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 compare two methods for assessing the performance of groupwise non-rigid registration (NRR) algorithms. The first approach utilizes a measure of overlap between ground-truth anatomical labels. The second, which is new, 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.

Both approaches detect 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 the two approaches and show that, as well as requiring no ground truth, *specificity* provides the most sensitive measure of misregistration.

The Two Approaches

Overlap-based:
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.
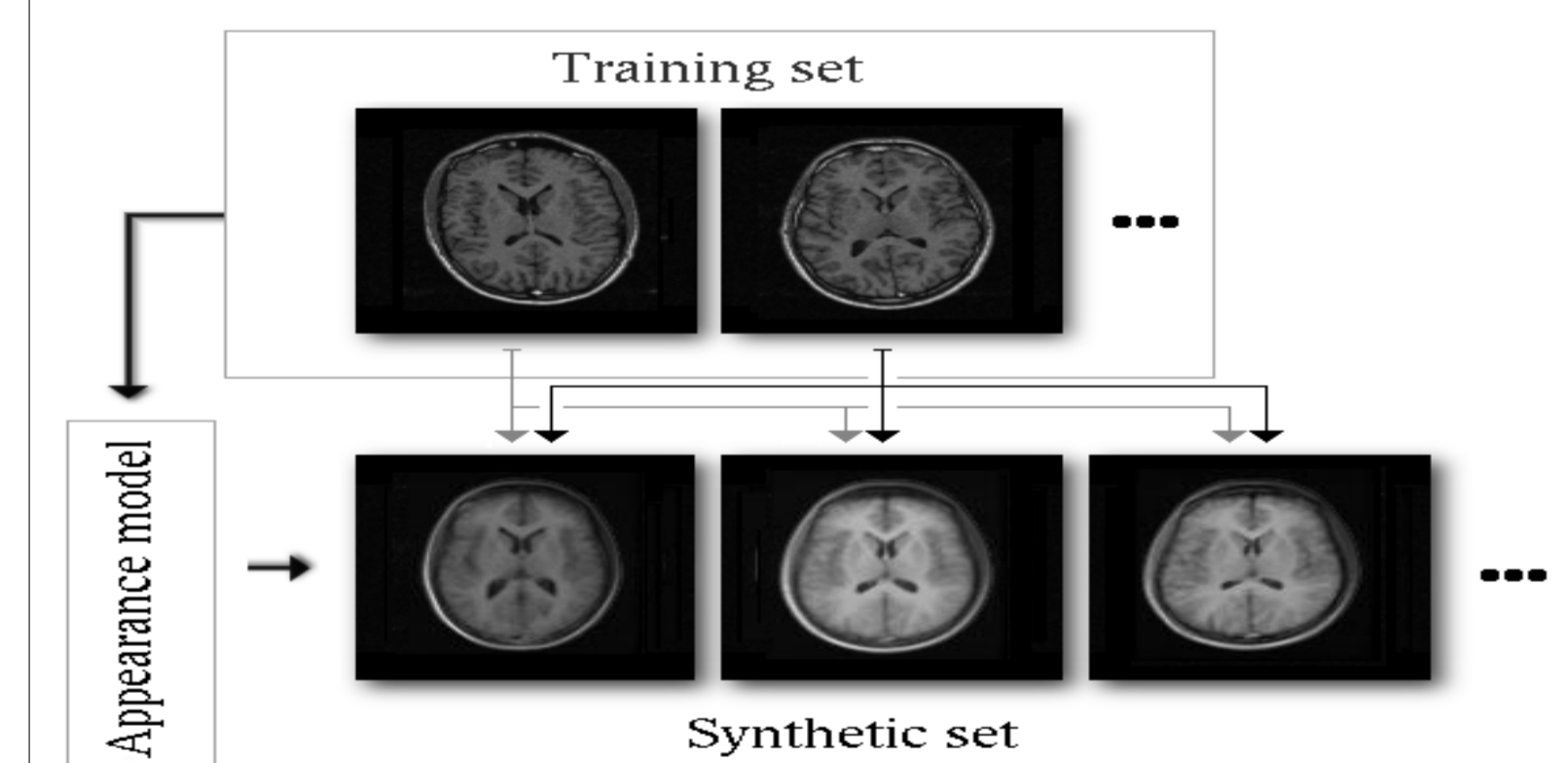
Model-based:
We automatically build an apperance model from the correspondences identified by a registration. We then synthesise images from the model and perform a cross-imageset comparison to assess model/registration quality.
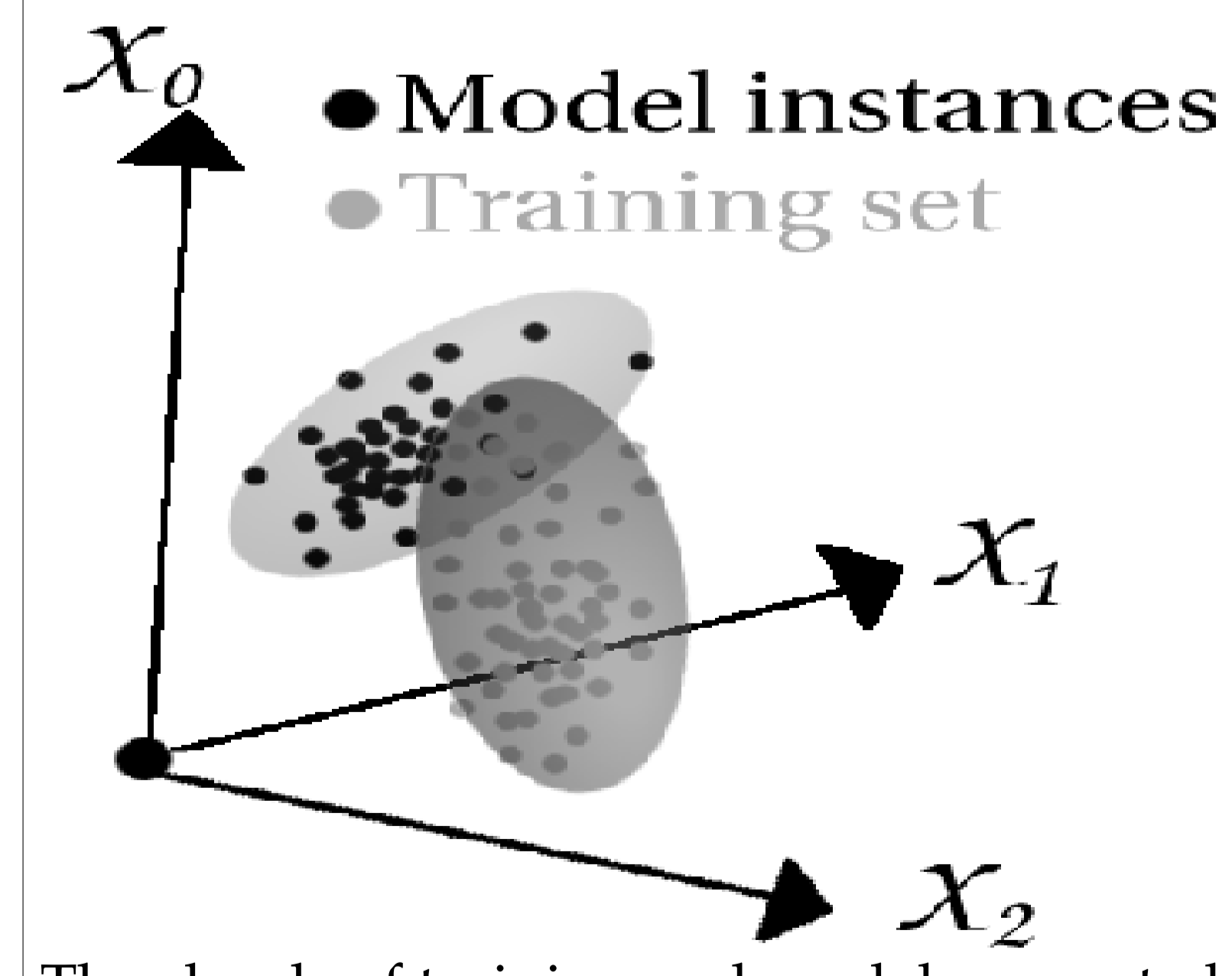
Overlap-based Assessment Algorithm

- 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

Model-based Assessment Method

- Use image correspondences
- Construct an appearance model (AM)
- Generate synthetic images from the AM
- Compare training images with sythetic images
    1. images can be vectorised
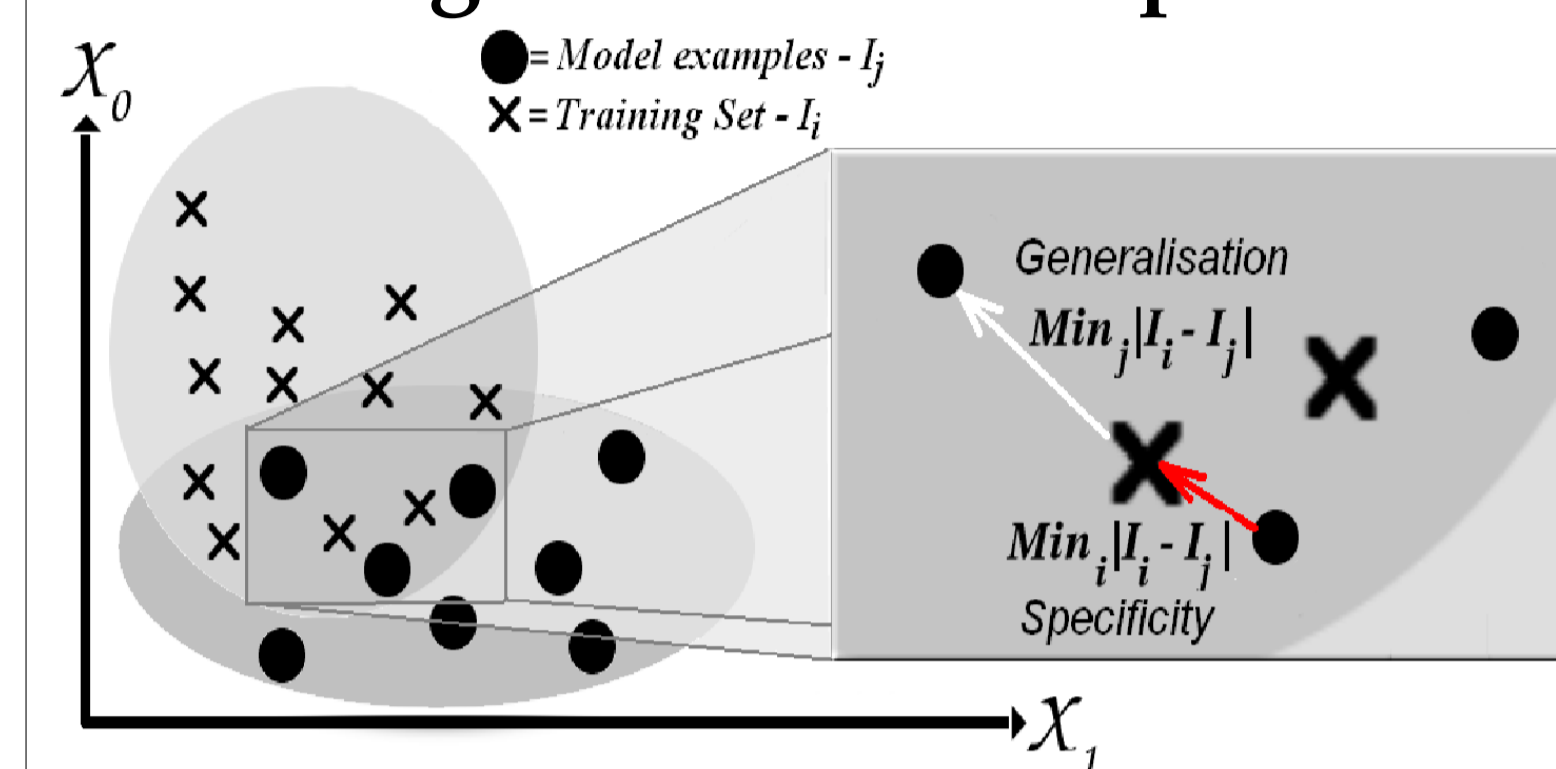   2. embed images in a high-dimensional space
    3. estimate clouds overlap

The model-based assessment framework.

The clouds of training and model-generated data estimated for their mutual overlap.

Assessing Clouds Overlap

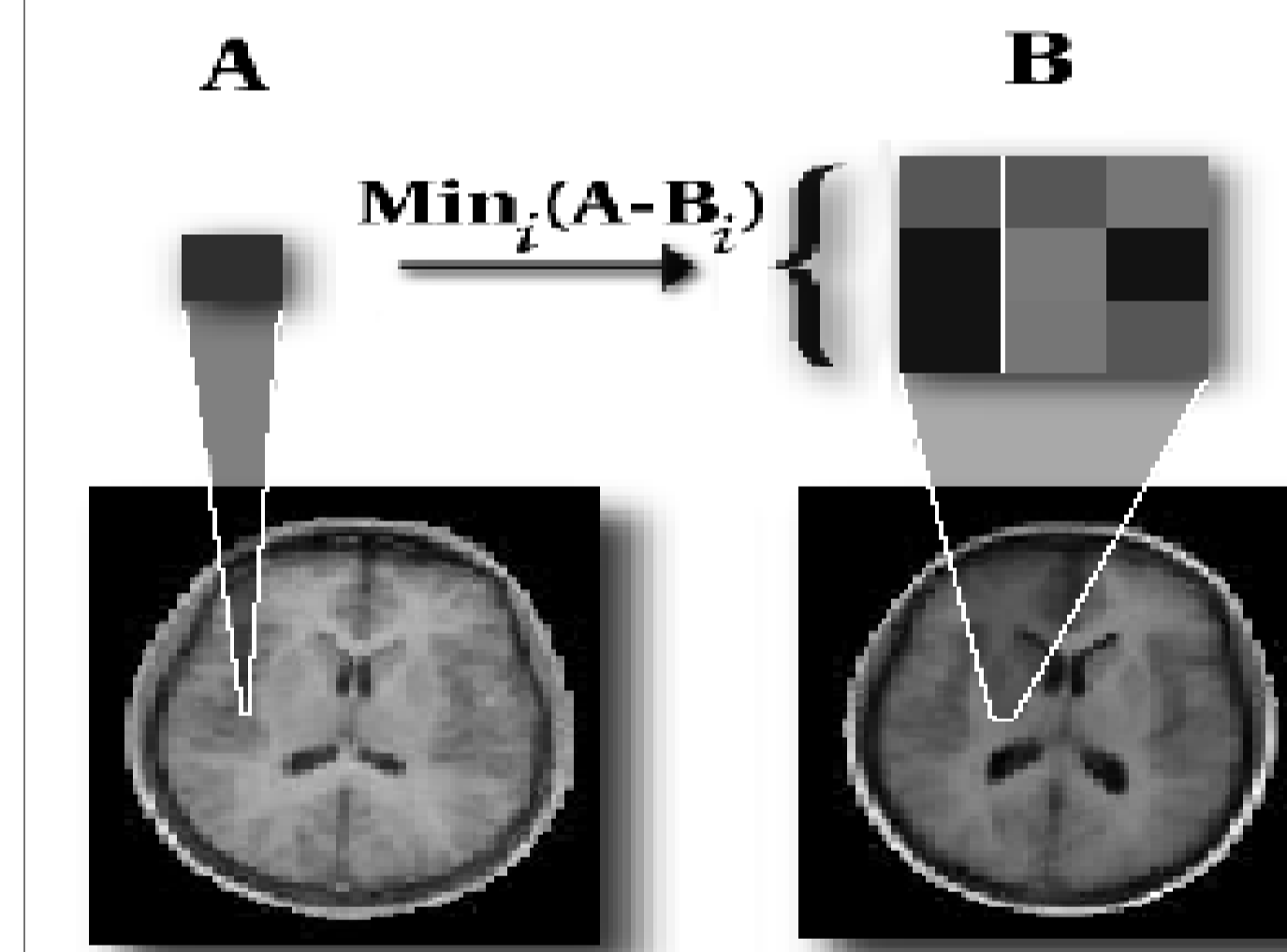

The calculation of clouds overlap.
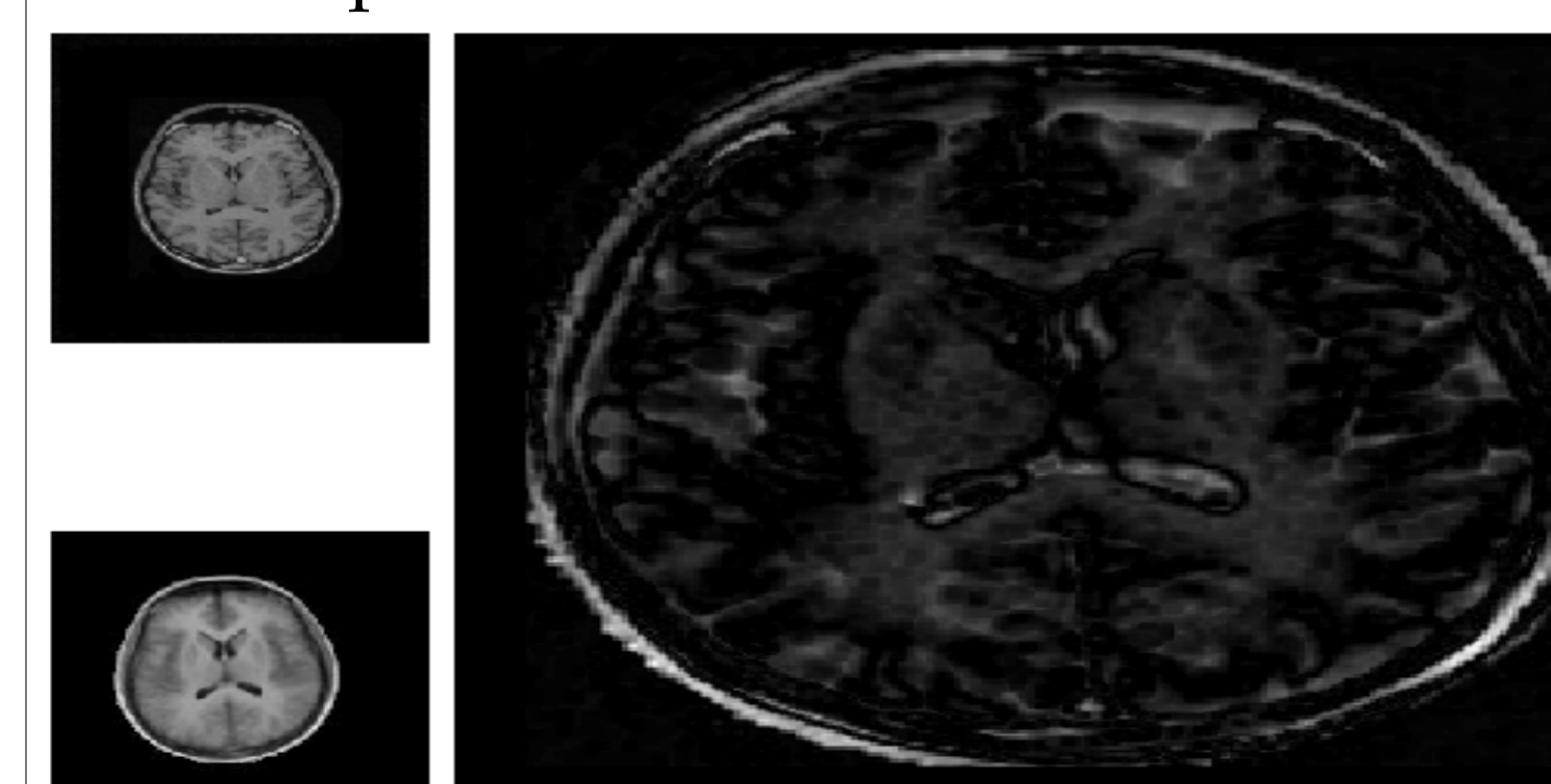
Measuring Inter-image distances

Euclidean Distance:
The simple method for comparing disgreement among two images is to take their per-pixel difference. This is sensitive to small mialignments though.
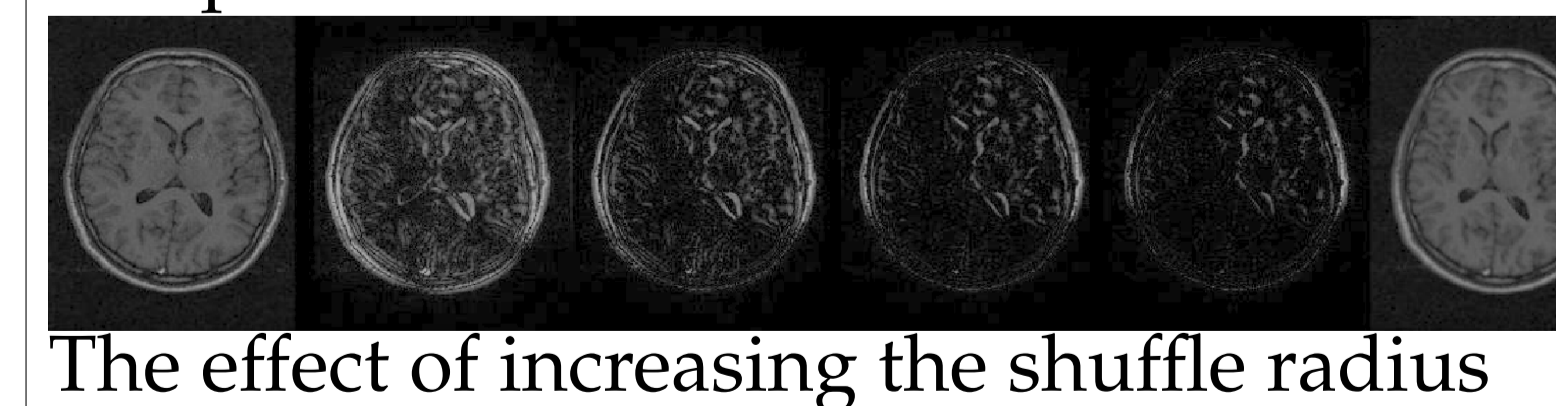
Shuffle Distance:
In our experiments we have defined image distance as the shuffle distance between two images, as illustrated below

$$Min_i(A\text{-}B_i)$$

The computation of the shuffle distance
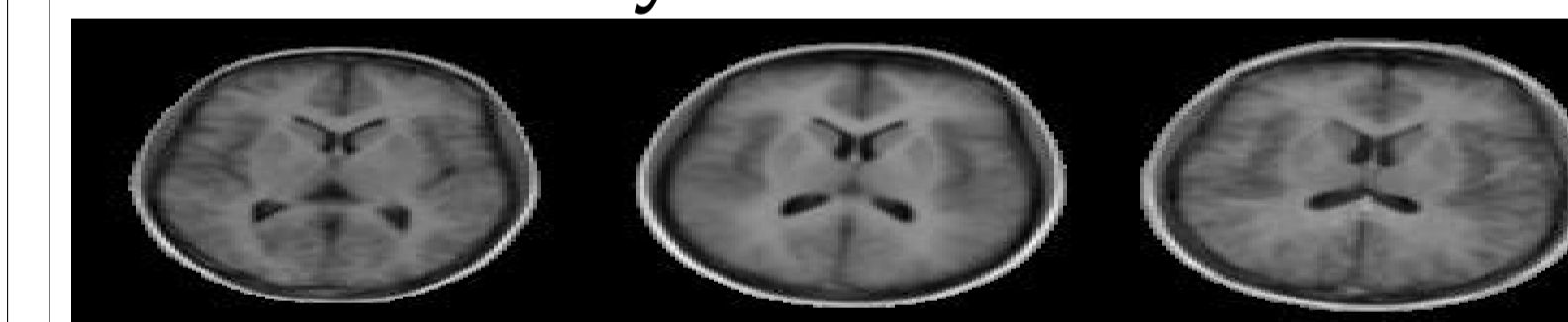

The product of a shuffle tranform

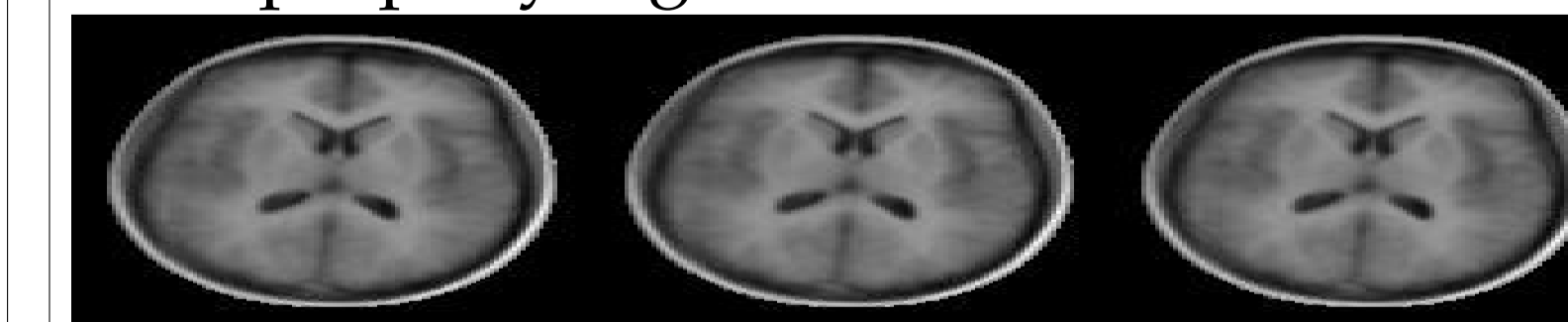The effect of increasing the shuffle radius

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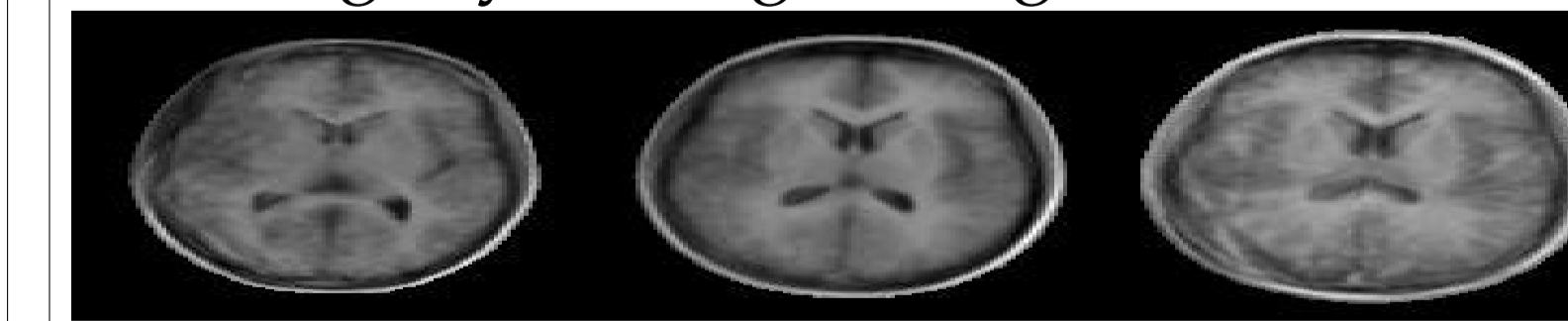
Automatically-Built Models

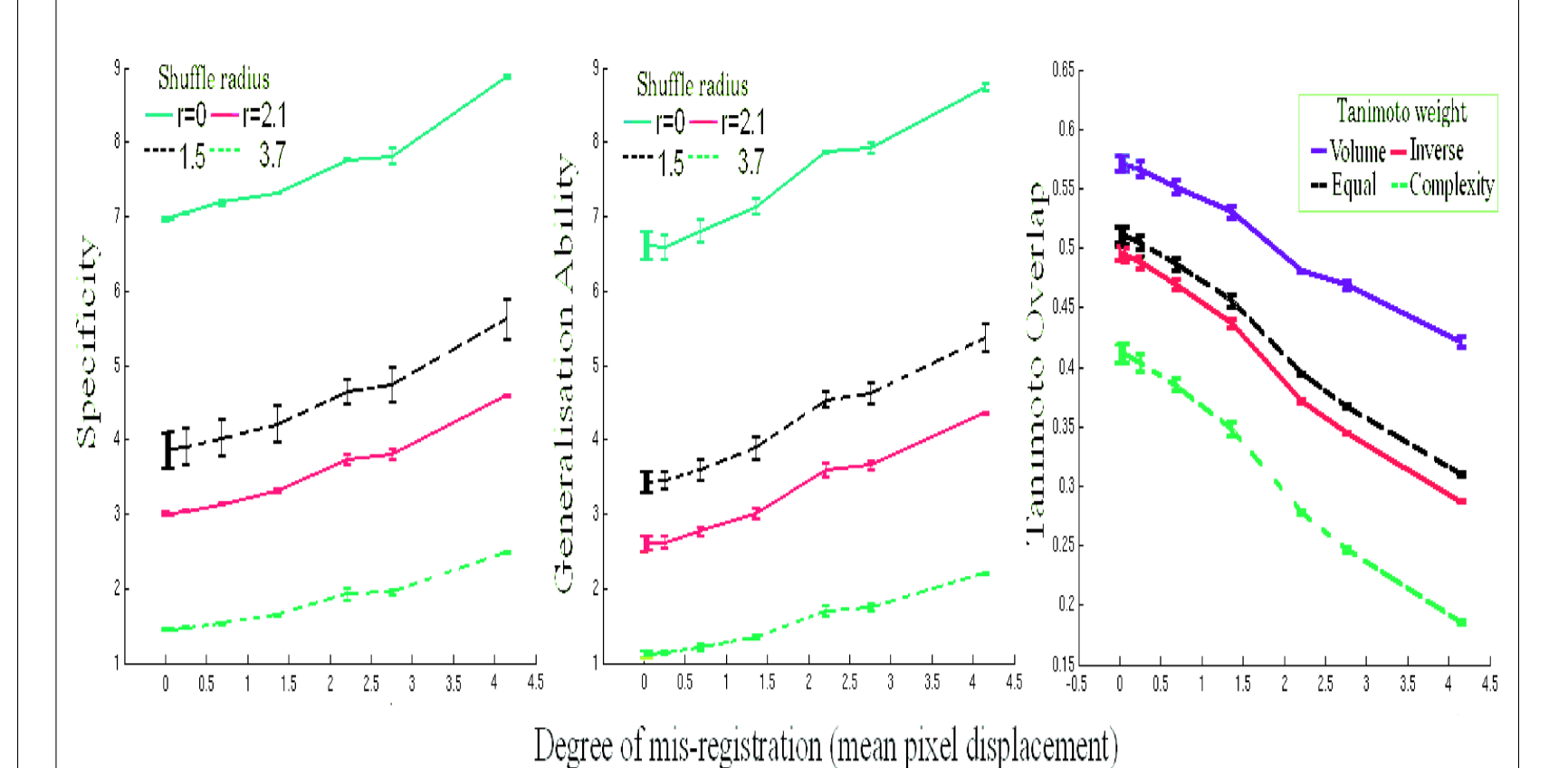First mode of an appearance model built from properly registered data

First mode of an appearance model built from slightly misaligned registered data
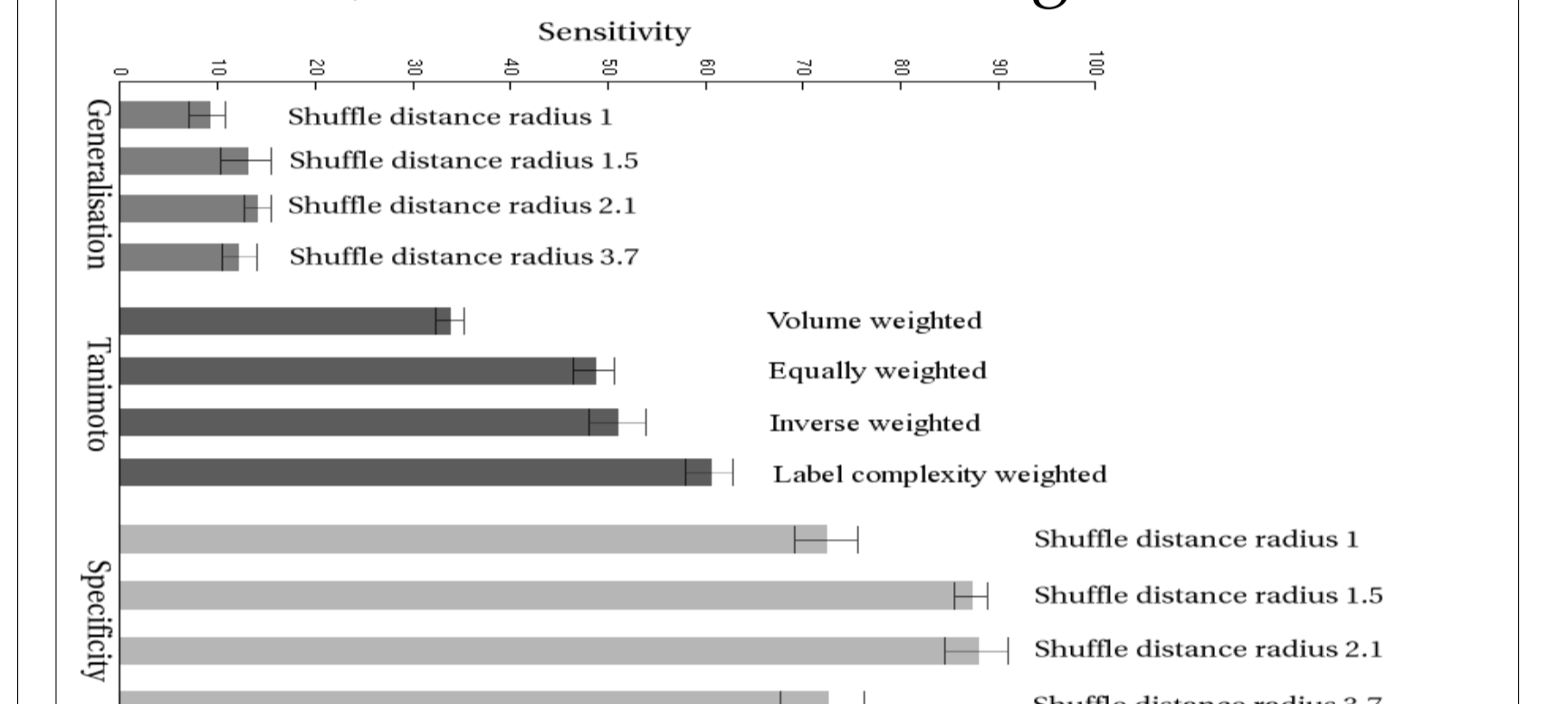

First mode of an appearance model built from highly misaligned data

Results

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

The sensitivity of the different assessment methods

Conclusions

Our results suggest that the performance of NNR algorithms can be compared objectively, using just the registered images they produce, and that the quality of registration can be assessed in routine applications of NRR, without the need for any additional information.

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.