

Voxel Similarity Measures for 3D Serial MR Brain Image Registration

Mark Holden¹, Derek L. G. Hill¹, Erika R. E. Denton²,
Jo M. Jarosz², Tim C. S. Cox³, and David J. Hawkes¹

¹ Radiological Sciences and Biomedical Engineering, The Guy's, King's and St
Thomas' School of Medicine, King's College London, London SE1 9RT, U.K.

Derek.Hill@kcl.ac.uk

² Radiology and Neuroimaging Depts, King's College Hospital, London SE5 9RS

³ Institute of Neurology, UCL, Queen's Square, London WC1N 3BG

Abstract. We investigated 7 different similarity measures for rigid body registration of serial MR brain scans. To assess their accuracy we used a set of 33 clinical 3D serial MR images, manually segmented by a radiologist to remove deformable extra-dural tissue, and also simulated brain model data. For each measure we determined the consistency of registration transformations for both sets of segmented and unsegmented data. The difference images produced by registration with and without segmentation were visually inspected by two radiologists in a blinded study. We have shown that of the measures tested, those based on joint entropy produced the best consistency and seemed least sensitive to the presence of extra-dural tissue. For this data the difference in accuracy of these joint entropy measures, with or without brain segmentation, was within the threshold of visually detectable change in the difference images.

1 Introduction

In this paper, we report the results of a systematic comparison of seven similarity measures for serial MR registration. We assess the accuracy of the measures using simulated MR brain images [2], and quantify consistency using images from a clinical study [3]. We compare the performance of the measures on the clinical data with, and without segmentation of extra-dural tissue. We interpret these results in the context of a blinded visual assessment study.

2 Methods

Our clinical data is from five growth hormone deficient adults undergoing therapy and six normal subjects [3]. Each subject was scanned 3 times at 3 monthly intervals. An additional normal subject was scanned twice on the same day, for assessing observer sensitivity to synthetic misregistration. All images were axial T1 weighted 3D spoiled gradient echo with 1x1x1.8mm voxels, including head and brain stem. A phantom was scanned to measure scaling errors [3, 7]. The clinical images were manually scalp segmented by a radiologist to eliminate deformable extra-dural tissue, using Analyze (Mayo Clinic, Rochester, MN, US).

Simulated MR Brain Image with Added Noise and Distortion Simulated data was based on the McGill full anatomical MR brain model image [2]. Two noiseless images were used, 1 with 40% RF inhomogeneity intensity distortion and 1 without. Noise in a modulus MR image is Rician distributed [4]. To simulate Rician noise a numerical complex random variable was added to each voxel of the noiseless (real) image and then the modulus was taken to produce a magnitude image. The random variable was constructed from 2 Gaussian distributed ones for the real and imaginary parts. The simulated Rician noise was parameterised by measuring the mean and standard deviation of intensities of an artefact free region of a clinical scan corresponding to air [6].

Similarity Measures and Registration Algorithm The ideal similarity measure would have one optimum at the point of registration. Viola states that for images that differ only by Gaussian noise, the χ^2 measure is optimal; with a linear intensity transformation the Pearson product moment measure is optimal, and where the intensity transformation is unknown joint entropy is likely to be optimal [11]. Two important properties of serial MR images that effect similarity measures are: intensity distortion (due to RF inhomogeneity and motion artefact) and deformation of extra-cranial tissue (approximately 20% of typical brain scans). We have implemented 3 measures used by other researchers in serial MR: (1) mean squared difference in intensities (chi) χ^2 [5]; (2) Pearson product-moment cross correlation (ncc) [8]; (3) ratio image uniformity (riu) [12]. We have also implemented 4 measures proposed for other medical image matching applications: (4) mutual information (mi) [9]; (5) normalised mutual information (nmi) [10]; (6) entropy of the difference image (edi) [1]; (7) pattern intensity, radius 1, $\sigma = 10$, (pi) [13]. The measures can be put into two groups: (a) those based on entropy: mi, nmi, edi and (b) those based on correlation: chi, ncc, pi, riu. Our algorithm optimises the measures using a multi-resolution strategy similar to Studholme [10].

Consistency of Two Transformations For two rigid-body transformations \mathbf{T}_1 and \mathbf{T}_2 in homogeneous form, $\mathbf{T}_2\mathbf{T}_1$ is the result of first applying \mathbf{T}_1 then \mathbf{T}_2 . Given two transformation estimates \mathbf{T}_a and \mathbf{T}_b , mapping points, $p(i)$, from image 1 to image 2, the difference between these transformations is the mean voxel displacement in the brain region I_0 , (containing N_0 voxels): $\langle dp \rangle = \frac{1}{N_0} \sum_{\forall i \in I_0} | \Delta(p(i)) |$. The RMS analogue is: $dp_{rms} = \frac{1}{N_0} \sqrt{(\sum_{\forall i \in I_0} | \Delta(p(i)) |^2)}$.

Consistency of 3 Transformations For N images there are $P(N, 2) = \frac{N!}{(N-2)!}$ possible transformations. So for 3 images there are 6 different transformations between image pairs. If we consider 3 transformations $\mathbf{T}_{12}, \mathbf{T}_{23}, \mathbf{T}_{31}$ between image pairs (\mathbf{T}_{12} transforms image 1 into image 2) then in the absence of error, $\mathbf{T}_{31}\mathbf{T}_{23}\mathbf{T}_{12}$ is the identity \mathbf{I} . Registration solutions, inevitably, have some error so: $\mathbf{T}_{31}\mathbf{T}_{23}\mathbf{T}_{12} = \mathbf{I} + \Delta\mathbf{T}$, i.e. $\Delta\mathbf{T} = \mathbf{I} - \mathbf{T}_{31}\mathbf{T}_{23}\mathbf{T}_{12}$ is the error (internal in-

consistency). Applying the error transformation to each voxel location, $p(i)$, and taking the modulus, the mean error over the image is: $\frac{1}{N_0} \sum_{v_i \in I_0} | \Delta \mathbf{T}(p(i)) |$.

Registration of Clinical Data and Measurement of Consistency All registration was rigid body (6 degrees of freedom) and 5 resolution levels. The search interval ranged from 4mm or degrees to 0.01mm or degrees. For all 11 subjects, the first image (baseline) was registered to the second, the second to the third and the third to the first, giving 33 transformations for unsegmented images and 33 for images where the target was segmented. A set of 66 registrations was performed with each similarity measure. The consistency of 33 transformation estimates obtained without segmentation and 33 with segmentation was calculated. The triangular (internal) consistency for 11 measurements with segmentation and 11 without were also determined. Each consistency measurement was expressed as the mean, RMS, and maximum brain voxel shift (μm).

Assessment of Difference Images from Clinical Data Three sets of difference images, derived from different groups of subjects, were used during assessment: the first was used to train radiologists, the second to test their abilities at detecting misregistration, and the third for assessment of misregistration differences between data registered with or without prior segmentation. For training, difference images were created with varying amounts of misregistration [3, 7]. For testing radiologist's ability to detect misregistration the two consecutive scans of the normal subject were used to eliminate the possibility of any anatomical change in subject or scanner calibration. The second image was registered to the first by maximising normalised mutual information [10] and transformed into the coordinate frame of the first by sinc interpolation (radius 6). The first image was then subtracted from the aligned second one to produce a difference image which corresponded to no added misregistration. Ten increasing amounts of misregistration were added synthetically by calculating successively scaled down versions of the original 6D transformation (corresponding to mean voxel shifts 50 – 500 μm in 50 μm steps). Difference images for the clinical study were produced by registering the second and third images to the first by maximising normalised mutual information (as above). For each subject the second and third images were then transformed into the coordinate frame of the first (as above) and the first image was subtracted to produce two difference images (2 – 1 and 3 – 1) from registration with segmented data and two from registration with unsegmented data. Radiologists were trained to recognise different amounts of misregistration using the training set. Then they rated the misregistration of each randomised difference image on a 7 point scale.

Registration of Simulated Data We used the noiseless brain model image and the noiseless brain model with 40% RF inhomogeneity from McGill University [2] to create four image pairs: (a) 2 identical images; (b) 2 images with added noise; (c) 1 noiseless image with RF inhomogeneity and one without; (d) 2 images with added noise 1 with, 1 without RF inhomogeneity.

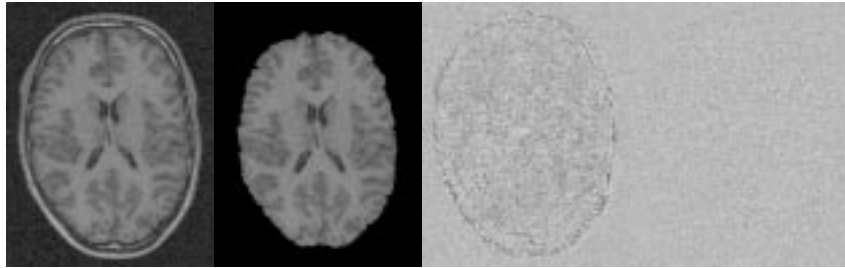


Fig. 1. Axial planes through clinical images: non-segmented (left), segmented, difference, McGill with simulated Rician noise (right)

3 Results

Three sets of results are given: (1) Consistency measurements for transformation estimates from registration of clinical data (segmented and not segmented) for the 7 measures. (2) Scores from radiologists' visual assessment of clinical difference images (segmented and not segmented). (3) Consistency measurements of the transformation estimates obtained from registration of the simulated data with the 7 measures. All consistency measurements correspond to the mean, RMS, and maximum voxel displacements over the segmented brain region and are given in μm , rounded to the nearest μm .

Registration Consistency for the 7 Similarity Measures The mean / standard deviation of 33 measurements of the mean voxel shift (μm) for registration solutions with and without segmentation of clinical data were: 122/46 (mi), 121/48 (nmi), 164/74 (ncc), 175/76 (chi), 8429/5316 (edi), 700/1503 (pi), 880/609 (riu). The smallest mean/standard deviation of 33 measurements of the maximum voxel shift were 223/96 (mi), 222/96 (nmi). Table 1 shows averaged consistency measurements for $\mathbf{T}_{31}\mathbf{T}_{23}\mathbf{T}_{12}$ with each of the 7 measures, for segmented and unsegmented data.

Visual Assessment of Difference Images Assessed misregistration was correlated with the added misregistration. For observer A the Spearman rank correlation coefficient (ρ) was 0.96 for observer B, ρ was 0.79. Inter-observer agreement was also tested and ρ was 0.85. These results suggest that radiologists are sensitive to misregistration in difference images corresponding to a mean, RMS and maximum voxel shift, over the brain, of: 195, 199, 299 microns respectively. There was no perceived difference in perceived misregistration with and without segmentation using the nmi measure ($p=0.35$).

Registration Consistency with Simulated Images Registration accuracy was measured by comparing the transformation estimate with the identity using

6 different starting estimates. The mean/standard deviation of voxel shift for the 6 starting transformations for the four image pairs was (microns): (a) less than 10/3 for all measures; (b) 127/2 (chi), 135/3 (edi), 121/2 (mi), 126/2 (ncc), 137/3 (nmi), 343/23 (pi), 416/74 (riu); (c) 8/2 (chi); 39/14 (edi); 51/1 (mi); 25/5 (ncc); 43/6 (nmi); 12/1 (pi); 30/7 (riu). There was a failure for riu which was omitted; (d) 203/3 (chi); 253/2 (edi); 163/5 (mi); 133/2 (ncc); 194/7 (nmi); 391/26 (pi); 402/57 (riu).

Table 1. Mean (standard deviation) of 11 consistency measurements ($\mathbf{T}_{31}\mathbf{T}_{23}\mathbf{T}_{12}$). Registration without prior segmentation (left) and with prior segmentation (right)

measure	unsegmented data			segmented data		
	mean	RMS	max	mean	RMS	max
chi	91 (28)	96 (31)	165 (55)	99 (31)	104 (33)	169 (62)
edi	1757 (1148)	1780 (1198)	2160 (2072)	66 (35)	69 (36)	117 (55)
mi	88 (23)	94 (26)	168 (53)	78 (29)	82 (31)	139 (59)
ncc	87 (37)	93 (40)	168 (73)	97 (25)	101 (27)	168 (63)
nmi	86 (32)	92 (35)	162 (70)	78 (29)	81 (30)	133 (57)
pi	1565 (2204)	1690 (2391)	3348 (4809)	145 (94)	154 (100)	278 (181)
riu	1221 (553)	1276 (549)	2100 (748)	258 (92)	276 (101)	531 (224)

4 Conclusion and Discussion

Table 1 shows that 4 of the 7 measures produced transformation estimates that were consistent to within 331 microns whether or not the data was pre-segmented and also had the best internal consistency ($\mathbf{T}_{31}\mathbf{T}_{23}\mathbf{T}_{12}$) for non-segmented data. For the joint entropy measures the mean of the maximum inconsistency between registrations with and without segmentation was 223 microns. The results from visual assessment of synthetically misregistered data indicated that the threshold for detecting misregistration corresponded to a mean and maximum inconsistency of about 200 and 300 microns respectively. These inconsistencies are larger than the averaged measured mean and maximum inconsistency suggesting that these inconsistencies are too small to be reliably detected by the visual inspection of difference images. For non-segmented data, there was little difference in the internal consistency of transformation estimates for those measures based on correlation (chi and ncc) and for those based on joint entropy (mi and nmi). However, registration results with and without prior segmentation were more self-consistent for those based on joint entropy. Results with the simulated images suggested that image noise had a significant effect on registration accuracy. However, the highest resolution matching was done with images at the original resolution without any filtering to reduce the impact of noise. It is possible that low pass filtering with intensity thresholding might improve performance of some measures. Our results show that the similarity measures based on mutual information are the most suitable for rigid body registration of serial MR images of the head. Using our optimisation strategy we achieve registration solutions with

and without extra-dural tissue segmentation that are consistent to within the threshold of observer discernibility (i.e. 200-300 μm). Our results apply under the conditions of typical scalp deformations and small scale anatomical change.

Acknowledgements

We thank Philips Medical Systems, EasyVision Advanced Development for funding, the McConnell Brain Imaging Centre, MNI, McGill University for use of the simulated brain data, and Dr. D. Russell-Jones, for the clinical data.

References

1. Buzug, T.M., Weese, J., Fassnacht, C., and Lorenz, C.: Using an Entropy Similarity Measure to Enhance the Quality of DSA Images with an Algorithm Based on Template Matching. VBC'96, SPIE Proc. **1808** (1996)
2. Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C.: Design and Construction of a Realistic Digital Brain Phantom. IEEE Transactions on Medical Imaging **17** (3) (1998) 463–468.
3. Denton, E.R.E., Holden, M., Jarosz, J.M., Studholme, C., Cox, T.C.S., Hill, D.L.G.: The identification of cerebral volume changes in treated growth hormone deficient patients using serial 3-D MR image processing. Radiology **209P** (SS) (1998) 1492
4. Gudbjartsson, H. and Patz, S.: The Rician Distribution of Noisy MRI data. Magnetic Resonance in Medicine **34** (1995) 910–914
5. Hajnal, J.V., Saeed, N., Soar, E.J., Oatridge, A., Young, I.R. and Bydder, G.M.: A Registration and Interpolation Procedure for Subvoxel Matching of Serially Acquired MR Images. JCAT **19** (2) (1995) 289–296
6. Henkelman, R.M.: Measurement of Signal Intensities in the Presence of Noise in MR Images. Medical Physics **12** (2) (1985) 232–233
7. Holden, M., Denton, E.R.E., Jarosz, J.M., Cox, T.C.S., Hawkes, D.J., Hill, D.L.G.: Detecting small anatomical change with 3D serial MR subtraction images. SPIE, Medical Imaging, **3661** (1999)
8. Lemieux, L, Wieshmann, U.C., et al.: The detection and significance of subtle changes in mixed-signal brain lesions by serial MRI scan matching and spatial normalisation. Medical Image Analysis **2** (3) (1998) 227–242
9. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration using mutual information. IEEE Transactions on Medical Imaging **16** (2) (1997) 187–198
10. Studholme, C, Hawkes, D.J, Hill, D.L.G: An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. Pattern Recognition **32** (1) (1999) 71–86
11. Viola, P.: Alignment by Maximization of Mutual Information. Ph.D. Thesis Massachusetts Institute of Technology (1995)
12. Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R., and Mazziotta, J. C.: Automated Image Registration: 1. General Methods and Intrasubject, Intramodality Validation. JCAT **22** (1) (1998) 139–152
13. Weese, J., Pennney, G.P., Desmedt, P., Buzug, T.M., Hill, D.L.G., and Hawkes D.J.: Voxel-Based 2-D/3-D Registration of Fluoroscopy Images and CT Scans for Image-Guided Surgery. IEEE Trans. on Info. Tech. in Biomedicine **1** (4) (1997) 284–293