A Minimum Description Length Objective Function for Groupwise Non-Rigid Image Registration

Stephen Marsland [a,*],Carole J. Twining [b,*],Chris J. Taylor [b]

[a] *Institute of Information Sciences, Massey University, Private Bag 11222, Palmerston North, New Zealand.*

[b] *Imaging Science and Biomedical Engineering (ISBE), Stopford Building, University of Manchester, Manchester, M13 9PL, U.K.*

Abstract

Non-rigid registration finds a dense correspondence between a pair of images, so that analogous structures in the two images are aligned. While this is sufficient for atlas comparisons, in order for registration to be an aid to diagnosis, registrations need to be performed on a *set* of images. In this paper we describe an objective function that can be used for this *groupwise* registration. We view the problem of image registration as one of learning correspondences from a set of examplar images (the registration set), and derive a Minimum Description Length (MDL) objective function.

We give a brief description of the MDL approach as applied to transmitting both single images and sets of images, and show that the concept of a reference image (which is central to defining a *consistent* correspondence across a set of images) appears naturally as a valid model choice in the MDL approach.

In this paper we demonstrate both rigid and non-rigid groupwise registration using our MDL objective function on two-dimensional T1 MR images of the human brain, and show that we obtain a sensible alignment. The extension to the multi-modal case is also discussed. We conclude with a discussion as to how the MDL principle can be extended to include other encoding models than those we present here.

Key words: image registration, non-rigid registration, groupwise registration, minimum description length (MDL)

* Corresponding authors and joint first authors.
Email addresses: s.r.marsland@massey.ac.nz (Stephen Marsland), carole.twining@manchester.ac.uk (Carole J. Twining), chris.taylor@manchester.ac.uk (Chris J. Taylor).

1 Introduction

Reliable registration of medical images has the potential to assist greatly in medical diagnosis. Medical images in 2D and 3D from imaging modalities as disparate as x-ray and MRI can be brought into alignment though a combination of affine and non-rigid image warps, and the resultant 'deformation fields' can be analysed to find patterns characteristic of certain diseases.

There are a myriad of methods of automatic non-rigid image registration of pairs of images (e.g., [1–4]); see [5] for a review of general image registration algorithms, not limited to medical images. Such algorithms typically involve two independent choices: the objective function, the extremum of which defines what is meant by the 'best' correspondence between the images, and the representation of the deformation field that defines the dense correspondence between the images. The choice of representation of the deformation field applies implicit constraints on the possible deformations, and hence on the possible correspondences. The objective function is typically a sum of several terms – a voxel-based similarity measure, and terms that assign a cost to each deformation.

In our approach to non-rigid registration, we assume that the inferences we make about the data do not depend on hypothetical data-generating processes. This effectively means that we are considering inter-subject registration, i.e., images of different subjects, where there is no underlying physical process that generates the data. Hence, 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 (the set of images).

In intra-subject registration 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 may well be one that closely models the underlying physical process, leading to physically-based registration algorithms (e.g., [6,7]), or physically-based models (e.g., [8]) that can be used to evaluate the results of non-rigid registration algorithms.

We are focussing here on providing an assisted diagnosis system based on images from many different subjects — providing a useable diagnosis system will require a classifier to be trained on a very large number of medical images, including examples of the known disease groups and normal patients. Current registration methods work only on pairs of images, and so the registration algorithm will have to be run many hundreds of times, once for each image

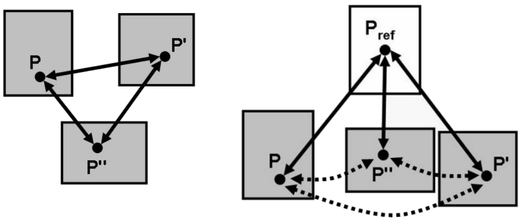

Fig. 1. *Left:* A consistent correspondence (solid arrows) between a set of images for a single point. *Right:* A consistent correspondence (dashed arrows) generated via correspondence to a reference (solid arrows).

against the chosen reference. Furthermore, the choice of reference image, which is used for the alignment of every image, is crucial. Pick a reference image that is not representative of the set, and not only will each registration take much longer, but the results will be biased.

There is another difficulty with this approach of using successive pairwise registration. The final classifier system will be based on statistical methods based on the distance between images. In previous work [9] we have shown that for these computed distances between images to be correct, the image warps all need to be based on the same set of knotpoints. Hence, it is not sufficient to use successive pairwise registrations, because the knotpoints computed in each image will be different. Thus, we believe that the only solution is to develop a groupwise registration algorithm that brings an entire group of images into registration simultaneously. In this way we can guarantee that the same knotpoints are used in each image, and that the reference image is correctly chosen to be representative of the full set.

The default assumption of non-rigid registration is that all structures present in any one image are present in all of the images, which means that the correspondence between any pair of images should be strictly one-to-one. For groupwise registration, this pairwise correspondence should be consistent across the whole set of images. While this is not necessarily true for multi-modal registration – something that is considered later in the paper – we assume that the differences are relatively small compared to the structures that do correspond. One way to represent consistent groupwise correspondence is by defining the correspondence between each image and some reference image, as shown in figure 1. The correspondence between any pair or set of images is that induced by this correspondence with the reference image, and is by definition consistent. This reference image need not be an image from the group, but if it is taken as an image from the group (or the mean of the currently-aligned images), then this consistency criterion gives us the naïve view of groupwise registration as successive pairwise registration, where the objective function is just the sum of the relevant pairwise objective functions. We consider a different approach here.

We present an objective function that is suitable for this problem of groupwise registration, and demonstrate its utility by integrating it with our existing image registration software [10]. In previous work [11] we have demonstrated that we can select common knotpoints across a set of images, and we now extend this with a groupwise objective function that is based on the Minimum Description Length (MDL) principle [12]. MDL is a model selection method that has been growing in popularity recently, including applications in the parameterisation of shapes defined by a set of points on their boundary [13] and statistical genetics [14]. However, the application of the method to images is certainly non-trivial, because there are several incommensurate terms that have to be amalgamated. This is one of the problems that is discussed in this paper. One benefit of using MDL is that all of the parameters of the non-rigid registration can be set out as modelling choices. This makes it a useful framework for the comparison of methods, and allows for the optimisation of the modelling choices within the algorithm, something that will be further developed as part of this research.

We begin the paper by discussing the links between modelling, correspondence, and image registration. This allows us to link image registration to shape and appearance modelling, since registration aims to compute a meaningful dense correspondence between a set of images, while shape and appearance modelling rely on such a dense correspondence being defined. We then introduce the MDL framework, beginning from first principles, before showing how MDL can be used to encode images. We consider the general framework of using MDL to encode images and then show how it can be used to encode individual images and sets of images. This provides us with the tools that we require to develop the full objective function, which is done in section 5. We then describe how the objective function can be implemented for rigid registration (section 5.2) and non-rigid registration (section 5.3), where the use of the objective function in complete non-rigid registration of groups of images is demonstrated. Section 5.4 gives the case where some parts of the images are obfuscated, so that successive pairwise registrations would not produce the correct results, but our groupwise registration does.

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) (as used in Active Shape Models [15]), or medial representations such as MREPS [16], 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 any affine/similarity transformation 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 using a multivariate Gaussian (which then gives a set of modes of variation). 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) [17]. 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.

Shape modelling is dependent upon the correspondence of the set of training shapes – altering the correspondence whilst maintaining the representation of the shapes will generate different shape models. Figure 2 gives a simple

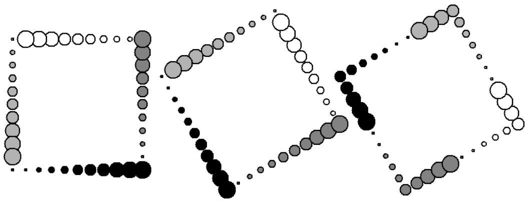

Fig. 2. **From the left:** Original shape, rotated shape with correct correspondence (as indicated by point size and colour), rotated shape with incorrect correspondence.

example. The sizes of the points and their shading indicate the correspondence, and so it can be seen that the correspondence between the first and second shapes is correct (a rotation of the points in the square of about 60 degrees clockwise has taken place). The alternative correspondence shown in the third image requires a much more complicated transformation, which will result in a significantly more complicated shape model being required to describe a set of such transformations. 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 [13], 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 Minimum Description Length (MDL)

The Minimum Description Length is a model-selection criteria. The MDL principle states that the best model to represent a set of given data is the one that requires the shortest message to transmit the data to some observer when the data is encoded using the specified model. The measure of message length that is used is the 'stochastic complexity' of the data [12]. In general, a complete message consists of two parts – the parameter values of the model, and the data encoded using the model. Code lengths (the length of the encoded message required to transmit one parameter, or one piece of data) are calculated using the fundamental result of Shannon [18] – 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:

$$\mathfrak{L}_i = -\ln p_i \text{ nats}, \tag{1}$$

where the 'nat' is the corresponding unit to the 'bit' when logarithms are taken to base e. The total message length/description length is then given by the sum of the parameter length and the data length:

$$\mathfrak{L} = \mathfrak{L}_{\text{para}} + \mathfrak{L}_{\text{data}}, \;\; \mathfrak{L}_{\text{data}} = \sum_i \mathfrak{L}_i, \tag{2}$$

where the parameter length $\mathfrak{L}_{\text{para}}$ is the sum of the code lengths for transmitting the set of parameter values of the model. The data length $\mathfrak{L}_{\text{data}}$ is minimised when the model probabilities $\{p_i\}$ exactly match the empirical distribution of the data. The MDL criterion minimises the description length $\mathfrak{L}$, balancing model complexity (as measured by $\mathfrak{L}_{\text{para}}$) against the degree of match between the empirical and model distributions.

Suppose that we wish to transmit a positive integer of the form $n = 2^k, k \in \mathbb{Z}^+$. Representing the number in binary form requires k bits, which can be also written in terms of n:

$$\mathfrak{L}_{\text{int}}(n) = k \text{ bits} = 1 + \text{int}\left(\log_2 n\right) \text{ bits}. \tag{3}$$

Converting to natural logarithms rather than base 2 for theoretical convenience, this then gives an approximate message length for transmission of an unsigned integer n of:

$$\mathfrak{L}_{\text{int}}(n) \approx \frac{1}{e} + \ln n \text{ nats}, \tag{4}$$

where we have converted to a continuum version of the function. 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.

Encoding a real number is not much more complicated, except that we have to include a parameter $\delta = 2^k$, $k \in \mathbb{Z}$ that describes the accuracy to which we wish to encode the real number. The description length for the real number and the accuracy δ are then given by:

$$\mathfrak{L}_{\text{real}}(x;\delta) = \mathfrak{L}_{\text{int}}\left(\frac{x}{\delta}\right) \approx \frac{1}{e} + \ln\left(\frac{x}{\delta}\right) \text{ nats}, \tag{5}$$

$$\mathfrak{L}(\delta) = 1 + \text{int}|\log_2(\delta)| \text{ bits} \approx \frac{1}{e}\left(1 + |\log_2(\delta)|\right) \text{ nats}. \tag{6}$$

It is important to note that in this final approximation, $\mathfrak{L}(\cdot)$ is *not* a continuous function – it is only defined for arguments $\delta = 2^k,\ k \in \mathbb{Z}$.

As an final example, and one that will be useful later, we will compute the description length of transmitting a quantized, pixellated grayscale image I with N pixels according to the image histogram of that image. We assume that the pixel-values $\{I(x) :\ x = 1, \ldots N\}$ are integers in the range $[1,\ M]$, and that there are n_m pixels in the image with pixel intensity m, with occupied bins situated at positions $\{m_\alpha\}$. Using this image histogram as the model, this gives the associated probability for each pixel being in histogram bin m as $p(m) = \frac{n_m}{N}$. The transmission then consists of the set of positions, $\{m_\alpha\}$, of the occupied bins (assuming a flat distribution over the allowed range, so that all pixel intensities are equally likely), the number of occupants of each bin, $\{n_{m_\alpha}\}$, (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 the combination of these three parts:

$$\begin{aligned}\mathfrak{L}_{\text{hist}} &= \sum_{\text{bins}} \mathfrak{L}_{\text{bin location}} + \sum_{\text{bins}} \mathfrak{L}_{\text{bin occupation}} + \mathfrak{L}_{\text{pixels}} \\ &= -\sum_{\alpha} \ln\left(\frac{1}{M}\right) + \sum_{\alpha} \mathfrak{L}_{\text{int}}(n_{m_\alpha}) - \sum_{x=1}^{N} \ln p(I(x)).\end{aligned} \tag{7}$$

4 Applying MDL to Images

In this section we consider how we can encode images within the MDL framework. Since MDL is a model-selection criteria, we can consider using a variety of different models, and then selecting the one that gives the shortest description length. We begin by investigating two different methods of encoding a single image, using either the image histogram, whose encoding was described in the previous section, or a single Gaussian. We then choose between these two encoding methods experimentally by encoding a series of different images. Following this, we examine how we can extend this to describe a set of images, either by using a reference image, or just transmitting the set of images. We will then be ready to describe our groupwise objective function in section 5.

4.1 Encoding a Single Image

Let us consider the simple case of transmitting one-dimensional ordered quantized[1] data $\{\hat{y}_i : i = 1, \dots N\}$, with quantization parameter Δ. We will suppose that the data is such that the mean is approximately zero, and that the receiver already knows this and knows the quantization parameter. However, we will not assume that the range of the data is known a priori.

The first model we will consider is one of the simplest parameterised models; we choose a Gaussian model, of zero mean and width $\hat{\sigma}$. The width is quantized using a parameter δ_σ, where δ_σ is restricted to the set of values $\{\delta_\sigma = 2^k : k \in \mathbb{Z},\ \delta_\sigma \leq \sigma\}$. Both δ_σ and $\hat{\sigma}$ have to be transmitted.

We will consider here just the case where[2] $\hat{\sigma} \gg \Delta$. The full set of model probabilities $\{p(\hat{y}) : \hat{y} = m\Delta,\ m \in \mathbb{Z}\}$ can then be approximated by:

$$p(\hat{y}) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} \int_{\hat{y}-\frac{\Delta}{2}}^{\hat{y}+\frac{\Delta}{2}} \exp\left(-\frac{\hat{y}^2}{2\hat{\sigma}^2}\right) \approx \frac{\Delta}{\sqrt{2\pi\hat{\sigma}^2}} \exp\left(-\frac{\hat{y}^2}{2\hat{\sigma}^2}\right) \tag{8}$$

$$\Rightarrow \quad \ln(p(\hat{y})) \approx \ln(\Delta) - \frac{1}{2}\ln(2\pi) - \ln(\hat{\sigma}) - \frac{\hat{y}^2}{2\hat{\sigma}^2}. \tag{9}$$

We hence obtain a complete description length of:

$$\begin{aligned} \mathfrak{L}_{\text{Gauss}}(\{\hat{y}_i\}) = \mathfrak{L}_{\text{int}}\left(\frac{\hat{\sigma}}{\delta_\sigma}\right) + \mathfrak{L}(\delta_\sigma) - \sum_{i=1}^{N} \log_2(p(\hat{y}_i)) \text{ bits} \\ \approx -N\ln\Delta + \frac{N}{2}\ln(2\pi) + \frac{1}{e} + N\ln(\hat{\sigma}) \\ + \ln\left(\frac{\hat{\sigma}}{\delta_\sigma}\right) + \mathfrak{L}(\delta_\sigma) + \sum_{i=1}^{N} \frac{\hat{y}_i^2}{2\hat{\sigma}^2} \text{ nats.} \end{aligned} \tag{10}$$

If we treat the quantized variable $\hat{\sigma}$ as a continuous variable σ (i.e., we take the limit $\delta_\sigma \to 0$) then, for fixed data, the optimum continuum value is given

[1] We will use $\hat{\cdot}$ to denote quantized variables.

[2] The case of Gaussian models where $\hat{\sigma} \approx \Delta$ is dealt with in [13], although only for the case of data where the range is known.

by:

$$\sigma_{\text{opt}}^2 = \frac{1}{N+1}\sum_{i=1}^{N} \hat{y}_i^2. \tag{11}$$

We would also like to be able to estimate the optimum value for δ_σ – that is, we wish to find a continuous function of δ_σ that approximates the discrete function given in (10). Note that there will be two types of terms involving δ_σ: the approximation of the $\log_2$ term, and terms arising from the quantization of σ_{opt}.

If $\delta_\sigma < 1$:

$$\mathfrak{L}\left(\delta_\sigma\right) \approx \frac{1}{e} - \ln\left(\delta_\sigma\right) \ \text{ nats.} \tag{12}$$

We know that:

$$|\hat{\sigma}_{\text{opt}} - \sigma_{\text{opt}}| \leq \frac{\delta_\sigma}{2}, \tag{13}$$

and that the data, and hence σ_{opt}, are *fixed*, whilst δ_σ, and hence $\hat{\sigma}_{\text{opt}}$, vary. We therefore take $\hat{\sigma}_{\text{opt}} = \sigma_{\text{opt}} + d_\sigma$, where we will assume that d_σ has a flat distribution within the range $|d_\sigma| \leq \frac{\delta_\sigma}{2}$. So, our *estimate* of functions $f(\hat{\sigma}_{\text{opt}})$ is:

$$f(\hat{\sigma}_{\text{opt}}) \approx \langle f(\sigma_{\text{opt}} + d_\sigma)\rangle_{d_\sigma} \approx f(\sigma_{\text{opt}}) + \frac{\delta_\sigma^2}{24} f''(\sigma_{\text{opt}}) + O\left(\delta_\sigma^4\right). \tag{14}$$

Then we find that:

$$\ln(\hat{\sigma}_{\text{opt}}) \approx \langle \ln(\sigma_{\text{opt}} + d_\sigma)\rangle_{d_\sigma} = \ln(\sigma_{\text{opt}}) - \frac{\delta_\sigma^2}{24\sigma_{\text{opt}}^2} + O\left(\delta_\sigma^4\right), \tag{15}$$

$$\frac{1}{\hat{\sigma}^2} \approx \left\langle \frac{1}{(\sigma_{\text{opt}} + d_\sigma)^2} \right\rangle_{d_\sigma} = \frac{1}{\sigma_{\text{opt}}^2} + \frac{\delta_\sigma^2}{4\sigma_{\text{opt}}^4} + O\left(\delta_\sigma^4\right). \tag{16}$$

Substituting from (12, 15, 16) into (10), we obtain:

$$\begin{aligned}\mathfrak{L}_{\text{Gauss}}(\{\hat{y}_i\};\delta_\sigma) \approx{}& \frac{2}{e} - N\ln(\Delta) + \frac{N}{2}\ln(2\pi) + (N+1)\ln(\sigma_{\text{opt}}) - \frac{(N+1)\delta_\sigma^2}{24\sigma_{\text{opt}}^2} \\ & - 2\ln(\delta_\sigma) + \frac{1}{2}\sum_{i=1}^{N}\hat{y}_i^2\left(\frac{1}{\sigma_{\text{opt}}^2} + \frac{\delta_\sigma^2}{4\sigma_{\text{opt}}^4}\right) + O\left(\delta_\sigma^4\right) \ \text{ nats.}\end{aligned} \tag{17}$$

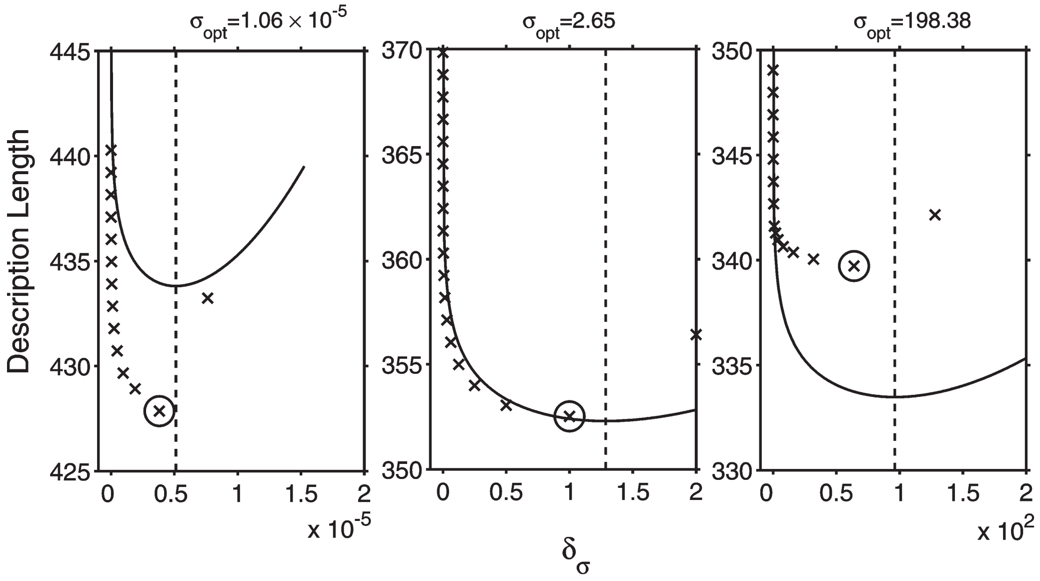

Fig. 3. Graphs showing description length as a function of δ_σ for 3 datasets with different variances, with $N = 50$. **Crosses:** the exact description length (equation (10)), with the minimum circled, **Solid line:** the continuum approximation (equation (19)), with the position of the minimum shown by the dashed line.

Hence, to lowest order, the optimum value of the parameter accuracy δ_σ is given by:

$$\delta_\sigma^2 = \frac{12\sigma_{\text{opt}}^2}{(N+1)}. \tag{18}$$

Substituting from (11) and (18) into (17) gives the final optimised form of the description length:

$$\begin{aligned}\mathfrak{L}_{\text{Gauss}}(\{\hat{y}_i\}) \approx & \frac{2}{e} - N\ln(\Delta) + \frac{N}{2}\ln(2\pi) + \frac{(N+3)}{2} \\ & + (N+1)\ln(\sigma_{\text{opt}}) - \ln\left(\frac{12\sigma_{\text{opt}}^2}{(N+1)}\right) \text{ nats.}\end{aligned} \tag{19}$$

In figure 3 we compare the exact expression for the description length (10) with the approximate continuous form (17) for 3 datasets of varying variance. Each dataset $\{\hat{y}_i\}$ consists of 50 quantized values randomly selected from a Gaussian distribution, with the mean being precisely zero. In each case, we use the calculated value of $\hat{\sigma}_{\text{opt}}$ or σ_{opt} as appropriate, since it was found that this gives an extremely good estimate of the true optimum value of σ, whatever the value of δ_σ. We can see from the figure that equation (18) gives a good order-of-magnitude estimate for the optimum value of δ_σ, across a range of values of σ_{opt} that covers 7 orders of magnitude, despite the fact that the approximate continuum expression was derived just for the case $\delta_\sigma < 1$, and despite the relatively small size of the dataset (i.e., $N = 50$). The minimum value for the description length given by equation (19) is also seen to be a reasonable estimate.

The Maximum Likelihood estimate for σ is given by optimising the sum of log probabilities from (9); the conventional error estimate for σ_{ML} is given by the Cramér-Rao-Frechet lower bound:

$$\sigma^2_{\text{ML}} = \frac{1}{N}\sum_{i=1}^{N} \hat{y}_i^2, \;\; \delta^2_{\text{CRF}} = \frac{\sigma^2_{\text{ML}}}{N}. \tag{20}$$

As we might have expected, the Gaussian model MDL estimate of σ_{opt} differs slightly from the Maximum Likelihood estimate, and ditto the estimates of the optimum value of δ_σ.

The second model we will consider consists of simply transmitting the histogram of the data as our model, with the bin widths given by the quantization scale of the data Δ. In terms of parameterised models, this is the most complex, since the model is exactly the empirical distribution of the data.

The set of occupied bin positions is given by $\{b_\alpha : \; b_\alpha = m_\alpha \Delta, \; m_\alpha \in \mathbb{Z}\}$, with occupancies $\{n_\alpha \geq 1\}$. Hence, the message length for transmitting all the parameters of the histogram (using equation (7)) is:

$$\begin{aligned}\mathfrak{L}_{\text{hist:param}} &= \sum_\alpha \left\{\frac{1}{e} + \mathfrak{L}_{\text{int}}(1+|m_\alpha|) + \mathfrak{L}_{\text{int}}(n_\alpha)\right\} \text{ nats} \\ &\approx \sum_\alpha \left\{\frac{3}{e} + \ln(1+|m_\alpha|) + \ln(n_\alpha)\right\} \text{ nats},\end{aligned} \tag{21}$$

giving a final description length of:

$$\begin{aligned}\mathfrak{L}_{\text{hist}} &= \mathfrak{L}_{\text{hist:param}} + \mathfrak{L}_{\text{hist:data}} \\ &= \mathfrak{L}_{\text{hist:param}} - \sum_\alpha n_\alpha \ln\left(\frac{n_\alpha}{N}\right).\end{aligned} \tag{22}$$

It should be noted that the data length $\mathfrak{L}_{\text{hist:data}}$ for transmitting data according to its empirical distribution is just the size of the dataset multiplied by the Shannon entropy $H(\{n_\alpha\})$ of the data histogram, where:

$$H(\{n_\alpha\}) = -\sum_\alpha \frac{n_\alpha}{N} \ln\left(\frac{n_\alpha}{N}\right). \tag{23}$$

So, in the context of MDL, the Shannon entropy is meaningful in its own right, and should *not* be considered as just a badly-behaved approximation to the differential entropy.

The description lengths for these 2 models (equations (19) and (22)) were then applied to a set of 8-bit ($=\frac{8}{e} \approx 2.943$ nats) greyscale images, with the data being centred before transmission. The description lengths per pixel for each image are shown in figure 4. The set of images consists of 3 images of ordinary objects, 2 medical images (a mammogram and a slice from a 3D MR image of a normal human brain), and an artificial image constructed from a set of independent Gaussian random variables. We can see that the description length per pixel is of the correct order compared to the greyscale resolution of the original images if we remember that the transmitted data has been centred – this then requires an upper bound of 2×8 bits ≈ 5.886 nats to transmit any such centred image without encoding. It can be seem that in all cases, even that of the Gaussian image, the increased parameter length for the model using the empirical distribution is more than compensated for by the exact fit to the data. This discrepancy is not due to any errors in approximating the optimum parameters for the Gaussian; the change in the Gaussian description length per pixel between taking the exact optimum value of δ_σ and equation (10), and that given by equation (19) was in all cases less than 3.0×10^{-5} nats.

As well as giving a smaller description length, encoding according to the empir-

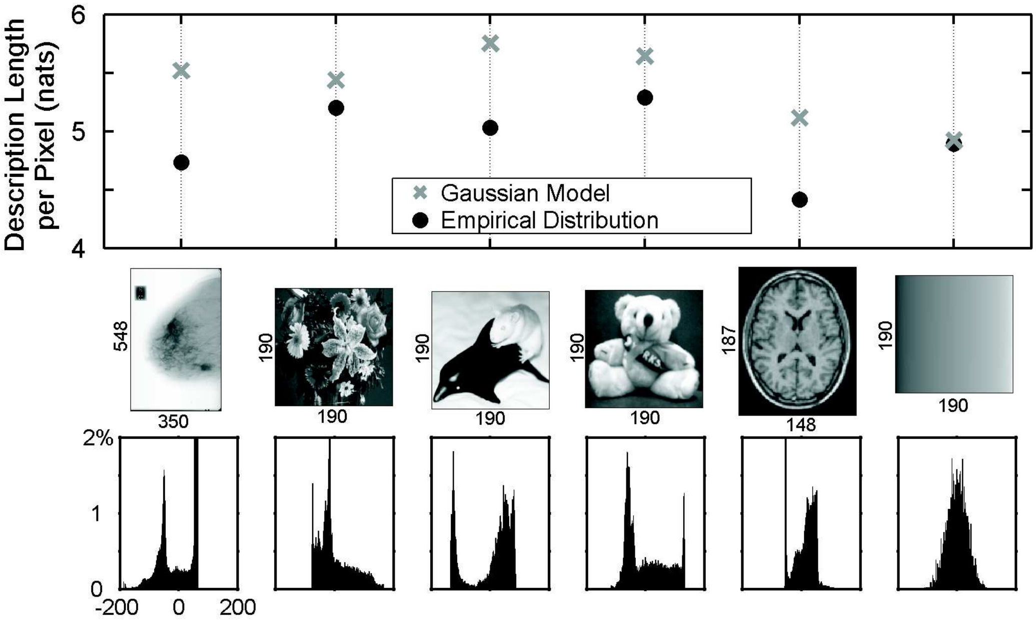

Fig. 4. *Top row:* The description lengths per pixel for a set of images, encoded using the 2 different models, **Optimised Gaussian:** grey crosses, **Empirical distribution:** black circles. *Middle row:* Thumbnails of the images with image dimensions in pixels, *Bottom row:* The centred image histograms, all to the same scale.

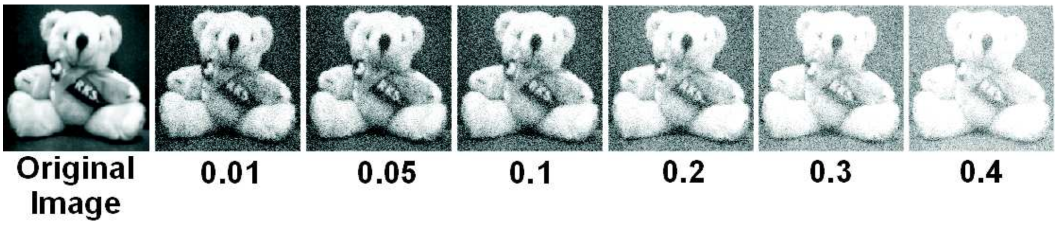

Fig. 5. The original image with various amounts of Gaussian white noise.

ical distribution potentially offers greater discrimination, given that the range of description lengths using this model (0.8743) is greater than the range obtained using the Gaussian model (0.8296). We therefore conclude that the appropriate description length for transmitting a single image is that given by the empirical distribution of the data.

4.2 Encoding a Set of Images

Having decided how to encode a single image, we next need to consider the problem of encoding a set of similar images. Conventional approaches to modelling represent a set of training examples as deformations of some reference example. This fits naturally 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. We will consider two different models: sending each image separately, and generating a reference image (the mean of the set) plus a set of 'discrepancy images' showing how each image in the set differs from the reference.

We investigate transmitting a set of aligned 8-bit greyscale images, and our experiments will use sets of images created by taking an original 190×190 pixel image and adding Gaussian white noise of varying variance (see figure 5 for examples). Images will be sent using the histogram encoding, except that we will now use the fact that the range of the image values is fixed, being $0 : 255$ for an ordinary greyscale image, and $-255 : 255$ for a discrepancy image. So, we take $\{n_\alpha \geq 1 : \alpha = 1, \ldots A\}$ as being the occupancies of the set of non-empty bins at positions $\{m_\alpha\}$, where M is now the width of the range of allowed bin positions. We take a flat distribution across this range when encoding the bin positions. This means that the expression for $\mathfrak{L}_{\text{hist:param}}$ from

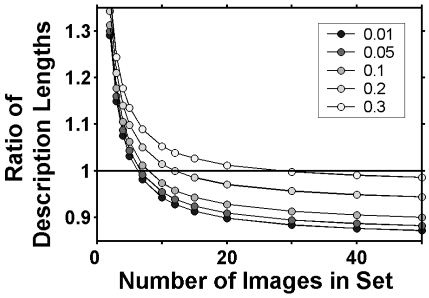

Fig. 6. The ratio of description lengths with and without an 8-bit reference image as a function of the number of images in the set, for varying values of the noise variance.

equation (21) now becomes:

$$\mathfrak{L}_{\text{hist:param}} = A \ln(M) + \sum_{\alpha=1}^{A} \left(\frac{1}{e} + \ln(n_\alpha) \right), \tag{24}$$

where the range M is 256 for greyscale images, and 512 for discrepancy images. $\mathfrak{L}_{\text{hist:data}}(\{n_\alpha\})$ is as given previously in equation (22).

The first question is whether sending a reference image and discrepancy images gives an advantage over sending the original images separately. To test this, we took sets of n_s noisy images, with the noise variance fixed. The reference image was taken as the mean of each set, with the same data resolution as the original images (i.e., 8-bit). We then computed the ratio of description lengths for transmission with and without a reference image, as a function of size of the set n_s, and as a function of the noise variance. The results are shown in figure 6. It can be seen that for all the values of noise variance considered, encoding using an 8-bit reference becomes advantageous (i.e., the ratio of description lengths is less than one) provided that the number of images in the set n_s is large enough. And, as we might have expected, the lower the noise variance, the lower the critical value of n_s.

In the approach described above, the reference image is considered as part of the model we are using to send the *set* of images. As the reference is taken to be the mean, it can be considered to consist of the information/structures that are common across the set. We therefore ask whether we are justified in using a full 8-bits (256 grey levels) to describe the reference. The question is answered by the graphs shown in figure 7. The variance of the noise was fixed at 0.2, and the number n_g of quantized grey levels in the reference was varied, whilst the range of the data was maintained. It can be seen that for all set sizes $n_s \geq 3$, there is an advantage to using a reference, provided n_g is suitable chosen. Furthermore, as the number of images in the set n_s increases, the optimum value of n_g also increases. However, for a set of 20 images, the optimal reference image encoding would take only 4 bits.

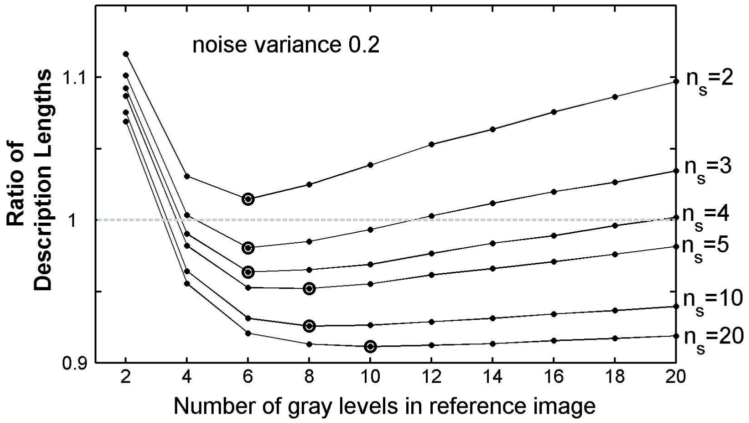

Fig. 7. The ratio of description lengths with and without a reference image as a function of the number of grey levels in the reference, for varying numbers of images in the set (n_s). The minimum point of each graph is circled.

5 An MDL Objective Function for Image Registration

5.1 Description of the Approach

We have shown in the preceding section that sending a set of (aligned) images encoded using a reference image of some type generally provides shorter description lengths than sending each image separately. This allows us to make the critical link between methods of image transmission and correspondence, as illustrated in figure 1: in some sense, the reference image contains the structures that are common to the set of *aligned* images, and it also allows us to define a consistent spatial correspondence across the image set, the generation of which is the aim of automatic non-rigid registration algorithms. If we employ a model of the general form described in the previous sections, then the total message consists of the following parts:

- The reference image
- The parameters of the model used to describe the set of deformations of the reference example
- The representation of each training example according to the model
- Any residual deformations

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

$$\mathfrak{L} = \mathfrak{L}_{\text{ref}} + \mathfrak{L}_{\text{params}} + \mathfrak{L}_{\text{data:model}} + \mathfrak{L}_{\text{residual}}. \tag{25}$$

The only additional factor we need to add is the spatial transformation between the original image planes/volumes and the reference image plane/volume,

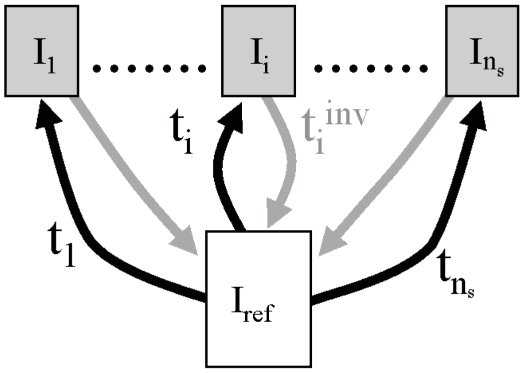

Fig. 8. The set of transformations between reference and image frames.

which is what is computed by the registration algorithm. The transformations between frames involved in the encoding and decoding processes are summarised in figure 8. We have a set of images $\{I_i : i = 1, \ldots n_s\}$ and a reference image I_{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. It is this set of transformations that defines the dense correspondence across the set of images, in the manner described in figure 1. Defining a transformation t_i also defines the pullback transformation t_i^{inv}. Note, however, that it is *not* strictly required that t_i^{inv} is the *exact* inverse of t_i, just 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}, we cannot find all the points corresponding to a point in image I_i.

Encoding the set of images then proceeds as follows. The transmitter first computes the initial reference image I_{ref}, as the mean of the image set, and computes image transformations $\{t_i\}$ to bring each the reference image into alignment with each image I_i. She then constructs the pullback mapping $\{t_i^{\text{inv}}\}$, 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_{ref} of the reference to give the image $\widetilde{I}_i(X_{\text{ref}})$ (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 $\{\widetilde{I}_i(X_{\text{ref}})\}$ is then averaged to create the *new* 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 $\widetilde{I}_{\text{ref}}(X_i)$, and the discrepancy image between the warped, resampled reference and the image I_i is computed, $I_i^{\text{disc}}(X_i) = I_i(X_i) - \widetilde{I}_{\text{ref}}(X_i)$. 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 $\{I_i^{\text{disc}}(X_i)\}$, which fits naturally into the scheme that was described previously.

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

$$\mathfrak{L} = \mathfrak{L}_{\text{params}}\left(\{t_i\}\right) + \mathfrak{L}\left(I_{\text{ref}}(X_{\text{ref}})\right) + \sum_{i=1}^{n_s} \mathfrak{L}\left(I_i^{\text{disc}}(X_i)\right), \tag{26}$$

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

The only free parameters of the encoding are the set of transformations $\{t_i\}$; a set of such transformations automatically defines the correspondence across the set of images. The optimum correspondence is then that given by the set of transformations that minimises the description length in equation (26). Finding these transformations is the task of the chosen registration algorithm. From an implementation point of view, it is important to note that we can optimise each transformation t_i individually; varying t_i alters the contribution of the i^{th} image to the mean, which alters the reference image, which hence alters the discrepancy images for all images in the set. So, although we can sequentially optimise the transformations, the effect is actually a fully groupwise one. Each iteration of the algorithm can thus correspond to optimising just one of the transformations $\{t_i\}$, which can significantly simplify the implementation, and the inclusion of our objective function into existing registration algorithms.

In the next sections we discuss suitable encoding schemes for transformations for both rigid and non-rigid registration algorithms, and provide a series of experimental results, demonstrating that the MDL objective function is suitable for the task of image registration.

5.2 Rigid Registration

To demonstrate the feasibility of the above scheme, we first consider the simplest case of a set of images produced by simply translating and resampling a single image. The transformations $\{t_i\}$ are then just translations, with parameters $\{x_i, y_i\}$. If we suppose that these are transmitted to an accuracy δ, and with some maximum modulus l, then the message length for the transformations is given by:

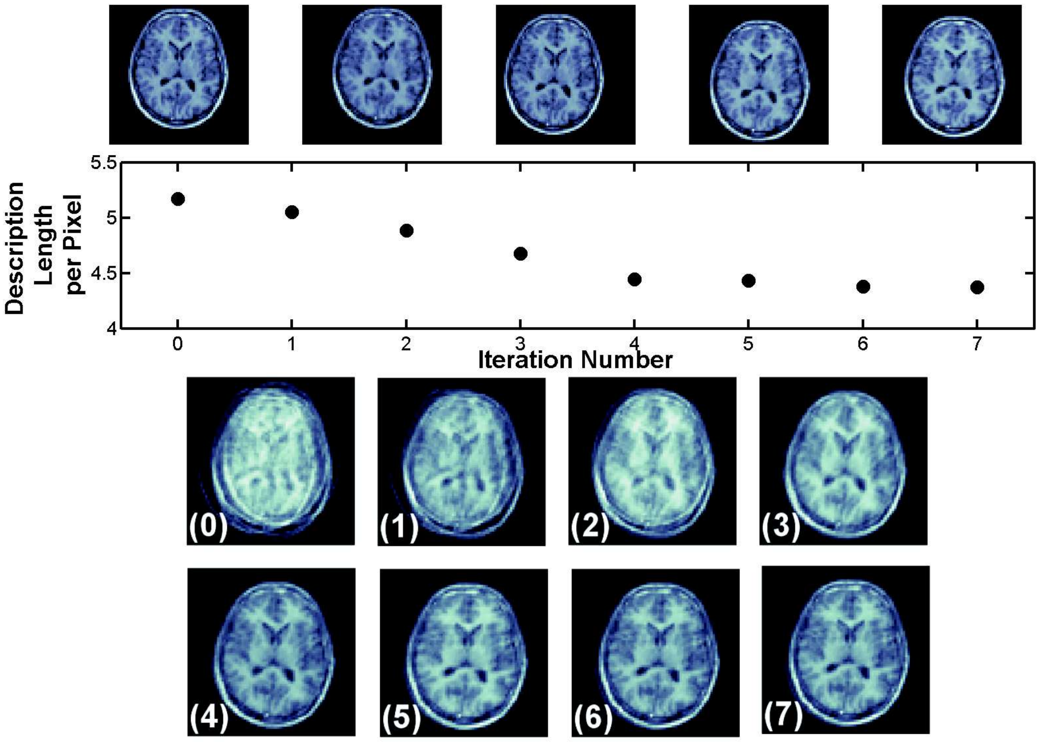

Fig. 9. *Rigid Groupwise Registration* **Top Row:** The group of 5 images to be aligned (translated versions of the same image). **Second Row:** The description length divided by the total number of pixels in the group of images as a function of iteration number, **Bottom Two Rows:** The mean/reference image at each iteration.

$$\begin{aligned}\mathfrak{L}_{\text{params}}\left(\{t_i\}\right) = {} & \underbrace{\left(\frac{1}{e} + |\ln(\delta)|\right)}_{\text{transmit } \delta} + \underbrace{\left(\frac{1}{e} + \ln\left(\frac{l}{\delta}\right)\right)}_{\text{transmit } l} \\ & + \underbrace{\sum_{i=1}^{n_s} 2\left[\frac{1}{e} + \ln\left(\frac{2l+1}{\delta}\right)\right]}_{\text{transmit } x_i, y_i} \text{ nats,} \qquad (27)\end{aligned}$$

and the images are individually transmitting using the histogram encoding described earlier (see equation (24)).

The results of such an optimisation for a set of $n_s = 5$ images are shown in figure 9. The images are 2D axial T1 MR slices of human brains. They are 8-bit grayscale images of size $N = 100 \times 100$. As we might have expected, the optimisation produces a good result after n_s iterations, i.e., one iteration of optimisation for each image. It is clear that the final reference image is exactly the generalization of the image set we would have expected, and that the algorithm converges to it despite the extremely poor quality of the initial reference image.

We now consider the case of full non-rigid registration. We choose as our parameterised set of transformations the polyharmonic Clamped-Plate splines (CPS) [9,19], 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 a the surface of a ball, which in our case (2D), we take to be the circumcircle of the images. We choose to use the biharmonic CPS.

We need to establish a spatial reference frame, which is equivalent to defining the knotpoint positions $\{x_\alpha^{\text{ref}}, y_\alpha^{\text{ref}} : \alpha = 1, \dots n_k\}$ on the reference image. Then the set of transformations $\{t_i\}$ is defined by specifying the knotpoint positions $\{x_\alpha^i, y_\alpha^i : i = 1, \dots n_s, \alpha = 1, \dots n_k\}$ on each image in the set. The description length for the parameters of the set of transformations $\{t_i\}$ is then:

$$\begin{aligned}\mathfrak{L}_{\text{params}}\left(\{t_i\}\right) = &\underbrace{\left(\frac{1}{e} + |\ln(\delta)|\right)}_{\text{transmit } \delta} + \underbrace{\left(\frac{1}{e} + \ln\left(\frac{l}{\delta}\right)\right)}_{\text{transmit } l} \\ &+ \underbrace{2(n_s+1)n_k\left[\frac{1}{e} + \ln\left(\frac{2l+1}{\delta}\right)\right]}_{\text{transmit } t_i} \text{ nats,}\end{aligned} \tag{28}$$

where, as before, l denotes the range of allowed values of the coordinates, and δ the accuracy, with the centre of the image circumcircle being the origin of coordinates. If we denote the CPS interpolant by $\omega\left(\{x_\alpha^{(0)}, y_\alpha^{(0)}\} \to \{x_\alpha^{(1)}, y_\alpha^{(1)}\}\right)$, then the transformations are given by:

$$t_i = \omega\left(\{x_\alpha^{\text{ref}}, y_\alpha^{\text{ref}}\} \to \{x_\alpha^i, y_\alpha^i\}\right), \quad t_i^{\text{inv}} = \omega(\{x_\alpha^i, y_\alpha^i\} \to \{x_\alpha^{\text{ref}}, y_\alpha^{\text{ref}}\}). \tag{29}$$

$t_i^{\text{inv}}$is **not** the exact inverse of t_i, but as mentioned previously, this does not matter! The CPS is not guaranteed diffeomorphic, however, we have found in practice that for these types of images, folding never occurs[3] .

In this and the following sections, we present examples of non-rigid registrations using our objective function. In all of the examples we use the same $N = 100 \times 100$ axial MR brain slices that were used in the rigid registration

[3] If images were used such that this became a problem, it should be noted that there is also a guaranteed *diffeomorphic* version of these splines [9]. This is discussed in section 6.

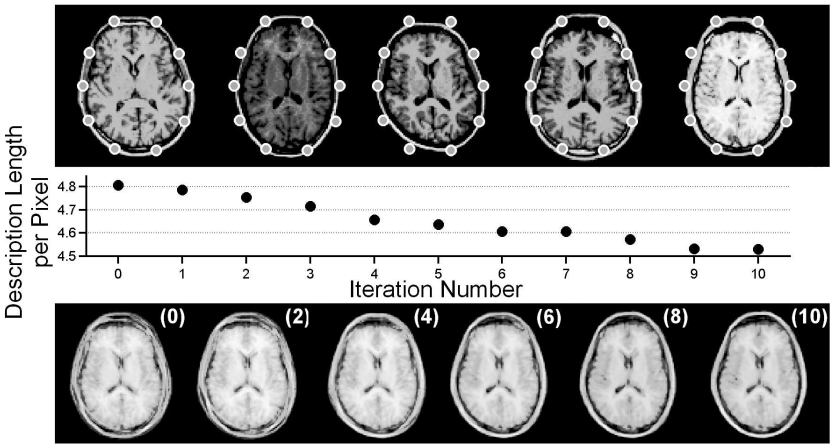

Fig. 10. *Non-rigid Registration* **Top Row:** The group of 5 images to be aligned, with the reference image knotpoints positions superimposed, **Second Row:** The description length divided by the total number of pixels in the group of images as a function of iteration number, **Bottom Row:** The mean/reference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

above, although in these experiments, the images have already been affinely aligned.

The CPS interpolates the motion of a set of knotpoints, so that the parameters of a transformation are the initial and final positions of those knotpoints (in the frame of the reference image). Transmitting a spatial deformation is then equivalent to transmitting the positions of a set of knotpoints. We quantize the knotpoint positions to an accuracy δ, 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 (7) with $M = 256$ 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 $M = 512$ for the discrepancy images. In future work we will examine whether it is possible to reduce the description length further by encoding the discrepancy images.

Our registration algorithm is based on that given in [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 $\{x^{\text{ref}}_\alpha, y^{\text{ref}}_\alpha\}$, 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 10), 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 10, 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.

5.4 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. [21] 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 11. It is obvious that using any of these training examples as the intensity reference (as in [21]) 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 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 12, 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

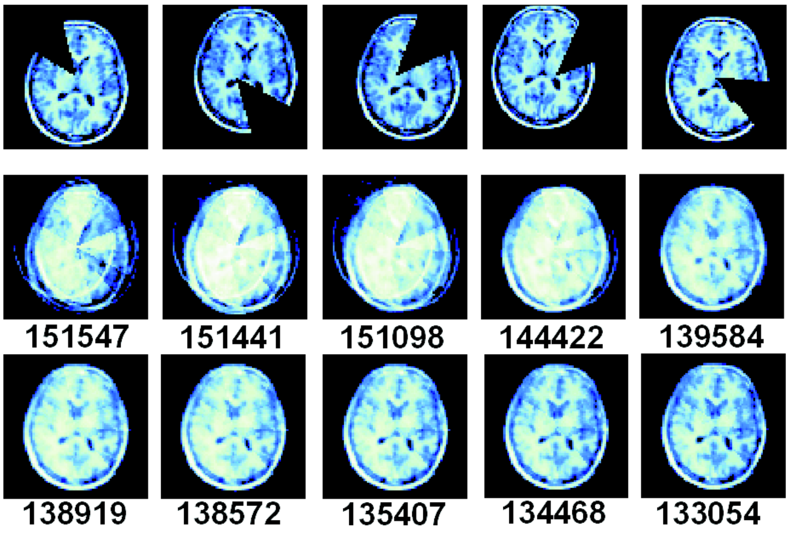

Fig. 11. **Top Row:** The set of training images, **Other Rows:** The reference image as the registration progresses, with the value of the objective function (the *total* description length for the set in nats).

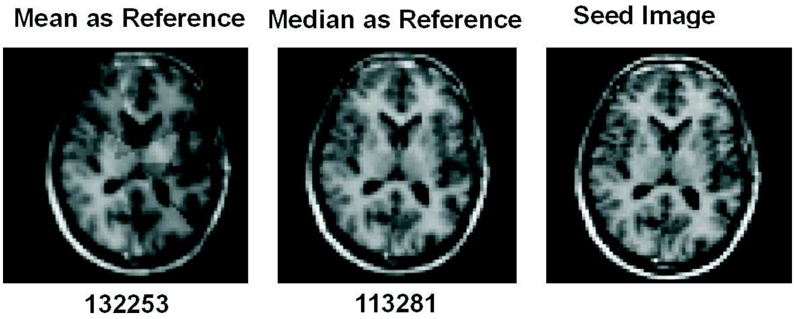

Fig. 12. The mean and median of the aligned training set from figure 11 compared to the seed (original) image. The value of the total description length for the two choices of reference is given below the image.

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 shows 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, which is something that will be considered in future work.

Continually computing the reference image as the mean is, however, a computationally expensive operation. We therefore investigated whether it is strictly necessary. Figure 13 shows a larger registration of $n_s = 7$ images. In this experiment, the reference image was not recomputed every iteration, but only

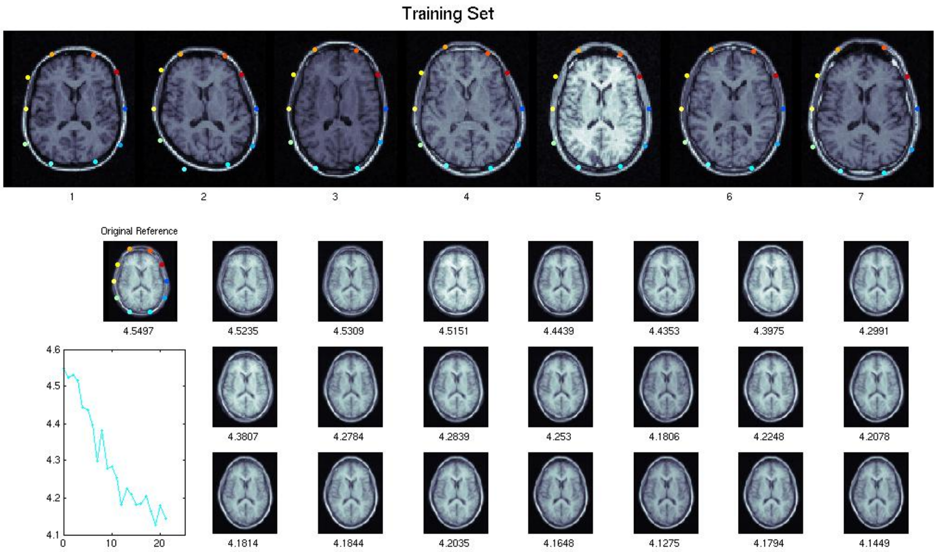

Fig. 13. *Non-rigid Registration* **Top Row:** The set of training images, **Other Rows:** The reference image as the registration progresses, with the value of the objective function (the *total* description length for the set in nats divided by the total number of pixels in all the images). The graph shows the objective function decreasing as the registration progresses.

after ever 3 iterations. This saves significant computational time, although it means that the optimisation is not always as efficient. One effect of this is that the objective function is not monotonic as a function of the iteration number, as can be seen in figure 13. Note that in this image set there are significant variation between the pixel intensities in the images, which would cause problems for an algorithm that only varies the spatial reference.

5.5 Comparing Different Classes of Model

In the examples given above, we used a single class of model, and showed that optimising the transformations $\{t_i\}$ 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.

For example, if our training set contains examples of different diseases as well as normals (with these images classified by an expert) then the MDL framework could be used to enable disease diagnosis from this, by finding suitable reference images from each different class. This could even be done as

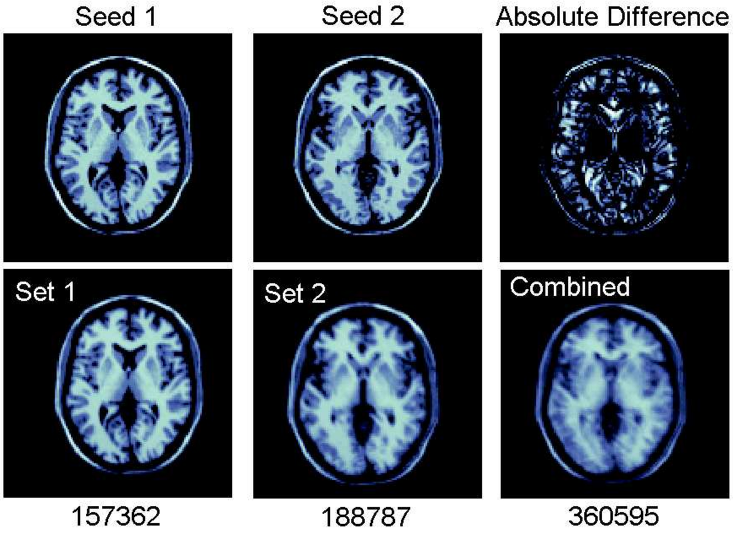

Fig. 14. **Top Row:** The 2 seed images, and the absolute difference between them. **Bottom Row:** The reference images for the two sub-sets, and the combined set, with the total description length in nats.

an unsupervised learning problem, so that the algorithm discovers the number of classes in the training set and identifies them. A simple example is illustrated in figure 14. We take two 2D $N = 129 \times 129$ seed images from the BrainWeb[4] 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 resampling each seed image as before, united them to create our final training set. The results shown in figure 14 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 lengths for the case of two reference images as opposed to just one, the summed 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. Investigating this further, and developing the unsupervised learning algorithm briefly mentioned above is an important part of our future work.

6 Discussion and Conclusions

We have described an objective function that is suitable for non-rigid registration of groups of images. The objective function comes from the MDL framework, which we have motivated by highlighting the links between image

[4] `http://www.bic.mni.mcgill.ca/brainweb/`

registration, where a correspondence between images is computed, and image modelling, where such a correspondence is assumed.

In this paper we have provided a proof-of-concept for this objective function for both rigid and non-rigid registration. We have used 2D T1-weighted MR images. The principal reason for this is that our current implementation is in Matlab, and a 3D experiments will require a compiled implementation of our algorithm. This is currently under development. However, the objective function that we have described, and the algorithm that we have used can be extended to 3D without any significant difficulties.

In addition, the extension to multi-modal images is also currently under investigation. The principal difficulty with multi-modal images is that there is not necessarily a one-one correspondence between the images. However, we have demonstrated in section 5.4 that our algorithm can deal with substantial differences between the images, even missing slices of the images. We have previously [22] developed a reference image/discrepancy image histogram encoding suitable for multi-modal images, thus, multi-modal image registration should not be a problem using this algorithm. As was identified in the previous section, we also plan to investigate whether the registration algorithm itself can be used as an unsupervised learning algorithm to cluster the images into different disease groups.

Another area that is still in need of further work is the optimisation. We currently use a line search method, which is inefficient. Investigations into a better optimisation scheme, possibly taking into account approximations to the gradient are also planned. This should significantly speed up the registration.

The principal motivation for our investigation of groupwise registration is the necessity for consistent landmarking across a group of images to allow image variation to be described numerically [9]. These computations are based on diffeomorphic warps between the images. In this paper we do not in fact compute diffeomorphic warps, the clamped-plate spline warps used in the registration algorithm are not guaranteed diffeomorphic. One possible solution is to instead use the geodesic interpolating clamped-plate spline that we have previously developed [9]. However, this would have significant computation costs as constructing the spline requires another optimisation with the optimisation of the final knotpoint positions. Instead, we believe that this analysis is something that should occur *after* the images have been brought into alignment. At this stage the full diffeomorphic warps can be computed, and the distances across the image set computed. Instead, we approximate the diffeomorphic warps using the clamped-plate spline, and we have not yet found examples in real images where this is a problem.

Acknowledgments

Our thanks to Andrew Fitzgibbon for suggesting the missing data experiments.

Stephen Marsland is funded by a Royal Society of New Zealand Marsden grant (04-MAU-015 MIS) (*"A Principled Approach to the Non-Rigid Registration and Structural Analysis of Groups of Medical Images"*), and Carole Twining is funded 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"*).

References

[1] M. Bro-Nielsen, C. Gramkow, Fast fluid registration of medical images, in: Proceedings of Visualization in Biomedical Computing (VBC), 1996, pp. 267–276.

[2] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, D. J. Hawkes, Non-rigid registration using free-form deformations: Application to breast MR images, IEEE Transactions on Medical Imaging 18 (8) (1999) 712–721.

[3] J. Gee, M. Reivich, R. Bajcsy, Elastically deforming 3D atlas to match anatomical brain images, Journal of Computer Assisted Tomography 17 (2) (1993) 225–236.

[4] C. Chefd'Hotel, G. Hermosillo, O. Faugeras, Variational methods for multimodal image matching, International Journal of Computer Vision 50 (3) (2002) 329 – 343.

[5] B. Zitová, J. Flusser, Image registration methods: A survey, Image and Vision Computing 21 (2003) 977 – 1000.

[6] M. Ferrant, S. K. Warfield, C. R. G. Guttmann, R. V. Mulkern, F. A. Jolesz, R. Kikinis, 3D image matching using a finite element based elastic deformation model, Lecture Notes in Computer Science 1679 (1999) 202–209.

[7] A. Hagemann, K. Rohr, H. S. Stiehl, U. Spetzger, J. M. Gilsbach, Biomechanical modelling of the human head for physically based, nonrigid registration, IEEE Transactions on Medical Imaging 18 (10) (1999) 875–884.

[8] J. A. Schnabel, C. Tanner, A. C. Smith, M. O. Leach, C. Hayes, A. Degenhard, R. Hose, D. L. G. Hill, D. J. Hawkes, Validation of non-rigid registration using finite element methods, Lecture Notes in Computer Science 2082 (2001) 344–357.

[9] S. Marsland, C. Twining, Constructing diffeomorphic representations for the groupwise analysis of non-rigid representations of medical images, IEEE Transactions on Medical Imaging 23 (8) (2004) 1006 – 1020.

[10] S. Marsland, C. J. Twining, Constructing data-driven optimal representations for iterative pairwise non-rigid registration, in: J. C. Gee, J. A. Maintz, M. W. Vannier (Eds.), Proceedings of the Second International Workshop on Biomedical Image Registration (WBIR), Vol. 2717 of Lecture Notes in Computer Science, Springer-Verlag, 2003, pp. 50–60.

[11] S. Marsland, C. J. Twining1, C. J. Taylor, Groupwise non-rigid registration using polyharmonic clamped-plate splines, Lecture Notes in Computer Science 2879 (2003) 771–779.

[12] J. Rissanen, Stochastic Complexity in Statistical Inquiry, World Scientific Press, Singapore, 1989.

[13] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, C. J. Taylor, 3D statistical shape models using direct optimisation of description length, Lecture Notes in Computer Science 2352 (2002) 3–20.

[14] P. Hraber, B. Korber, S. Wolinsky, H. Erlich, E. Trachtenberg, T. Kepler, HLA and HIV infection progression: Application of the minimum description length principle to statistical genetics, Santa Fe Institute Working Paper 03-04-023.

[15] T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham, Active shape models – their training and application, Computer Vision and Image Understanding 61 (1) (1995) 38–59.

[16] S. M. Pizer, D. Eberly, D. S. Fritsch, B. S. Morse, Zoom-invariant vision of figural shape: The mathematics of cores, Computer Vision and Image Understanding 69 (1) (1998) 55–71.

[17] T. F. Cootes, G. J. Edwards, C. J. Taylor, Active appearance models, in: H. Burkhardt, B. Neumann (Eds.), Proceedings of the European Conference on Computer Vision (ECCV), Vol. 1407 of Lecture Notes in Computer Science, Springer, 1998, pp. 484–498.

[18] C. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (1948) 379–423,623–656.

[19] C. J. Twining, S. Marsland, C. J. Taylor, Measuring geodesic distances on the space of bounded diffeomorphisms, Proceedings of the British Machine Vision Conference (BMVC), Cardiff, September 2002 2 (2002) 847–856.

[20] C. J. Twining, S. Marsland, Constructing diffeomorphic representations of non-rigid registrations of medical images, in: C. J. Taylor, J. A. Noble (Eds.), Proceedings of the 18^{th} International Conference on Information Processing in Medical Imaging (IPMI), Vol. 2732 of Lecture Notes in Computer Science, Springer-Verlag, 2003, pp. 413–425.

[21] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, D. Rueckert, Consistent groupwise non-rigid registration for atlas construction, ISBI (2004) 908–911.

[22] C. Twining, S. Marsland, C. Taylor, A unified information-theoretic approach to the correspondence problem in image registration, in: International Conference on Pattern Recognition (ICPR), Vol. 3, 2004, pp. 704 – 709.