Information-Theoretic Unification of Groupwise Non-Rigid Registration and Model Building.

Carole J. Twining[a], T.F. Cootes[a], S. Marsland[b], V. Petrovic[a], R. Schestowitz[a], and C.J. Taylor[a]

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

Abstract. There is a feature common to both non-rigid registration of a group of images and building a model of a group of images: a dense, consistent correspondence across the group. The former aims to find such a correspondence, whilst the latter 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 using the concepts of generalisability and specificity, which provide an independent metric for comparing various registration algorithms. 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

Non-rigid registration (NRR) is being increasingly used as a basis for medical image analysis, with applications that include structural analysis, atlas matching and change analysis. There are well-established methods for pairwise image registration(for a review, see e.g., [1]), 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. Our approach is to consider NRR and modelling a group of images [3] as complementary problems: the aim of NRR is to find a meaningful dense correspondence across the group, whereas modelling requires it. Building on the optimal shape model approach of Davies et al [4], we define a minimum description length (MDL) criterion for image model quality, and show that a unique groupwise correspondence can be defined by explicit minimisation of this criterion. NRR and modelling have been combined previously [6], however this required an initial manual labelling of every image. As regards groupwise non-rigid registration, several authors have considered the problem of choosing the best reference image(e.g., [2, 5]). 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 unifies registration and modelling within a *single* coherent theoretical framework.

In this paper, we present a brief description of our framework for groupwise registration, including the MDL objective function, the method of optimisation, and metrics for evaluating different groupwise correspondences. We 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 correspondence across the group has to be consistent, and one way to ensure this is to define all correspondences w.r.t. a spatial reference frame. We hence 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 $I_{\mathcal{R}}(X_0)$ is the values of a function $I_{\mathcal{R}}$ at the positions X_0.
- The set of N training images $\{I_{\mathcal{T}_i}(X_0) : i = 1, \dots N\}$. Image i, function $I_{\mathcal{T}_i}$ in spatial frame $\mathcal{T}_i$.

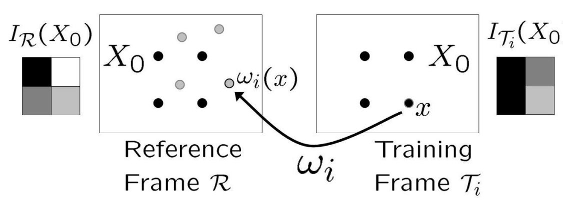

Figure 1: A spatial warp ω_i from training frame $\mathcal{T}_i$ to reference frame $\mathcal{R}$. X_0 (black filled circles) is the set of regular voxel positions, with the grey filled circles being the warped voxel positions $\omega_i(X_0)$.

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:

The push-forward: $\omega_i : I_{\mathcal{T}_i} \mapsto I_{\mathcal{T}_i}^{\omega_i} \doteq \omega_i(I_{\mathcal{T}_i})$, $I_{\mathcal{T}_i}^{\omega_i}(\omega_i(x)) \doteq I_{\mathcal{T}_i}(x)$

The pullback: $\omega_i^* : I_{\mathcal{R}} \mapsto I_{\mathcal{R}}^* \doteq \omega_i^*(I_{\mathcal{R}})$, $I_R^*(x) \doteq I_{\mathcal{R}}(\omega_i(x))$

*carole.twining@manchester.ac.uk

Because of re-sampling issues for regular and irregular pixel grids, the pullback ω_i^* is easier to compute than the push-forward mapping, so that 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. The ability to map images between frames means we can compare images. We will denote a general discrepancy-image by ΔI. In the example above, a discrepancy image in the frame $\mathcal{T}_i$ is: $\Delta I_{\mathcal{T}_i}(X_0) = I_{\mathcal{T}_i}(X_0) - I^*_{\mathcal{R}}(X_0) \Longrightarrow (\Delta I_{\mathcal{T}_i} \circ \omega_i^*) I_{\mathcal{R}}(X_0) \equiv I_{\mathcal{T}_i}(X_0)$ 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 $\Delta I_{\mathcal{T}_i}$. 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. 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 that given previously, but with the intermediate image $I_{\mathcal{M}_i}$ taking the place of the reference image $I_{\mathcal{R}}$.

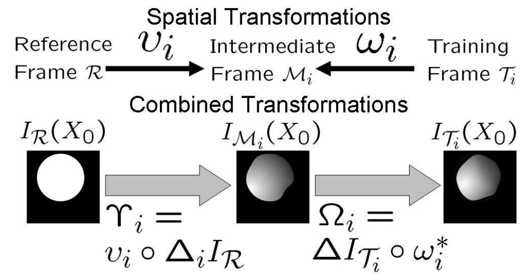

Figure 2: **Top:** The spatial transformations (black arrows) between reference, intermediate and training image frames. **Bottom:** The corresponding combined (spatial and voxel-intensity) transformations (broad grey arrows) between images.

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 [8]. The basic idea behind MDL is that we consider transmitting our dataset to a receiver, encoding the dataset using some model. 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. The total description length can hence be decomposed thus:

$$\mathfrak{L}_{\text{total}} = \underbrace{\mathfrak{L}_{\mathcal{R}}(\mathcal{R}, I_{\mathcal{R}})}_{\text{Reference frame \& reference image}} \; \underbrace{+\mathfrak{L}_{\text{params}}}_{\text{Parameters of groupwise model}} \; \underbrace{+\mathfrak{L}_{\text{group}}(\{\Upsilon_i\})}_{\text{Encoded using groupwise model}} \; \underbrace{+\mathfrak{L}_{\text{residuals}}(\{\Omega_i\})}_{\text{Encoded residuals}} \tag{1}$$

Actual description lengths are computed using the fundamental result of Shannon [10] – 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: $\mathfrak{L}_A = -\ln p_A$ nats.[1] See [11–13] for details as to the explicit calculation of description lengths.

4 The Algorithmic Framework

4.1 Initialisation

In [11], 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 $\{\omega_i^{-1}\}$. 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 [11], 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:

$$I_{\mathcal{M}_i}(X_0) = \frac{1}{N-1} \sum_{j \neq i} [\omega_j^{-1*}(I_{\mathcal{T}_j})](X_0), \tag{2}$$

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 $\{\Delta_i I_{\mathcal{R}} \mapsto \varnothing\}$.

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

Algorithm 1 : MDL NRR Initialisation

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

The true description length is estimated thus :

$$\mathfrak{L}_{\text{init}}(\{\omega_i\}) = \underbrace{\frac{1}{N}\sum_i \mathfrak{L}_{\text{Hist}}(I_{\mathcal{M}_i}(X_0))}_{\text{Estimate of } \mathfrak{L}_{\text{Hist}}(I_{\mathcal{R}}(X_0))} + \underbrace{\sum_i \mathfrak{L}(\omega_i)}_{\text{Spatial Warps}} + \underbrace{\sum_i \mathfrak{L}(\Delta I_{\mathcal{T}_i}(X_0))}_{\text{Discrepancy Images}}. \tag{3}$$

The pseudocode for the initialisation algorithm is given in Alg. 1.

4.2 Groupwise Models

We have shown how to initialise the registration algorithm, we now have to consider the explicit groupwise model. 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.

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. (1)**
    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.(1)**
$\upsilon_i^{\text{new}} \Leftarrow \omega_i^{-1} \circ \upsilon_i$ **%: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

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.

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. 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 knotpoints are also chosen in a data-driven manner (e.g., image features such as edges, or places of high discrepancy – see [7, 9] for further examples of such data-driven techniques). The optimisation scheme for the knotpoints 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. [4], 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 $\{I_{\mathfrak{a}}(X_0) : \mathfrak{a} = 1, \dots \mathfrak{N}\}$ be some large set of images, generated by the groupwise model, and having a distribution which is the model distribution. We then define the following quantitative measures:

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

where the distance $|\cdot|$ is a measure of the distance between two images, such as the Euclidean or shuffle distance.

7 Experiments: Evaluation of Pairwise & Groupwise Registration and Models

We have previously performed experiments to validate our MDL objective function and model evaluation criteria, see [13] for details. Here we investigate the performance of several different 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. 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. 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 and Specificity S 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:

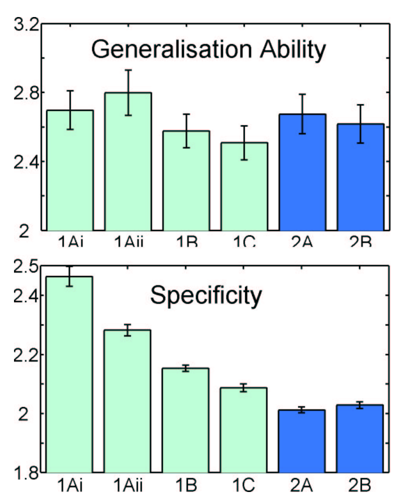

Figure 3: Generalisation ability and Specificity for the strategies listed – dark bars groupwise, light bars pairwise.

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 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. 3.

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 registration algorithm, hence the registration also produces an optimal appearance model. We have given a brief description of a practical implementation of the basic ideas. The key results are those summarised in Fig. 3. 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. Barbara Zitová and Jan Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:977 – 1000, 2003.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific Press, 1989.
9. 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.
10. C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423,623–656, 1948.
11. 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.
12. 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.
13. C.J. Twining, Tim Cootes, Stephen Marsland, Vladimir Petrovic, Roy Schestowitz, and Chris J. Taylor. A Unified Information-Theoretic Approach to Groupwise Non-Rigid Registration and Model Building In *Proceedings of the 19th International Conference on Information Processing in Medical Images (IPMI'05)*, volume 3565 of Lecture Notes in Computer Science, pages 1-14. Springer, 2005.