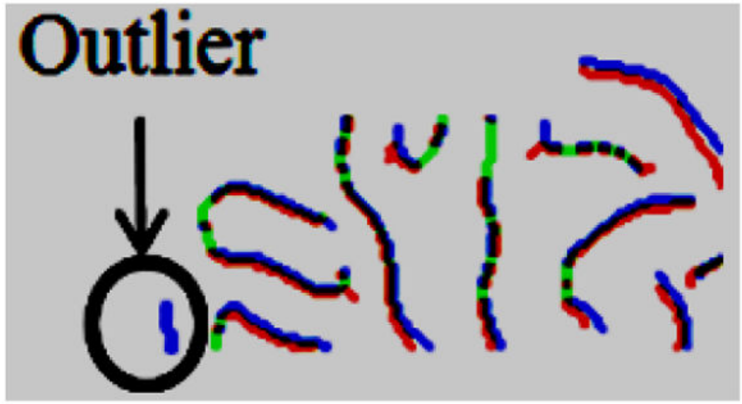


Fig. 4.
Illustration of an example outlier (Case 1), that occurred during alignment error assessment using our edge-based HD metric. The Canny edges of the images being compared are overlaid on each other. The green coloured portion represents overlapping edges, the blue coloured part identifies non-overlapping edges of the first image, and the red coloured part identifies non-overlapping edges of the second image (used as ground truth).

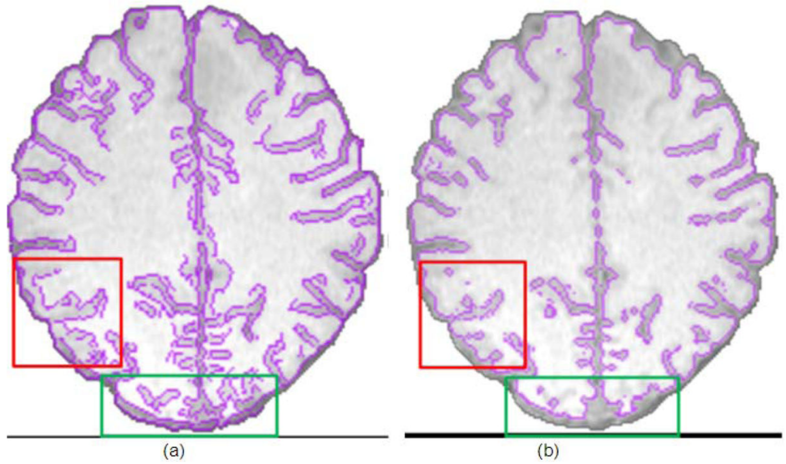


Fig. 5.
Comparison between boundaries (contours) detected using brain tissue segmentation and the Canny edges (Case 1). (a) The edges extracted by Canny edge detection method are disrupted, or the boundaries belonging to white or grey matters falsely connected to each other (red rectangle). Further, the boundaries of the skull residues might be detected as the tissue boundaries (green rectangle). (b) The grey matter boundaries detected by the tissue segmentation are more accurate with fewer disruptions (red rectangle). In addition, the skull residues are removed so that the soft tissue boundary can be extracted easily (green rectangle).

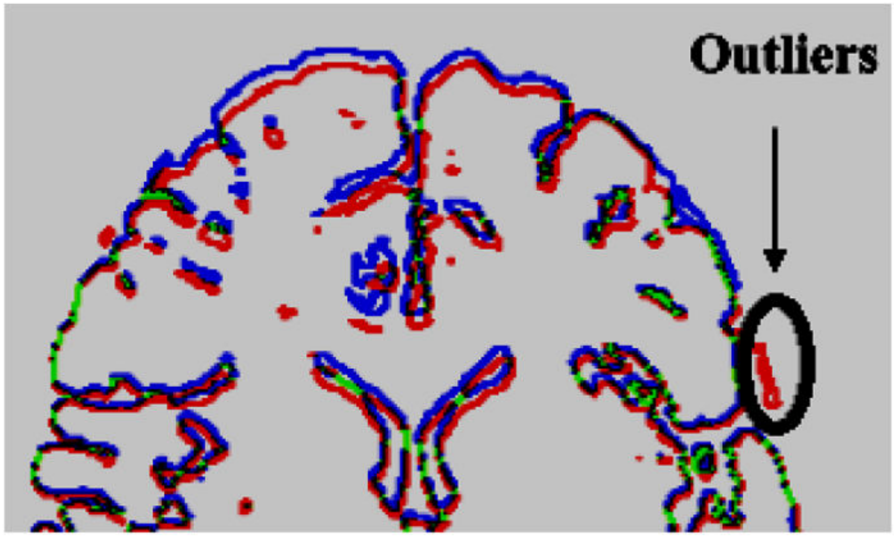


Fig. 6.
Illustration of outliers (Case 1), that occurred during alignment error assessment using the RHD metric. The contours of the images being compared are overlaid on each other. The green coloured portion represents overlapping contours, the blue coloured part identifies non-overlapping contours of the first image, and the red coloured part identifies non-overlapping contours of the second image.

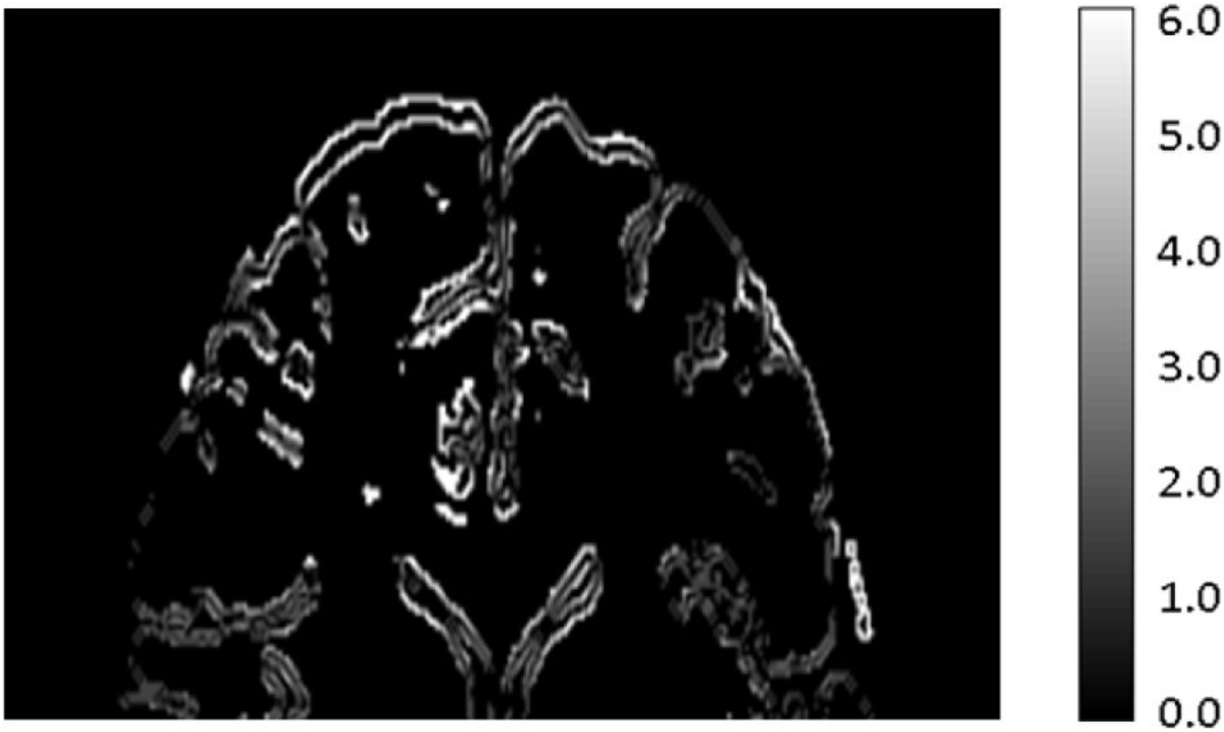


Fig. 7.
The alignment errors corresponding to pixels associated with the overlaid contours computed using the RHD metric. The brightness of non-zero pixels in the image is proportional to the computed RHD value.

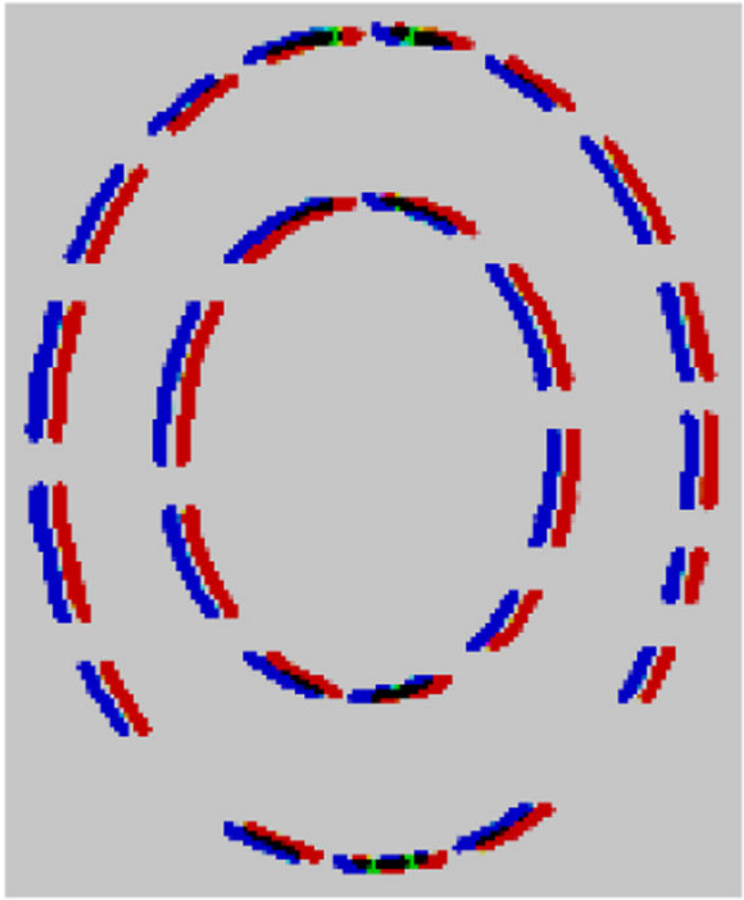


Fig. 8.

Verification of our EBHD metric using simple brain-like object images. The translated and original brain-like object images overlaid on each other. The green coloured portion represents overlapping edges, the blue coloured part identifies non-overlapping edges of the original image, and the red coloured part identifies non-overlapping edges of the translated image.

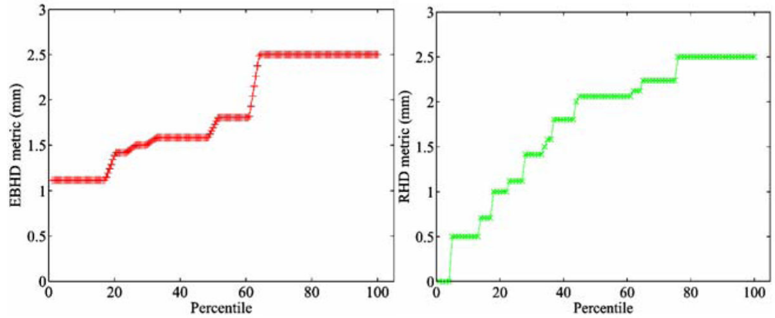


Fig. 9.
Percentile-EBHD and Percentile-RHD metric plots corresponding to rigid translation by 2.5 mm, for the simple brain-like object.

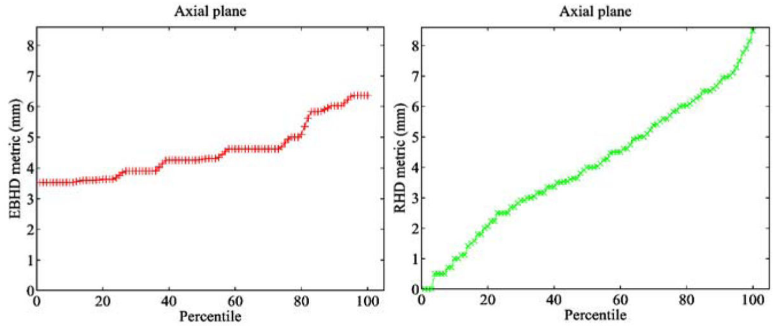


Fig. 10.

Percentile-EBHD and Percentile-RHD metric plots corresponding to rigid translation by 7 mm, for the ellipsoid in the axial plane.

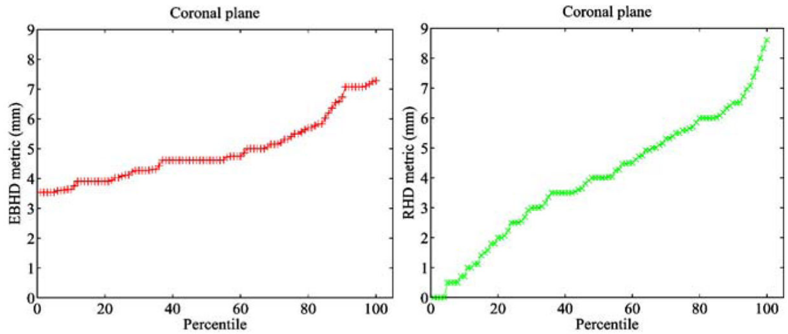
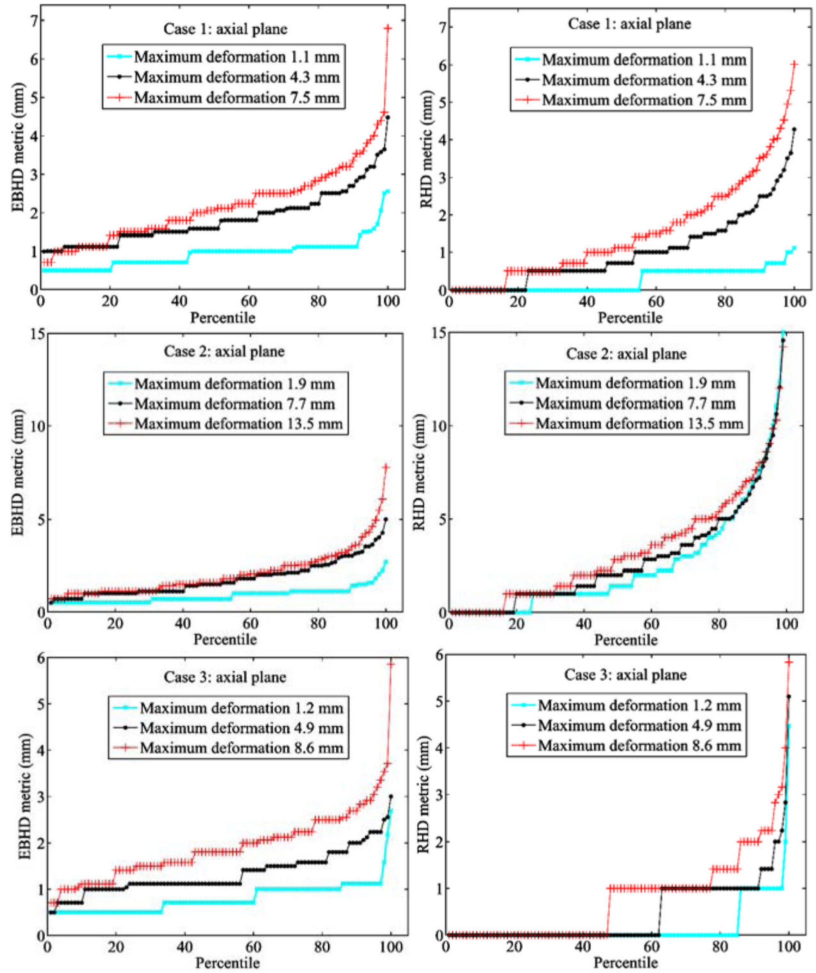


Fig. 11.

Percentile-EBHD and Percentile-RHD metric plots corresponding to rigid translation by 7 mm, for the ellipsoid in the coronal plane.


Case 1: axial plane
Maximum deformation 1.1 mm
Maximum deformation 4.3 mm
Maximum deformation 7.5 mm
EBHD metric (mm)
RHD metric (mm)
Percentile
Case 2: axial plane
Maximum deformation 1.9 mm
Maximum deformation 7.7 mm
Maximum deformation 13.5 mm
Case 3: axial plane
Maximum deformation 1.2 mm
Maximum deformation 4.9 mm
Maximum deformation 8.6 mm

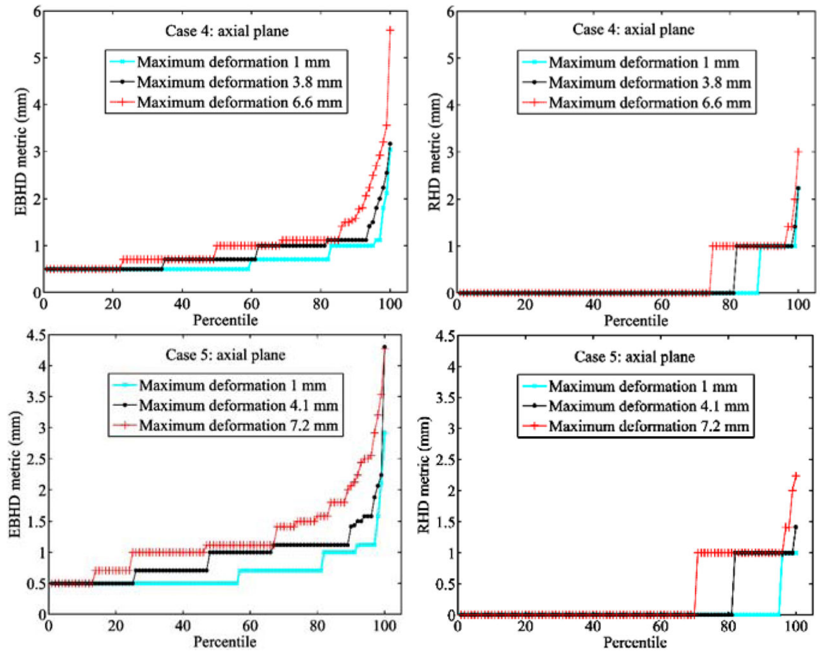
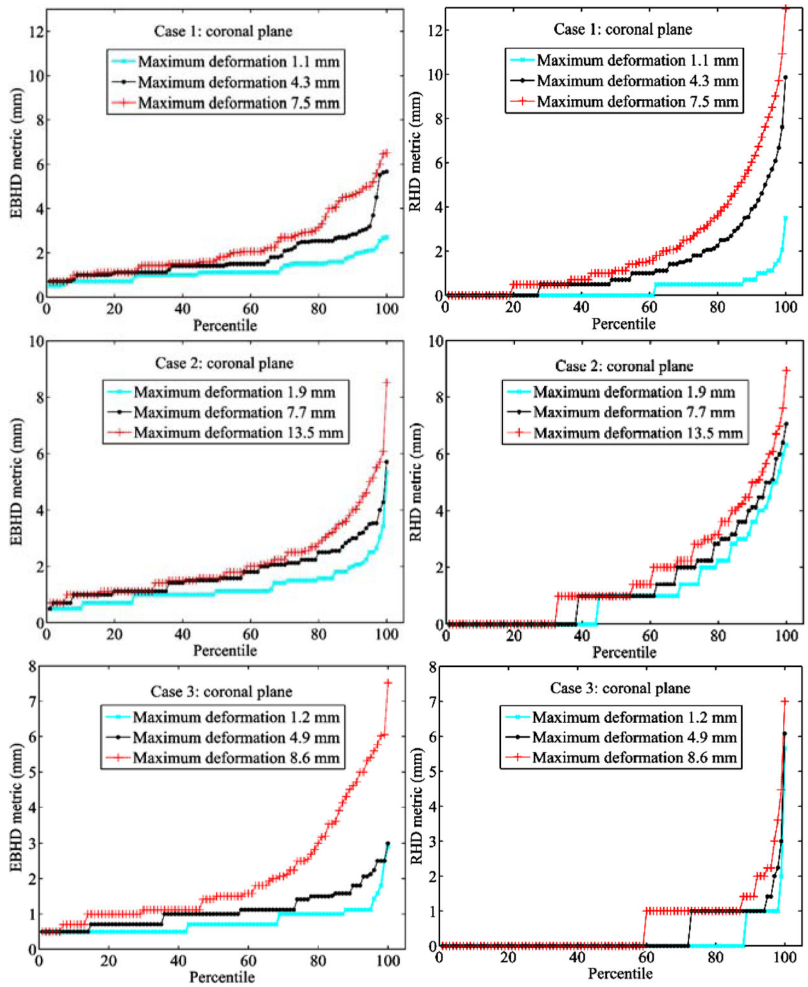


Fig. 12.

Percentile-EBHD metric and percentile-RHD metric plots corresponding to axial plane, for five real brain MR images.



Case 1: coronal plane
Maximum deformation 1.1 mm
Maximum deformation 4.3 mm
Maximum deformation 7.5 mm
EBHD metric (mm)
RHD metric (mm)
Percentile
Case 2: coronal plane
Maximum deformation 1.9 mm
Maximum deformation 7.7 mm
Maximum deformation 13.5 mm
Case 3: coronal plane
Maximum deformation 1.2 mm
Maximum deformation 4.9 mm
Maximum deformation 8.6 mm

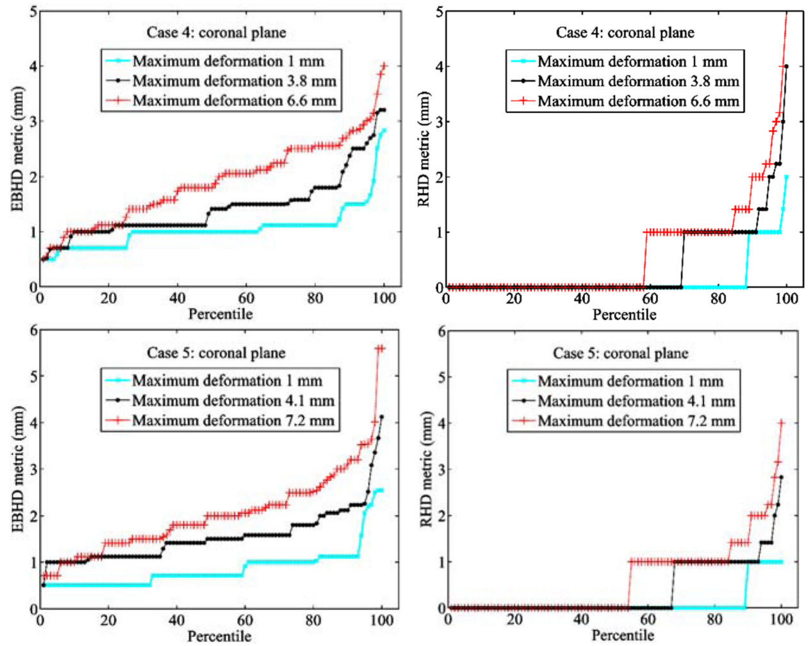


Fig. 13.

Percentile-EBHD and percentile-RHD metric plots corresponding to coronal plane, for five real brain MR images.

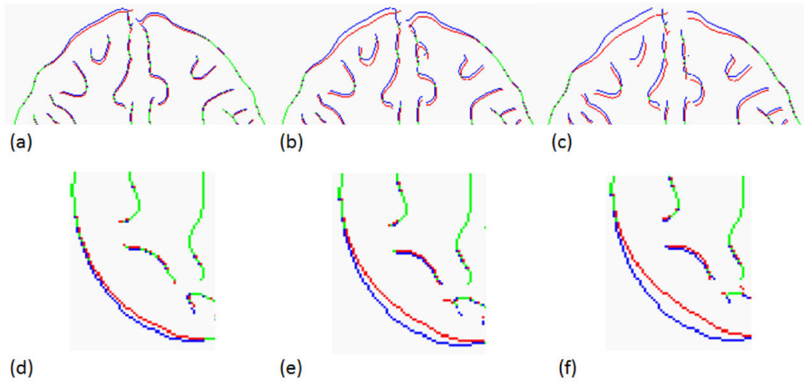


Fig. 14.

Overlaid Canny edges of warped pre-operative images and the corresponding pre-operative images in the axial and coronal planes for various deformation fields. The maximum deformations associated with the images in the axial plane, i.e., (a), (b) and (c) are 1.1 mm, 4.3 mm and 7.5mm respectively. The maximum deformations associated with the images in the coronal plane, i.e., (d), (e) and (f) are 1.1 mm, 4.3 mm and 7.5mm respectively. The green coloured portion represents overlapping edges, the blue coloured part identifies non-overlapping edges of the warped pre-operative image, and the red coloured portion identifies non-overlapping edges of the pre-operative image.

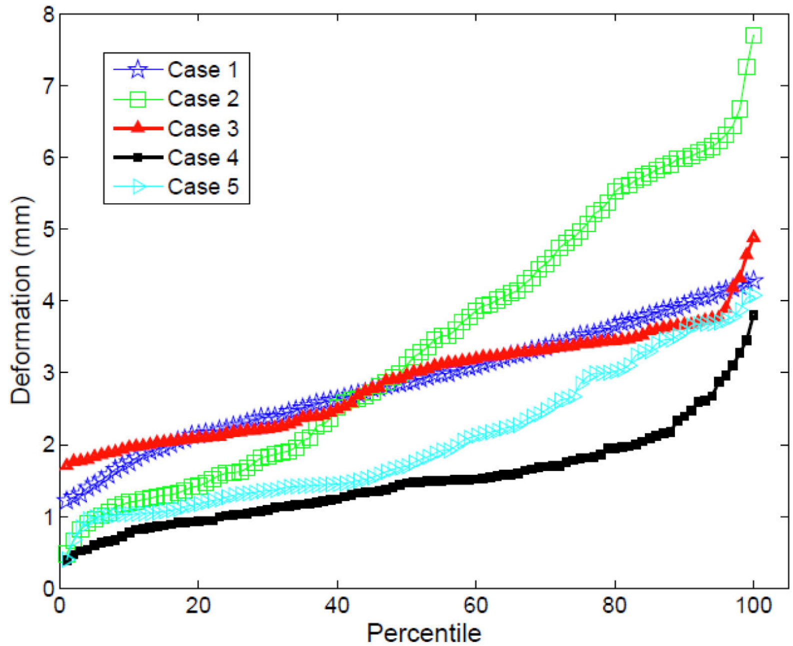


Fig. 15.
The deformation ranges for Cases 1–5 in the craniotomy region.

Table 1

Intensity-based similarity metric results for real brain images.

	Maximum deformation (mm)	MI	NMI	ECC	KLD	CR
Case 1	**1.1**	1.22	1.82	0.90	0.76	0.997
	4.3	1.12	1.71	0.83	0.85	0.968
	7.5	1.08	1.68	0.81	0.87	0.930
Case 2	**1.9**	0.68	1.82	0.90	0.53	0.996
	7.7	0.61	1.69	0.82	0.58	0.994
	13.5	0.59	1.65	0.78	0.61	0.970
Case 3	**1.2**	1.28	1.82	0.90	0.94	0.999
	4.9	1.15	1.68	0.81	1.07	0.994
	8.6	1.09	1.63	0.77	1.13	0.986
Case 4	**1.0**	1.68	1.83	0.91	1.03	0.999
	3.8	1.58	1.74	0.85	1.13	0.998
	6.6	1.52	1.69	0.82	1.19	0.996
Case 5	**1.0**	1.48	1.91	0.95	0.80	0.999
	4.1	1.37	1.78	0.88	0.91	0.997
	7.2	1.31	1.72	0.84	0.97	0.993