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

We introduce a method for assessing the performance of non-rigid registration (NRR). The method requires no ground truth. We compare it to another, ground-truth-dependent method and use it assess registration algorithms.

Motivation

The main aim is to gain ability to assess and compare different approaches to NRR.

Broad Contributions

- NRR without ground truth
- Method validation, based on ground truth
- Comparion with assessment which is based on ground truth
- Evaluation of various 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, so we define measures of model *specificity* and *generalisation* – based on comparing synthetic images sampled from the model, with those in the original image set – that can be used to assess model/registration quality.

Results

The method detects the loss of registration accuracy as the alignment of a set of correctly registered MR images of the brain is progressively perturbed. We compare the sensitivities of several approaches and show that, as well as requiring no ground truth, *specificity* provides the most sensitive measure of misregistration.

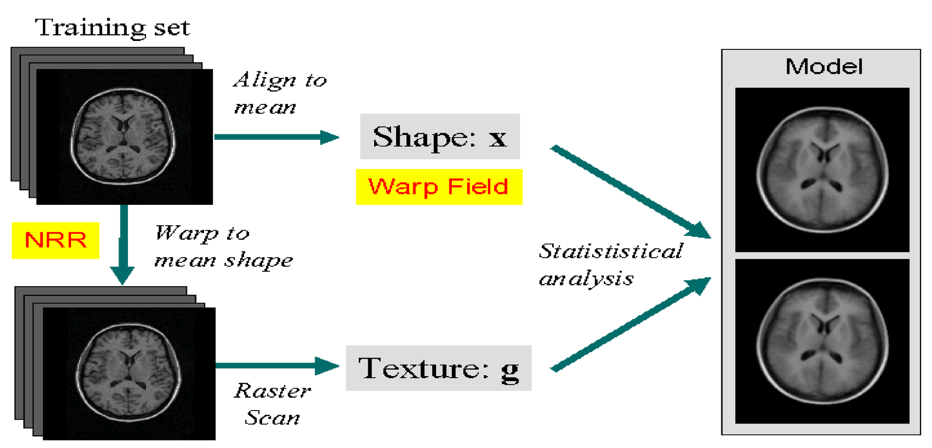
Lastly, we use the model-based assess method to demonstrate that groupwise registration is better than its pairwise counterpart.

Model-based Method

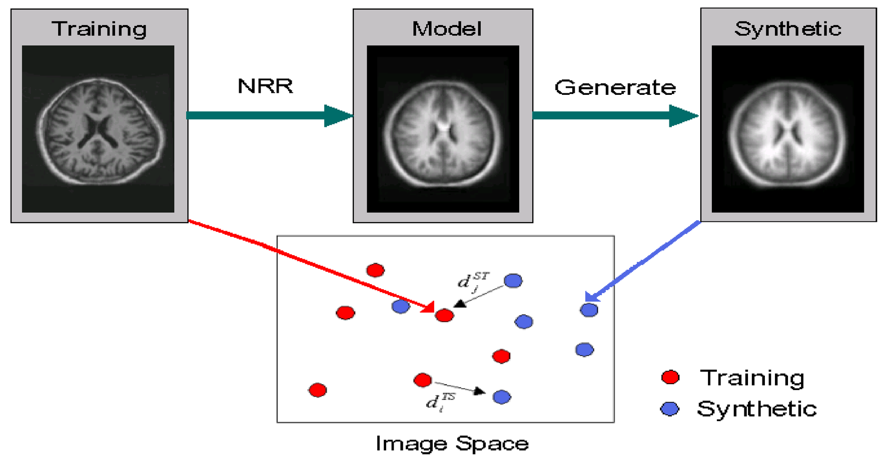
We automatically build an appearance model from the correspondences identified by a registration. We then synthesise images from the model and perform a cross-imageset comparison to assess model/registration quality.

Building Models

- Use image correspondences
- Construct an appearance model (AM)



Assessing Clouds Overlap



- Generate synthetic images from the AM
- Compare training images with synthetic images
  - images can be vectorised
  - embed images in a high-dimensional space
  - estimate clouds overlap

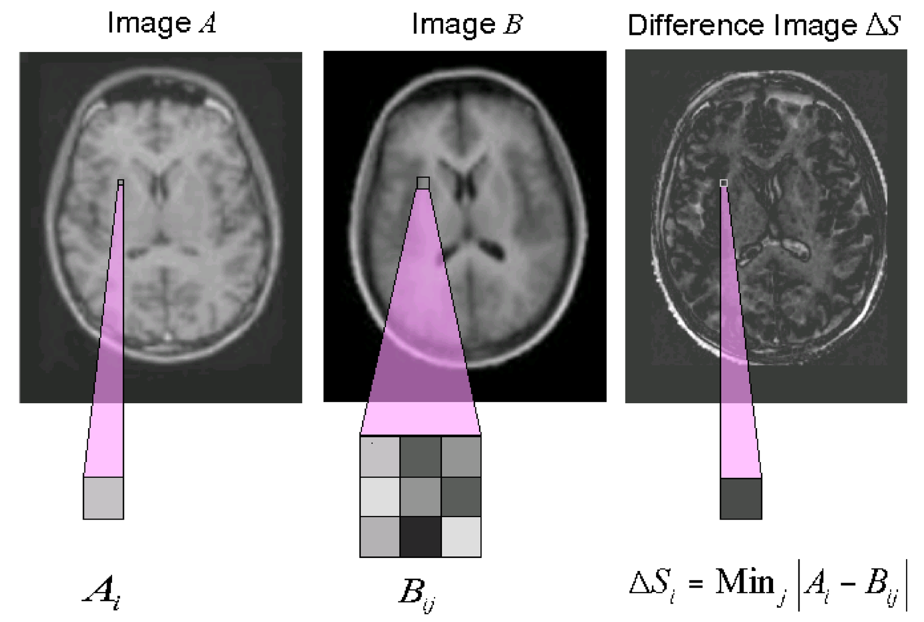
Measuring Inter-image distances

Euclidean Distance:

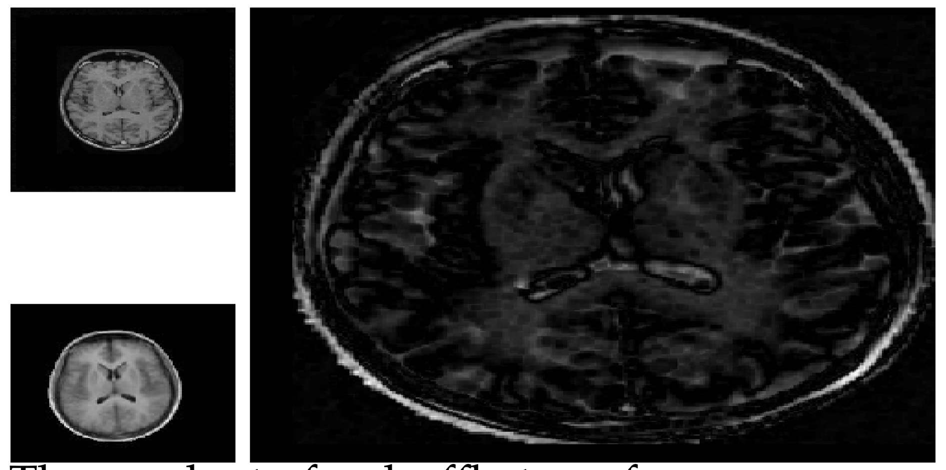
The simple method for comparing disagreement among two images is to take their per-pixel difference. This is sensitive to small misalignments though.

Shuffle Distance:

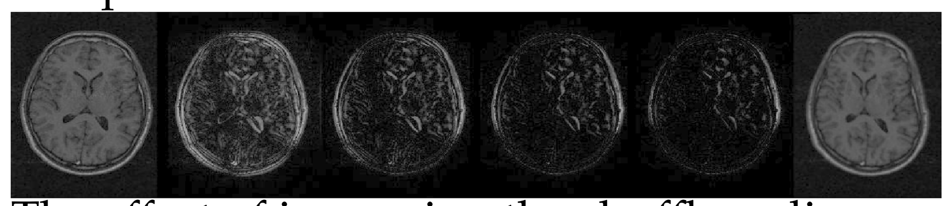
In our experiments we have defined image distance as the shuffle distance between two images, as illustrated below



The computation of the shuffle distance



The product of a shuffle transform



The effect of increasing the shuffle radius

Overlap-based Assessment

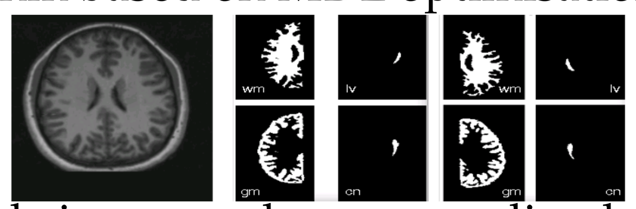
The approach utilizes a measure of overlap between ground-truth anatomical labels. It exploits a generalisation of Tanimoto's spatial overlap measure [1] . We start with a manual mark-up of each image, providing an anatomical/tissue label for each voxel, and measure the overlap of corresponding labels following registration.

- Each label is represented using a binary image
- Becomes fuzzy after interpolation
- Combined in a generalised overlap score
  1. implicitly volume weighted
  2. inverse label volume
  3. inverse label volume squared
  4. label complexity

Experiments

Validation:

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. The images were brought into alignment using an NRR algorithm based on MDL optimisation [2].



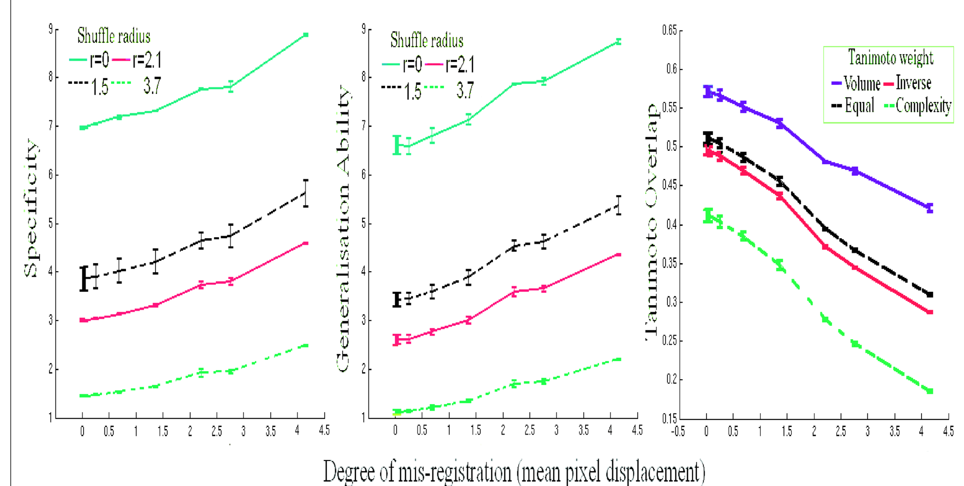
Example image and corresponding labels

Perturbation:

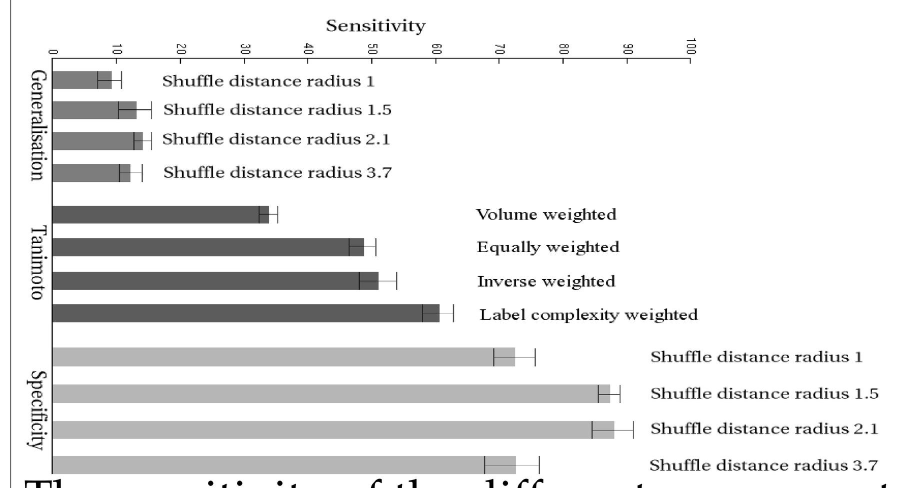
A set of different mis-registrations was then created by applying smooth pseudo-random spatial warps to the registered images. These warps were based on biharmonic Clamped Plate Splines. Ten different warp instantiations were generated for each image at each of seven progressively increasing values of average pixel displacement. Registration quality was measured, for each level of registration degradation, using several variants of each of the proposed assessment methods.

Examples of data with increased perturbation magnitude

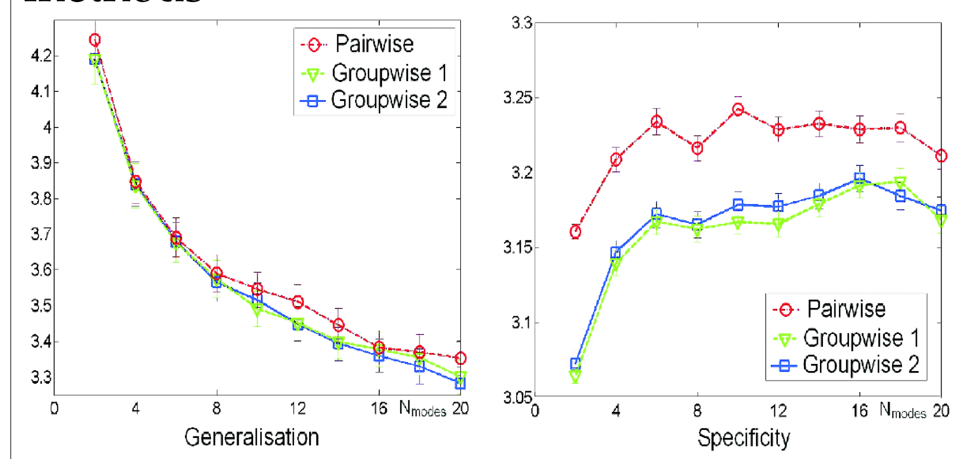
Results



The three non-rigid registration assessment methods, as function of misalignment



The sensitivity of the different assessment methods



Practical application: registration methods comparison

Conclusions

- Both approaches sensitive to subtle misregistration
- Overlap and model-based approaches 'equivalent'
- Overlap provides 'gold standard'
- Specificity is a good surrogate
  - monotonically related
  - no need for ground truth
  - more sensitive
  - only applies to groups (but any NRR method)

Acknowledgements:

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