Groupwise Non-Rigid Registration of Medical Images: The Minimum Description Length Approach

Carole J. Twining[1*], Stephen Marsland[2*], and Chris Taylor[1]
[1] ISBE, University of Manchester, U.K.
[2] Institute of Information Sciences, Massey University, N.Z.

Abstract. The aim of non-rigid registration as applied to a group of images is to find a 'meaningful' dense spatial correspondence across the set. There are many methods available for finding such a correspondence given a *pair* of images, but viewing the groupwise case as just successive pairwise is rather naïve. The principled non-rigid registration of *groups* of images hence requires a fully groupwise objective function. Statistical analysis of the spatial and pixel-value deformations across the set (as defined by the found correspondence), means that these deformations have to be defined with respect to a common spatial and pixel value reference.
We show how the optimal groupwise correspondence can be defined using the Minimum Description Length (MDL) principle, where the definition of the spatial and pixel-value reference is also part of the optimisation. We demonstrate the use of such an objective function as applied to non-rigid registration of a set of 2D T1-weighted images of the human brain. As regards constructing the optimal reference image, we show that even in the case when substantial portions of the images are missing, the algorithm not only converges to the correct solution, but also allows meaningful integration of image data across the training set, allowing the original image to be reconstructed as the reference image.

1 Introduction

The aim of a non-rigid registration algorithm as applied to a set of medical images is to find a 'meaningful' dense correspondence across the whole set of images. As regards *pairwise* registration, there are many methods available (for a review, see [16]). Such pairwise registration is obviously sufficient for some applications (e.g., comparison to an atlas [2]). However,in any application where the statistical analysis of the resulting deformation fields is required, such as the modelling of biological variability, or of assisting in disease diagnosis across the population, performing repeated pairwise registrations over the set of images is, at best, naïve. In order to facilitate useful statistical analysis, the registration of the group of images needs to be considered as a single problem, so that the parameters of the warps on all of the images lie in a common manifold. We have previously [7] considered a method of non-rigid registration that ensures that there is a common set of knotpoints that define the warps across all of the images. In this paper, we extend that work by considering a groupwise objective function for non-rigid registration.

There is an important distinction to be made between **intra**-subject as opposed to **inter**-subject registration. In intra-subject registration there is often some actual physical process determining the observed deformation (e.g., tissue deformation due to patient position, the needle insertion or organ motion). Alternatively, the deformation may be viewed as the result of some long-term biological process (e.g., natural growth or tumour growth, or atrophy as in dementia). The most suitable choice of registration algorithm is hence one that closely models the underlying process, leading to physically-based registration algorithms (e.g., [4, 5]), or physically-based models (e.g., [9]) that can be used to evaluate the results of non-rigid registration algorithms. However, in inter-subject registration there is no longer a direct underlying physical or biological process that generates the observed data. We therefore contend that in the absence of expert anatomical knowledge (i.e., for the case of purely *automatic* registration), the meaning of correspondences should be derived purely from the available data (i.e., the set of images). Further, any statistical inferences that we make about the data should not depend on hypothetical data-generating processes; an assumption that underlies parameter estimation techniques such as maximum likelihood. The Minimum Description Length (MDL) [8] and Minimum Message Length (MML) [15] principles are closely related approaches [1] to model-selection and statistical inference that satisfy these restrictions.

The MDL principle has previously been shown to give excellent results when applied to the correspondence problem in shape modelling [3]. However, naïve attempts at extending the methods described there to images have not been successful [11]. This paper describes the application of the ideas developed in [13, 14].

2 The MDL Principle applied to Groupwise Registration: A Brief Overview

We give here a brief overview of the MDL principle as applied to image registration; for further details, see [13,14]. The key idea is to consider transmitting the full set of quantized images to a receiver, where this image set has been encoded using some type of model. For the case of non-rigid registration, this transmission can be taken to consist of the following parts:

*Joint first authors. Contact: **carole.twining@man.ac.uk, S.R.Marsland@massey.ac.nz**

- The reference example or template – *this defines the common spatial and pixel-value reference for the set.*
- A set of spatial and pixel/voxel value deformations. These can be separated into:
  - A explicitly modelled part – *model parameters and the data as represented by the model.*
  - Any residual deformations.

The model can be either an explicit parametric statistical model (such as a multivariate gaussian), or the empirical distribution of the actual data (that is, the histogram of the data values). In either case, the receiver reconstructs the reference image, and then applies the specified spatial and pixel-value deformations to this reference so as to reconstruct *exactly* each image in the training set.

The total description length $\mathfrak{L}$ can then be written as a sum of corresponding terms thus:

$$\mathfrak{L} = \mathfrak{L}_{\text{reference}} + (\mathfrak{L}_{\text{params:model}} + \mathfrak{L}_{\text{data:model}}) + \mathfrak{L}_{\text{residual}}. \tag{1}$$

2.1 Computing Description Lengths

The actual description lengths for the transmission of one parameter or one piece of data are computed using the fundamental result of Shannon [10] – if there are a set of possible, discrete events $\{i\}$ with associated model probabilities $\{p_i\}$, then the optimum code length required to transmit the occurrence of event i is given by:

$$\mathfrak{L}_i = -\ln p_i \ \text{nats}, \tag{2}$$

where the *nat* is the analogous unit to the *bit*, but using a base of e rather than base 2, so that e bits $\equiv 1$ nat. So, for the case of a statistical model, the probability is that given by the model.

The other case we consider is where we wish to transmit an unbounded, quantized data value, or an integer. The two are equivalent, as a quantized data value can always be reduced to an integer. The approximate description lengths are as follows:

$$\textbf{Unsigned Integer: } n \in \mathbb{Z}^+, \ \mathfrak{L}_{\mathbb{Z}^+}(n) \approx \frac{1}{e} + \ln(n) \ \text{nats}, \ \textbf{Signed Integer: } n \in \mathbb{Z}, \ \mathfrak{L}_{\mathbb{Z}}(n) \approx \frac{2}{e} + \ln(n) \ \text{nats}. \tag{3}$$

We can hence compute the description length for transmitting a quantized, pixellated grayscale image I with N_I pixels according to the image histogram of that image. Suppose that the pixel-values $\{I(A) : A = 1, \ldots N_I\}$ are integers in the range $[1\,N]$, and that there are N_m pixels in the image with the value m, with occupied bins situated at positions $\{m_\alpha\}$. Using this image histogram as the model, this gives the associated probability $p(m) = \frac{N_m}{N_I}$. The transmission then consists of the positions of the occupied bins (assuming a flat distribution over the allowed range), the occupation numbers of each bin, and then finally the ordered set of actual pixel values in the image, encoded using the histogram as model. The description length is hence:

$$\mathfrak{L}_{\text{histogram}} = -\sum_{\alpha} \ln\left(\frac{m_\alpha}{N}\right) + \sum_{\alpha} \mathfrak{L}_{\mathbb{Z}^+}(N_{m_\alpha}) - \sum_{A=1}^{N_I} \ln p(I(A)), \tag{4}$$

which is a form of image encoding that we will use later on.

2.2 A Simple MDL Algorithm for Image Registration

We here describe a simple MDL algorithm for image registration. This algorithm has the advantage that it can bootstrap itself from the assumption of the identity transform between image frames, and hence can be used to initialize other more complicated algorithms. We take a set of training images $I_1, \ldots I_{n_s}$ and a reference image I_{ref}, where for each training image we have the spatial transformation t_i as shown in Figure 1 between the reference and image planes. A set of such transformations $\{t_i\}$ is sufficient to define a consistent dense correspondence across the set of images, and in this formulation, these transformations are the only free parameters of the encoding. For a given set of transformations, the reference image is taken as the average of the set of training images, pulled-back into the frame of the reference. We also transmit the discrepancy images, given by calculating the discrepancies between each training image and the reference image pushed-forward into the frame of each training image. The description length for transmitting the set of transformations $\{t_I\}$, the reference image, and the set of discrepancy images is then computed, and used as the objective function for defined the optimal set of such transformations. See [13] for further details.

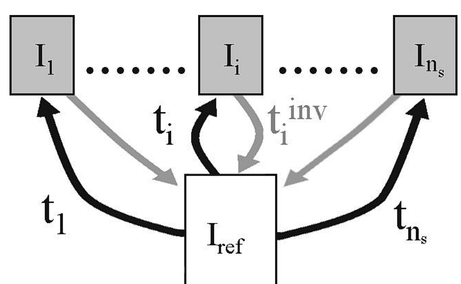

Figure 1. Spatial transformations between reference and image frames.

3 Experiments

3.1 Non-Rigid Registration

As an example, we take as our training set a set of $n_s = 5$ 2D axial T1 MR slices of human brains, which have already been affinely aligned. The images are 8-bit grayscale images of size $N_I = 100 \times 100$. We take as our

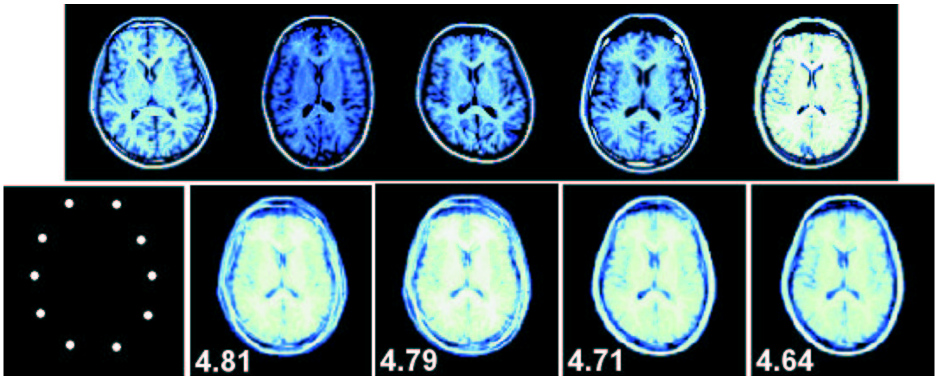

Figure 2. Top: The training set. **Bottom:** The fixed knotpoint positions in the reference frame, and the mean of the aligned images after 0, 2, 6 and 10 iterations, with the transmission length per pixel shown.

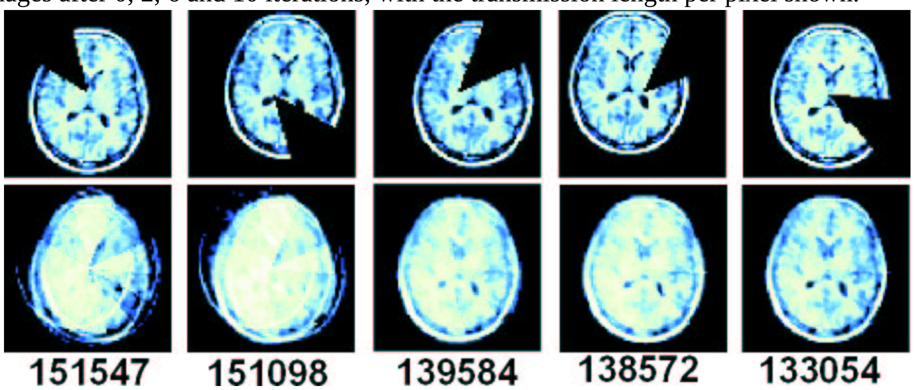

Figure 3. Top Row: The set of training images, **Bottom Row:** The reference image as the registration progresses, with the value of the objective function (the *total* description length for the set in nats).

parameterised set of transformations the biharmonic Clamped-Plate splines (CPS) [12]. The CPS interpolates the motion of a set of knotpoints, hence the parameters of a transformation t_i are the initial and final positions of those knotpoints. The bounding ball for these splines was taken to be the circumcircle of the images. The reference image and discrepancy images are all encoded using the histogram encoding given earlier (4) with $N = 256$ for the reference, and $N = 512$ for the discrepancy images. Following [6], we first generate a set of $n_k = 10$ equi-angularly spaced knotpoints around the skull for each image, these being averaged to give the reference image knotpoint positions which remain fixed, and provide us with our spatial reference. For the purposes of illustration, the image knotpoint positions were initialised to the reference knotpoint positions, so that the transformation starts at the identity. Optimising the set of transformations $\{t_i\}$ then corresponds to optimising the set of knotpoint positions on each image. It can be seen in Figure 2, as the optimisation proceeds, the reference image sharpens. We see that the structures in the vicinity of the knotpoints are aligned, giving a clear distinction in the reference image between skull, CSF, and the brain surface. The brain structures far from the knotpoints (i.e., the ventricles and sulci) are only approximately aligned, as we would expect. Note also that the final reference does not have the same skull shape as any of the originals. In these results we have only shown the first stage in the registration – as in [6], the registration would be refined by adding more knotpoints, and then re-optimising.

3.2 Optimising the Reference Image

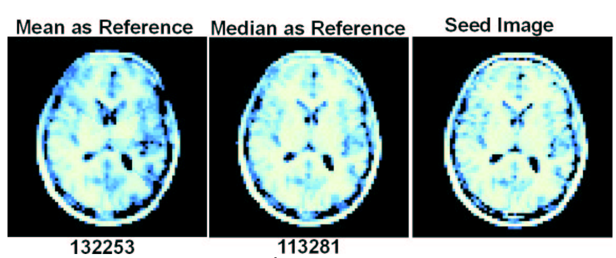

Figure 4.

We could have used one of the training examples itself as the reference image – however it is well known that changing the choice of reference can greatly change the final results when it comes to atlas construction. Bhatia et al. [2] perform groupwise registration to a varying spatial reference, yet use a fixed example from the training set as the intensity reference. The problem with such a fixed choice of intensity reference is illustrated in the following example. We take a seed image of a brain slice, and generate a training set of transformed versions of this seed image by translating and re-sampling. We then obscure part of the brain in each training example, as is shown in Figure 3. It is obvious that using any of these training examples as the intensity reference (e.g., as in [2]) will give poor results, since none of the training examples contain all the structures present in the seed image. However, as can be seen from the Figure, aligning to the continually-updated mean produces good results, with all the examples being brought into the correct relative alignment. Note, however, that the MDL formulation is not limited just to the choice of the mean of the aligned images as the intensity reference – the values of the reference image are a part of the model, and so could theoretically be optimised over. This is illustrated in Figure 4, where we take the set of transformations given in the previous Figure, but rather than computing the mean, we instead compute the median of the aligned training examples. As can be seen, this not only gives a much smaller description length, but also gives a reference image that is much closer to the original seed image.

4 Discussion & Conclusions

This paper has described a novel objective function that enables true groupwise non-rigid registration. The objective function is based on the Minimum Description Length (MDL) principle, and in previous work [14] we have shown that all of the common objective functions used for image registration can be described as modelling choices within the MDL framework. Thus, although in our examples we have demonstrated results using only one imaging modality, the extension to multi-modal images involves merely a change of modelling choice. Similarly, the extension to non-scalar valued images is also possible.

In the experiments that we present in this paper the reference image is chosen to be the average image of the aligned training set (we consider the mean and median averages). We could also have refined this reference image using the MDL objective function, which may have further improved the results – this will be investigated in future work. The experiments that we have presented have clearly demonstrated the power of the method in that even when different regions of the images are masked off, the algorithm still converges to the correct answer, since there is sufficient information in the entire group. This is only possible because the reference image for both spatial and intensity information is a function of *all* of the images in the group, in contrast to work such as [2]. This paper has demonstrated a successful proof-of-concept for the groupwise objective function that we propose. Demonstrating the method on multi-modal images and in 3D does not provide any theoretical difficulties, and will be followed-up in the future.

Acknowledgements. Our thanks to A. Fitzgibbon for suggesting the missing data experiment. This research was supported by the MIAS IRC project, EPSRC grant number GR/N14248/01.

References

1. R.A. Baxter and J.J. Oliver. MDL and MML: Similarities and differences. Technical report TR 207, Department of Computer Science, Monash University, Australia, 1994.
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 ISBI 2004*, pages 908–911, 2004.
3. Rh. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. 3D statistical shape models using direct optimisation of description length. *LNCS*, 2352:3–20, 2002.
4. M. Ferrant, S. K. Warfield, C. R. G. Guttmann, R. V. Mulkern, F. A. Jolesz, and R. Kikinis. 3D image matching using a finite element based elastic deformation model. *Lecture Notes in Computer Science*, 1679:202–209, 1999.
5. A. Hagemann, K. Rohr, H. S. Stiehl, U. Spetzger, and J. M. Gilsbach. Biomechanical modelling of the human head for physically based, nonrigid registration. *IEEE Transactions on Medical Imaging*, 18(10):875–884, 1999.
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. S. Marsland, C. J. Twining, and C. J. Taylor. Groupwise non-rigid registration using polyharmonic clamped-plate splines. *Lecture Notes in Computer Science*, 2879:771–779, 2003.
8. J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific Press, 1989.
9. J. A. Schnabel, C. Tanner, A. Castellano Smith, M. O. Leach, C. Hayes, A. Degenhard, R. Hose, D. L. G. Hill, and D. J. Hawkes. Validation of non-rigid registration using finite element methods. *Lecture Notes in Computer Science*, 2082:344–357, 2001.
10. C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423,623–656, 1948.
11. H.H. Thodberg. Minimum description length shape and appearance models. *Lecture Notes in Computer Science*, 2732, 2003.
12. C. J. Twining, S. Marsland, and C. J. Taylor. Measuring geodesic distances on the space of bounded diffeomorphisms. *Proceedings of the British Machine Vision Conference (BMVC)*, 2:847–856, 2002.
13. C. J. Twining, S. Marsland, and C. J. Taylor. Groupwise non-rigid registration: The minimum desription length approach. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2004.
14. C.J. Twining, S. Marsland, and C.J. Taylor. A unified information-theoretic approach to the correspondence problem in image registration. In *Proceedings of International Conference on Pattern Recognition (ICPR)*, 2004.
15. C. S. Wallace and P. R. Freeman. Estimation and inference by compact coding. *Journal of the Royal Statistical Society B*, 54(3):240–265, 1987.
16. Barbara Zitová and Jan Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:977 – 1000, 2003.