A Unified Information-Theoretic Approach to Groupwise Non-Rigid Registration and Model Building.

Carole J. Twining[1], Tim Cootes[1], Stephen Marsland[2], Vladimir Petrovic[1], Roy Schestowitz[1], and Chris J. Taylor[1]

[1] Imaging Science and Biomedical Engineering (ISBE), Stopford Building, University of Manchester, Manchester, U.K.
[2] Institute of Information Sciences, Massey University, Private Bag 11222, Palmerston North, New Zealand.

Abstract. The non-rigid registration of a group of images shares a common feature with building a model of a group of images: a dense, consistent correspondence across the group. Image registration aims to find the correspondence, while modelling requires it. This paper presents the theoretical framework required to unify these two areas, providing a groupwise registration algorithm, where the inherently groupwise model of the image data becomes an integral part of the registration process. The performance of this algorithm is evaluated by extending the concepts of generalisability and specificity from shape models to image models. This provides an independent metric for comparing registration algorithms of groups of images. Experimental results on MR data of brains for various pairwise and groupwise registration algorithms is presented, and demonstrates the feasibility of the combined registration/modelling framework, as well as providing quantitative evidence for the superiority of groupwise approaches to registration.

1 Introduction

Over the past few years, non-rigid registration has been used increasingly as a basis for medical image analysis. Applications include structural analysis, atlas matching and change analysis. There are well-established methods for pairwise image registration(for a review, see e.g., [12]), but often it is necessary to register a group of images. This can be achieved by repeatedly applying pairwise registration, but there is no guarantee that the solution is unique – depending on the choice of reference image, representation of warp, and optimisation strategy, many different results can be obtained for the same set of images. Clearly, this does not form a satisfactory basis for analysis.

In this paper we consider non-rigid image registration as a complementary problem to that of modelling a group of images [2]. A statistical model of a group of images requires that a dense correspondence is defined across the group, which is precisely what non-rigid image registration provides. The key idea explored in this paper is that the best correspondence is that which generates the best model of the data. Building on the optimal shape model approach of Davies et al [3], we define a minimum description length (MDL) criterion for image model

quality. We show that a unique correspondence can be defined across a group of images by minimising, explicitly, an MDL objective function.

The combination of non-rigid image registration with modelling was shown previously by Frangi et al. [5], who used non-rigid registration to automatically construct 3D statistical shape models of the left and right ventricles of the heart. However, their method did require an initial manual labelling of every image in the training set. As regards groupwise non-rigid registration, several authors have considered the problem of choosing the best reference image. For instance, Bhatia et al [1] use a fixed intensity reference picked from the training set, but select the spatial frame of the reference so that the sum of deformations from this spatial reference frame is zero. Davis et al [4] concentrate specifically on deriving the most representative template image for a group of images, using sum-of-squared difference on the space of image discrepancies, and a metric on the diffeomorphism group of spatial deformations. Each of these approaches involve defining a series of independant criteria for what constitutes image matching, how image deformation is weighted against spatial deformation and so on. The advantage of our approach is that we use a single criterion – minimum description length – which can in principle determine not just the groupwise correspondence across the set of images, but also the optimal spatial reference frame, the optimal reference image and, potentially, the optimal model parameters (e.g., number of modes of the model retained). It hence combines registration and modelling within a *single* framework.

In this paper, we present a full description of our framework for groupwise registration, defining the MDL objective function and showing how the optimisation can be performed in a principled way by moving between different frames of reference. We validate the MDL objective function experimentally, using a set of annotated 2D MR brain slices. We also address the problem of evaluating different groupwise correspondences, by defining the generalisabilty and specificity of the resulting models. Again, we validate these measures using annotated data. We use these measures to evaluate the performance of a range of pairwise and groupwise approaches to registering a set of brain images, and show that the groupwise approach gives quantitatively better performance than pairwise.

2 Spatial & Pixel/Voxel-Value Transformations

The aim of non-rigid registration is to define a consistent spatial correspondence across a set of training images. One way to ensure a consistent correspondence is to define all correspondences w.r.t. a spatial reference frame – the origin of the space of spatial deformations. We define the following basic notational conventions , taking as our example the simplest case of a spatial warp directly between a training image frame and a reference frame (see Fig. 1):

- X_0 is the regular grid of pixel/voxel positions on which each of our images is defined.
- $\mathcal{R}$ is the spatial frame of the reference. A reference image $I_{\mathcal{R}}(X_0)$ consists of the set of values of a function $I_{\mathcal{R}}$, taken at the set of positions X_0.
- The set of N training images is denoted by $\{I_{\mathcal{T}_i}(X_0) : i = 1, \dots N\}$, where $\mathcal{T}_i$ is the spatial frame of image i, with associated image function $I_{\mathcal{T}_i}$.

The dense correspondence between a training image frame $\mathcal{T}_i$ and the reference frame $\mathcal{R}$ is defined by a spatial warp $\omega_i : x \in \mathcal{T}_i \mapsto \omega_i(x) \in \mathcal{R}$. The warp ω_i also induces a mapping between the function spaces (that is, it warps images between frames). Mathematically, there are two such mappings:

$$\begin{aligned} \textbf{The push-forward:} \quad & \omega_i : I_{\mathcal{T}_i} \mapsto I_{\mathcal{T}_i}^{\omega_i} \doteq \omega_i(I_{\mathcal{T}_i}), \quad I_{\mathcal{T}_i}^{\omega_i}(\omega_i(x)) \doteq I_{\mathcal{T}_i}(x) \\ \textbf{The pullback:} \quad & \omega_i^* : I_{\mathcal{R}} \mapsto I_{\mathcal{R}}^* \doteq \omega_i^*(I_{\mathcal{R}}), \quad I_R^*(x) \doteq I_{\mathcal{R}}(\omega_i(x)) \end{aligned}$$

The pullback ω_i^* is easier to compute, since we resample $I_{\mathcal{R}}$ in $\mathcal{R}$ from the regular grid X_0 to the irregular, warped grid $\omega_i(X_0)$ to obtain $I_R^*(X_0)$ in $\mathcal{T}_i$, whereas the push-forward mapping entails resampling $I_{\mathcal{T}_i}^{\omega_i}$ in $\mathcal{R}$ from the irregular grid $\omega_i(X_0)$ to the regular grid X_0, which is computationally more expensive. So, in what follows, we will use the **pullback** mapping wherever possible, where the direction of flow of image information is in the **opposite** direction to that of the spatial mapping.

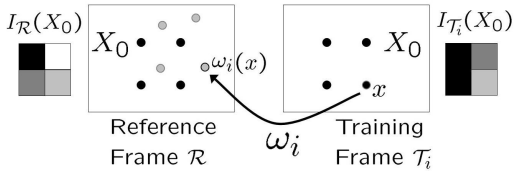

Fig. 1: A spatial warp ω_i from training frame $\mathcal{T}_i$ to reference frame $\mathcal{R}$. X_0 (black filled circles) is the set of regular voxel positions, with the grey filled circles being the warped voxel positions $\omega_i(X_0)$.

Once we can map images between frames, we can compare images. We will denote a general image-difference/discrepancy-image by ΔI. So, in the example above, if we define a discrepancy image in the frame $\mathcal{T}_i$:

$$\Delta I_{\mathcal{T}_i}(X_0) = I_{\mathcal{T}_i}(X_0) - I_{\mathcal{R}}^*(X_0) \implies (\Delta I_{\mathcal{T}_i} \circ \omega_i^*) I_{\mathcal{R}}(X_0) \equiv I_{\mathcal{T}_i}(X_0), \quad (1)$$

where $(\Delta I_{\mathcal{T}_i} \circ \omega_i^*)$ is taken to denote the composition of a pullback mapping ω_i^* and a voxel-value deformation $\Delta I_{\mathcal{T}_i}(X_0)$. The pixel/voxel-value deformation in this case is defined such that when applied to the warped reference image $I_{\mathcal{R}}^*(X_0)$ it exactly recreates the training set image $I_{\mathcal{T}_i}(X_0)$. It is important to note that in general these two classes of transformations **do not** commute. We now have a general class of image deformations, composed of a spatial part and a discrepancy image part – we will denote such a general combined deformation by capital greek letters (e.g., Ω_i).

A more complicated situation is shown in Fig. 2. This shows the reference image being transformed into a training image $I_{\mathcal{T}_i}$, by a sequence of two combined transformations Υ_i then Ω_i. We take this approach since, if we are to model combined transformations across the group of images, we need them to be applied in a **common** frame. So, the spatial transformations $\{\upsilon_i\}$ and the discrepancy images $\{\Delta_i I_{\mathcal{R}}\}$ are all applied in the reference frame $\mathcal{R}$, hence can be modelled across the group. However, the direction of the spatial warp υ_i is now in the **same** direction as the combined warp Υ_i (the direction of flow of image information), which means that Υ_i no longer has the simple form given above, but is given by:

$$\Upsilon_i = \upsilon_i \circ \Delta_i I_{\mathcal{R}}, \quad (2)$$

which uses the push-forward mapping υ_i as applied to images, rather than the easier-to-compute pullback. The spatial warp ω_i is now just from the training

frame $\mathcal{T}_i$ to the intermediate frame $\mathcal{M}_i$, the corresponding combined warp Ω_i being constructed using the pullback ω_i^* and the discrepancy image $\Delta I_{\mathcal{T}_i}$, which is calculated in a manner analogous to (1), but with the intermediate image $I_{\mathcal{M}_i}$ taking the place of the reference image $I_{\mathcal{R}}$. This second combined transformation is included because in general the groupwise-modelled transformation will not completely represent the total required transformation.

3 The Objective Function

As we explained in the Introduction, we have chosen to define the optimal groupwise non-rigid registration as that which minimises an objective function based on the minimum description length (MDL) principle [7].

The basic idea behind MDL is that we consider transmitting our dataset to a receiver, encoding the dataset using some model[1]. Using the structure and notation defined in the previous section, the data we have to transmit is the reference image $I_{\mathcal{R}}$ and the set of combined deformations $\{\Upsilon_i, \Omega_i\}$ that enable us to exactly reconstruct each training image. Optimising the description length means in principle finding:

- The optimal reference image $I_{\mathcal{R}}(X_0)$ and optimal reference frame $\mathcal{R}$.
- The optimal set of combined transformations $\{\Upsilon_i, \Omega_i\}$ via:
  - The optimal groupwise encoding of the deformations that act in a common frame, that is, the optimal groupwise model of the set $\{\Upsilon_i\}$,
  - Encoding of the residual deformations $\{\Omega_i\}$, which do not act in a common frame.

The total description length can hence be decomposed thus:

$$\begin{array}{lllll} \mathfrak{L}_{\text{total}} = & \mathfrak{L}_{\mathcal{R}}(\mathcal{R}, I_{\mathcal{R}}) & +\mathfrak{L}_{\text{params}} & +\mathfrak{L}_{\text{group}}(\{\Upsilon_i\}) & +\mathfrak{L}_{\text{residuals}}(\{\Omega_i\}) \\ & \text{Reference frame} & \text{Parameters of} & \text{Encoded using} & \text{Encoded residuals} \\ & \text{\& reference image} & \text{groupwise model} & \text{groupwise model} & \end{array} \tag{3}$$

[1] Note that in this paper, we use 'model' in two senses – in terms of an encoding model, which can be something very simple, such as a flat distribution over a known range, and in terms of a groupwise model, explicitly constructed to fit the data.

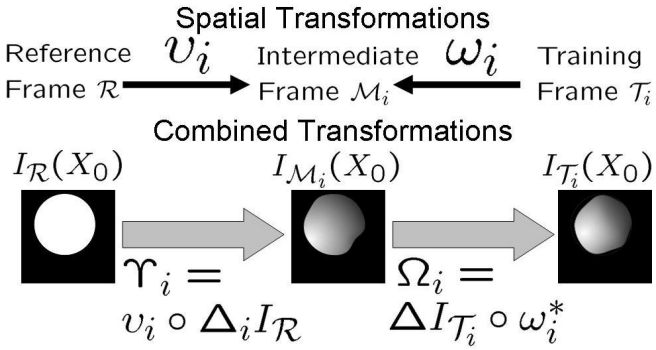

Fig. 2: **Top:** The spatial transformations (black arrows) between reference, intermediate and training image frames for one image i in the training set. **Bottom:** The corresponding combined (spatial and voxel-intensity) transformations (broad grey arrows) between images.

Actual description lengths are computed using the fundamental result of Shannon [9] – if there are a set of possible, discrete events $\{A\}$ with associated encoding-model probabilities $\{p_A\}$, then the optimum code length required to transmit the occurrence of event A is given by:

$$\mathfrak{L}_A = -\ln p_A \text{ nats}^\star \tag{4}$$

The encoding lengths for unsigned and signed integers are calculated thus:

$$\mathfrak{L}_{\mathbb{Z}^+}(n) = \frac{1}{e} + \ln(n) \text{ nats}, \; n \in \mathbb{Z}^+, \;\; \mathfrak{L}_{\mathbb{Z}}(n) = \frac{2}{e} + \ln(n) \text{ nats}, \; n \in \mathbb{Z}. \tag{5}$$

As an example, consider the description length for transmitting a discrepancy image $\Delta I(X_0)$ according to the image histogram. The $N_I = \text{size}(X_0)$ voxels of the image are taken to be integers in the range $[-N_{\text{range}}, \ldots N_{\text{range}}]$, N_m voxels having the value m. The associated model probability is then $p(m) = \frac{N_m}{N_I}$. The description length is :

$$\mathfrak{L}_{\text{Hist}}(\Delta I) = \underbrace{-\sum_{m,\, N_m>0} \ln\left(\tfrac{1}{2N_{\text{range}}+1}\right)}_{\text{Positions of occupied bins}} + \underbrace{\sum_{m,\, N_m>0} \mathfrak{L}_{\mathbb{Z}^+}(N_m)}_{\text{Bin Occupancies}} \; \underbrace{-\sum_{x\in X_0} \ln p(\Delta I(x))}_{\text{Encoded Data}}. \tag{6}$$

See [11, 10] for further details.

4 The Algorithmic Framework

4.1 Initialisation

In [10], an algorithm was presented to find an initial correspondence using MDL. The structure of the algorithm followed that shown in Fig. 1. The free variables were the set of spatial warps $\{\omega_i\}$, initialised to the identity $\mathbb{I}$, and the reference image was taken to be the mean of the training images, pulled-back using the inverses $\{\omega_i^{-1}\}$:

$$I_{\mathcal{R}}(X_0) = \frac{1}{N}\sum_{i=1}^{N}\left[\omega_i^{-1^*}(I_{\mathcal{T}_i})\right](X_0). \tag{7}$$

This algorithm was fully groupwise, in that changes to any of the $\{\omega_i\}$ change the reference, hence change the description length for all of the images in the set. However, the calculation of the inverse warps (or alternatively the push-forward mappings generated by $\{\omega_i\}$) is computationally expensive.

We propose here a computationally cheaper initialisation algorithm, within the structure shown in Fig. 2. We keep the idea from the algorithm presented in [10], of initial image estimates based on averages of pushed-forward training images, but instead choose to populate the intermediate images, using the leave-one-out means:

$$I_{\mathcal{M}_i}(X_0) = \frac{1}{N-1}\sum_{j\neq i}[\omega_j^{-1^*}(I_{\mathcal{T}_j})](X_0), \tag{8}$$

with $\{\upsilon_i = \mathbb{I}\}$. We do not explicitly assign a value to the reference image. But we would expect the intermediate images to mutually converge as the algorithm progresses and the images are brought into alignment, so that $\{\Delta_i I_{\mathcal{R}} \mapsto \varnothing\}$.

---

* The **nat** is the analogous unit to the **bit**, but using a base of e rather than base 2.

Algorithm 1 : MDL NRR Initialisation

1: $\{\omega_i = \mathbb{I},\, i = 1, \ldots N\}$ **%:Initialize warps to the identity.**
2: **Repeat**
3: Randomize the order of the set of training images $I_{\mathcal{T}_i}(X_0)$, indexed by i.
4: **For** $i = 1$ to N **do**
5: Optimise $\mathfrak{L}_{\text{init}}(\{\omega_k\})$ w.r.t. spatial warp ω_i.
6: Update Intermediate Images $\{I_{\mathcal{M}_j}(X_0) : j \neq i\}$. **%:Using equation (8).**
7: **End**
8: **Until** convergence

So, we estimate the true description length thus:

$$\mathfrak{L}_{\text{init}}(\{\omega_i\}) = \underbrace{\tfrac{1}{N}\sum_i \mathfrak{L}_{\text{Hist}}(I_{\mathcal{M}_i}(X_0))}_{\text{Estimate of } \mathfrak{L}_{\text{Hist}}(I_{\mathcal{R}}(X_0))} + \underbrace{\sum_i \mathfrak{L}(\omega_i)}_{\text{Spatial Warps}} + \underbrace{\sum_i \mathfrak{L}(\Delta I_{\mathcal{T}_i}(X_0))}_{\text{Discrepancy Images}}. \quad (9)$$

The pseudocode for the initialisation algorithm is given in Alg. 1. Note that the update of the Intermediate images $\{I_{\mathcal{M}_i}(X_0)\}$ (line 6) can be carried out less-frequently than at every training image, if required.

4.2 Groupwise Models

We have shown how to initialise the registration algorithm, within the structure shown in Fig. 1. However, when it comes to building groupwise models, we have the structure shown in Fig. 2. One method would be to build some default generative model of the set of deformations $\{\Upsilon_i\}$, and then search within the space of this model. However, this approach suffers from two drawbacks; firstly, the use of a default model (such as a gaussian) would bias the results, since it would tend to force the deformations to have a gaussian distribution, rather than finding the best deformations. The second drawback is computational – if we alter Υ_i, we have to then re-calculate Ω_i so that the combined deformation does indeed re-create our target training image $I_{\mathcal{T}_i}(X_0)$. This means that we have to re-calculate the intermediate image $I_{\mathcal{M}_i}(X_0)$, which means either calculating a pushforward mapping via υ_i, or a pushback via υ_i^{-1}, both of which are computationally expensive.

We take an alternative approach, which is to optimise the $\{\omega_i\}$. As in Alg. 1, this only involves computing the pullback ω_i^*. So, after we have optimised the set $\{\Omega_i\}$, we then transfer of much of this combined deformation as possible from the intermediate frame $\mathcal{M}_i$ to the equivalent deformation applied in the reference frame $\mathcal{R}$. We can then construct a model in the reference frame. The proposed algorithm is given in Alg. 2. Lines 1-5 are just the initialisation stages, which run the previous initialisation algorithm. The transfer between $\{\Omega_i\}$ and $\{\Upsilon_i\}$ is given in lines 2-3 of the function TEST-MODEL. An important point to note is in line 4 of that function – we maintain the spatial correspondence that we have previously found, despite moving spatial warps between frames. We then build a model of the set of combined deformations $\{\Upsilon_i = (\upsilon_i \circ \Delta_i I_{\mathcal{R}})\}$ and the reference image $I_{\mathcal{R}}(X_0)$. The modelled deformations are not necessarily the same as the input deformations to the modelling process, which is the reason for the resetting in line 5. We then accept this model provided that it decreases the total description length.

Algorithm 2 : MDL NRR & Groupwise Model Building

Run Algorithm 1 **%:Output is** $\{I_{\mathcal{M}_i}(X_0), \omega_i, \Delta I_{\mathcal{T}_i}(X_0)\}$
$\upsilon_i \Leftarrow \mathbb{I}$ **%:Initial Shared frame for all Intermediate Images**
$I_{\mathcal{R}}(X_0) \Leftarrow \frac{1}{N}\sum_i I_{\mathcal{M}_i}(X_0)$ **%:Estimate Reference as Mean**
$\Delta_i I_{\mathcal{R}} \Leftarrow I_{\mathcal{M}_i}(X_0) - I_{\mathcal{R}}(X_0)$ **%:Maintain Intermediate Images**
Build & Test groupwise model of $\{\Upsilon_i \equiv \upsilon_i \circ \Delta_i I_{\mathcal{R}}\}$
$(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, \upsilon_i, \omega_i, I_{\mathcal{M}_i}, \Delta I_{\mathcal{T}_I}\}) \Leftarrow$ **TEST-MODEL**$(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, \upsilon_i, \omega_i\})$
Main Loop
Repeat
  Repeat
    Randomize the order of the set of training images $I_{\mathcal{T}_i}(X_0)$, indexed by i
    Optimise warps ω_i
    For $i = 1$ to N **do**
      Optimise $\mathfrak{L}_{\text{total}}$ w.r.t. spatial warps ω_i. **%:$\mathfrak{L}_{\text{total}}$ calculated from eq. (3)**
    End
  Until convergence
  Re-Build Model
  $(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, \upsilon_i, \omega_i, I_{\mathcal{M}_i}, \Delta I_{\mathcal{T}_I}\}) \Leftarrow$ **TEST-MODEL**$(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, \upsilon_i, \omega_i\})$
Until convergence

Function TEST-MODEL: Build & Test Groupwise Model

$\mathfrak{L}_{\text{old}} \Leftarrow \mathfrak{L}_{\text{total}}(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, \upsilon_i, \omega_i\})$ **%:Description Length $\mathfrak{L}$ before modelling, eq.(3)**
$\upsilon_i^{\text{new}} \Leftarrow \omega_i^{-1} \circ \upsilon_i$ **%:Put all spatial warp into υ_i**
Build Model
$(I_{\mathcal{R}}^{\text{new}}, \{\Delta_i^{\text{new}} I_{\mathcal{R}}, \upsilon_i^{\text{new}}\}) \Leftarrow$ **MODEL**$(I_{\mathcal{R}}, \{\Delta_i I_{\mathcal{R}}, \upsilon_i^{\text{new}}\})$
$\omega_i^{\text{new}} \Leftarrow \upsilon_i^{\text{new}} \circ (\upsilon_i^{-1} \circ \omega_i)$ **%:Reset ω_i^{new} to maintain spatial correspondence**
$\mathfrak{L}_{\text{new}} \Leftarrow \mathfrak{L}_{\text{total}}(I_{\mathcal{R}}^{\text{new}}, \{\Delta_i^{\text{new}} I_{\mathcal{R}}, \upsilon_i^{\text{new}}, \omega_i^{\text{new}}\})$ **%:Description Length after modelling**
If $\mathfrak{L}_{\text{new}} \leq \mathfrak{L}_{\text{old}}$ **then**
  $\omega_i \Leftarrow \omega_i^{\text{new}}$, $\upsilon_i \Leftarrow \upsilon_i^{\text{new}}$, $I_{\mathcal{R}} \Leftarrow I_{\mathcal{R}}^{\text{new}}$, $\Delta_i I_{\mathcal{R}} \Leftarrow \Delta_i^{\text{new}} I_{\mathcal{R}}$ **%:Accept new values**
  $I_{\mathcal{M}_i}(X_0) \Leftarrow (\upsilon_i \circ \Delta_i I_{\mathcal{R}}) I_{\mathcal{R}}(X_0)$ **%:Reset Intermediate Images**
  $\Delta I_{\mathcal{T}_i}(X_0) \Leftarrow I_{\mathcal{T}_i}(X_0) - [\omega_i^*(I_{\mathcal{M}_i})](X_0)$ **%:Reset discrepancies in Training frame**
End

5 Implementation Issues

Consider the relation of spatial frames for the groupwise algorithm (e.g., see Fig. 2 and Alg. 2) – it is clear that we require a description of spatial warps $\{\omega_i, \upsilon_i\}$ that allows us to efficiently invert and concatenate warps, as well as a description which allows us to represent a set of warps (i.e.,$\{\upsilon_i\}$) within a common representation for the purposes of modelling. Such a description is provided by spline-based formulations which interpolate the movement of general points from the movement of a set of nodes/knotpoints, where the knotpoints can take **arbitrary** positions. In the experiments which follow, we use both the clamped-plate spline, and an efficient spline based on the piecewise-linear interpolation of movements across a tesselated set of knotpoints in either 2D or 3D.

The advantages of such a knotpoint based scheme is that it can be applied in both a multi-resolution and a data-driven fashion. Successive optimisations of the set $\{\omega_i\}$ in Alg. 2 are calculated by adding knotpoints to the previously-optimised set (hence increasing the resolution of the spatial warp). These knot-

points are also chosen in a data-driven manner (e.g., image features such as edges, or places of high discrepancy – see [6, 8] for further examples of such data-driven techniques). This not only increases the computational efficiency of our implementation but, as will be shown later, also leads to quantitatively better models. We use a coarse-to-fine strategy during the optimisation – at a coarse spatial resolution, node movements can be large, and it is sufficient to use a low-resolution version of the image. As the spatial resolution of the warps increases, so does the spatial resolution of the image used. The optimisation scheme for the nodes is a simple gradient descent – points are moved singly to estimate the gradient direction for the objective function, but moved all at once using a line search.

6 Model Evaluation Criteria

In order to compare different algorithms for non-rigid registration and model building, we need to have some quantitative measures of the properties of a given model. Following Davies et al. [3], we use two measures of model performance:

- **Generalisability:** the ability to represent unseen images which belong to the same class as images in the training set.
- **Specificity:** the ability to only represent images similar to those seen in the training set.

Let $\{I_{\mathfrak{a}}(X_0) : \mathfrak{a} = 1, \ldots \mathfrak{N}\}$ be some large set of images, generated by the groupwise model, and having a distribution which is the model distribution. Then we define the following:

$$G = \frac{1}{N} \sum_{i=1}^{N} \min_{\text{w.r.t } \mathfrak{a}} \left(|I_{\mathcal{T}_i}(X_0) - I_{\mathfrak{a}}(X_0)| \right), \text{ Generalisability,} \tag{10}$$

$$S = \frac{1}{\mathfrak{N}} \sum_{\mathfrak{a}=1}^{\mathfrak{N}} \min_{\text{w.r.t. } i} \left(|I_{\mathcal{T}_i}(X_0) - I_{\mathfrak{a}}(X_0)| \right), \text{ Specificity,} \tag{11}$$

where the distance $|\cdot|$ is a measure of the distance between two images. This could be taken as the Euclidean distance between images, but this is likely to be very sensitive to quite small shape changes or misalignments, and thus not provide a useful measure of image difference. To deal with this problem, we have used shuffle distance, calculating, for each pixel in one image, the minimum intensity difference to any pixel/voxel within a radius r of the corresponding pixel/voxel in the other image. The shuffle distance is then defined as the sum across all voxels/pixels of the absolute intensity differences, since this is more robust to outliers than sum-of squares. Note that our definition of G is not that used in [3], but a form which is symmetric as regards the form of S; G measures how close each training image is to images in the modelled distribution, whereas S measures how close each model-generated image is to the training data. Standard errors for S and G can be defined similarly to Davies et al.

7 Experiments

We have performed experiments to validate our MDL objective function and model evaluation criteria and investigate the performance of several different

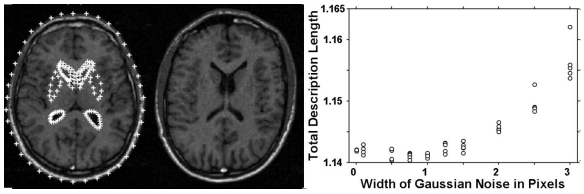

Fig. 3: **Left:** Two examples of marked-up brains, showing annotation. **Right:** Total description length for this dataset as a function of the size of the perturbation of the points.

non-rigid registration methods, including that presented in this paper. Although all the methods we have described can be used in 3D, it was impractical to run the very large set of experiments required in the time available, thus we present results for 2D images of the brain.

7.1 Behaviour of the MDL Objective Function

The first question to be answered is whether the total description length has a suitable minimum as regards correspondence across a set of images. To investigate this, we took a dataset which consisted of 2D MR image slices; this dataset had been expertly annotated with 163 points around the skull, ventricles, the caudate nucleus and the lentiform nucleus (see Fig. 3). The clamped-plate spline warp between these points then defined dense image correspondence. We applied a perturbation to the point positions on all the images (independent Gaussian noise of width σ, 5 trials for each value of σ, with 10 images in the dataset). For each value of σ, we constructed the corresponding shape and texture models, the discrepancy between the actual images and the model representations, and hence calculated the total description length. As can be seen from the Figure, there is a general trend that as the perturbation increases, so does the total description length, indicating that the description length does indeed have a minimum in the vicinity of the annotated correspondence.

7.2 Behaviour of the Model Evaluation Criteria

To validate our Generalisability G (10) and Specificity S (11) criteria, we took the same dataset and markup as above, but now with 36 examples. As before, we perturbed the point positions and built the corresponding shape and texture

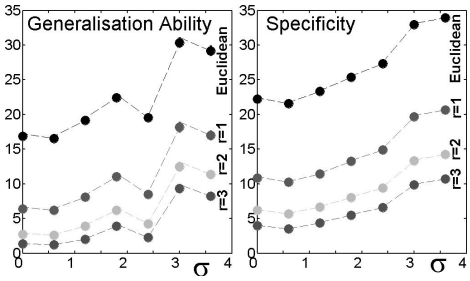

Fig. 4: **Left:** Generalisation Ability and **Right:** Specificity as a function of the size of the perturbation on the points, for various radii of shuffle distance plus Euclidean distance. Standard errorbars smaller than markers in all cases.

models. We then generated 1000 examples sampled from each model p.d.f., and calculated G and S. The results for various values of the perturbation width σ, and different shuffle distances (r), are shown in Fig. 4. It can be seen that both Generalisability and Specificity increase (get worse) as σ is increased, indicating that they provide useful independent measures of model quality.

The useful range of response is greater for larger shuffle distances (e.g., the slope of the 5×5 ($r = 2$) shuffle distance curve is lower than that of the Euclidian distance curve). In the automatic model building experiments described below we used the 5×5 shuffle distance to calculate G and S.

7.3 Evaluation of Pairwise & Groupwise Registration and Models

To evaluate different methods of non-rigid registration we used a dataset consisting of 104 2D MR slices of brains taken from normals; the initial 3D data set was affinely-aligned, and then the corresponding slice extracted from each example. Fig. 5 shows examples of the slices. In order to compare different registration strategies, for each technique we registered the entire set of 104 images and built the statistical models of shape and appearance given by the found correspondence, using the nodes/knotpoints used during the registration. We then computed the Generalisability G (10) and Specificity S (11) for each model (generating 1000 model examples in each case, and using a 5-pixels square sample region for the shuffle distance), enabling a quantitative comparison of the registration strategies from which each model was derived. The strategies tested were:

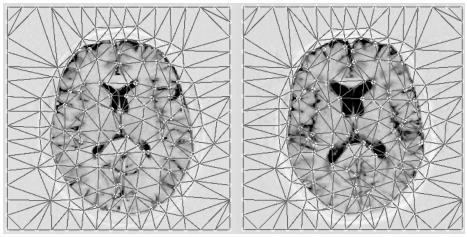

Fig. 5: Example images from the brain slice training set, showing the tessellation.

1 Pairwise Registration:

A Image from training set chosen as reference & 16×16 regular grid of nodes:
- **i** Residuals calculated in reference frame.
- **ii** Residuals calculated in training frame.

B As above, but removing points from the grid in regions of low texture variance.

C Ditto, but moving points to nearby strong edges.

2 Groupwise Registration:

A Registering to Intermediate Images estimated as the leave-one-out means (Alg. 1).

B Registering to Intermediate Images estimated using the leave-one-out models.

Note that for **1**, we tried a selection of images from the training set as the reference, and choose that which gave the best results in terms of the evaluation criteria. Strategy **2B** can be viewed as an approximation to the full algorithm given in Alg. 2; in the same way that in the initialisation algorithm (Alg. 1) we estimate the Intermediate Images $\{I_{\mathcal{M}_i}\}$ using the leave-one-out mean, in this case we estimate them by finding the closest fit to the training image $I_{\mathcal{T}_i}$ using the shape model built from all the other examples and the current best estimate of their correspondence. We then optimise the description length of the shape

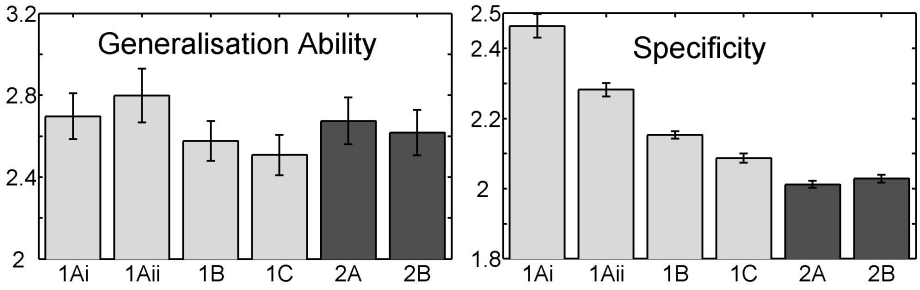

Fig. 6: Generalisation ability and Specificity for the strategies listed in §7.3 – dark bars groupwise, light bars pairwise.

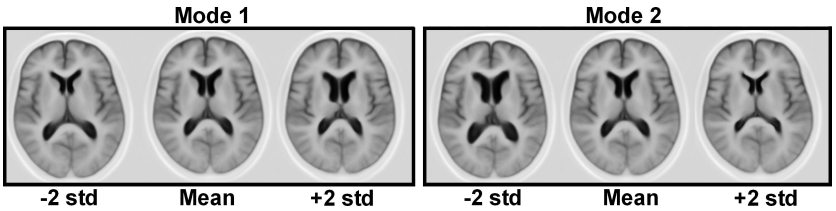

Fig. 7: The first two modes of the shape model built using the results of groupwise registration, acting upon the mean of the texture model.

and texture discrepancies between this model estimate and the training image. Note that we do not model the texture at this intermediate stage – this is because in the inner loops of Algs. 1&2, the warps $\{\omega_i\}$ at each spatial resolution are fully optimised, hence can then be modelled, whereas the texture discrepancy is merely continually reduced. The results of this comparison are given in Fig. 6.

8 Discussion & Conclusions

We have presented a principled framework for groupwise non-rigid registration, based on the concept of minimum description length. A groupwise model of shape and appearance is an integral part of the regsitration algorithm, hence the registration also produces an optimal appearance model. We have given a full description of a practical implementation of the basic ideas. Another important contribution is the introduction of objective criteria for evaluating the results of non-rigid registration, based on the properties of the resulting appearance model. The results summarised in Fig. 4 show that the method of evaluation we propose provides a practical method of comparing the quality of different non-rigid registrations. The results summarised in Fig. 3 show that our MDL objective function behaves as expected, with a minimum for a groupwise correspondence close to that given by expert manual annotation. the key results are those summarised in Fig. 6. These show that our groupwise approach achieves better Specificity than several different pairwise approaches. They also show the importance of measuring errors in the correct frame of reference. Further work is required to implement more sophisticated versions of our groupwise approach, and to provide a more comprehensive set of comparisions to alternative approaches. Our initial results are, however, extremely encouraging.

Acknowledgements This research was supported by the MIAS IRC project, EPSRC grant No. GR/N14248/01, UK Medical Research Council Grant No. D2025/31 (*"From Medical Images and Signals to Clinical Information"*). S. Marsland was supported by the Marsden Fund grant `MAU0408`, Royal Society of New Zealand, *"A principled approach to the non-rigid registration and structural analysis of groups of medical images"*.

References

1. K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. *Proceedings of the IEEE Symposium on Biomedical Imaging (ISBI)*, pages 908–911, 2004.
2. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:681–685, 2001.
3. R. H. Davies, C. J. Twining, P. D. Allen, T. F. Cootes, and C. J. Taylor. Shape discrimination in the hippocampus using an MDL model. In *Proceedings of IPMI 2003*, pages 38–50, 2003.
4. B. Davis, P Lorenzen, and S. Joshi. Large deformation minimum mean squared error template estimation for computational anatomy. *Proceedings of the IEEE Symposium on Biomedical Imaging (ISBI)*, pages 173–176, 2004.
5. A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modelling. *IEEE Transactions on Medical Imaging*, 21(9):1151–1166, 2002.
6. S. Marsland and C. J. Twining. Constructing data-driven optimal representations for iterative pairwise non-rigid registration. *Lecture Notes in Computer Science*, 2717:50–60, 2003.
7. J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific Press, 1989.
8. J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, F. A. Gerritsen, D.L.G. Hill, and D. J. Hawkes. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *Proceedings of MICCAI 2001*, number 2208 in Lecture Notes in Computer Science, pages 573 – 581, 2001.
9. C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423,623–656, 1948.
10. C. J. Twining, S. Marsland, and C. J. Taylor. Groupwise non-rigid registration: The minimum description length approach. In *Proceedings of the British Machine Vision Conference (BMVC)*, volume 1, pages 417–426, 2004.
11. C.J. Twining, S. Marsland, and C.J. Taylor. A unified information-theoretic approach to the correspondence problem in image registration. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, 2004.
12. Barbara Zitová and Jan Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:977 – 1000, 2003.