

# Groupwise Non-Rigid Registration: The Minimum Description Length Approach

Carole J. Twining<sup>1</sup> and Stephen Marsland<sup>1</sup>

## Abstract

The principled non-rigid registration of *groups* of images requires a fully groupwise objective function. We consider the problem as one of finding the optimal dense correspondence between the images in the set, where optimality is defined using the Minimum Description Length (MDL) principle, that the transmission of a model of the data, together with the parameters of that model, should be as short as possible. We demonstrate that this approach provides a suitable objective function by applying it to the task of non-rigid registration of a set of 2D T1-weighted MR images of the human brain. Furthermore, 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.

## 1 Introduction

There are many methods available for the non-rigid registration of pairs of images (for a review, see [18]). For some of the target applications of non-rigid registration, such as comparing to an atlas [4], it is sufficient to consider only pairwise registration. 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 [1] 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.

In intra-subject registration of medical images there is often some actual physical process determining the observed deformation, for example, tissue deformation due to patient position, the insertion of an external object such as a needle, or patient and organ motion. Alternatively, the deformation may be caused by atrophy, such as in dementia, or growth, as in a tumour. In either case, the most suitable choice of registration algorithm is one that closely models the underlying physical process, leading to physically-based registration algorithms (e.g., [8, 9]), or physically-based models (e.g., [13]) 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 process that generates the observed data. We therefore contend that in the absence

---

<sup>1</sup>Joint first authors. Email: carole.twining@man.ac.uk, s.r.marsland@massey.ac.nz

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) [12] and Minimum Message Length (MML) [17] principles are closely related approaches [3] 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 [7]. However, naïve attempts at extending the methods described there to images have not been successful [15]. This paper develops the ideas described in [2], and applies them to the complete problem of groupwise non-rigid registration of a set of images. The fundamental problem of groupwise non-rigid registration is to find a set of matching points that are defined across all of the images and that can be used to bring the images into better alignment. This is precisely the correspondence problem.

In this paper, we develop the application of the MDL principle to complete images by considering the transmission of a reference image (initially the mean of the set of aligned images) together with the parameters of the model that describes the set of deformations required to transform the reference image into each image in the set, together with any residual deformations. We begin by considering in detail the link between modelling and correspondence, and then describe how MDL can be used to provide an objective function for the non-rigid registration of images. This is followed by experiments showing the application of the method, and demonstrating that groupwise non-rigid registration can deal with significant absences in the data, where parts of the images have been removed.

## 2 Modelling and Correspondence

We will first consider the case of shape modelling. The scenario is that we are given a training set of shape examples, and we wish to represent all of these shape examples as specific instantiations of some parametric shape model. Conventional approaches such as the Statistical Shape Model (SSM) [6], or medial representations such as MREPS [11], represent the shapes in the training set as deformed examples of a single reference shape. This means that we have an explicit, consistent correspondence across all the shapes in the training set. In the case of point-based representations such as the SSM, the initial correspondence is provided by means of a set of manually-placed landmarks on each shape in the training set: this suffers from the problem that it is a time-consuming and subjective process, as well as being extremely difficult to perform for 2D shapes (surfaces) in 3D. In volumetric representations such as MREPS, the correspondence is implicit in the medial representation. Given a consistent correspondence across the training set, the reference shape is then conventionally defined as the mean shape across the training set, using an appropriate metric. Given the correspondence and the reference shape, the remaining part of the shape model is the set of deformations between the reference shape and the training set.

It is usual to first factor out the affine part of the deformations through the use of some alignment algorithm (e.g., Procrustes Analysis). The remaining non-rigid part of the set of deformations is then usually represented in some convenient dimensionally-reduced fashion – for example, in the case of the SSM, the set of shape deformations is represented as a multivariate Gaussian. The final shape model then consists of the

reference shape, the parameterised set of non-rigid deformations of this reference shape, and the set of affine deformations. To allow for the fact that there may be some mismatch between the actual training shapes and their representation by the shape model, we also allow a set of residual deformations, which represent this discrepancy between the model representation and the actual shape.

This modelling approach can be extended to regions of interest in a set of training images, in approaches such as the Active Appearance Model (AAM) [5]. As previously, the model-building starts from a set of manually-placed landmarks on the boundary and interior of the region of interest. The reference now consists of a reference shape, and the image appearance (pixel values) within the reference shape. The deformations of this reference required to reproduce each training example now include both a spatial deformation of the reference and a pixel-value deformation of the reference appearance. The required deformations can be combined into a single statistical model, which allows for correlations between shape change and appearance change. Note that the sensible combination of the incommensurate quantities of spatial deformation and pixel-value deformation into a single model is only possible because we know the correspondence; the scaling between spatial and pixel-value deformation can be chosen so that both parts have equal variance. As in the shape case, the model-building process starts from a user-defined correspondence across the set of training images.

It should be clear from the previous description of shape modelling that altering the correspondence across the set of training shapes (whilst maintaining the representation of the shapes) will produce slightly different shape models. Therefore, if we have an objective function that allows us to compare models, we can – by varying the correspondence – find the optimal shape model, and hence the optimal correspondence for a particular set of training shapes. This was the approach taken in the Minimum Description Length (MDL) [12] approach to shape modelling [7], where it was found that the resulting models had improved performance compared to models built using other methods. The application of the MDL principle to model selection is described in the next section.

In summary, we see that the conventional approaches to both shape modelling and shape-and-appearance modelling rest on a definition of a dense correspondence across a set of training examples. In contrast, the aim of automatic non-rigid registration algorithms is to find a meaningful dense correspondence across a set of training images. This suggests that we should view groupwise non-rigid registration as a modelling problem, where the sought-for dense correspondence across the training set of images is one that produces the optimal model. How this model of images is constructed, and the criterion used to define the optimal model is the subject of the next section.

### **3 The MDL Principle for Model Selection**

We have seen that conventional approaches to modelling represent a set of training examples as deformations of some reference example. This naturally fits into the MDL approach to statistical inference when we consider transmitting a dataset (our training set) to a receiver. Rather than transmitting the data directly, we attempt to reduce the total length of the transmission by encoding the data using some model. If our data is quantized, this can obviously be done using a message of some finite length. The optimal encoding of the data is then defined to be the encoding that has the shortest total transmission length, which is the *description length*. If we employ a model of the general form described in the previous section, then the total message consists of the following parts:

- The reference example (*For an SSM, this would be just the mean shape.*)
- The parameters of the model used to describe the set of deformations of the reference example (*SSM: this would be the set of modes of shape variation.*)
- The representation of each training example according to the model (*SSM: the set of shape vectors.*)
- Any residual deformations (*SSM: the affine transformations and any residual deformation.*)

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

$$\mathcal{L} = \mathcal{L}_{\text{ref}} + \mathcal{L}_{\text{params}} + \mathcal{L}_{\text{data:model}} + \mathcal{L}_{\text{residual}}. \quad (1)$$

### 3.1 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 [14] – 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:

$$\mathcal{L}_i = -\ln p_i \text{ nats}, \quad (2)$$

where the *nat* is the analogous unit to the *bit*, but using a base of  $e$  rather than base 2, so that  $e \text{ bits} \equiv 1 \text{ nat}$ . If our quantized data to be transmitted is encoded according to some parametric statistical model, this is equivalent to saying that the model assigns a non-zero, normalised probability to *every* possible quantized data value. Hence, the probability used in the above equation is the probability of the occurrence of this particular quantized data value according to 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 description length for the transmission of an unsigned integer can be approximated as follows. Consider a positive integer of the form  $n = 2^k$ ,  $k \in \mathbb{Z}^+$ , that, in binary, contains  $k$  bits. Hence:

$$\mathcal{L}_{\mathbb{Z}^+}(n) = k \text{ bits} = 1 + \text{int}(\log_2 n) \text{ bits} \approx \frac{1}{e} + \ln(n) \text{ nats}, \quad n \in \mathbb{Z}^+, \quad (3)$$

$$\mathcal{L}_{\mathbb{Z}}(n) = 2 + \text{int}|\log_2 n| \text{ bits} \approx \frac{2}{e} + \ln(n) \text{ nats}, \quad n \in \mathbb{Z}, \quad (4)$$

where the second expression for the description length for signed integers contains an extra bit for the sign.

As an example, we will consider 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  $\{I(A) : A = 1, \dots, N_I\}$  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 (which allows the receiver to construct the full image histogram), and then finally the ordered set of actual pixel values in the image, encoded using the histogram as model. The description length is hence:

$$\mathcal{L}_{\text{hist}} = -\sum_{\alpha} \ln\left(\frac{m_{\alpha}}{N}\right) + \sum_{\alpha} \mathcal{L}_{\mathbb{Z}^+}(N_{m_{\alpha}}) - \sum_{A=1}^{N_I} \ln p(I(A)), \quad (5)$$

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

### 3.2 The MDL Algorithm for Image Registration

The aim of non-rigid registration is to define a consistent spatial correspondence across the image set. We thus need to consider the spatial transformation between the original image planes/volumes and that of the reference image. The transformations between frames involved in the encoding and decoding processes are summarised in the diagram in Figure 1. We have a set of training images  $I_1, \dots, I_{n_s}$  and a reference image  $I_{\text{ref}}$ . There is also a set of diffeomorphic transformations  $\{t_i\}$  between the image plane/volume of the reference image and the image plane of each image in the set.

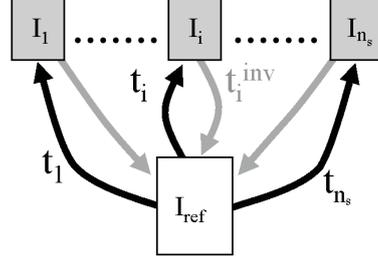


Figure 1: The set of transformations between reference and image frames.

It is this set of transformations that defines the consistent dense correspondence across the set of images. Defining a transformation  $t_i$  also defines the pullback transformation  $t_i^{\text{inv}}$ . Note however that it is *not* strictly required that  $t_i^{\text{inv}}$  is the *exact* inverse of  $t_i$ , only that it is also diffeomorphic, and that the transmitter and receiver both use the same algorithm to compute the set  $\{t_i^{\text{inv}}\}$  from the set  $\{t_i\}$ . The set  $\{t_i\}$  on its own is enough to define a consistent correspondence across the set, allowing us to find, for each point in the reference, the set of corresponding points across all the images. However, without an *exact* inverse,  $t_i^{-1}$  say, we cannot find all the points corresponding to a point in image  $I_i$ .

So, encoding the set of images proceeds as follows. The transmitter first decides on a set of transformations  $\{t_i\}$ . She then constructs the set  $\{t_i^{\text{inv}}\}$ , and maps each image  $I_i$  into the plane/volume of the reference. The image values from  $I_i$  are then resampled onto the regular grid  $X_{\text{ref}}$  of the reference to give the image  $\tilde{I}_i(X_{\text{ref}})$  (we assume that transmitter and receiver have previously agreed on a resampling scheme). The full set of resampled images in the frame of the reference  $\{\tilde{I}_i(X_{\text{ref}})\}$  is then averaged to create the reference image  $I_{\text{ref}}(X_{\text{ref}})$ . This reference image is transformed to the image plane of each image  $I_i$  in turn, and resampled onto the regular image grid  $X_i$  to give the image  $\tilde{I}_{\text{ref}}(X_i)$ , and the discrepancy image between the warped, resampled reference and the image  $I_i$  is computed,  $I_i^{\text{disc}}(X_i) = I_i(X_i) - \tilde{I}_{\text{ref}}(X_i)$ . The transmission then consists of the reference image  $I_{\text{ref}}(X_{\text{ref}})$ , the set of parameterised transformations  $\{t_i\}$ , and the set of discrepancy images  $\{I_i^{\text{disc}}(X_i)\}$ .

To decode the  $i^{\text{th}}$  image, the receiver decodes the reference image  $I_{\text{ref}}(X_{\text{ref}})$ , the warp  $t_i$ , and the  $i^{\text{th}}$  discrepancy image  $I_i^{\text{disc}}(X_i)$ . She then applies the transformation to the reference image, and resamples the warped reference on the regular image grid of image  $I_i$  (which we will assume is the same size as the grid of the discrepancy image) to create the image  $\tilde{I}_{\text{ref}}(X_i)$ . Adding the discrepancy image  $I_i^{\text{disc}}(X_i)$  to the image  $\tilde{I}_{\text{ref}}(X_i)$  then allows her to reconstruct the original image  $I_i(X_i)$  *exactly*. The description length for this encoding is given symbolically by:

$$\mathcal{L} = \mathcal{L}_{\text{trans}}(\{t_i\}) + \mathcal{L}(I_{\text{ref}}(X_{\text{ref}})) + \sum_{i=1}^{n_s} \mathcal{L}(I_i^{\text{disc}}(X_i)), \quad (6)$$

where  $\mathcal{L}_{\text{trans}}(\{t_i\})$  is the message length for transmitting the set of quantized parameters of the transformations, plus the set of quantization scales.

So, in this formulation, the only free parameters of the encoding are the set of transformations  $\{t_i\}$ . The optimum correspondence is hence that given by the set of transfor-

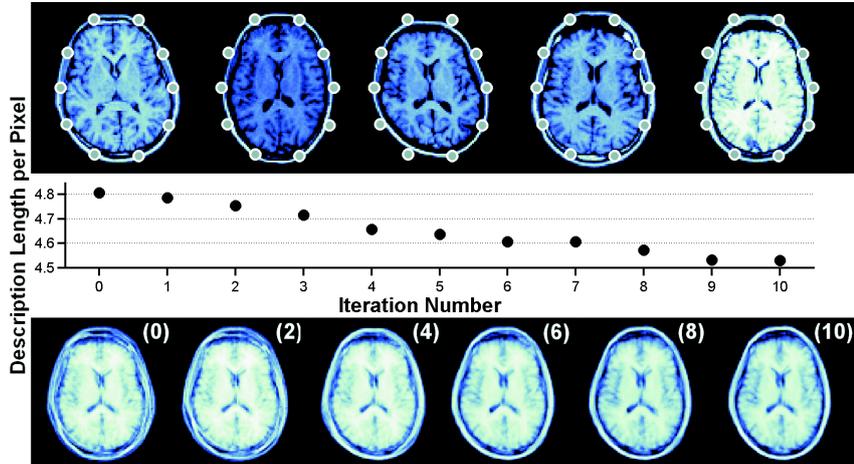


Figure 2: **Top Row:** The group of 5 images to be aligned, with the reference image knotpoints positions superimposed, **Second Row:** The description length divided by the total number of pixels in the group of images as a function of iteration number, **Bottom Row:** The mean/reference image at the start, and at the  $2^{nd}$ ,  $4^{th}$ ,  $6^{th}$ ,  $8^{th}$ , and  $10^{th}$  iterations.

mations that optimise the description length given in equation (6). As regards performing this optimisation, we note that it is possible to optimise the transformation for each image in the training set separately. This might appear to be just successive pairwise optimisations to the reference. However, it should be noted that the MDL objective function is truly a groupwise objective function; changing the transformation for one image in the set changes the reference image, and hence changes the discrepancy images for all the images in the set. This is a convenient method of optimisation, since the contributions of the other images to the reference remain unchanged. As only a single transformation  $t_i$  is varied, we need only re-calculate the pullback mapping  $t_i^{inv}$  for this particular transformation, which reduces the computation time. In the examples that follow, we will use the optimisation scheme that only one training image transformation at a time is optimised, where the particular training image is chosen at random from the set.

## 4 Experiments

### 4.1 Non-Rigid Registration

As an example, we take 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 parameterised set of transformations the polyharmonic Clamped-Plate splines (CPS) [16], which have successfully been used in non-rigid registration [10]. The CPS interpolates the motion of a set of knotpoints, hence the parameters of a transformation are the initial and final positions of those knotpoints. The boundary conditions on these splines are that the transformation vanishes smoothly on the surface of a ball, which in our case (2D), we take to be the circumference of the images. We choose to use the biharmonic CPS. Transmitting a spatial deformation is then equivalent to transmitting the positions of a set of knotpoints. We quantize the knotpoint positions to an accuracy  $\delta$ , with a range of possible positions equal to the size of the image, and a flat distribution over

this range; this then comprises the probabilistic model for the encoding of the knotpoint positions. We encode the reference image using the histogram encoding given earlier (5) with  $N = 256$  since we have 8-bit grayscale images. Because the number of training examples is small, we do not assume any relation between the discrepancy images for different training examples, and instead we transmit each discrepancy image according to its own histogram, shifting the data so that  $N = 512$  for the discrepancy images.

Following [10], we first generate a set of  $n_k = 10$  equi-angularly spaced knotpoints around the skull for each image. We then take the average positions of these points across the set as our reference image knotpoint positions  $\{x_\alpha^{\text{ref}}, y_\alpha^{\text{ref}}\}$ , 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 (as is shown in Figure 2), so that the transformation starts at the identity. We take each image in turn, and then take one knotpoint at a time, and optimise its position on this image. We use a fixed position accuracy of  $\delta = 0.05$  pixels.

As can be seen in Figure 2, as the optimisation proceeds, the reference image sharpens – after 10 iterations (that is, 2 passes through each image), we see that the skulls are aligned, giving a clear distinction in the reference image between skull, CSF, and the brain surface. These are the structures in the vicinity of the knotpoints. 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 [10], the registration would be refined by adding more knotpoints, and then re-optimising.

## 4.2 Optimising the Reference Image

Our choice to use the continually-updated mean as the reference image was initially motivated by the analogy that we drew with the standard approaches to the reference in shape-modelling. 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. [4] 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 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 aligned. Note that the final spatial reference is not fixed, but will vary depending on the order in which the transformations of the training examples are optimised.

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

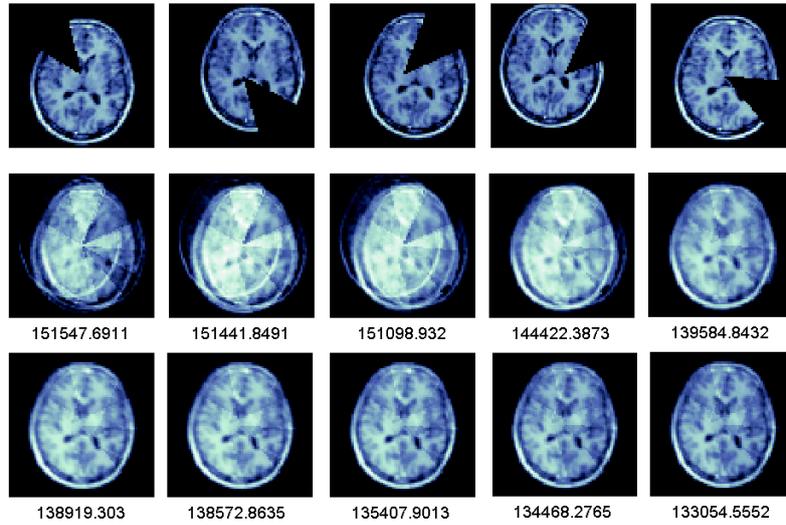


Figure 3: **Top Row:** The set of training images, **Other Rows:** The reference image as the registration progresses, with the value of the objective function (the *total* description length for the set in nats).

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. We would not necessarily expect to be able to reconstruct the reference image exactly, since the re-sampling will introduce some blurring. This result for the refined reference image show

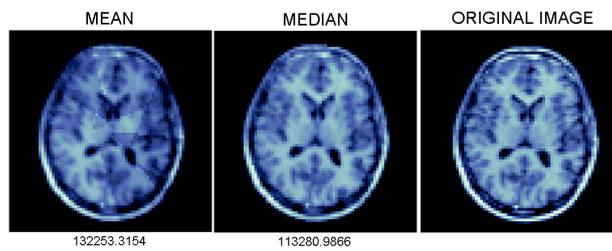


Figure 4: The mean and median of the aligned training set from Figure 3 compared to the seed (original) image. The value of the total description length for the two choices of reference is given below the image.

not only that we are able to correctly align a set of images, despite missing structures in each of the images, but also that the same MDL approach has allowed us to correctly extract the *union* of structures from the training set, not just the *commonality* of structure. This suggests possible links to the problem of super-resolution, but space does not permit this point to be pursued further here, although it will be considered in future work.

### 4.3 Comparing Different Classes of Model

In the examples given above, we used a single class of model, and showed that optimising the transformations  $\{t_l\}$  gave us a reasonable registration, whilst also optimising the

pixel-values of the reference image enabled us to integrate information across the training set. Both of these results can be seen as specific examples of optimising the parameter values for a given class of model. However, the MDL approach also allows us to compare different classes of model, since the description lengths can be compared directly.

This is illustrated in Figure 5. We take two 2D seed images from the BrainWeb database, chosen to be slices that are close together, so that they show the same structures. We generated two subsets of images by translating and re-sampling each seed image as before, united them to create our final training set. The results shown in Figure 5 compare describing the whole training set using a single reference to describing each subset separately, using the same registration algorithm as earlier. If we compare the description

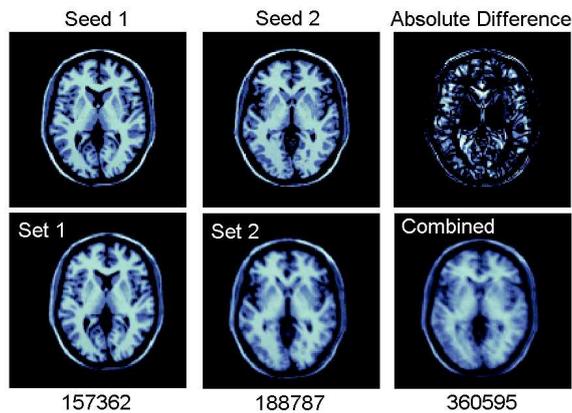


Figure 5: **Top Row:** The 2 seed images, and the absolute difference between them. **Bottom Row:** The reference images for the two sub-sets, and the combined set, with the total description length in nats.

lengths for the case of two reference images as opposed to just one, the cost with two reference images is 4% lower than the cost for the combined transmission – this is as expected, since the combined training set really only contains two independent images, that is, the original seed images.

## 5 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 [2] 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 other imaging modalities is merely a change of modelling choice.

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 con-

verges 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 [4]. 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 his useful suggestion.

## References

- [1] Anonymised. 2003.
- [2] Anonymised. 2004.
- [3] 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.
- [4] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert. Consistent groupwise non-rigid registration for atlas construction. *ISBI*, pages 908–911, 2004.
- [5] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *Lecture Notes in Computer Science*, 1407:484–498, 1998.
- [6] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models – their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] S. M. Pizer, D. Eberly, D. S. Fritsch, and B. S. Morse. Zoom-invariant vision of figural shape: The mathematics of cores. *Computer Vision and Image Understanding*, 69(1):055–071, 1998.
- [12] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific Press, 1989.
- [13] 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.
- [14] C.E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.
- [15] H.H. Thodberg. Minimum description length shape and appearance models. *Lecture Notes in Computer Science*, 2732, 2003.
- [16] C. J. Twining, S. Marsland, and C. J. Taylor. Measuring geodesic distances on the space of bounded diffeomorphisms. *British Machine Vision Conference (BMVC)*, 2:847–856, 2002.
- [17] 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.
- [18] Barbara Zitová and Jan Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:977 – 1000, 2003.