A Unified Framework for Atlas Matching using Active Appearance Models

T.F. Cootes, C. Beeston, G.J. Edwards and C.J. Taylor

Imaging Science and Biomedical Engineering,
University of Manchester,
Manchester M13 9PT, U.K.
t.cootes@man.ac.uk
http://www.wiau.man.ac.uk

Abstract. [1]
We propose to use statistical models of shape and texture as deformable anatomical atlases. By training on sets of labelled examples these can represent both the mean structure and appearance of anatomy in medical images, and the allowable modes of deformation. Given enough training examples such a model should be able synthesise any image of normal anatomy. By finding the parameters which minimise the difference between the synthesised model image and the target image we can locate all the modelled structure. This potentially time consuming step can be solved rapidly using the Active Appearance Model (AAM). In this paper we describe the models and the AAM algorithm and demonstrate the approach on structures in MR brain cross-sections.

1 Introduction

It has been recognised for some time that the ability to match an anatomical atlas to individual patient images provides the basis for solving several important problems in medical image interpretation. Once the atlas has been matched to a particular image, structures of interest can be labelled and extracted for further analysis. Matching to an atlas also defines the registration between different images of the same patient - allowing information obtained at different times or from different imaging modalities to be combined - and the non-rigid registration of images of different patients - allowing population studies to be analysed in a common frame of reference. Same-patient data fusion is sometimes approached directly as a rigid registration problem (particularly in the brain) but the atlas matching approach is more general.

Given its central importance, the atlas-matching problem has received considerable attention. Two main approaches can be identified: landmark-based - in which key points or surfaces in image and atlas are brought into alignment; and image-based - in which an atlas image is allowed to deform to achieve as close a match as possible between corresponding pixel/voxel intensity values in

[1] This paper appears in Proc IPMI 1999, (LNCS 1613), pub. Springer. pp. 322-333

the deformed atlas and patient image. In either case, a dense correspondence is established between atlas and image, allowing labels and image values to be transferred between the two frames of reference. The landmark-based approach relies on extracting landmark points/surfaces on the basis of local image structure, then establishing correspondences between atlas and image landmarks. We have shown previously that this approach can be made efficient and robust if a statistical model of shape (representing the possible spatial arrangements of landmarks) is used to constrain the solution to the correspondence problem via an Active Shape Model [13]. Once landmark correspondences have been established a dense correspondence is obtained by interpolation. The image-based approach has the advantage that all the data are used in establishing the dense correspondence. To set against this is the disadvantage that shape statistics cannot easily be used in establishing the match - typically, arbitrary elastic or viscous regularisation terms are used to limit the degree of deformation allowed. Recently Wang and Staib [25] have described a method of incorporating statistical shape information into an image-based elastic matching algorithm. Although they show that this leads to more accurate results, shape and intensity matching are combined in an ad hoc way and the method is slow. In this paper we describe a unified approach to matching an atlas to patient images using both shape and intensity information. We show how a statistical appearance model (atlas), describing allowable variation in shape and intensity, can be constructed from a set of example images. We also describe an efficient Active Appearance Model (AAM) algorithm for matching the model to new images by minimising pixel/voxel intensity differences, subject to statistical constraints captured by the model. We illustrate the method applied to 2-D MR images of the brain, using an atlas containing all the important sub-cortical structures, and present quantitative results demonstrating that our method achieves accurate matching in a few seconds on a modern PC.

2 Background

The inter- and intra-personal variability inherent in biological structures makes medical image interpretation a difficult task. In recent years there has been considerable interest in methods that use deformable models, or atlases, to interpret images. One motivation is to achieve robust performance by using the atlas to constrain solutions to be valid examples of the structure(s) modelled. Of more fundamental importance is the fact that, once an atlas and patient image have been matched - producing a dense correspondence - anatomical labels and intensity values can be transferred directly. This forms a basis for automated anatomical interpretation and for data fusion across different images of the same individual or across similar images of different individuals. For a comprehensive review of work in this field there are recent surveys of image registration methods and deformable models in medical image analysis [19][18]. We give here a brief review covering some of the more important points.

Bajcsy *et. al.* describe an image-based atlas that deforms to fit new images by minimising pixel/voxel intensity differences [2]. Since this is an under-constrained problem, they regularise their solution by introducing an elastic deformation cost. Christensen *et. al.* describe a similar approach, but use a viscous flow rather than elastic model of deformation, and incorporate statistical information about local deformations [9] [8]. This results in more accurate matching, but is computationally expensive. Both approaches require good initialisation to converge to a satisfactory solution, since the deformations allowed are not constrained to be anatomically plausible. Landmark-based methods involve three steps: locating the landmarks, establishing correspondences, and warping the image or atlas to align the corresponded landmarks. Bookstein describes an elastic matching approach based on the use of thin plate splines [3] - he assumes that landmarks have been identified and corresponded manually. Subsol *et. al.* [23] extract crest-lines, which they use to establish landmark-based correspondence. They use these to perform morphometrical studies and to match images to atlases.

We have previously shown that a statistical deformation model can be used to simultaneously locate landmarks and establish image-atlas correspondences [12]. We obtain a parameterised statistical model of the domain of 'legal' shape variation from a set of training images. An Active Shape Model (ASM) is used to search for local image structure consistent with each of the landmarks, whilst constraining the configuration of landmarks using the statistical shape model. Typically, landmarks are closely spaced around the boundaries of structures of interest. A dense image-atlas correspondence can be established using thin plate splines. The original scheme was described in 2-D - it has been extended to 3-D by Hill *et. al.* [16] and Szekely *et. al.* [24].

None of the approaches outlined above is ideal. The use of a statistical deformation model allows rapid, reasonably robust matching and provides a principled basis for constraining deformation during matching. The ASM algorithm does not, however, use the image evidence particularly efficiently - only the intensity data in the vicinity of landmark points affects the final solution. The image-based approaches of Bajcsy *et. al.* [1] and Christensen *et. al.* [9] use the image evidence more efficiently, but allow arbitrary deformations. Wang and Staib [25] have recently attempted to incorporate statistical shape information into an image-based elastic matching approach. They do this by using a method very closely related to an ASM to find boundary landmarks in the image. An additional elastic matching term is added to the matching criterion, to encourage the image boundaries to coincide with the atlas boundaries. This is a rather *ad hoc* approach and the method is computationally expensive. In this paper we seek to unify the image-based and statistical modelling approaches in a principled way, leading to a method that is fast, robust and makes optimal use of both the image data and prior knowledge of the variability present in the class of images to be analysed.

The Active Appearance Model (AAM) approach that we describe also draws on other previous work. Cootes *et. al.* describe a model of the position-intensity surface, allowing full synthesis of the appearance of objects that are variable in

shape and intensity [11]. They do not, however, describe a plausible matching algorithm. Nastar *et. al.* describe a related model combining physical and statistical modes of deformation [20]. Although they describe a matching algorithm it requires very good initialisation. Jones and Poggio use a model capable of synthesizing faces and describe a stochastic optimisation method to match the model to new face images [17]. The method is slow but can be robust because of the quality of the synthesized images. Edwards *et. al.* also describe models of the combined shape and intensity appearance of faces [14]. They describe how the models can be matched to new images using an ASM; the method is fast, but does not make full use of the image data. Our new AAM approach is an extension of this idea, using all the information in the combined appearance model to match to the image. Sclaroff and Isidoro describe Active Blobs for tracking [22]. Their approach is similar to our AAM, though an Active Blob is derived from a single image rather than a training set of images. The example is used as a template, allowing low energy shape deformations and simple intensity variation. In contrast, AAMs learn what are valid shape and intensity variations from a training set.

3 Active Appearance Models

This section describes our statistical appearance models and outlines the basic AAM matching algorithm. A more comprehensive description is given in [10]. An AAM contains two main components: A parameterised model of object appearance, and an estimate of the relationship between parameter errors and induced image residuals.

3.1 Appearance Models

An appearance model can represent both the shape and texture variability seen in a training set. The training set consists of labelled images, where key landmark points are marked on each example object. For instance, to build a model of the sub-cortical structures in 2D MR images of the brain we need a number of images marked with points at key positions to outline the main features (Figure 1).

Given such a set we can generate a statistical model of shape variation by applying Principal Component Analysis (PCA) to the set of vectors describing the shapes in the training set (see [13] for details). The labelled points, $\mathbf{x}$, on a single object describe the shape of that object. Any example can then be approximated using:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_s \mathbf{b}_s \tag{1}$$

where $\bar{\mathbf{x}}$ is the mean shape vector, $\mathbf{P}_s$ is a set of orthogonal *modes of shape variation* and $\mathbf{b}_s$ is a vector of shape parameters.

To build a statistical model of the grey-level appearance we warp each example image so that its control points match the mean shape (using a triangulation

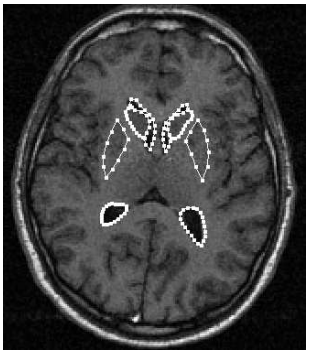


Fig. 1. Example of MR brain slice labelled with 123 landmark points around the ventricles, the caudate nucleus and the lentiform nucleus

algorithm). We then sample the intensity information from the *shape-normalised* image over the region covered by the mean shape. To minimise the effect of global lighting variation, we normalise the resulting samples.

By applying PCA to the normalised data we obtain a linear model:

$$\mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \mathbf{b}_g \tag{2}$$

where $\bar{\mathbf{g}}$ is the mean normalised grey-level vector, $\mathbf{P}_g$ is a set of orthogonal *modes of intensity variation* and $\mathbf{b}_g$ is a set of grey-level parameters.

The shape and appearance of any example can thus be summarised by the vectors $\mathbf{b}_s$ and $\mathbf{b}_g$. Since there may be correlations between the shape and grey-level variations, we concatenate the vectors, apply a further PCA and obtain a model of the form

$$\begin{pmatrix} \mathbf{W}_s \mathbf{b}_s \\ \mathbf{b}_g \end{pmatrix} = \mathbf{b} = \begin{pmatrix} \mathbf{Q}_s \\ \mathbf{Q}_g \end{pmatrix} \mathbf{c} = \mathbf{Q}\mathbf{c} \tag{3}$$

where $\mathbf{W}_s$ is a diagonal matrix of weights for each shape parameter, allowing for the difference in units between the shape and grey models, $\mathbf{Q}$ is a set of orthogonal modes and $\mathbf{c}$ is a vector of *appearance* parameters controlling both the shape and grey-levels of the model. Since the shape and grey-model parameters have zero mean, so does $\mathbf{c}$.

Note that the linear nature of the model allows us to express the shape and grey-levels directly as functions of $\mathbf{c}$

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}_s \mathbf{W}_s^{-1} \mathbf{Q}_s \mathbf{c} \ , \ \mathbf{g} = \bar{\mathbf{g}} + \mathbf{P}_g \mathbf{Q}_g \mathbf{c} \tag{4}$$

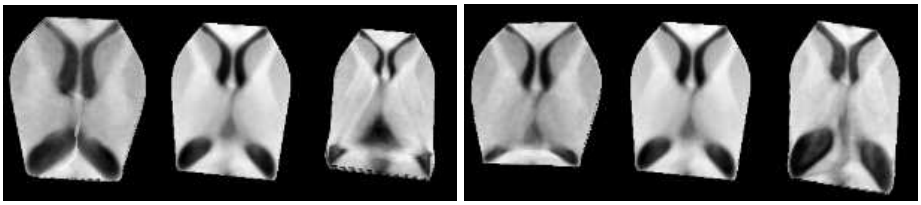
An example image can be synthesised for a given $\mathbf{c}$ by generating the shape-free grey-level image from the vector $\mathbf{g}$ and warping it using the control points described by $\mathbf{x}$.

For instance, Figure 2 shows the effects of varying the first two shape model parameters, b_{s1}, b_{s2}, of a model trained on a set of 72 2D MR images of the

brain, labelled as shown in Figure 1. Figure 2 shows the effects of varying the first two appearance model parameters, c_1, c_2, which change both the shape and the texture component of the synthesised image.

b_{s1} varies by ± 2 s.d.s b_{s2} varies by ± 2 s.d.s

Fig. 2. First two modes of shape model of part of a 2D MR image of the brain



c_1 varies by ± 2 s.d.s c_2 varies by ± 2 s.d.s

Fig. 3. First two modes of appearance model of part of a 2D MR image of the brain

3.2 Active Appearance Model Matching

We treat matching as an optimisation problem in which we minimise the difference between a new image and one synthesised by the appearance model.

Given a set of model parameters, $\mathbf{c}$, we can generate a hypothesis for the shape, $\mathbf{x}$, and texture, $\mathbf{g}_m$, of a model instance. To compare this hypothesis with the image, we use the suggested shape to sample the image texture, $\mathbf{g}_s$, and compute the difference, $\delta\mathbf{g} = \mathbf{g}_s - \mathbf{g}_m$. We seek to minimise the magnitude of $|\delta\mathbf{g}|$.

This is potentially a very difficult optimisation problem, but we exploit the fact that whenever we use a given model with images containing the modelled structure the optimisation problem will be similar. This means that we can learn how to solve the problem off-line. In particular, we observe that the pattern in the difference vector $\delta\mathbf{g}$ will be related to the error in the model parameters.

During a training phase, the AAM learns a linear relationship between $\delta\mathbf{g}$ and the parameter perturbation required to correct this, $\delta\mathbf{c}$,

$$\delta \mathbf{c} = \mathbf{A}\delta \mathbf{g} \tag{5}$$

The matrix $\mathbf{A}$ is obtained by linear regression on random displacements from the true training set positions and the induced image residuals (See [10] for details).

We can use (5) in an iterative matching algorithm. Given the current estimate of model parameters, $\mathbf{c}$, and the normalised image sample at the current estimate, $\mathbf{g}_s$, each iteration proceeds as follows:

- Evaluate the error vector $\delta \mathbf{g} = \mathbf{g}_s - \mathbf{g}_m$
- Evaluate the current error $E = |\delta \mathbf{g}|^2$
- Compute the predicted displacement, $\delta \mathbf{c} = \mathbf{A}\delta \mathbf{g}$
- Set $k = 1$
- Let $\mathbf{c}' = \mathbf{c} - k\delta \mathbf{c}$
- Sample the image at this new prediction, and calculate a new error vector, $\delta \mathbf{g}'$
- If $|\delta \mathbf{g}|^2 < E$ then accept the new estimate, $\mathbf{c}'$,
- Otherwise try at $k = 0.5$, $k = 0.25$ etc.

This is repeated until no improvement is made to the error, $|\delta \mathbf{g}|^2$, and convergence is declared.

We use a multi-resolution implementation, in which we iterate to convergence at each level before projecting the current solution to the next level of the model. This is more efficient and can converge to the correct solution from further away than search at a single resolution.

For example, Figure 4 shows an example of an AAM of the central structures of the brain slice converging from a displaced position on a previously unseen image. The model could represent about 10000 pixels and had 30 $\mathbf{c}$ parameters. The search took about a second on a modern PC. Figure 5 shows examples of the results of the search, with the found model points superimposed on the target images.

4 Results of Experiments

We have applied our approach to 2D slices taken from similar positions in 28 3D MR images of the brain. The in-slice resolution is 1mm and the between slice resolution 1.5mm. A total of 72 slices were used, two or three from each brain image. Ground truth for the structures of interest (ventricles, caudate nucleus and lentiform nucleus) was annotated by hand using expert radiologist input.

A set of 'leave-one-brain-out' experiments were performed to test the performance of our appoach.

We trained a model using all the examples except those from one brain, then ran the AAM to convergence on each of the excluded slices. We measured the quality of fit of the texture model, and the errors in the model point positions compared to the original labelling. We missed out each brain in turn, and averaged the results.. Table 1 summarises the results. It includes the results of

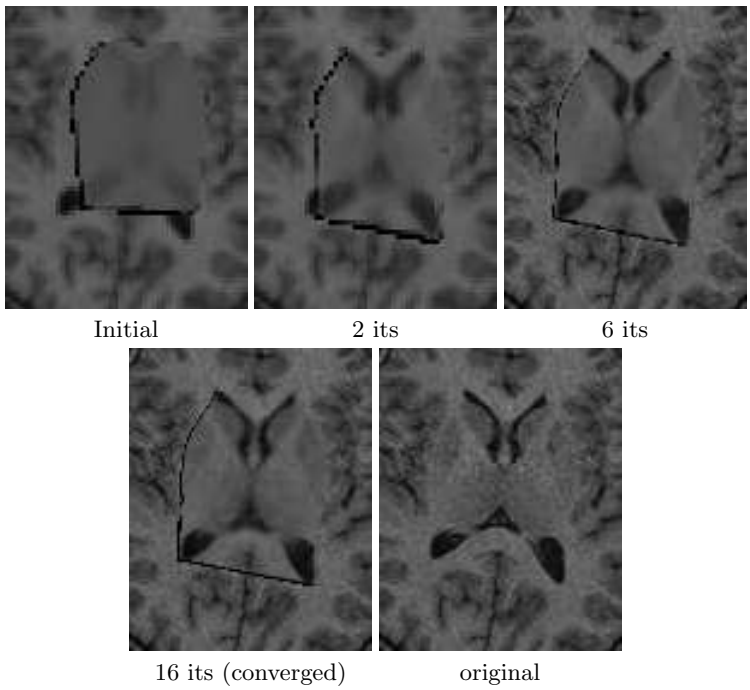


Fig. 4. Multi-resolution AAM search from a displaced position

'leave-all-in' experiments for comparison, in which the model was used to search the training set. This gives an upper bound on performance.

In addition we give the errors obtained when the model is fit directly to the labelled points - the 'best fit' column. This gives a measure of the best possible model fit.

The texture difference is given as the RMS difference between the intensities synthesised by the model and those in the target image over the modelled region. The units are those of grey-level. The full range of grey-levels in the image was about 140 units, with noise of about 7 units (s.d.). Notice that in the miss-1-out experiments the texture error found by search is better than that when fitting to the hand labelled points. This is because the search is able to compromise point position in favour of reducing texture error.

The point error is given as the mean distance between corresponding model and image label points (Pt-Pt) and as the mean distance between model points and the labelled image boundary (Pt-Bnd). Close examination of the hand labelled points suggests there is noise in their placement which may contribute considerably to the measured results.

The code was written in C++ and run on a 166MHz Pentium II under Linux. The mean time per model match was about five seconds for a 30 parameter, 10000 pixel model. This would take around one second on a modern PC.

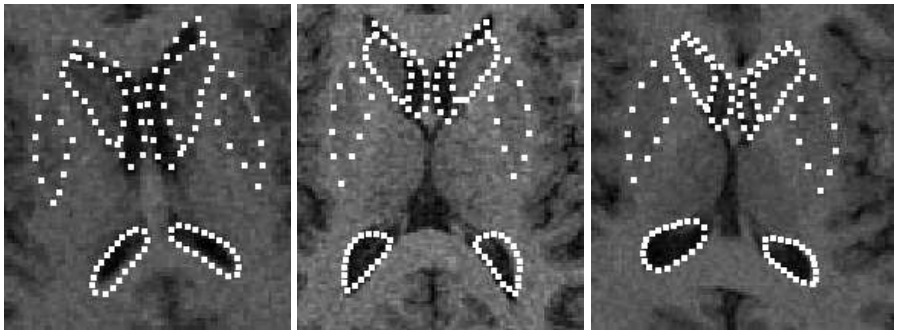


Fig. 5. Results of AAM search. Model points superimposed on target image

Measure	Miss-1-Out		Leave-all-in	
	Search	Best Fit	Search	Best Fit
Texture Error	12.8 (±3.1)	14.6 (±2.8)	10.9 (±2.2)	8.4 (±1.5)
Pt-Pt Error (pixels)	2.4 (±0.7)	0.9 (±0.3)	1.7 (±0.4)	0.4 (±0.07)
Pt-Bnd Error (pixels)	1.2 (±0.3)	0.6(±0.2)	0.9 (±0.2)	0.3(±0.05)

Table 1. Performance of AAM at matching brain model to images (± s.d.)(See Text)

4.1 Examples of Failure

Figure 6 shows two examples where the AAM has failed to locate boundaries correctly on unseen images. In both cases the examples show more extreme shape variation from the mean, and it is the outer boundaries that the model cannot locate. This is because the model only samples the image under its current location. There is not always enough information to drive the model outward to the correct outer boundary. One solution is to model the whole of the visible structure (see below). Alternatively it may be possible to include explicit searching outside the current patch, for instance by searching along normals to current boundaries as is done in the Active Shape Model [12]. This is the subject of current research. In practice, where time permits, one can use multiple starting points and then select the best result (the one with the smallest texture error).

5 Discussion and Conclusions

We have demonstrated that a deformable anatomical atlas can be built using statistical models of shape and appearance. Both the shape and the appearance of the atlas can vary in ways observed in the training set. Arbitrary deformations are not allowed. Matching to a new image involves minimising the difference between the synthesised atlas image and the target. This can be achieved rapidly using the Active Appearance Model matching algorithm.

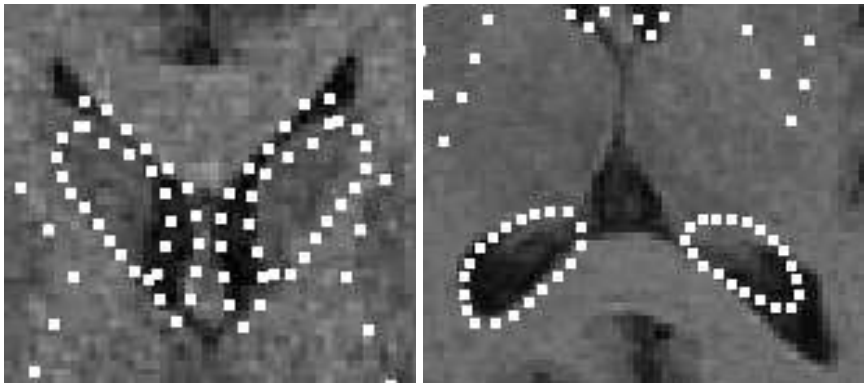


Fig. 6. Detail of examples of search failure. The AAM does not always find the correct outer boundaries of the ventricles (see text).

The AAM may not always give optimal results, but it would be straightforward to use a general purpose optimiser (eg Simplex or Powell [21]) to 'polish' the final fit.

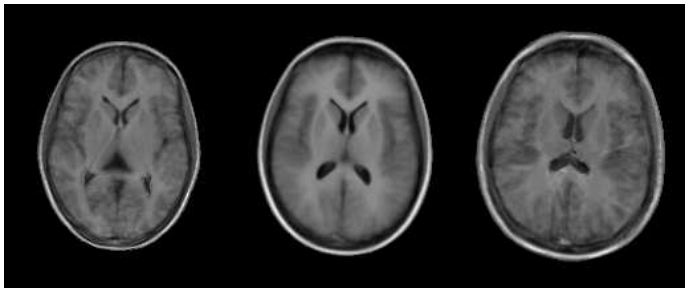
Though we only demonstrated on the central part of the brain, models can be build of the whole cross-section. Figure 7 shows the first two modes of such a model. This was trained from the same 72 example slices as above, but with additional points marked around the outside of the skull. The first modes are dominated by relative size changes between the structures.

The appearance model relies on the existence of correspondence between structures in different images, and thus on a consistent topology across examples. For some structures (for example, the sulci), this does not hold true. An alternative approach for sulci is described by Caunce and Taylor [7, 6].

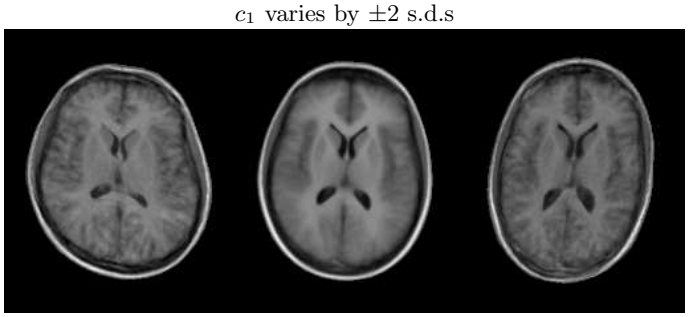
The approach has been demonstrated in 2D, but is extensible to 3D. The main complications are the size of the model and the difficulty of obtaining well annotated training data. Each mode of the texture model is the same size as an image - if many modes are used, the model could be rather large. Obtaining good (dense) correspondences in 3D images is difficult, and is the subject of current research [4, 5] [24] [15].

We hope to be able to match the models to different modalities by maximising mutual information, rather than minimising intensity errors. During search we would form an 'information difference' image, measuring the areas in the target image not well predicted by the model, and use this to update the current parameters.

We have shown how statistical models of appearance can represent both the mean and the modes of variation of shape and texture of structures appearing in medical images. Such models act as deformable anatomical atlases, in which the allowed deformation is learnt from a training set. The Active Appearance Model algorithm gives a fast method of matching the atlas to new images.



c_1 varies by ± 2 s.d.s



c_2 varies by ± 2 s.d.s

Fig. 7. First two modes of appearance model of full brain cross-section from an MR image

Acknowledgements

Dr Cootes is funded under an EPSRC Advanced Fellowship Grant. The brain images were generated by Dr Hutchinson and colleagues in the Dept. Diagnostic Radiology. They were marked up by Dr Hutchinson, Dr Hill and K. Davies and Prof. A.Jackson(from the Medical School, University of Manchester) and Dr G. Cameron (from Dept. Biomedical Physics, University of Aberdeen).

References

1. Bajcsy, and Kovacic, A: Multiresolution elastic matching. Computer Graphics and Image Processing **46** (1989) 1–21
2. Bajcsy, R., Lieberson, R., and Reivich, M: A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. J. Comput. Assis. Tomogr. **7** (1983) 618–625
3. Bookstein, F. L: Shape and the infromation in medical images: A decade of the morphometric synthesis. Computer Vision and Image Understanding **66**, 2 (1997), 97–118
4. Brett, A. D., and Taylor, C. J: A method of automatic landmark generation for automated 3D PDM construction. In: British Machine Vison Conference **2**, (1998) 914–923
5. Brett, A. D., and Taylor, C. J: A framework for automated landmark generation for automated 3D statistical model construction. In: IPMI (1999)

6. Caunce, A., and Taylor, C: Using local geometry to build 3D sulcal models. In: IPMI (1999).
7. Caunce, A., and Taylor, C. J: 3D point distribution models of the cortical sulci. In: British Machine Vision Conference. (1997) 550–559
8. Christensen, G. E., Joshi, S. C., and Miller, M: Volumetric transformation of brain anatomy. IEEE Trans. Medical Image **16** (1997), 864–877
9. Christensen, G. E., Rabbitt, R. D., Miller, M. I., Joshi, S. C., Grenander, U., Coogan, T. A., and Essen, D. C. V: Topological Properties of Smooth Anatomic Maps. Kluwer Academic Publishers, (1995), 101–112
10. Cootes, T., Edwards, G. J., and Taylor, C. J: Active appearance models. In: ECCV (1998), 484–498
11. Cootes, T., and Taylor, C: Modelling object appearance using the grey-level surface. In: British Machine Vision Conference, (1994) 479–488
12. Cootes, T. F., Hill, A., Taylor, C. J., and Haslam, J: The use of active shape models for locating structures in medical images. Image and Vision Computing **12** (1994), 276–285
13. Cootes, T. F., Taylor, C. J., Cooper, D. H., and Graham, J: Active shape models - their training and application. Computer Vision and Image Understanding **61**, (1995), 38–59.
14. Edwards, G. J., Taylor, C. J., and Cootes, T: Learning to identify and track faces in image sequences. In: British Machine Vison Conference (1997) 130–139
15. Fleute, M., and Lavallee, S: Building a complete surface model from sparse data using statistical shape models: Application to computer assisted knee surgery. In: MICCAI (1998) 878–887
16. Hill, A., Cootes, T. F., Taylor, C. J., and Lindley, K: Medical image interpretation: A generic approach using deformable templates. Journal of Medical Informatics **19** (1994) 47–59
17. Jones, M. J., and Poggio, T: Multidimensional morphable models. In: ICCV (1998) 683–688.
18. Maintz, J. B. A., and Viergever, M. A: A survey of medical image registration. Medical Image Analysis **2** (1998) 1–36
19. McInerney, T., and Terzopoulos, D: Deformable models in medical image analysis: a survey. Medical Image Analysis **1** (1996) 91–108
20. Nastar, C., Moghaddam, B., and Pentland, A: Generalized image matching: Statistical learning of physically-based deformations. In: ECCV Vol. 1 (1996) 589–598.
21. Press, W., Teukolsky, S., Vetterling, W., and Flannery, B: Numerical Recipes in C (2nd Edition). Cambridge University Press, (1992)
22. Sclaroff, S., and Isidoro, J: Active blobs. In: ICCV (1998) 1146–53
23. Subsol, G., Thirion, J. P., and Ayache, N: A general scheme for automatically building 3D morphometric anatomical atlases: application to a skull atlas. Medical Image Analysis **2** (1998) 37–60
24. Szekely, G., Kelemen, A., Brechbuhler, C., and Gerig, G: Segmentation of 2-D and 3-D objects from mri volume data using constrained elastic deformations of flexible fourier contour and surface models. Medical Image Analysis **1** (1996) 19–34
25. Wang, Y., and Staib, L. H: Elastic model based non-rigid registration incorporating statistical shape information. In MICCAI (1998) 1162–1173