ASSESSING THE ACCURACY OF NON-RIGID REGISTRATION WITH AND WITHOUT GROUND TRUTH

R. S. Schestowitz, C. J. Twining, T. F. Cootes, V. S. Petrović, and C. J. Taylor – *University of Manchester*
W. R. Crum – *University College London*

Overview

Non-rigid registration (NRR) is used increasingly routinely in medical image analysis. There are, however, many different approaches to NRR, each leading to different results. We wish to find an objective method for evaluating the quality of NRR, so that different approaches can be compared and the quality of the NRR step in real applications can be be assessed. The main contributions of the work are:

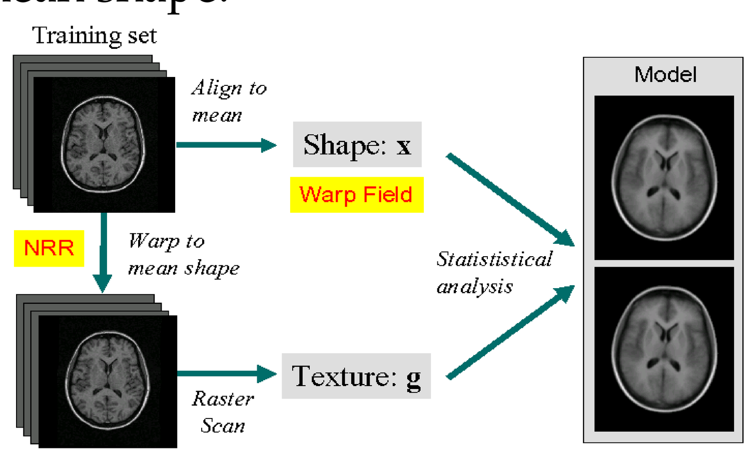
- objective evaluation of NRR;
- no ground-truth required;
- validation using perturbed NRR;
- comparison with ground-truth based approach;
- practical comparison of NRR algorithms.

Key Idea

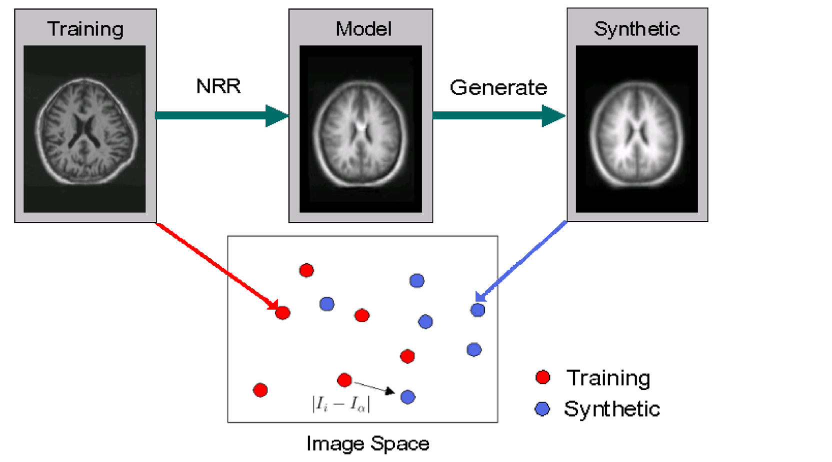
Our method exploits the fact that, given a set of non-rigidly registered images, a generative statistical model of appearance can be constructed. The quality of this model depends on the quality of the registration. We define a measure of model quality – *specificity* – that can be used to assess model, and thus registration, quality.

Method

Given a set of non-rigidly registered images, we can build a generative appearance model that captures both the shape and intensity variation across the set, by performing a joint statistical analysis of the warp fields x resulting from the NRR and the shape-free intensity patterns g measured in the frame of the mean shape.



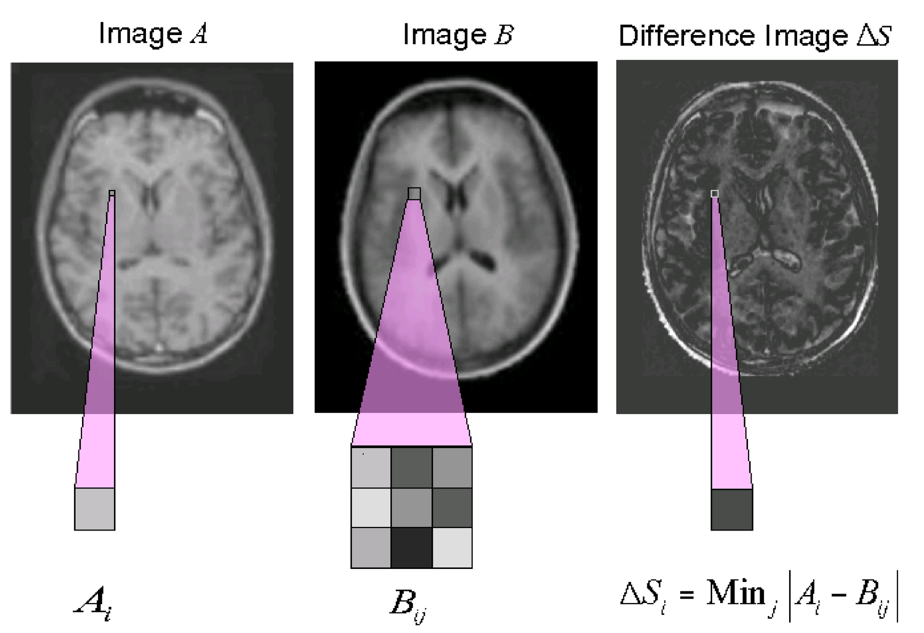
We can use the generative property of the model to synthesise a large set of images, $\{I_\alpha : \alpha = 1, \ldots m\}$ which, if the model is good, should form a cloud that overlaps the cloud of training images.



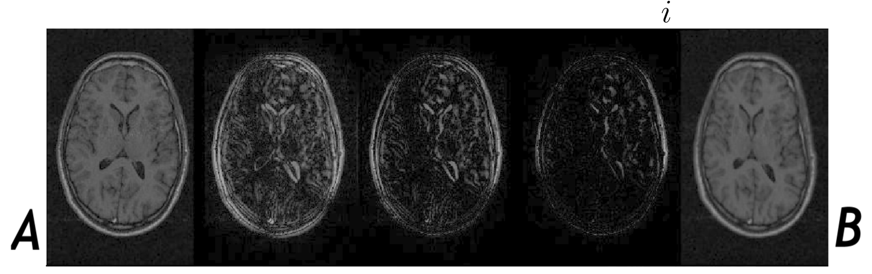
We define the Specificity S of the model as $S = \frac{1}{m}\sum_{\alpha=1}^{m} \min_i |I_\alpha - I_i|$ where $|\cdot|$ is a measure of distance between images, I_i is the i^{th} training image, and $\min_i$ is the minimum over i (the set of *training* images). S will be small if the two clouds fully overlap.

Measuring Inter-image Distance

We could simply take $|\cdot|$ as the Euclidean distance between images. This measure is, however, extremely sensitive to small misalignments, so we also investigated the use of shuffle difference images as defined in the diagram below.



The shuffle differences between two images are shown below, for different sizes of the shuffle neighbourhood B_{ij}. The shuffle distance between two images is $\sum_i \triangle S_i$.



Overlap-based Assessment

We have compared our approach with a 'gold standard' method of assessment, which uses a generalisation of Tanimoto's spatial overlap measure [1].

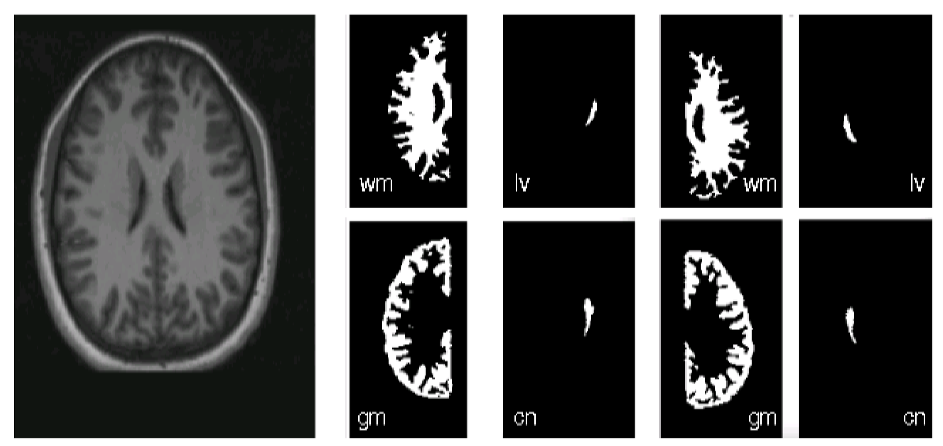
A manual mark-up of each image, is used to provide an anatomical/tissue label for each voxel. The overlap of corresponding labels following registration is defined as

$$\mathcal{O} = \frac{\sum_{\text{pairs},k} \sum_{\text{labels},l} \alpha_l \sum_{\text{voxels},i} MIN(A_{kli}, B_{kli})}{\sum_{\text{pairs},k} \sum_{\text{labels},l} \alpha_l \sum_{\text{voxels},i} MAX(A_{kli}, B_{kli})} \quad (1)$$

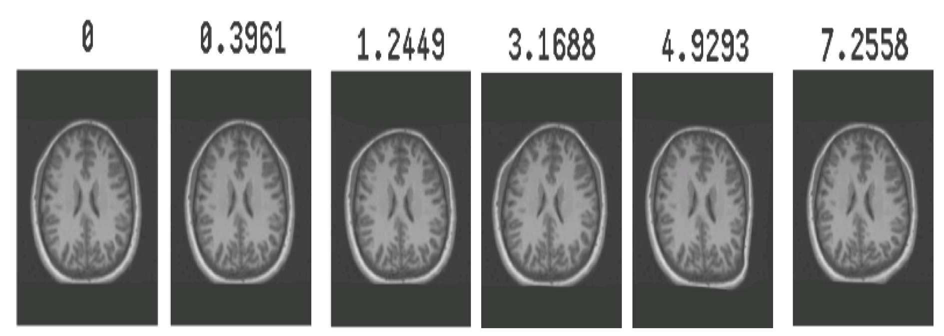
where i indexes voxels in the registered images, l indexes the label and k indexes image pairs and α_l is a label weight. A_{kli} and B_{kli} represent voxel label values in a pair of registered images and are in the range [0, 1]. The $MIN()$ and $MAX()$ operators are standard results for the intersection and union of fuzzy sets.

Validation Experiments

The overlap-based and model-based approaches were validated and compared, using a dataset consisting of 36 transaxial mid-brain slices, extracted at equivalent levels from a set of T1-weighted 3D MR scans of different subjects. Eight manually annotated anatomical labels were used as the basis for the overlap method: L/R white matter, L/R grey matter, L/R lateral ventricle, and L/R caudate, as shown below.

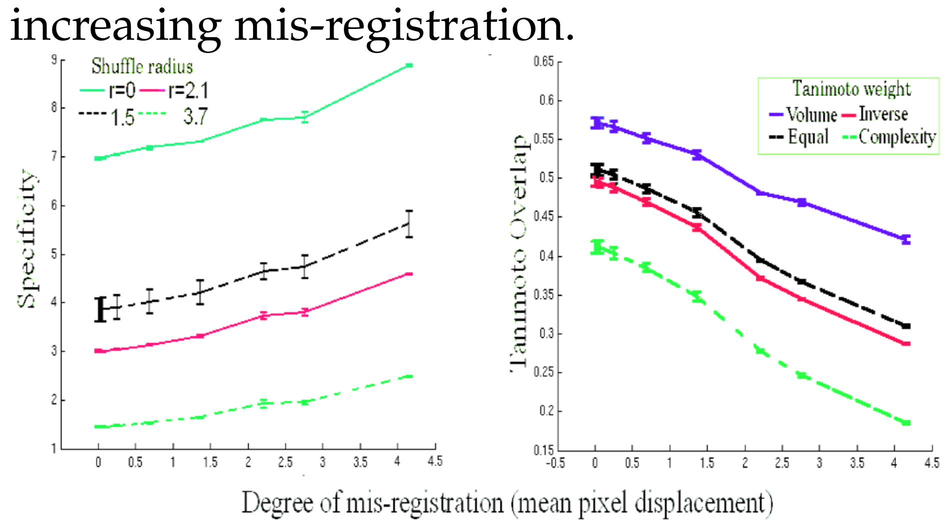


The images were initially registered using an NRR algorithm based on MDL optimisation [2]. The images were then systematically misregistered by applying smooth pseudo-random spatial warps to the registered images. Ten different warp instantiations were generated for each image at each of seven progressively increasing values of average pixel displacement. Registration quality was measured, for each level of registration degradation, using several variants of each of the proposed assessment methods.



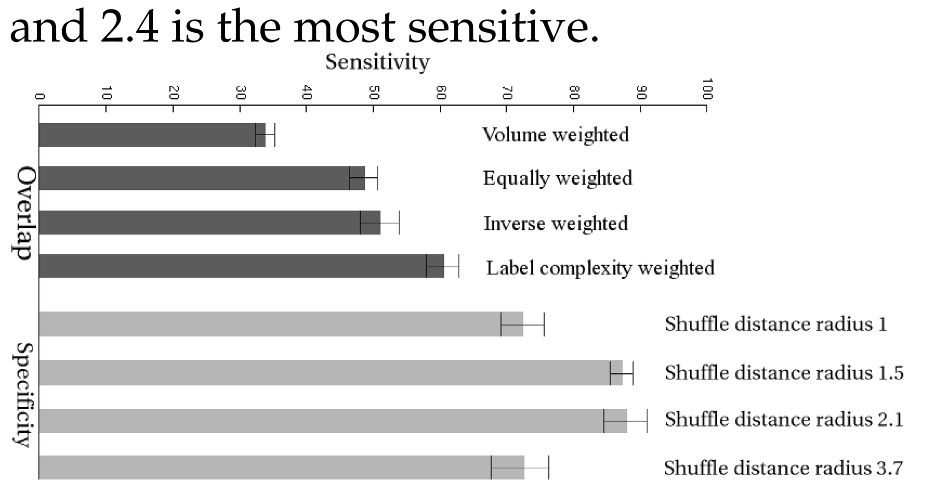
Increasing perturbation in mean pixel shift

The results below show that all variants of both the overlap and model-based quality measures change systematically with increasing mis-registration.



Sensitivity

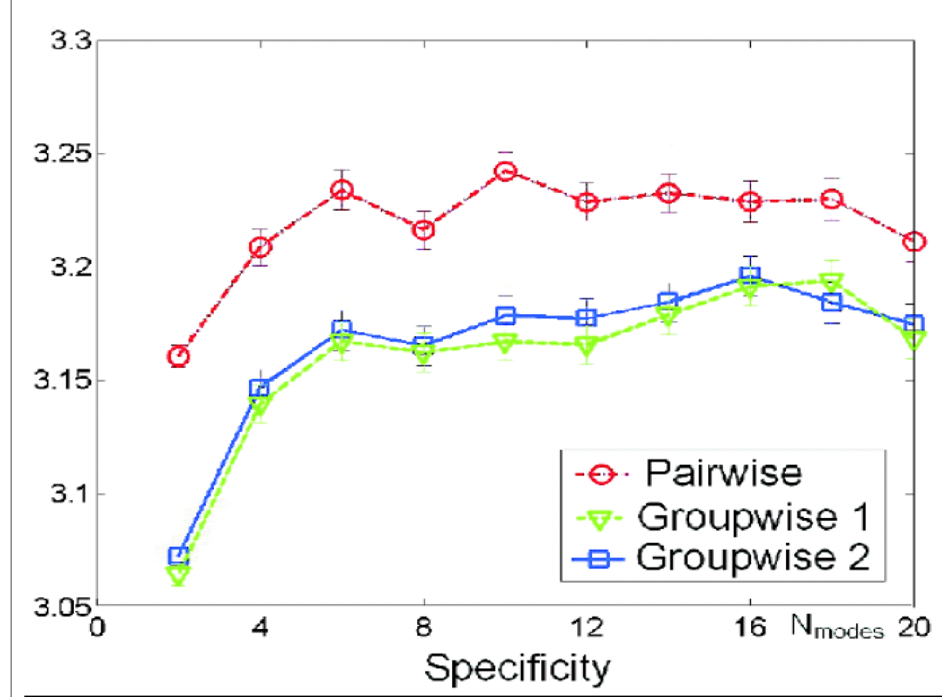
To compare the different quality measures in more detail, we define sensitivity as the mean slope of a measure over the range investigated, divided by the mean uncertainty in the measure – that is the smallest misregistration that can just be detected. The sensitivities of the different methods evaluated are plotted below, demonstrating that the model-based approach with a shuffle radius between 1.5 and 2.4 is the most sensitive.



Practical Application

We present results for the comparison of three methods of NRR of the set of original images described above, one based on pairwise registration to a reference image, the others two variants of a groupwise algorithm, based on minimising description length [2].

The plots below show that, whatever number of modes (active dimensions) are retained in the model, the two groupwise approaches are better than the pairwise approach, but were not distinguishable from each other.



Conclusions

- Overlap method provides 'gold standard'
- Model-based method a good surrogate
- Model-based method more sensitive
- Ground-truth-free evaluation a reality

Acknowledgements

Many thanks to David Kennedy of the Centre for Morphometric Analysis at MGH for segmented brain data. The work was supported by the EPSRC/MRC-funded Medical Image and Signal IRC (GR-/N14248/01), Integrated Brain Image Modelling (EPSRC GR-/S82503/01) and Modelling, Understanding and Predicting Structural Brain Change (EPSRC GR/S48844/01).

References

[1] W. R. Crum, O. Camara, D. Rueckert, K. Bhatia, M. Jenkinson, and D. L. G. Hill. Generalised overlap measures for assessment of pairwise and groupwise image registration and segmentation. In *Proceedings of MICCAI*, 3749:99-106, 2005.
[2] C. J. Twining, T.F. Cootes, S. Marsland, S. V. Petrovic, R. S. Schestowitz, and C. J. Taylor. A unified information-theoretic approach to groupwise non-rigid registration and model building. In *Information Processing in Medical Imaging*, 3565:1-14, 2005.