Groupwise Diffeomorphic Non-rigid Registration for Automatic Model Building

T.F. Cootes, S. Marsland, C.J. Twining, K. Smith, and C.J. Taylor

Department of Imaging Science and Biomedical Engineering
University of Manchester, Manchester M13 9PT, UK
t.cootes@man.ac.uk

Abstract. We describe a framework for registering a group of images together using a set of non-linear diffeomorphic warps. The result of the groupwise registration is an implicit definition of dense correspondences between all of the images in a set, which can be used to construct statistical models of shape change across the set, avoiding the need for manual annotation of training images. We give examples on two datasets (brains and faces) and show the resulting models of shape and appearance variation. We show results of experiments demonstrating that the groupwise approach gives a more reliable correspondence than pairwise matching alone.

1 Introduction

We address the problem of determining dense correspondences across a set of images of similar but varying objects, a key problem in computer vision. Given such a set of images and their correspondences, an annotation of one image can be propagated to all of the others, and statistical shape models of the appearance and variations of the set of images can be built. Furthermore, a method of *automatically* determining the correspondences leads to a system capable of learning statistical models of appearance in an entirely unsupervised fashion.

Registration of pairs of images has been extensively studied for medical images, with many different non-rigid registration algorithms being proposed to deform one image until it matches a second, see for example [11]. These methods can be extended to finding correspondences across a *set* of images by registering each image in the set to a chosen reference image using pairwise methods [18]. However, only by examining a whole set of images of a class of objects can one learn which are the important features. We this cannot be determined from pairwise approaches alone. Following recent work on landmark correspondence for sets of shapes [6], we propose that the groupwise correspondence problem should explicitly optimise functions that measure the quality of the correspondence across the whole set of images simultaneously.

We believe that for non-rigid registration the warping functions should be continuous, smooth and invertable, so that every point in image A maps to exactly one point in image B, and vice-versa. Such smooth, invertable functions are known as *diffeomorphisms*. The mathematics of diffeomorphism groups is

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3024, pp. 316–327, 2004.
© Springer-Verlag Berlin Heidelberg 2004

complex and mysterious – they form infinite dimensional groups and smooth manifolds, yet are not Lie groups. However, usefully complex sets of diffeomorphic functions can be constructed by the composition of simple basis functions (see section 3.2). In cases where structures appear or disappear between one image and the next, these should be explicitly modelled as creation or destruction processes – such processes will not be addressed in this paper.

This paper proposes a general framework for computing diffeomorphisms that define dense correspondences across a set of images so as to minimise a groupwise objective function based on ideas of minimum message length. We will first introduce a novel pairwise algorithm capable of achieving a diffeomorphic mapping between a pair of images, and then generalise this to the groupwise case. We present results of applying the algorithm to sets of brain images and face images, show the statistical models of shape and appearance constructed from the correspondences and demonstrate that the groupwise method gives more reliable results than an equivalent pairwise approach.

2 Background

Finding mappings between structures across a set of images can facilitate many image analysis tasks. One particular area of importance is in medical image interpretation, where image registration can help in tasks as diverse as anatomical atlas matching and labelling, image classification, and data fusion. Statistical models may be constructed based on these mappings, and have been found to be widely applicable to image analysis problems [5,4]. However, variability in anatomy and in capture conditions – both inter-patient and intra-patient – means that identifying correspondences is far from straightforward.

The same correspondence problem is found in computer vision tasks (for example, for stereo vision tasks) and in remote sensing. This wide range of applications mean that many researchers have investigated image registration methods and the use of deformable models, for overviews see for example [22, 14,11,21].

Many algorithms algorithms have been proposed which are driven by the maximisation of some measure of intensity similarity between images, such as sum-of-squared-intensity differences, or mutual information [10,19].

Methods of propagating the deformations across the image include elastic deformations [1], viscous fluid models [3], and splines [18,12]. The method that is most similar to that proposed here is that of Lötjönen and Mäkelä [9], who describe an elastic matching approach in which spherical regions of one image are deformed so as to better match the other. However, the deformations do not have continuous derivatives at the border, and are therefore not diffeomorphic – in the work described in this paper we use a similar representation of deformation, but use fully diffeomorphic deformations within ellipsoidal regions. There are also similarities with the work of Feldmar and Ayache [7], who used local affine transformations to deform one surface onto another.

Davies *et al.* [6] directly addressed the problem of generating optimal correspondences for building shape models from landmarked data, noting improvements in model quality when a 'groupwise' cost function was used that was based on Minimum Description Length. Spherical harmonic parameters can also be used to directly optimise the shape parameterisation [15].

Some early work on groupwise image registration based on the discrepancy between the set of images and the reference image has been performed [13], and a groupwise model-matching algorithm that represents image intensities as well as shape has also been proposed [8]. It is also possible to consider building an appearance model as an image coding problem [2]. The model parameters are iteratively re-estimated after fitting the current model to the images, leading to an implicit correspondence defined across the data set.

3 Pairwise Non-rigid Registration

In this section we describe our approach to registering a pair of images based on the repeated composition of local diffeomorphic warps. The extension to registering groups of images is detailed in Section 4.

3.1 Overview

We first consider the registration of two images I_1 and I_2. This requires finding a non-linear deformation function that transforms (warps) image I_1 until it is as similar as possible (as measured by some objective function) to I_2.

More formally, we define the following:

Image functions $I_{1,2}$**.** We assume that the image functions are originally defined only on some dense set of points (pixels or voxels) with positions $\mathbf{X}_{1,2}$, but that we can interpolate such functions to obtain $I_{1,2}(\mathbf{x})$ at any point
A warp function $W(\cdot;\boldsymbol{\Phi})$ with parameters $\boldsymbol{\Phi}$ that acts on sets of points $\mathbf{X} \to W(\mathbf{X};\boldsymbol{\Phi})$
A sampled set of values $I(\mathbf{X})$ from an image at a set of points $\mathbf{X}$
An objective function $F_{pair}\left(I(\mathbf{X}), I'(\mathbf{X}')\right)$ that computes the 'similarity' between any two equi-sized samples
The 'cost' of a deformation $G_{pair}(W)$ for deformation $W(\cdot,\boldsymbol{\Phi})$

The task of image registration can then be considered as the task of finding parameters $\boldsymbol{\Phi}$ of the warping function $W(\cdot,\boldsymbol{\Phi})$ that minimise the combined objective function:

$$\boldsymbol{\Phi}_{\text{opt}} = \arg\min_{\boldsymbol{\Phi}} \Big(F_{pair}\big(I_1(\mathbf{X}_1), I_2\left(W(\mathbf{X}_1,\boldsymbol{\Phi})\right)\big) + G_{pair}\big(W(\cdot,\boldsymbol{\Phi})\big)\Big) \quad (1)$$

The form of $G_{pair}(\cdot)$ is typically chosen to penalise more convoluted deformations, and acts as a regularisation term.

3.2 Diffeomorphic Warps

We assume that each image in a set should contain the same structures, and hence there should be a unique and invertible one-to-one correspondence between all points on each pair of images. This suggests that the correct representation of warps is one that will not 'tear' or fold the images, we therefore choose to select warps from the diffeomorphism group. For more discussion of this point, see [20]. For any two diffeomorphisms $f(\mathbf{x})$, $g(\mathbf{x})$, their composition $(f \circ g)(\mathbf{x}) \equiv f(g(\mathbf{x}))$ is also a diffeomorphism. We can thus construct a wide class of diffeomorphic functions by repeated compositions of a basis set of simple diffeomorphisms.

In the Appendix we describe how such sets of *bounded* diffeomorphisms may be constructed. Boundedness is a useful property when performing numerical optimisation, as we can take advantage of the fact that only a subset of the image samples (those within the area of effect of the warp) will change. Our basis warps are parameterised by the movement of the centre point of the ellipsoid affected, and the size, position and orientation of the ellipsoidal region. We will denote the i^{th} such warp by $f_i = f(\cdot, \boldsymbol{\phi}_i)$, where $\boldsymbol{\phi}_i$ is the set of parameters for the i^{th} warp. The total warp is then $W(\cdot, \boldsymbol{\Phi}) = f_n \circ f_{n-1} \circ, \ldots f_2 \circ f_1 \circ A$, where $A(\cdot, \boldsymbol{\phi}_A)$ is an affine transformation with parameters $\boldsymbol{\phi}_A$, and the parameters of the total warp are $\boldsymbol{\Phi} = (\boldsymbol{\phi}_A, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n)$.

3.3 Optimisation Regime

The representation of complex warps requires many parameters, so that it is not feasible to optimise over all the parameters at once. We therefore adopt a sequential strategy in which we start with relatively simple warps and incrementally compose and optimise additional warps. In practice, we do the following:

- Optimise the affine registration parameters $\boldsymbol{\phi}_A$
- Construct the affinely-warped points $\mathbf{X}^{(0)} = A(\mathbf{X}_1, \boldsymbol{\phi}_A)$, the zeroth-order estimate of the warp and the associated warped points
- For each non-linear diffeomorphism $f_i = f(\cdot, \boldsymbol{\phi}_i)$, $i = 1, \ldots, n$:
  - For each given set of parameters $\boldsymbol{\phi}'_i$, apply the local warp $f(\cdot, \boldsymbol{\phi}'_i)$ to the current estimate $\mathbf{X}^{(i-1)}$ of the warped points, and sample from I_2 at these new points. This gives estimates of the fully-optimised warp $W(\mathbf{X}_1, \boldsymbol{\Phi}_{opt})$ and the true optimal sample $I_2(W(\mathbf{X}_1, \boldsymbol{\Phi}_{\text{opt}}))$
  - Find the particular parameters $\boldsymbol{\phi}_i$ that minimise the objective function given in Eq. (1), recalculating the estimate of the warp at each stage
  - Update the estimate of the full warp, $\mathbf{X}^{(i)} = f_i(\mathbf{X}^{(i-1)}; \boldsymbol{\phi}_i)$
- Output $\mathbf{X}^{(n)}$, the estimate of the true global optimum warp $W(\mathbf{X}_1, \boldsymbol{\Phi}_{\text{opt}})$

In our implementation we have used downhill simplex in the early stages and simple gradient descent in the later stages of the optimisation, although any non-linear optimiser could be used. We use a multi-resolution approach to give better robustness. The search regime is then defined by the positioning of the effective regions of the local warps. In the experiments presented here the regions to be warped were disks of randomly chosen position and radii. This approach is similar to that described in [9], but their local warps were not smooth at the edges of the regions of effect, nor did their construction guarantee diffeomorphisms or even C^1 differentiability, so that their total warp was not necessarily smooth or invertible.

4 Groupwise Non-rigid Registation

4.1 Groupwise Objective Functions

Suppose that instead of two images we now have a set of N images, I_i. We wish to register these images into a common coordinate frame. Following Davies *et al.* [6] we treat the problem of global registration as one of optimising a groupwise objective function that essentially measures the compactness of a statistical model built from correspondences resulting from the registration process. For registration, a suitable function for optimisation combines both shape (position) and intensity components.

The correspondence between a set of images explicitly defines those structures that should be treated as analogous. The contention is that statistical modelling of variations between analogous parts of structures should, in some sense, be 'simpler' than modelling variations between non-analogous parts. This idea of 'simplicity', or of appropriateness of the model, is expressed in the Minimum Description Length (MDL) framework [17] in terms of the length of an encoded message; this message transmits the whole set of examples, encoded by using the statistical model defined by the correspondence. Inappropriate choice of correspondence then leads to a non-optimal encoding of the data, and a greater length of the message. Note that this definition of optimal correspondence is explicitly concerned with the whole set of images, rather than correspondences defined between pairs of examples. We show below how we are able to construct an information-theoretic objective function for groupwise non-rigid registration.

Let $\mathbf{X}_i$ be the points on the i^{th} image obtained by applying the current estimate of the set of warp parameters for the warp between example I_i and the reference image I_1. That is, $\mathbf{X}_i = W(\mathbf{X}_1; \boldsymbol{\Phi}_i)$. Suppose also that $\mathbf{s}_i = I_i(\mathbf{X}_i)$ is the vector of image intensities sampled at those points. We then seek to find the set of full warp parameters $\{\boldsymbol{\Phi}_i\}$ that minimise some objective function

$$C_N(\langle \mathbf{s}_1, \mathbf{X}_1, \rangle \ldots, \langle \mathbf{s}_i, \mathbf{X}_i \rangle, \ldots, \langle \mathbf{s}_N, \mathbf{X}_N \rangle), \tag{2}$$

which is chosen to measure the appropriateness of the groupwise correspondence, potentially a challenging problem given the dimensionality of both the data set and the parametrisation of the warps.

In the work described here we will treat the shape and texture independently (although, ideally, correlations between shape and texture should also be considered). That is, we will use a function of the simplified form:

$$C_N(\langle \mathbf{s}_1, \mathbf{X}_1, \rangle \ldots, \langle \mathbf{s}_N, \mathbf{X}_N \rangle) = F_N(\mathbf{s}_1, \ldots, \mathbf{s}_N) + G_N(\mathbf{X}_1, \ldots, \mathbf{X}_N). \tag{3}$$

One information-theory based approach to the construction of such an objective function, which is suitable for problems where sequential optimisation is the only feasible optimisation strategy, is that of Minimum Message Length [17]. Each image example is left out in turn, and a model is built using the other $N-1$ examples and their current correspondences. The length of the message required to transmit the left-out example using this model is then calculated. We can then optimise this message length by manipulating the correspondence for the missing example.

To be specific, let $P_i^{(X)}(\mathbf{X})$ be an estimate of the model probability density function computed from the vectors of all corresponding points leaving out example i; that is, all the sets of points $\mathbf{X}_j$, $j \neq i$. Similarly, let $P_i^{(s)}(\mathbf{s})$ be an estimate of the model density function computed from all the texture sample vectors $\mathbf{s}_j = I_j(\mathbf{X}_j)$, $j \neq i$.

The estimated message length for transmitting example i with the correspondence defined by $\mathbf{X}_i = W(\mathbf{X}_1, \boldsymbol{\Phi}_i)$ then leads to the objective function:

$$C_i(\boldsymbol{\Phi}_i) = -\log P_i^{(X)}\Big(W(\mathbf{X}_1, \boldsymbol{\Phi}_i)\Big) - \lambda \log P_i^{(s)}\Big(I_i(W(\mathbf{X}_1, \boldsymbol{\Phi}_i))\Big), \quad (4)$$

where λ represents the relative weighting given to the shape and texture parts. By manipulating the warp parameters $\boldsymbol{\Phi}_i$, we manipulate the correspondence for image I_i relative to the rest of the examples, and can hence optimise this correspondence for this example by minimising the value of $C_i(\boldsymbol{\Phi}_i)$. This single-example warp optimisation is performed in an analogous fashion to the pairwise example given previously. We next describe the full groupwise optimisation strategy.

4.2 Groupwise Optimisation Algorithm

In what follows, we will use the term 'correspondence' to denote the correspondence between images induced by a warp $W(\cdot, \boldsymbol{\Phi})$; manipulating the set of warp parameters $\boldsymbol{\Phi}$ manipulates the correspondence. Since our warps are diffeomorphic by construction, all correspondences so defined are one-to-one and invertible. The groupwise optimisation algorithm is:

- Initialisation: perform pairwise non-rigid registrations between each image and I_1, giving initial estimate $\mathbf{X}_i^{(0)}$ for each.
- **REPEAT**
  - For each $i = 2, \ldots, N$
    - (Re)compute the model p.d.f.s $P_i^{(X)}(\mathbf{X})$ and $P_i^{(s)}(\mathbf{s})$, leaving out example i from the model building process
    - Find the optimal set of warp parameters $\boldsymbol{\Phi}_i$ that minimise $C_i(\boldsymbol{\Phi}_i)$
    - Update estimate of correspondence for example i using these optimal warp parameters, $\mathbf{X}_i \to W(\mathbf{X}_1, \boldsymbol{\Phi}_i)$
- **UNTIL CONVERGENCE**

5 Results of Experiments

We have applied the groupwise model building approach to examples of faces and 2D MR brain images. For the intensity part our chosen objective function is a sum-of-absolute-differences (implying an exponential PDF), a more robust statistic that sum-of-squares. In the examples given we have similar ranges of intensity across the set, so do not need to further normalise the intensity ranges. The shape component of the objective function is a sum of squares second derivatives of the deformation field (evaluated at each warped grid point $\mathbf{X}_i$) – this discourages excessive bending. In the experiments below we heuristically choose the factor λ that weights the shape objective function relative to the intensity measure. We are currently examining ways of automatically selecting suitable values.

We initialise $\mathbf{X}_1$ to a grid covering the reference image, and as before let $\mathbf{s}_i$ be the result of sampling image i at the current warped points. Let $d\mathbf{X}_i$ be the vector concatenation of all the second derivatives evaluated at each of the warped grid points. The pairwise objective function we use is

$$F_{pair}(i) = \sum_k |s_{ik} - s_{1k}| + \lambda_{pair}|d\mathbf{X}_i|^2. \quad (5)$$

During the groupwise stage, when optimising on image i, we use

$$F_{group}(i) = \sum_k \frac{|s_{ik} - \hat{s}_k|}{w_k} + \lambda_{group} d\mathbf{X}_i^T \mathbf{W}_i^{-1} d\mathbf{X}_i, \quad (6)$$

where $\hat{s}_k$ is the mean of the k^{th} sample across the other members of the set, w_k is the mean absolute difference from the mean, and $\mathbf{W}_i$ is a diagonal matrix describing the variances of the elements of $d\mathbf{X}_j$ for $j \neq i$. This simple gaussian model of the distribution is a natural groupwise extension of the commonly used bending energy term. It allows more freedom to deform in areas in which other images exhibit larger deformations.

5.1 Corresponding MR Brain Slices

We applied the method to 16 MR brain slices, each from an image of a different person, with approximately corresponding axial slices being chosen. The optimisation regime for the groupwise algorithm first requires finding the best affine transformation, before composing 1500 randomly sized and centred warps during the pairwise stage and a further 3000 randomly sized and centred warps during the groupwise stage. The algorithm is implemented in C++ [1] and the optimisation took about 15 minutes on a 2.8GHz PC.

Figure 1 shows the resulting deformation of one of the brains. We took a hand annotation of the reference image and used the acquired warps to propogate this to the other images. We then constructed a linear statistical shape and appearance model [4] from the resulting annotations. Figure 2 shows the two largest modes of shape deformation, while Figure 3 shows the two largest modes of combined shape and texture variation. Note that the shape model is built from the points of the projected annotation only, not on the dense grid of points used in the correspondence process. This allows us to use a sparser representation of the key features only, potentially leading to more compact models.

[1] Using the VXL computer vision library: `www.sourceforge.org/projects/vxl`

Such models give a compact summary of the variation used in the set, and can be used to match to further images using rapid optimisation algorithms such as the Active Appearance Model [4]. Note that the linear model does not enforce diffeomorphisms.

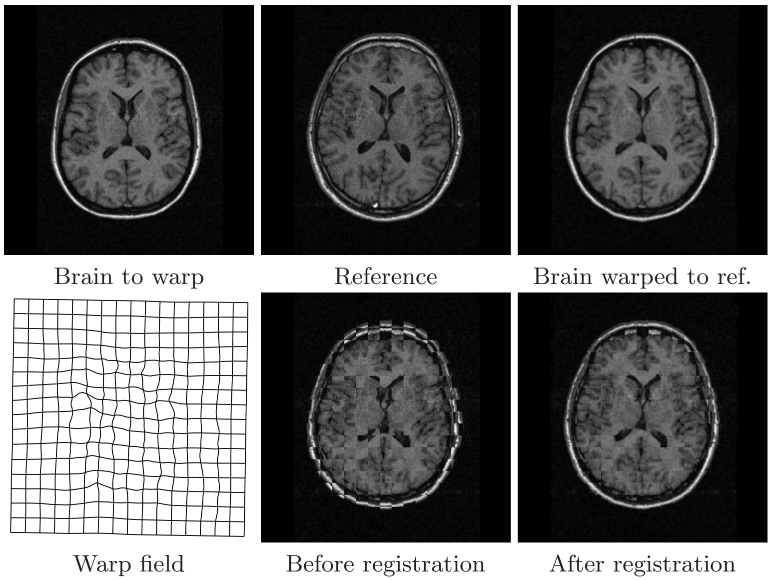


Fig. 1. Example MR slices before and after groupwise registration

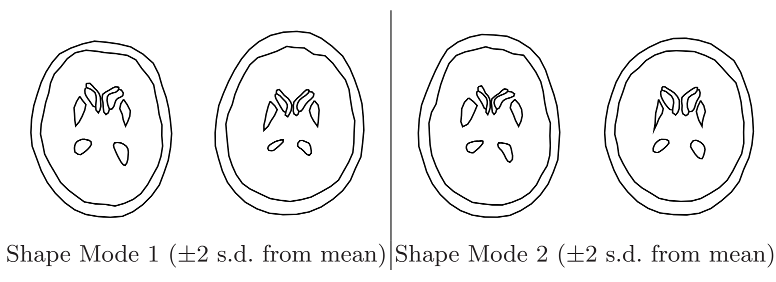
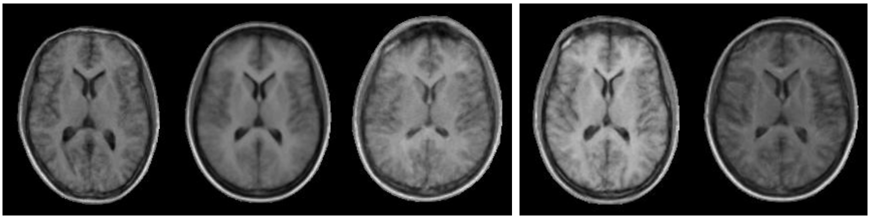


Fig. 2. Two largest modes of shape variation of a model built from 2D brain slices

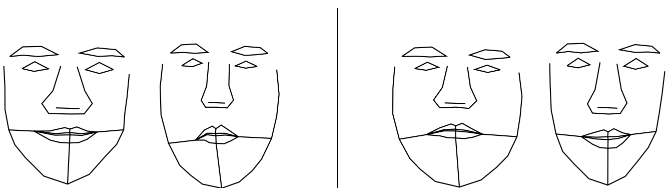


Mode 1 (−2 s.d.,mean,+2 s.d.) Mode 2 (±2 s.d. from mean)

Fig. 3. Two largest modes of appearance variation (model built from 2D brain slices)

5.2 Corresponding Face Images

We took 51 face images, each of a different person, from the XM2VTS face database [16] [2] We applied the groupwise registration to find correspondences, which took about 30 minutes on a 2.8GHz PC. As before, we propogated an annotation of the reference image to the rest of the set and constructed a linear model of appearance. Figure 4 shows the two largest modes of shape deformation, while Figure 5 shows the two largest modes of combined shape and texture variation. The crispness of the resulting appearance model demonstrates that an accurate correspondence has been achieved.



Shape Mode 1 (±2 s.d. from mean) Shape Mode 2 (±2 s.d. from mean)

Fig. 4. Two largest modes of shape variation of a model built from 51 face images

Mode 1 (−2 s.d.,mean,+2 s.d.) Mode 2 (±2 s.d. from mean)

Fig. 5. Two largest modes of appearance variation of a model built from 51 face images

[2] We selected the first 51 people without glasses or facial hair. Such features, which appear or dissappear from one image to another, break the assumptions of diffeomorphic correspondence in the process.

Table 1. Point-curve errors after registration of 51 face images. (The faces are approximately 100 pixels wide)

	Errors (pixels)		
	Mean	SD	Max
After initial pairwise	2.0	1.5	11.4
After full groupwise	2.0	1.0	7.1
Pairwise (same regime)	2.1	1.6	11.4

In order to evaluate the performance of the system, we compared the point positions obtained by transfering landmarks with those from a manual annotation of the 51 images. We measured the mean absolute difference between the found points and the equivalent curve on the manual annotation (see Figure 4). The results are summarised in Table 1. After the initial pairwise stage of the search (1100 warps) we obtain a mean accuracy of 2.0 pixels with an s.d. of 1.5 pixels. Completing the groupwise phase (a further 2000 warps) does not improve the mean but tightens up the distribution considerably, reducing both the variance and the maximum error. For comparison we ran a purely pairwise registration with the same number and distribution of additional random warps - the additional warps make little difference to the original pairwise result.

6 Discussion

We have presented a framework for establishing dense correspondences across groups of images using diffeomorphic functions and have demonstrated its application to two different domains. We have shown that in the case of the faces the groupwise method produces a more reliable registration than a purely pairwise approach.

We have described one example of objective functions, warping functions and optimisation regime, which appear to give good results. There is considerable research to be done investigating alternatives for each component of the framework. For instance, the groupwise function used above assumed diagonal covariance and may be improved with a full covariance matrix. Alternatively a statistical model of position, rather than derivatives, may lead to better results.

Similarly, the relative weighting between shape and intensity terms, λ, is somewhat arbitrary. For the pairwise case it is hard to select by anything other than trial and error. However, in the groupwise case, if the terms in the functions are related to log probabilities, it is possible to select λ_{group} more systematically – in the experiments we used a value of $\frac{1}{3}$ (there are 3 terms in the derivative vector for each element in the texture vector, and each term is normalised by its standard deviation).

The methods described above extend directly into three (and higher dimensions). The diffeomorphic warps of disks become warps of spheres (see Appendix A). We have used the techniques to register 3D MR images of the brain, and are currently evaluating the performance of the algorithms.

The general framework gives a powerful technique for registering images and for unsupervised shape and appearance model building. We anticipate it will have applications in many domains of computer vision.

Acknowledgements. The work described in the paper was done under the aegis of the MIAS Interdisiplinary Research Collaboration funded by EPSRC and MRC. The authors would like to thank other collaborators on the project for many useful discussions. The brains were gathered and annotated by C. Hutchinson.

References

1. R. Bajcsy, R. Lieberson, and M. Reivich. A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. *J. Comput. Assis. Tomogr.*, 7:618–625, 1983.
2. S. Baker, I. Matthews, and J. Schneider. Image coding with active appearance models. Technical Report CMU-RI-TR-03-13, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, April 2003.
3. G. E. Christensen, S. C. Joshi, and M. Miller. Volumetric transformation of brain anatomy. *IEEE Trans. Medical Image*, 16:864–877, 1997.
4. T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In H.Burkhardt and B. Neumann, editors, 5^{th} *European Conference on Computer Vision*, volume 2, pages 484–498. Springer, Berlin, 1998.
5. T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995.
6. R. Davies, C.Twining, T. Cootes, and C. Taylor. An information theoretic approach to statistical shape modelling. In 7^{th} *European Conference on Computer Vision*, volume 3, pages 3–20. Springer, 2002.
7. J. Feldmar and N. Ayache. Locally affine registration of free-form surfaces. In *CVPR94*, pages 496–501, 1994.
8. M. J. Jones and T. Poggio. Multidimensional morphable models : A framework for representing and matching object classes. *International Journal of Computer Vision*, 2(29):107–131, 1998.
9. J. Lötjönen and T. Mäkelä. Elastic matching using a deformation sphere. In *MICCAI*, pages 541–548, 2001.
10. F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, April 1997.
11. J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.
12. S. Marsland and C. Twining. Constructing data-driven optimal representations for iterative pairwise non-rigid registration. In *Biomedical Image Registration*, Lecture Notes in Computer Science, 2003.
13. S. Marsland, C. Twining, and C. Taylor. Groupwise non-rigid registration using polyharmonic clamped-plate splines. In *MICCAI*, Lecture Notes in Computer Science, 2003.

14. T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–108, 1996.
15. D. Meier and E. Fisher. Parameter space warping: Shape-based correspondence between morphologically different objects. *IEEE Trans. Medical Image*, 21:31–47, 2002.
16. K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSdb: The extended m2vts database. In *Proc. 2nd Conf. on Audio and Video-based Biometric Personal Verification*. Springer Verlag, 1999.
17. J. Rissanen. *Stochastic Complexity in Statistical Inquiry*, volume 15 of *Series in Computer Science*. World Scientific, Singapore, 1989.
18. D. Rueckert, A. Frangi, and J. Schnabel. Automatic construction of 3D statistical deformation models using non-rigid registration. In *MICCAI*, pages 77–84, 2001.
19. C. Studholme, C. Hill, and D. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32:71–86, 1999.
20. C. Twining, S. Marsland, and C. Taylor. Measuring geodesic distances on the space of bounded diffeomorphisms. In P.L.Rosin and D. Marshall, editors, 13^{th} *British Machine Vison Conference*, volume 2, pages 847–856. BMVA Press, Sept. 2002.
21. Y. Wang and L. H. Staib. Elastic model based non-rigid registration incorporating statistical shape information. In *MICCAI*, pages 1162–1173, 1998.
22. B. Zitová and J. Flusser. Image registration methods: A survey. *Image and Vision Computing*, 21:977–1000, 2003.

Appendix A: Bounded Diffeomorphisms

A useful class of bounded diffeomorphisms in arbitrary dimensions can be constructed using the following equation, which warps space only within the unit ball, based on the displacement of the centre by $\mathbf{a}$,

$$f(\mathbf{x};\mathbf{a}) = \begin{cases} \mathbf{x} + g(|\mathbf{x}|)\mathbf{a} & (|\mathbf{x}| < 1) \\ \mathbf{x} & \text{otherwise}, \end{cases} \tag{7}$$

where $\mathbf{a}$ is the position to which the origin is warped ($|\mathbf{a}| < 1$) and $g(r)$ is a smooth function satisfying the following properties: $g(0) = 1$, $g(1) = 0$, $g'(0) = 0$, $g'(1) = 0$. $f(\mathbf{x};\mathbf{a})$ is diffeomorphic providing that $|\mathbf{a}| < 1/d_{max}$, where $d_{max} = \max_{0<r<1}|g'(r)|$. This function is bounded, so that it deforms space only within the unit disc (2D) or unit sphere (3D). In the 2D case, if $g(r) = 1-r^2+r^2\log(r^2)$, then this is a Clamped Plate Spline with a single control point at the origin [20]. This function is guaranteed diffeomorphic provided that $|\mathbf{a}| < 0.25e$, providing a family of bounded diffeomorphisms parameterised by the point to which the origin is warped ($\mathbf{a}$).

The simplest polynomial form for $g(r)$ is $g(r) = (1 - r^2)^2$, which leads to an efficient implementation of the function in arbitrary dimensions. In this case we require $|\mathbf{a}| < 3\sqrt{3}/8 = 0.650$ for a diffeomorphism. By combining with a suitable affine transformation we can generate diffeomorphisms that only affect a particular ellipsoidal region of space.